2006 Generation Challenge Programme Competitive and Commissioned Research Project Mid-Year Reports

1 August 2006

Correct Citation: Generation Challenge Programme, 2006. *2006 Generation Challenge Programme Competitive and Commissioned Research Project Mid-Year Reports.* Mexico D.F.: Generation Challenge Programme.

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 4
3. Identifying the Physiological and Genetic Traits that Make Cassava One of the Most Drought Tolerant Crops 9
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 10
5. Unlocking the Genetic Diversity in Peanut's Wild Relatives with Genomic and Genetic Tools 12
6. Marker Development and Marker-assisted Selection for Striga Resistance in Cowpea 15
7. Measuring Linkage Disequilibrium across Three Genomic Regions in Rice 17
8. Targeted Discovery of Superior Disease QTL Alleles in the Maize and Rice Genomes 20
9. Development of Low-Cost Technologies for Pyramiding Useful Genes from Wild Relatives of Cassava into Elite Progenitors 24
10. Exploring Natural Genetic Variation: Developing genomic resources and introgression lines for four AA genome rice relatives 28
11. Functional Genomics of Cross-species Resistance to Fungal Diseases in Rice and Wheat (CEREALIMMUNITY) 40
12. Drought Tolerant Rice Cultivars for North China and South/Southeast Asia by Highly Efficient Pyramiding of QTLs from Diverse Origins 48
13. Development of Informative DNA Markers through Association Mapping in Maize to Improve Drought Tolerance in Cereals 54
14. Characterisation of Genetic Diversity of Maize Populations: Documenting global maize migration from the centre of origin 56
15. Determination of a Common Genetic Basis for Tissue Growth Rate under Water-limited Conditions across Plant Organs and Genomes 58
16. Isolation and Characterisation of Aluminium Tolerance Genes in the Cereals: An Integrated Functional Genomic, Molecular Genetic and Physiological Analysis 64
17. Allele Mining Based on Non-Coding Regulatory SNPs in Barley Germplasm 69

SP1 Commissioned Grants 72
2004—SP1CL2: SSR Analysis of Musa Germplasm 72
2005-01a: Completing Genotyping of Composite Germplasm Set of Barley 73
2005-01b: Completing Genotyping of Composite Germplasm Set of Wheat 75
2005-01c: Completing Genotyping of Composite Germplasm Set of Sorghum 76
2005-01d: Completing Genotyping of Composite Germplasm Set of Chickpea 77
2005-01e: Genotyping of Composite Germplasm Set, Tier 1, Maize 80
2005-01f: Genotyping of Composite Germplasm Set, Tier 1, Wheat 90
2005-01g: Genotyping of Composite Germplasm Set, Tier 1, Rice 92
2005-01j: Genotyping of Composite Germplasm Set, Tier 1, Common Bean 93
2005-01k: Cowpea Genotyping 94
2005-01l: Genotyping of Composite Germplasm Set, Tier 1, Cassava 95
2005-01n: Genotyping of Composite Germplasm Set, Tier 1, *Musa* 98

2005-02: Supporting Distribution of Reference Germplasm (CIP) 99
2005-02: Support and Distribution of Reference Germplasm, *Musa* (IPGRI) 100
2005-03c: Application of Molecular Markers for Gene Pool Division and Heterosis Estimation under Drought Stress Conditions in Sweetpotato 103
2005-03d: Molecular Characterisation of Tier 2 (orphan) Crops – Yam 104
2005-03e: Molecular Characterisation of Tier 2 (orphan) Crops – Lentil 104
2005-03f: Molecular Characterisation of Groundnut (*Arachis hypogea* L.) Composite Collection 105
2005-03g: Molecular Characterisation of Tier 2 Crops: Coconut 108
2005-04: Validation of Diversity Arrays Technology (DArT) as a Platform for Whole-Genome Profiling in Orphan Crops 110
2005-05: Assessing Ecotilling as a Methodology for Targeted Genotyping and SNP Discovery 113
2005-06: Supporting Emergence of Reference Drought Tolerance Phenotyping Centres 115
2005-07: Whole Plant Physiology Modeling of Drought Tolerance in Cereals 126
2005-08: Population Structure, Phenotypic Information and Association Studies in Long-generation Crops 129
2006-01: Developing Strategies for Allele Mining within Large Collections 136
2006-02: A Dataset on Allele Diversity at Orthologous Candidate Genes in GCP Crops 136
2006-03: SNP Analysis of the Genetic Diversity along the Rice Genome (HAPLORYZA) 139
2006-04: Phenotyping in the Field: Global capacity accessible to the GCP – Inventory of phenotyping resources and capacity for the CGP 140
2006-05: Development and Genotyping of a Faba Bean Composite Collection 141
2006-06: Genotyping of Composite Collection of Finger Millet 141
2006-29: Preparing IITA-Cassava Reference Germplasm for Distribution and Association Mapping 142
2006-30: Development and Genotyping of a Foxtail Millet Composite Collection 143
2006-31: Development and Genotyping of a Pearl Millet Composite Collection 144
2006-32: Molecular Characterisation of a Representative Pigeonpea Germplasm Sample 145
2006-33: Development and Genotyping of a Composite Germplasm Sample of Potato 146

SP2 Commissioned Grants 147
2004—SP2CL3: Sequencing of Normalised Cassava cDNA Libraries 147
2005-09: Systematic Evaluation of Rice Mutant Collections for Conditional Phenotypes with Emphasis on Stress Tolerance 148
2005-10: 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 152
2005-11: Legume Mutant Resource Development 152
2005-12: A Saturated Potato Mutant Population for Functional Genomics among Solanaceae and Tuber Crops 154
2005-13: Crop Gene Expression Profiles and Stress-gene Arrays 155
2005-14: Stress Response-enriched EST Resources for Targeted Species 160
2005-15: *Musa* Genome Frame-map Construction and Connection with the Rice Sequence 161
2005-16: Validation of Conserved Orthologous Markers 165
2005-17: Comparative QTL Mapping for Drought Tolerance 172

2005-35: Sequencing Multiple and Diverse Rice Varieties: Connecting Whole-Genome Variation with Phenotype 175

SP3 Commissioned Grants 176
2005-18: Development of Low-Cost Gene-Based Trait Assay Technologies in Cereals 176
2005-19: Evaluation and Deployment of Transgenic Drought-Tolerant Varieties—Potato 183
2005-19: Evaluation and Deployment of Transgenic Drought-Tolerant Varieties—IRRI 184
2005-20: Optimising Marker-assisted Breeding Systems for Drought Tolerance in Cereals through Linkage of Physiological and Genetic Models 187
2005-21: Planning for Effective Product Development, Delivery, and Use 189

SP4 Commissioned Grants 193
2004—SP4CL2: Establishing Fingerprinting/Allele Data Repositories for All GCP Partners 193
2005-22: Development of GenerationCP Domain Models 194
2005-23: Implementation of Web Services Technology in the GCP Consortium 195
2005-24: Application and Development of Web Services Technology 197
2005-25: Creation and Maintenance of Data Templates 198
2005-26: Management of the Generation CP Central Registry 198
2005-27: Integration of the High Performance Computing (HPC)-Facilities in the GenerationCP Toolbox 200
2005-31: Development of Ortholog-Function Display Tools 202
2005-32: Development of Crop Gene Expression Database and Data Mining Tools 203
2005-33: Development of an Integrated Decision Support System for Marker-assisted Plant Breeding 204
2005-34: GenerationCP Software Engineering and Collaboration Platform 205
2006-08: Data Analysis Support for Existing Projects in SP2 with Emphasis on Integrating Results across Gene Expression and QTL Mapping Experiments 206
2006-16: Development of an Integrated GCP Information Platform 206
2006-17: GenerationCP Data Quality Improvement and Assurance 208
2006-18: Creation of Institutional Bioinformatics Capacity (CIAT)-2006 208
2006-19: Creation of Institutional Bioinformatics Capacity (CIMMYT) 209
2006-20: Creation of Institutional Bioinformatics Capacity (CIP) 209
2006-21: Creation of Institutional Bioinformatics Capacity (ICARDA) 209
2006-22: Creation of Institutional Bioinformatics Capacity (ICRISAT) 210
2006-23: Creation of Institutional Bioinformatics Capacity (IITA) 211
2006-24: Creation of Institutional Bioinformatics Capacity (IPGRI) 211
2006-25: Creation of Institutional Bioinformatics Capacity (IRRI) 213
2006-34: Installation and Implementation of the ICRISAT LIMS at the Biosciences Eastern and Central Africa (BecA) Facility and IITA-Ibadan 213
2006-35: Data Analysis Support for Existing Projects in SP1 with Emphasis on Sampling Germplasm (DASSP1) 214

SP5 Commissioned Grants 217
2005-CB01: Training Materials for Molecular Genetic Diversity Analysis 217
2005-CB02: Development of Training Materials for a Course in Genomics and Comparative Genomics, and Design of Course Curriculum 218
2005-CB04: Online Bioinformatics Course 218

2005-CB05: Publication and Dissemination of Microsatellite Kits for Crops Genotyped in the First Tier 219
2005-CB08: Functional Genomics to Improve African Crops 219
2005-CB09: Molecular Markers for Allele Mining: A Workshop 220
2005-CB13: The Institute for Genomic Diversity's Interactive Resource Centre 220
2005-CB15: Distant Policies: A distance learning module for scientists on genetic resource policies and their implications for freedom-to-operate 221
2005-CB16: IP Matters: An Intellectual Property and Access & Benefit Sharing Helpdesk and on-line resource for the GCP community, partners, and stakeholders 222
2005-CB17: Reporting for Product Distribution: An asset inventory system for the Generation Challenge Programme 223
2005-CB23: Genotyping Support Service 223
2006-09: Training Course on Phenotyping 224
2006-11: Establishment of training materials for a course in association study/linkage disequilibrium mapping (TM_AS) 225
2006-15: Fellowships and Travel Grants 226
2006-28: Regional PGR Courses 227

Publisher's Note

This publication has been published in black and white and without appendices. To review color images or appendices for any of these reports, please contact the appropriate Principal Investigator directly or email Adriana Santiago at GCP Headquarters at asantiago@cgiar.org.

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
A.G. Condon, CSIRO
L. McIntyre, CSIRO
J. Bennett, IRRI
K. McNally, IRRI
R. Serraj, IRRI
S. Kikuchi, NIAS
Kouji Satoh, NIAS
R.C. Babu, Tamil Nadu Agricultural University, India
Zhengqiang Ma, Nanjing Agricultural University, China

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 exsertion 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 exserted 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).

Panicle exsertion

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 equalise 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. We found in pot grown that it is possible to produce similar levels of stress in rice and wheat under controlled conditions. Rice and wheat showed similar yield reductions with stress at vegetative stages, but the early reproductive stage was relatively more susceptible in wheat whilst late reproductive stage was more susceptible in rice. Our results indicate that carefully controlled environmental experiments can provide meaningful samples for comparative genomic studies, but differences in crop adaptation must be considered. The relevance of these findings to field-grown plants remains to be established.

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 mobilised 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 also examining the impact of the supply of exogenous sucrose on the rice plant under stress. When 1M sucrose is injected daily into the space between the penultimate leaf sheath and internode of half the tillers of drought-stressed plants, the decline in water status of all tillers of the plant is delayed, peduncle elongation is enhanced and OsCIN2 expression is maintained, but spikelet fertility is not improved. We are examining the mechanism by which exogenous sucrose affects all tillers.

We are using proteomics and microarray analysis to characterise 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 hybridised with rice and wheat RNA. Hybridisations 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. Of particular interest is the sharp down-regulation by stress and up-regulation by re-watering of the expression of specific genes concerned with cell division and cellular elongation. 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.

We have identified a set of 8 wheat varieties that show a good contrast in terms of drought tolerance. For each tolerant variety we have sought to match a sensitive variety that matches its breeding pedigree as closely as possible. The seeds of these varieties were grown and clonally propagated for further experiments. We decided to focus initially on two varieties: Sundor (drought senstive) vs. Sunstar (drought tolerant). These varieties were chosen firstly to reduce background variability; both are Australian semi-dwarf derivatives from the same original variety (Condor). The most important reason however is that they show differences in sensitivity to ABA treatment (Blum and Sinmena, 1995). Our previous work on cold-induced pollen sterility in rice has shown that ABA is a critical regulator of cold tolerance. Sundor and Sunstar plants were grown in the phytotron and staging experiments were carried out to determine the growth parameters for different reproductive stages (pollen and ovule meiosis, anthesis, grain-filling). In parallel, we have initiated establishing conditions for the reliable induction of water stress conditions in the reproductive organs of the plants. The severity and length of treatment has to coincide with the stage of reproductive development we are interested in.

Drought stress significantly decreased exsertion of panicles in rice cultivars IR64 (susceptible) and Apo (tolerant) around heading. Stressed IR64 showed poor panicle exsertion all stages. There was a significant difference between cultivars ($p<0.01$), water levels ($p<0.01$), stages ($p<0.01$), and a significant interaction between cultivar x water level ($p<0.05$). Even in control plants IR64 had shorter peduncles than Apo. At all stages control Apo showed 100% panicle exsertion. The tolerant wheat cultivar Weebil showed greater osmotic adjustment in the flag leaves than the sensitive cultivar Seri at all stages.

Floret fertility

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. Two rice and two wheat varieties which showed contrasting floret fertility were selected for microarray analysis. Using oligo arrays from the Beijing Genomics Institute, we investigated the gene expression profiles in rice panicles and wheat spikes under mild and severe drought stress and after re-watering. We are conducting a comparative analysis of differentially expressed genes. The gene expression data from the oligo arrays is being compared with data obtained using 9K panicle EST arrays printed at IRRI. We also checked the effect of exogenous ethylene on floret fertility of IR64, a drought susceptible rice variety, under drought stress and control conditions, with the corresponding changes in gene expression being monitored with BGI oligo arrays.

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 hybridisation to identify which cell-wall invertases and hexose transporters operate in each of the tissues of rice and wheat showing drought-sensitive development. We have developed multiple RNAi lines in the rice cultivar Nipponbare to examine the effect of down-regulating these genes.

A segregating population of ~900 lines has been developed to the F5 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 fertilisation 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.

We have used RT-PCR using wheat anther RNA to clone an anther-specific wheat cell wall invertase gene. We isolated a cDNA clone of the wheat IVR1 invertase gene that was previously shown to be repressed by drought stress. Cell wall invertase is repressed by stresses such as drought and cold and this leads to repression of sugar supply, abortion of pollen development and pollen sterility. Invertase gene expression can therefore be used as an indicator of drought tolerance and we are in the stage of collecting sufficient anther material from control and drought-stressed wheat plants of the drought-tolerant and drought-sensitive model varieties in order to investigate IVR1 gene expression in response to drought stress. This anther material will also be used for isolation of RNA for microarray screening and for construction of cDNA libraries. Preliminary experiments have shown that IVR1 expression is anther-specific and is not expressed in the ovules. We will use the cDNA libraries to isolate genes encoding the wheat ovule-specific cell wall invertase, as well as the ABA biosynthetic gene NCED. At the same time we are isolating wheat ovules from the meiosis and grain-filling stage (control and drought-stressed) for microarray experiments and gene expression studies.

Tangible outputs delivered:
Publication:

Ji X.M., Raveendran M., Oane R, Ismail A, Lafitte R, Bruskiewich R, Cheng S.H., Bennett J. 2005. Tissue-specific expression and drought responsiveness of cell-wall invertase genes of rice at flowering. Plant Mol Biol. 59:945-964.

Liu, J.X. Liao D.Q., Oane R., Estenor L., Yang X.E., Li Z.C., Bennett J. 2006. Genetic variation in the sensitivity of anther dehiscence to drought stress in rice. Field Crops Res.97:87-100 Kathiresan A., Lafitte H.R., Chen J.X., Mansueto L., Bruskiewich R., Bennett J. 2006. Gene expression microarrays and their application in drought stress research. Field Crops Res. 97:101-110

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:
David J. Mackill, IRRI
Mike Thomson, IRRI
Sigrid Heuer, IRRI
Xiaochun Lu, IRRI
Glenn Gregorio, IRRI
R. K. Singh, IRRI
Matthias Wissuwa, JIRCAS
Eduardo Blumwald, University of California, Davis
Emmanuel Delhaize, CSIRO
Zeba I. Seraj, Dhaka University, Bangladesh

Masdiar Bustamam, ICABGRRD, Indonesia
Collaborating Scientists:
Timothy J. Close, University of California, Riverside
Ghasem H. Salekdeh, ABRII, Iran
Massahiro Yano, NIAS

Mid-year Report

Objective 2: Development and validation of genetic markers for MAS

Markers designed and validated

***Saltol*:** At IRRI, three populations derived from the crosses 'BRRI dhan40' (susceptible)/ 'IR61920-3B-22-2-1' (highly tolerant); 'BRRI dhan28' (highly susceptible)/ 'IR50184-3B-18-2B-1' (moderately tolerant); and 'Kajalsail' (tolerant)/ 'IR52713-2B-8-2B-1-2' (tolerant) were used to test the effect of the *Saltol* QTL and to validate markers across different populations. Targeted mapping of the chromosome region containing *Saltol* (49.6 to 87.1 cM) on chromosome 1 was conducted using 20 SSR and two EST markers. The *Saltol* QTL was only detected for the 'BRRI dhan40'/ 'IR61920-3B-22-2-1' population. The SSR marker RM8094 was the most tightly-linked marker ($P<0.001$); while four other markers, RM1287, RM3412, RM493 and CP03970 were also significantly associated with salinity tolerance ($P<0.05$). An F_3 population of the cross 'BRRI dhan40'/ 'IR61920' was used to reconfirm this result. This was interesting because the tolerant parent in this population was not related to the tolerant parent (Pokkali) used for the original mapping population. QTLs were not detected at the *Saltol* locus for either of the other two populations. This was consistent with the phenotypes of the parents used to construct these populations, and indicates that the *Saltol* QTL may only be effective in specific populations.

At Dhaka University, F_3 progeny from tolerant BC_3F_2 NILs between Pokkali and recurrent parent IR29 were tested for seedling tolerance, and 90% of them had a SES score of 3-4 (on a scale of 1 to 9, with 1 being highly tolerant). DNA from 80 of these tolerant progeny from NIL families 9-5-5, 9-5-10 and 9-10-8 were tested for co-segregation of markers with the Pokkali allele. Percent co-segregation of markers between tolerant BC_3F_3 and Pokkali allele was found to be 60%, 55% and 39% at 11.45, 12.28 and 12.32 million base pairs (see Appendix, Table 1). Marker co-segregation at 12.67 Mbp, shown to be important in BC_3F_2 was only 11%. Marker co-segregation with the Pokkali allele at flanking positions of 11.25 and 12.73 Mbp was >50% for about 20 of the progenies tested so far. The rest (60 progenies) are currently being tested. BC_3F_3 progeny from tolerant BC_3F_2 are currently being screened for seedling salinity tolerance and DNA from tolerant 9-5-2, 9-5-8, 9-5-9, 9-10-2, 9-10-9 and 11-6-5 families will be tested for marker co-segregation with the Pokkali allele. A new marker at 11.45 mbp, designed from the Nipponbare sequence near OsHKT8 identified from Nona Bokra (Ren et al., 2005), shows >60% co-segregation with the Pokkali allele. These markers may be used for breeding saltol into recipient varieties like BR28, BR45 and BR11.

***Pup1*:** At IRRI, intron spanning primers were designed for Pup1 candidate genes with introns of about 1 kb for the development of molecular markers that will facilitate marker assisted introgression of *Pup1* into intolerant varieties. This was based on the hypothesis that introns are less conserved and markers are more likely to be polymorphic between tolerant and intolerant accessions. Out of a total of 23 genes for which intron spanning primers were designed, eleven showed polymorphism between Kasalath and Nipponbare (i.e no amplification in Nipponbare). These primers were tested in a range of tolerant and intolerant accessions and associate well with presence of the *Pup1* locus and P-uptake efficiency (Appendix, Table 2). Plants with high P-uptake rates predominantly posses the Kasalath allele of the respective genes. For the genes #50, #59 and #64 all accessions classified as tolerant carry exclusively the Kasalath alleles, whereas all intolerant accessions carry the Nipponbare allele.

Recipient cultivars identified

***Pup1*:** Seeds of a number of released upland rice varieties from Indonesia were sent to IRRI for use as susceptible parents, although the MTA is still in progress. The susceptibility of those upland varieties to P deficiency had been evaluated at two screening sites (Sitiung, West Sumatra, and Jasinga West Java), as well as in the greenhouse using soil samples from the field and in nutrient solution. The Sitiung screening site seems to be appropriate for field evaluation of breeding materials to test for tolerance to P-deficiency. The field test in Jasinga, West Java indicated that this location is not reliable due to inadequate uniformity of soil fertility (a new location needs to be identified to test breeding materials in more then one location). The data from the greenhouse study using the soil sample from Sitiung will allow at least two varieties to be recommended for introgression of the *Pup*1 gene.

MAS system developed and applied

***Saltol*:** At IRRI, a marker assisted backcrossing (MAB) system is being developed by selecting polymorphic SSRs at the Saltol locus and evenly spaced SSRs across the genome using a variety of potential donor and recipient parents. At the same time, a number of crosses have been made to start transferring *Saltol* into popular varieties in conjunction with Project 7 of the Challenge Programme for Water and Food. BC_1 seeds using the RIL FL478 (IR66946-3R-178-1-1) as the donor for the Pokkali *Saltol* allele in combination with four popular varieties (BR28, Swarna, Samba Mahsuri, and IR64) are available.

Additional crosses have also been made as part of a combined conventional and MAS breeding programme for salinity tolerance at IRRI. Six backcrosses were made for transferring *Saltol* into improved background / NPT, and they are currently at the BC_2F_1 generation. The donors are FL478 and Pokkali while the recurrent parents are IR64680-81-2-2-1-3, IR69727-37-2-1-3-2, IR72875-94-3-3-2 and IR73885-1-4-3-2-1-10. The BC_1F_1 plants were crossed with the recurrent parents based on the phenotypic as well as genotypic selection for 3 SSR markers linked to *saltol* locus (RM8094, RM3412 and RM493). Two single crosses and one double cross were made to combine the *Saltol* and *Sub1* in one genetic background as for most of the coastal saline areas both submergence and salinity tolerance are needed. The crosses were Cheriviruppu x IR82809-237; IR82810-407 x IR66946-3R-178-1-1 and the double cross involving these two single crosses. Three more single crosses were made to transfer the *saltol* into NPT / improved varieties (IR66946-3R-178-1-1 x IR71606-1-1-4-2-3-1-2 ; Pokkali x IR71676-90-2-2 and Pokkali x IR73717-46-1-3-3).

***Pup1*:** At IRRI, to prepare for MAS for Pup1, SSR markers on chromosome 12 were tested and polymorphic markers between Nipponbare and Kasalath were identified (Appendix, Figure 1). A large scale survey with the gene specific markers and the SSR markers of diverse upland and lowland germplasm selected by IRRI breeders is currently underway to further evaluate the markers and to test for presence of the *Pup1* locus in advanced breeding lines.

Objective 3: Gene identification and validation

Identification and expression analysis of candidate genes

***Saltol*:** At UC Davis, Northern blot analyses were performed to confirm the expression of 17 candidate genes that were found to be differentially expressed using DNA microarray analysis. RNA was isolated from the different rice lines that were subjected to salt stress at 100mM NaCl for 3 and 7 days. Out of 17 genes, nine were confirmed by RNA blot analysis. The analysis showed that while 8 genes were upregulated, one gene encoding a WD 40 repeat-containing protein was down-regulated (Appendix, Figure 2). Another Northern blot experiment was done using four rice cultivar, Pokkali, IR29, FL478 and one NIL. Salt stress was applied gradually, with 50 mM NaCl for three days and then the salt concentration was increased to 100 mM and maintained for 14 days. At this point RNA was isolated and Northern blots performed. The overall results suggested that 6 of the genes tested (alkaline invertase, Chloroplast splicing factor,

Myb transcription factor, pectinesterase, and serine threonine kinase and dual phosphatase) were not expressed differentially at the gradual application of salt. Clearly, a more quantitative approach is needed in order to assess changes in gene expression and corroborate the results obtained by DNA microarrays. We are now attempting semi-quantitative RT- PCR to analyse the expression of these genes. In addition, other genes in the QTL region such as a putative potassium transporter and a cation/chloride co-transporter will also be studied. RT-PCR sets of primers for the detection of 23 genes have been already designed and the optimisation is being finalised.

***Pup1*:** At IRRI, 47 candidate genes at the *Pup1* locus were identified in the 195 kb region on chromosome 12 based on the published genomic sequence of Nipponbare. Expression analysis of these genes by RT-PCR of RNA from plants grown hydroponically under low and high P conditions revealed differential expression of two genes (gene 66P28 and #25). Gene 66P28 was induced by P-deficiency only in Kasalath, whereas gene #25 was induced by P-deficiency in Kasalath, but was constitutively expressed in Nipponbare. The genes showed some sequence similarity to a hypothetical protein (66P28) and CF1 ATPase beta subunit (gene #25), respectively.

Subsequent sequencing of the Kasalath *Pup1* locus revealed major differences compared to the Nipponbare sequence (Appendix, Figure 3). The Kasalath *Pup1* region comprises about 350 kb and is therefore much larger than the respective Nipponbare region. Assessment of Kasalath *Pup1* candidate genes was performed and 69 genes were identified, including putative retro-transposon related and repeat sequences (Appendix, Table 3). Of the 69 putative Kasalath *Pup1* genes about 50% are putative transposable elements likely to be responsible for the observed major structural differences between the Nipponbare and Kasalath *Pup1* locus. At least four Kasalath genes seem to be absent from the Nipponbare and the 93-11 genome and are therefore unique to Kasalath (genes 39, 44, 50, 64; Table 3). Most genes show limited sequence similarity to known genes and some are present on different chromosomes in Nipponbare. Furthermore, major structural differences between Nipponbare and Kasalath predicted genes were observed (e.g. Kasalath genes 4 and 5 align to one 10 kb gene in Nipponbare) making detailed gene prediction analyses necessary. Gene specific primers for most genes were designed based on the Kasalath sequence information. Available primers were tested in PCR analyses of genomic DNA and in RT-PCR analysis of root and shoot tissues of plants grown under high and low P conditions, both hydroponically and under P-limited field conditions in Japan. First analysis of hydroponics root samples in a *Pup1* near isogenic line (NIL-C443) showed that some of the analysed genes are up-regulated under low P conditions, whereas others are down-regulated (Appendix, Figure 4).

In Iran, the physiological response of plants to P deficiency was analysed and the genotypes showed significant differences in root and shoot dry weight at third sampling during the booting stage (Appendix, Figure 5). We selected Nipponbare, pup1 and 6-4 for proteome analysis of root samples. We have been analysing samples from normal and stress condition in Nipponbare. We have identified several proteins present only at P=100 (spots 8,9,10,11,12) or up/down regulated in response to phosphorous deficiency (Appendix, Figure 6 and 7). The analysis in underway and we shall be able to present all responsive proteins in the next report. We have just run gels from other genotypes and will include them in analysis.

Expression analysis using contrasting NILs

Saltol: At UC Davis, the characterisation of new sets of BC_3F_5 NILs (provided by IRRI) is presently in progress. Nine NILs were selected based on the following criteria: (i) presence and absence of the Saltol region; (ii) Na/K concentration ratio of shoots; and (iii) scoring of seedling salt-tolerance. These NILs together with Pokkali, IR29 and Fl478 will be characterised physiologically and the expression of the candidate genes selected from the *Saltol* region of rice chromosome 1 determined.

Objective 4: Functional confirmation and impact assessment

Sets of contrasting NILs tested by NARES partners:

***Saltol*:** At IRRI, BC_3F_4 near-isogenic lines were developed from *indica* parents IR29 and Pokkali. These NILs composed of three families (BC_3F_3-11-6-5, BC_3F_3-9-5-2 and BC_3F_3-9-10-9) were generated via rapid generation advance (RGA) method. The population shows wide variation in response to different levels (12, 16 and 18 d Sm^{-1}) of salinity stress in the IRRI phytotron. The frequency distribution of the salinity response among the near isogenic lines at 12 d Sm^{-1} and 18 d Sm^{-1} salinity levels showed skewed behaviour towards the tolerant and sensitive parents respectively while at 16 d Sm^{-1}, the distribution was continuous and nearly normal. The background of these NILs will be checked with SSR markers before distributing to the NARES partners for impact assessment.

Objective 5: Evaluation of candidate genes

Overexpression and heterologous expression of candidate genes

***Saltol*:** At UC Davis, cloning of major component genes of the QTL region controlling low Na^+/K^+ for salt tolerance from the Pokkali cultivar is in progress to prepare for transformation experiments. There are two putative cation transporters in the *Saltol* region of chromosome 1. One is the potassium transporter SKC1 (OsHKT8) that was recently isolated from Nona Bokra by Ren et al. (2005). The second transporter is a cation chloride co-trasporter. The SKC1 gene from Pokkali was cloned using the Gateway technology. Based on the sequence analysis, Pokkali SKC1 is identical to the similar gene in Nona Bokra, and based on multiple alignments, SKC1 from Pokkali has 32% and 53.7% identity with the published OsHKT1 and Oshkt2 (Horie et al.,2001), respectively. The Pokkali SKC1 is 1665 bp and encodes a 555 amino acid protein. The translated sequence contains the cation transporter domain and a Myb binding domain. The SKC1 gene is now in a destination vector for its heterologous expression and physiological and biochemical characterisation and also for plant transformation. The cation chloride transporter is presently being cloned using the Gateway system. The transcript length of cation chloride is 2.7 kb and the genomic structure contains 12 intron and 13 exons. Our cloning strategy is the isolation of the core sequence first and to perform RACE PCR in order to obtain the full length cDNA. The 13^{th} exon containing 1094 bp has been already isolated. The isolation of the other exons is in progress. We have already made constructs of the 17 genes that are now available for transformation experiments, either by overexpression in the susceptible genotype IR29 or silencing in the tolerant genotype Pokkali.

Objective 6: Capacity building

Currently two students are pursuing their PhD training through this project (Ms. Ellen Tumimbang at UCD/IRRI and Mr. Joko Prasetiyono, Bogor Agricultural University Indonesia/IRRI).

Tangible outputs delivered:

Intermediate outputs as described above are being developed and further tested. Tangible outputs under development include an optimised MAS system for *Saltol* and *Pup1*, in addition to transgenic and MAS-developed varieties with salinity and p-deficiency tolerance.

Deviations from the work plan:

The sets of contrasting NILs for *Saltol* and *Pup1* were scheduled to be tested by the NARES partners in the current reporting period. However, the backgrounds of these NILs are being checked at IRRI using additional SSR markers and phenotype tests are ongoing to select the best NILs to be sent to our NARES partners.

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

Principal Investigator:
Alfredo Alves, EMBRAPA/CNPMF
Co-Principal Investigators:
Hernán Ceballos, CIAT
Martin Fregene, CIAT
Morag Ferguson, IITA
Tim Setter, Cornell University

Mid-year Report
Activities accomplished:

1. *In vitro* micropropagation of drought tolerant cassava contrasting varieties at Embrapa, Brazil. The *in vitro* plants imported from CIAT were not able to go directly to the field due to the lost of many plants during delayed importing process.

2. Establishment of a field trial with 56 drought tolerant contrasting cassava varieties in Petrolina, Brazil, to evaluate drought tolerance related traits.

3. Establishment of a crossing block at CIAT, Colombia, for pair-wise crosses between the contrasting genotypes to produce segregating populations for genetic mapping of traits associated with drought tolerance.

4. Establishment of a trial with 36 and 34 drought tolerant genotypes in ICA-Nataima and El Guamo (Department of Tolima), Colombia, respectively to evaluate drought related traits

5. Field establishment of a trial with 10 contrasting cassava genotypes in Tamale, Ghana. These varieties were selected from the SARI current germplasm and from farmers who indicated the drought tolerant genotypes. Growth analysis data were collected.

6. Field establishment of a trial with 9 drought tolerant contrasting cassava varieties at Hombolo, Dodoma, Tanzania. These 9 varieties were identified in two previous field evaluated trials. Preliminary data on sprouting (establishment), disease (cassava mosaic disease-CMD and cassava brown streak disease-CBSD) and pest (cassava green mites-CGM) incidence have been collected.

7. Preliminary studies to assess the genetic distance amongst contrasting drought tolerant cassava varieties at IITA, Nairobi, Kenya. Ten varieties, selected from IITA drought tolerant mapping population, were screened at eight SSR loci and data analysed using NTSYS and the algorithm of Nei's similarity.

Tangible outputs delivered:
1. Drought tolerant contrasting cassava varieties identified
2. *In vitro* plants of the drought tolerant contrasting varieties produced
3. Crossing blocks of drought constrasting cassava varieties established
4. Field trails in the target sites (Brazil, Colombia, Ghana, and Tanzania), to evaluate drought tolerance related traits, established

Deviations from the work plan:
The following constraints have been faced in this project: 1) Delay for releasing the plant materials (imported from CIAT) from Brazilian Ministry of Agriculture to CENARGEN (for quarantine); 2) Delay from CENARGEN to CNPMF; 3) Damages occured in the imported *in*

vitro plants; and 4) Delay for acquisition of equipments for evaluation of physiological parameters in the field.

These constraints have caused significant delay of the original workplan and can be minimised by an 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:
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
Marcos Malosetti, WUR
Ky Mathews, CSIRO

Mid-year Report

The planned tasks for the period December 2005 - May 2006 were (continuation of):

1. Data base construction
2. Genotypic and environmental characterisation
3. Single cross GxE and QTLxE analyses

Close by deliverables (planned for the end of September 2006) are 1) workable data base with genotypic, phenotypic, and environmental information; 2) draft papers for wheat and maize on useful GxE and QTLxE methodology and results of the application of this methodology.

Maize

Tasks 1, 2 and 3 in the proposed schedule have been largely completed for three important CIMMYT maize populations (C1, C4, C6).

Task 1: The database for maize is operational and ready to use in GxE and QTLxE analysis. All phenotypic and genotypic information has been collated and is ready-to-use, as well as most of the environmental information. Work on selection and inclusion of climatic information has started. Values for some drought stress indices need to be calculated.

Task 2: Inspection of which drought stress indicators to use has been initiated. For both tasks 1 and 2, extensive communication with the data owners/producers took place, leading to a better understanding of the characteristics of the maize (and wheat) data sets.

Task 3: The C1 population has been analysed using a mixed model approach. GxE and QTLxE single trait models were constructed and applied to the C1 population. Furthermore, a multi-trait multi-environment QTL analysis was developed and applied (see ppt slides). Results of these analyses were reported at the "Workshop Gene-Plant-Crops-Relations" held in Wageningen, 23rd to 26th of April, 2006, and two publications in the form of Proceedings chapters were submitted.

Two other populations are presently being analysed with mixed models for GxE and QTLxE, one in Wageningen by Marcos Malosetti (C4) and a second one at CIMMYT by José Crossa and Mateo Vargas (C6).

Wheat

Task 1: Data included in database:
Molecular data: 587 markers run across the SeriBabax RIL population prior to the commencement of this project: 74 SSR, 249 AFLP, 264 DArT. A map was constructed following best practices learnt at Wageningen UR by Ky Mathews (CSIRO). The final map contained 39 linkage groups representing 20 of the 21 wheat chromosomes.

Phenotypic data: 10 trials (8 Australian and 2 Mexico) have been collated and stored in a database. Four trials are currently being conducted and data from these will be available in 2006. Data from a further 8 trials, conducted in Mexico from 1999 to 2001 will be available in mid-2006.

Task 2: Genotypic characterisations have been finished and data were imported into the data base. Potential environmental stress indices and covariables to be included in database and analyses have been discussed. Final decisions on this issue need to be taken the coming half year.

Task 3: Prior to commencing QTLxE analyses it is important to have an understanding of the relationships between traits and between environments. This requires identifying target traits and describing the extent of their genotype by environment interaction, at the phenotypic level. Target traits may be those which have a direct relationship with yield, such as thousand grain weight and grain size. However, when investigating drought adaptation, traits may be physiologically driven such as water soluble carbohydrate and canopy temperature. GxE analyses have been undertaken for part of the identified target traits (see ppt slides). Further analyses follow soon.

The mixed model approach to QTLxE analysis described in Malosetti et al (2004) is being implemented and preliminary results are available (see ppt slides).

Annual project meeting was held in Wageningen on April 27. Attendants were Fred van Eeuwijk, Scott Chapman, Matthew Reynolds, José Crossa, Marcos Malosetti and Ky Mathews. Jean Marcel Ribaut send a message that he could not attend the meeting. Minutes of the meeting are available on request.

Tangible outputs delivered:
Presentations and papers at "Workshop Gene-Plant-Crops-Relations", Wageningen, April 23-26, 2006:

Fred A. van Eeuwijk, Marcos Malosetti and Martin P. Boer (2006) Modelling The Genetic Basis Of Response Curves Underlying Genotype By Environment Interaction. Proceedings Workshop Gene-Plant-Crop Relations, April 23-26 2006, Wageningen, The Netherlands. Eds. Paul Struik & Gon van Laar. Springer Verlag. In press.

Marcos Malosetti, Jean-Marcel Ribaut, Mateo Vargas, José Crossa , Martin P. Boer and Fred A. van Eeuwijk (2006) Multi-trait multi-environment QTL modelling for drought stress adaptation in maize. Proceedings Workshop Gene-Plant-Crop Relations, April 23-26 2006, Wageningen, The Netherlands. Eds. Paul Struik & Gon van Laar. Springer Verlag. In press.

Deviations from the work plan:
The project is following well the schedule as described in the Amended project proposal.

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

Genetic maps

Aliquots of primer pairs for microsatellite markers have been exchanged between ICRISAT, EMBRAPA and UCB. Aliquots of anchor marker primers have been sent from Aarhus to UCB and ICRISAT.

UCB/EMBRAPA/Aarhus

A further improved version of the published map of the AA genome of *Arachis* (Moretzsohn et al 2005) has been produced. The map now has 184 SSRs, 65 anchor markers, 15 resistance gene analogues and 24 other markers, placed in the correct number of 10 linkage groups. Electronic mapping of marker homologues in the genome of *Lotus japonicus* now gives 50 points of contact between the two genomes. Synteny between *Arachis* LG3 and *Lotus* LGs4 and 3 and *Arachis* LG6 and *Lotus* LG1 is particularly clear.

575 SSRs have now been screened for the BB (*A.ipaensis* and *A.magna*) genetic map, 151 being polymorphic. 108 have mapped in 10 linkage groups. Clear synteny between three linkage groups of the AA and BB maps are visible.

450 SSRs have been screened in the tetraploid population 138 are polymorphic, 74 for the genome AA and 32 for the genome BB.

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

F3 seed from the tetraploid mapping population is being harvested, we anticipate that a bioassay for disease resistance will be possible in the next Brazilian rainy season.

ICRISAT

All available SSR marker in groundnut have been screened in the parents of a population developed for transpiration efficiency (TAG24 and ICGV86031). So far 74 markers have shown polymorphism. These polymorphic markers are being screened on the RIL progenies.

Aarhus/EMBRAPA/UCB/ICRISAT

The development of further markers for *Arachis* continues

Datamining and marker development from candidate genes

EMBRAPA/UCB

From our *Arachis stenosperma* ESTs, 73 candidate genes, mostly for for disease resistance and drought tolerance, have been selected. Amplification and sequencing of regions of 21 these candidate genes has been done in the parentals of the AA genome mapping population. So far, clear well characterised SNPs have been detected in 14 of these candidate genes.

Physical mapping.

IBONE/EMBRAPA/UCB

Analysis of the repetitive elements present in *Arachis* is continuing. For the best characterised elements we now have pseudocontigs of about 6kb and 14kb. Flourescent *in-situ* hybridisation has shown a dispersed euchromatic location for five repetitive element sequence fragments. Several right-hand borders of LTRs (suitable for marker development) have been identified. Isolation of new repetitive elements continues

Production of BAC libraries.

CIRAD/EMBRAPA

Methods for the extraction of nuclei and the purification of high molecular from *Arachis* have been optimised. A BAC library from *A.duranensis* with an average insert size of more than 120kb and approximately 80,000 recombinants has been obtained (about 6.5x coverage). Validation of the AA BAC library and production of the BB BAC library is ongoing,

Drought tolerance assays in wild *Arachis* .

ICRISAT

A preliminary bioassay for component traits of drought tolerance has been done and has revealed a large variation of response to a progressive exposure to water deficit among different species. In particular, one accession of *A. hoehnei* had a decline in transpiration that occurred at a very low level of soil moisture (opportunistic behavior), compared to one accession of *A. duranensis* in which transpiration declined at very high levels of soil moisture (conservative behavior).

Amphidiploids

The amphidiploid (*A.ipaënsis* x *A.duranensis*)^{c} ,supplied by EMBAPA has been successfully germinated in ICRISAT. The amphidiploid [(*A.diogoi x A.cardenasii) x A.batizocoi*]^{c} has been sent by Prof. Charles Simpson to ICRISAT.

The amphidiploid (*A.ipaënsis* x *A.duranensis*)^{c} ,supplied by EMBAPA has been successfully multiplied in CERAAS. Seeds of the amphidiploid [(*A.diogoi x A.cardenasii) x A.batizocoi*]^{c}, sent by Prof. Charles Simpson, have arrived in CERAAS.

EMBRAPA

A new amphidiploid (*A. hoehnei* x *A. simpsonii*)^{c} and a new diploid hybrid *A. magna* KG 30097 x *A. stenosperma* V 15076 have been produced. *A. stenosperma* V15076 was selected because it was collected from coastal sand dunes. Multiplication of existing amphidiploids is being done and the construction of new amphiploids is ongoing.

ICRISAT

Wild *Arachis* crossing has been initiated to develop new amphidiploids. So far a number of hybrid seeds have been produced. These seeds will be germinated and used for colchicine treatment and amphidiploid production.

Proof of principle disease resistance introgression into peanut

Four F_2 BC1 [*A.hypogaea* cv. BR1 x (*A.ipaensis* KG 30076 x *Aduranensis* V 14167)^{c}] plants were selected in 2005, for resistance against leaf spots and rust. The F3 from these plants are being field assayed and the plants backcrossed with BR-1.

Tangible outputs delivered:
1) Exchange of molecular markers between EMBRAPA, ICRISAT and Aarhus.

2) A gene-rich microsatellite-based genetic map for the AA genome of *Arachis*, with anchor markers and an initial comparison with the genomes of *Lotus*, *Medicago,* and the BB genome of *Arachis.*

3) A first draft microsatellite-based map of the BB genetic map of *Arachis*

4) A bank of 3500 assembled ESTs from *A.stenosperma* assembled, and with access in an easy-to-use Web-based interface.

5) Development of optimised methods for the extraction of high molecular weight DNA from *Arachis.*

6) A BAC library with approximately 6.5x coverage for the AA genome of *Arachis* (*A.duranensis*).

7) Initial analysis of the large variability in the response to drought among different wild *Arachis* species

8) Supply of two amphidiploids to CERAAS and ICRISAT: (*A.ipaënsis* x *A.duranensis*)^{c} from EMBRAPA, and [(*A.diogoi x A.cardenasii) x A.batizocoi*]^{c}, from Prof. Charles Simpson.

9) Production of a new amphidiploid (*A. hoehnei* x *A. simpsonii*)^{c}.

10) Four lines derived from amphidiploid x cultivated cross selected for disease resistance.

Deviations from the work plan:
Setting up contracts, transfer of money to partners, the transfer of the amphidiploid *A.ipaensis* KG30076 x *A.duranensis* V14167)^{c} to partners in SENEGAL and ICRISAT-India took longer than anticipated. This has delayed work.

Multiplication of wild *Arachis* seed has taken longer than anticipated, and the drought assays use more seeds than initially anticipated; this will delay the evaluation of wild *Arachis* for drought tolerance.

Transfer of any material derived from Brazilian germplasm outside Brazil is now subject to a new provisional National Law. Whilst the exchange of this type of germplasm may be possible in the future, for the moment it is probably not possible.

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
Ndiaga Cissé, CERAAS, Senegal
Moctar Wade, CNRA
Satoru Muranaka, IITA
Christian Fatokun, IITA
Adebola Raji, IITA
Ivan Ingelbrecht, IITA

Mid-year Report

The present project seeks to develop molecular markers and establish MAS for *Striga* resistance in cowpea. The period between November 2005 and April 2006 is generally an off-season period for most cowpea growing regions. There were no field studies during this period.

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

To improve the efficacy of the 61R primers were redesigned based upon the sequences of the 61R SCAR, and the improved 61R marker (MAHSE2) was tested on three different RIL populations previously phenotyped for *Striga* resistance. The improved marker is much cleaner and clearer in its correlation to SG3 and SG1 and was found to be also loosely linked (~20 cM) to SG5. This key finding suggests that SG1, SG3, and SG5 are all in the same vicinity in LG1. To develop additional more tightly linked markers for resistance to *S. gesnerioides* from Cameroon (SG5), the RIL population derived from IT84S-2246 X Tvu-14676 was screened in pots against SG5, and resistant and susceptible individuals identified. Bulked segregant analysis (BSA) was used to evaluate selective primer combinations that amplified polymorphic DNA bands between the parental lines. One of the markers AGG/CTT 200B showed polymorphic fragments of ~150 and 300 bp that were linked to susceptible phenotype. The PCR products of AGG/CTT 200B were isolated and their nucleotide sequence determined and new primers designed for use in additional rounds of BSA. One of the primer combinations, AGG/CTT 200BR/50R showed a single band present in IT84S-2246 and segregating with the susceptible phenotype in bulks prepared from both SG5 and SG3.

Development of molecular markers for identification of cowpea *Striga* races

We previously reported the identification of a series of AFLP primers that discriminates among the different races of *S. gesnerioides* parasitic on cowpea. Based on these initial studies, we are converting these markers to SCARs to facilitate rapid identification of pathogen diversity in the field.

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

In order to evaluate and screen different cowpea varieties and breeding lines against known races of cowpea *Striga* and possible identification of new races, multi-location trials involving different countries in West Africa were conducted in 2005. The *Striga* `hotspots` location were; Kano and

Borno States in Nigeria; Toumnia in Niger Republic; Zakpota in Benin Republic; Cinzana and Koporo in Mali; Maroua in Cameroon, Ngalbane in Senegal and Ouagadougou in Burkina Faso. There was no single variety or breeding line that was resistant to *Striga* across all the locations, however, some lines showed resistant to *Striga* in more locations than others. For example, B301 was resistant to *Striga* in most locations except in Ngalbane, Senegal and Zakpota, Benin Republic. We know from previous reports that B301 is not resistant to *Striga* race 4 found in Zakpota, Benin Republic, however, the observation that B301 is also not resistant to *Striga* found in Senegal further cements the recent evidence that there is *Striga* race 6 prevalent in Senegal.

Previous reports indicate that the *Striga* race 1 prevalent in Nigeria is similar to the one found in Niger Republic. The results of this study seem to agree with the previous reports with an exception of one line (IT00K-1217) that was resistant in Nigeria but susceptible in Niger Republic. Some genotypes (Aloka local, TVU7778 and TVX3236) were susceptible to *Striga* in all the locations tested. This trial will be conducted again in 2006 to ascertain the results of 2005.

Testing of markers and development of MAS protocols

A total of 50 genotypes with different genetic background and an F6 population (111 individuals), created from a cross between a *Striga* resistant IT97K-499-35 and *Striga* susceptible IT97K-461-4, were used in this study. The markers tested were: 61R, a SCAR marker associated with resistance to SG3 and SG1 on LG1. A second SCAR (SEACTMCAC83/85) linked to SG3 on LG1 (Boukar *et al*., 2004) and 8 AFLP primer combinations were also tested. The results obtained with SEACTMCAC83/85 SCAR marker were inconclusive and no further tests will be done using this marker.

A total of 26 out of the 50 genotypes that were tested were positive for the 61R marker. A positive result was more common in those genotypes (83%) that also showed resistant to *Striga* race 3 in pot screening. Testing of an F6 population using the 61R SCAR showed 1:1 segregation ratio for resistant (positive) to susceptible (negative) lines. Unfortunately the results obtained from the testing of these lines under artificial infestation with Striga race 3 to determine correlation between 61R and actual resistance to *Striga* race 3 were mixed up. The experiment will be repeated in June 2006. From the evidence collected so far it appears that 61R is an excellent predictor of SG3 and could be used in MAS this year.

Evaluation of drought tolerance in an RIL population Dan Ila x TVu7778

To evaluate drought tolerance of an RIL population (DT1) obtained from drought tolerant (DT) cultivar Dan Ila and susceptible (DS) cultivar TVu7778 an experiment was conducted in the field at IITA Ibadan, Nigeria. Plants in the stressed plots were irrigated weekly for the first three weeks after planting after which watering was carried out once in two weeks. The non-stressed plots were irrigated twice every week until pods showed signs of physiological maturity. There was wide variation among 120 RILs and also between the two parents with regards to biomass yield loss and grain yield loss under control (irrigated) condition. Dan Ila (DT) showed smaller biomass yield loss under drought condition than drought susceptible TVu7778. The fact that some RILs showed smaller yield loss than Dan Ila (DT) and larger yield loss than TVu7778 (DS) means the mechanisms related to drought tolerance in the field is not controlled by single major gene but rather QTLs. Further studies are being conducted to identify QTLs that contribute to drought tolerance in cowpea. There are two strategies for plant to perform well under drought condition; drought escape and drought resistance. Drought escape is to complete their life cycle before severe drought, however, no relation between flowering date and biomass loss under drought condition was observed in this experiment and the RILs which showed smaller biomass loss tend to maintain higher grain yield production. These results indicate that some DT1 RILs have ability to maintain their biomass production under drought condition and it is not related to drought escape. Further studies are being conducted to elucidate the mechanisms behind the difference of drought resistance among DT1 RILs.

Capacity building
Ms. Charlotte Dolou Tonessia, a graduate student from Cote d'Ivore (working with Jean Francois Remi at CERAAS in Senegal joined the UVA, USA laboratory in January 15, 2006 and will remain until July 15, 2006. Dr. Moctar Wade from ISRA CNRA, Senegal, will visit the laboratory between May 15, 2006 and June 20, 2006. While at UVA, the trainees participate in various courses, colloquia, and seminars offered by the University of Virginia in addition to mentored in-lab training

Tangible outputs delivered:
- Development of a SCAR marker, 61R – a good predictor of SG1 and SG3 resistance gene on LG1.
- AFLP markers linked to *A. vogelii* resistance gene in cowpea
- A number of AFLP markers linked to SG1, SG3 and SG5 resistance genes in cowpea
- Series of AFLP primers that discriminates among the different races of *S. gesnerioides* parasitic on cowpea
- A number of advanced populations for both *Striga* resistance and drought tolerant studies
- Response of different cowpea genotypes to different *Striga* races in selected West African countries
- Training of NARS partners

7. Measuring Linkage Disequilibrium across Three Genomic Regions in Rice

Principal Investigator:
Susan McCouch, Cornell University
Co-Principal Investigators:
Michael Thomson, IRRI
Endang Septiningsih, IRRI

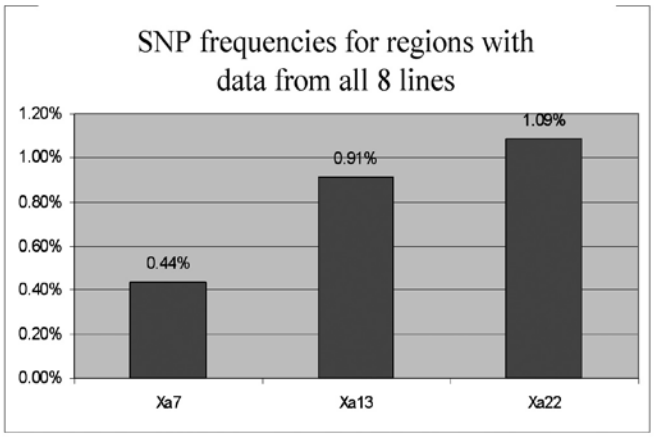
Mid-year Report
This is a one-year project with 3 primary activities:
1) SNP discovery and marker development across 3 genomic regions associated with resistance to bacterial blight (using 8 diverse Indonesian rice accessions)
2) Evaluation of SNP diversity in 96 diverse Indonesian accessions and measuring Linkage Disequilibrium (LD) in the 3 target regions
3) Technology transfer and capacity building

During the first 6 months of this project, we addressed the first and third objectives of this project. A scientist from Bogor (Fatimah Suwardjo) spent 4 months at Cornell (April 6 to July 31, 2005) working closely with a Cornell research assistant (Nicholas Polato) to complete the work outlined in the previous report (November 2005).

During that time, 300 PCR primers were designed across the 3 target regions of interest surrounding the bacterial blight resistance genes, Xa7 (chr. 6), xa13 (chr. 8) and the Xa4/Xa22/Xa26 cluster (chr. 11) and 250 primer pairs were selected for sequencing based on confirmed amplification in two rice accessions (a *tropical japonica* and an *indica* cultivar). The 250 primers were used to amplify products from 8 diverse, Indonesian accessions (4 *tropical japonica* and 4 *indica* varieties) and these products were sequenced in both directions as the basis for SNP detection. SNPs were called and the data was used to compare the frequency and distribution of SNPs within and between the *japonica* and *indica* sub-populations. The results of this work were summarised in the last report and will not be repeated here.

The frequency of SNPs in this sample was 1 SNP every 132 bp, but because the SNPs were not distributed evenly across the 3 regions (Fig. 1), we decided to do additional SNP discovery using a targeted approach to try to fill in the gaps.

Figure 1.



Our objective was to identify a SNP approximately every 40 kb across each of the target regions. To do this without additional sequencing, we analysed the Nipponbare and 9311 sequence data available in the Gramene database and targeted SNPs discovered by comparing these genomes. Primers were designed to bracket SNPs using the cv Nipponbare (*japonica)* genomic sequence (TIGR version 3 pseudomolecules) as template. The primers were used to generate PCR products which were sequenced in both directions (by Genaissance Pharamceuticals, Inc in New Haven Connecticut, using the Sequenom MassARRAY platform) using DNA from the 8 Indonesian accessions described previously. Sequence information from these diverse accessions was aligned and analysed for the presence of SNPs.

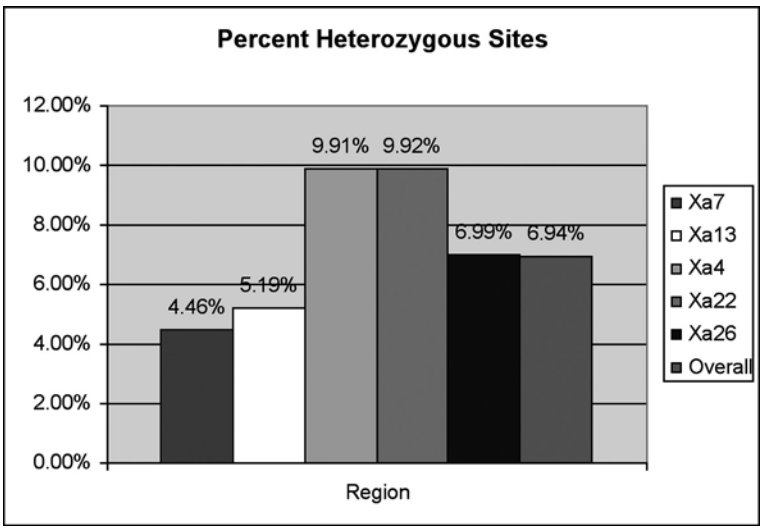
Upon completion of SNP detection in the target regions for the subset of 8 Indonesian varieties, a final set of SNP target sites was selected based on the recommendations of Genaissance and on the SNP site frequency spectrum in the material we analysed. SNPs were selected to be evenly spaced across each of the ~1 Mb target regions, resulting in a SNP density of one SNP every 40 Kb on average. We designed primers to be used for the SNP detection assay in the 96 diverse DNA samples based on the sequence data we had obtained for the preliminary set of 8 Indonesian varieties. SNPs were selected from the 25 sequenced fragments we had from the *Xa7* and *Xa13* regions, and the 75 fragments in the Xa4*/Xa22*/*Xa26* cluster on chromosome 11. Primers were selected and information was prepared according to Genaissance's specifications. This information along with the 96 DNA samples (extracted at ICABIOGRAD in Indonesia) was transferred to Genaissance as the basis for genotyping.

SNP detection primers were then designed by the automated system at Genaissance for each of the 125 target SNPs in our dataset. Using the Sequenom MassARRAY system, genotypes were generated for the 125 SNP loci in each of the 96 varieties. An initial subset of the genotypic data was sent to us by Genaissance as of May 3, 2006. This included data for 64 of the 125 SNPs applied to the 96 samples. There were 5846 genotypes in total, with an average coverage of 95.15%. Genotypes for the remaining samples are expected to be received in the coming 3-4 weeks. For 18 of the target SNPs, assays could not be designed due to sequence features that caused them to be rejected.

We are currently waiting for Genaissance to send the rest of the data (61 SNP targets on 96 accessions). Upon receipt of the complete dataset, we will apply the analysis techniques learned over the last 2 months to characterise the extent of Linkage Disequilibrium in the three target regions and we will compare levels of LD in the two sub-populations found in Indonesia: *tropical japonica* and *indica*.

Interestingly, analysis of the preliminary dataset shows a relatively high level of heterozygosity at the SNP loci. Of the 5,846 genotypes generated so far, 406 of them are heterozygous (6.94% overall). High levels of heterozygosity are consistent with the fact that the samples represent traditional landrace varieties rather than modern, improved varieties. It is also interesting that the levels of heterozygosity appear to be positively correlated with SNP frequencies in the three target regions. As can be seen by comparing Fig. 1 (above) with Fig. 2 (below), levels of heterozygosity were lowest in the Xa7 region and highest in the Xa4/Xa22 region.

Figure 2.



Request for an additional no-cost extension until October 2006
Due to the outsourcing of the genotyping work to Genaissance Pharmaceuticals we do not yet have the complete genotype dataset. As such additional time will be necessary to perform all of the LD analysis. Currently we have received only an incomplete preliminary sample of the data. While we expect the remainder of the data to arrive within the coming weeks, we expect that an additional 3 to 4 months may be necessary to complete all of the analyses.

Technology transfer and capacity-building. From February 16 – May 12, 2006, Kurniawan Tri Jatmiko ("Rudi"), a scientist from ICABIOGRAD in Bogor, visited Cornell University for training in Bioinformatics, linkage disequilibrium (LD) analysis, and DNA sequencing techniques. He worked closely with Nick Polato learning how to analyse LD and evaluate levels of genetic diversity (using SNP data from another project). During this time, a variety of software packages were evaluated for use in these analyses, including TASSEL, DNAsp, and Powermarker. Based on this experience, we have developed a collaboration with Carlos Bustamante's group in the Dept. of Biological Statistics and Computational Biology at Cornell to develop new statistical approaches for analysing population genetic parameters of inbreeding species such as *O. sativa.*

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) All of the sequences from the 8 rice accessions have been aligned and SNPs have been called for all of them. Information about SNP frequency and location is available for these regions.

4) 125 SNP detection assays have been designed and carried out on 96 diverse Indonesian accessions at Genaissance Pharmaceuticals. Genotype information for these assays will be available shortly

Future work:
We have selected 25 high frequency SNP targets for the *Xa7* region, 25 for the *Xa13* region and 75 SNP targets in the *Xa4*/*Xa22*/*Xa26* region. SNP assays for each locus have been designed and carried out by Genaissance Pharmaceuticals. Upon receipt of the complete genotype dataset from Genaissance we will begin analysis of LD in the three target regions.

Once the data is received from Genaissance, researchers at Cornell (Nick Polato and Carlos Bustamante's group), ICABIOGRAD (Fatimah and Kurniawan) and at IRRI (Mike Thomson and Endang Septiningsih) will work together to analyse the data. We will evaluate the rate of LD decay in the three target regions, we will compare the level of genetic diversity in the *tropical japonica* and the *indica* sub-populations of rice found in Indonesia and we will attempt to trace the ancestry of the bacterial blight resistance genes in each of the target regions.

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 scientists at ICABIOGRAD in Bogor, Indonesia and at IRRI 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, to get to know Indonesian researchers and to establish meaningful long-term collaborations with scientists around the world.

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
Germplasm analysis

- A survey of a 300-line diverse set of maize germplasm, conducted at NCSU over two years and two sites, indicated there was significant correlation between resistance to gray leaf spot (GLS) and southern corn leaf blight (SLB). This correlation was high, even after taking maturity into account. On the other hand, preliminary results indicated that major QTL for GLS and SLB did not significantly co-localise in 4 RIL populations.
- To generate near-isogenic lines using the "heterogeneous inbred family" strategy (HIF) in rice, the IRRI group previously identified five families from bulk DNA analysis as

heterozygotes for several candidate gene loci. Two hundred individuals from each F_6 families have been planted and currently being analysed for heterozygous individuals. Heterozygous individuals will be selfed and progenies will be analysed for individuals homozygous to the candidate gene allele. These near-isogenic lines will be evaluated for their resistance to fungal pathogens of rice.

- We are analysing a panel of more than 100 rice germplasm accessions with moderate resistance to sheath blight based on results of repeated screening in the field of over 40,000 germplasm evaluated over 10 years at IRRI. The current replicated set-up with closer spacing (15x15 cm distance) and higher relative humidity (80-90% RH) under screenhouse condition is in progress. Forty-five day old plants were inoculated with *Rhizoctonia solani* in 1:1 ratio of rice grain and rice hull; disease assessment is in progress.
- More than 1,500 germplasm accessions previously analysed for inheritance of resistance to one or a few races of bacterial blight were phenotyped in two replications with 12 strains of 10 races of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*). More than 90% of the germplasm evaluated were inferred to carry *Xa4* gene based on their resistance to *Xoo* race 1, and susceptibility to race 2. Resistance to more than five *Xoo* races were observed on 66% of the germplasm. Accession #s 332, 8065, 15232, 15177, 19501, 19300, 17350 showed differential reaction to races 3B and 3C while Accession #s 27005, 27495, 27982, 31837, 33351, 53069, 16461, 24924 and 26369 showed differential reaction to races 9a and 9b, which were previously identified to be in the same race group due to lack of differential cultivars. Confirmation of the strong interactions between these differential races will be done so that inheritance study can be performed for these accessions. Race 9b is the dominant pathogen population in the Philippines, and therefore new genes for resistance can be identified that can be useful for managing the disease.

Mapping, genetic analysis and candidate genes

- Several near isogenic pairs (NIPs) that differ for SLB and GLS resistance have been identified at NCSU. We have genotyped some of these pairs and have identified the loci that distinguish them from each other. These NIPs are being used for fine mapping and physiological characterisation of the resistance loci.
- Four pairs of NIPs differing for resistance to northern corn leaf blight (NCLB) have been identified at Cornell from lines derived from the CML52/B73 cross using the HIF strategy. CML52 has resistance to NCLB, GLS, SLB and ear rot. A group of 95 F5 lines in 19 families were provided by NCSU. NIPs were identified in a survey of 65 SSR markers in 39 bins prioritised based on the QTL synthesis. The NIPs show phenotypic contrasts associated with bins 1.08, 5.03, 6.05, and 8.03.
- A maize population that had been subjected to recurrent selection was previously analysed by selection mapping. A second population has now been analysed and the results of the two populations were compared. For the 124 SSR loci analysed for both populations, an average of 7.7 alleles was identified per locus. Of these, an average of 3.5 alleles was found in only one of the two populations. Several of the loci found to be significantly shifted in allele frequency between the first and last cycles were shared. One of the loci of interest was shown to be associated with quantitative disease resistance in inoculation tests.
- The five inbred line GLS sources were crossed to the two recurrent parents and eight F1 hybrids generated at KARI. The F1 were planted to test for adaptability and potential in April at Kiboko and Mtwapa (Kenya) under natural infestation. The generation of heterogeneous inbred families was initiated using the sources identified. The BC1F1 were made involving the eight F1 hybrids and will be harvested in May. These will be advanced to BC2F1.
- A complete genome scan of 84 Vandana/Moroberekan F_6 families selected for resistance to rice blast as well as 24 introgression lines for drought tolerance showed significant correlation ($P<0.001$) by single-marker analysis of the candidate gene (oxalate oxidase and peroxidase) alleles from Moroberekan to yield under blast as well as panicle blast in Almora, a blast hotspot in India. Thaumatin was correlated with lower neck blast incidence in Cavinti,

Philippines. Significant two-gene interactions (P<0.001) were observed in seedling blast (HSP90 x thaumatin and thaumatin x oxalate oxidase) and neck blast (chitinase x thaumatin and thaumatin x oxalate oxidase-like protein). Of the lines selected for blast resistance, 23 yielded more than Vandana under natural drought stress. These include two lines (IR78221-19-6-7-B-B and IR78221-19-6-99-B-B) that are also high-yielding under high blast pressure in Almora. IR78221-19-6-7-B-B has alleles of oxalate oxidase, peroxidase, and HSP90 from Moroberekan. IR78221-19-6-7-B-B has only oxalate oxidase allele from Moroberekan.

- We then looked into the expression of oxalate oxidase in selected Vandana/Moroberekan progenies. *In silico* analysis of the oxalate oxidase gene in the rice genome indicated four tandemly-repeated genes in chromosome 3 -- LOC_Os03g48750, LOC_Os03g48760, LOC_Os03g48770 and LOC_Os03g48780. We designated these genes as OsOXO1, OsOXO2, OsOXO3 and OsOXO4, respectively. Our results indicate induction of the OsOXO4 in resistant Moroberekan as well as in resistant lines IR78221-19-6-7-B-B, IR78221-19-6-56-B-B and IR78221-19-6-99-B-B at 24 hours post-inoculation with blast isolate PO6-6 (Slide 1). Genome scan of these lines show that they possess chromosomal segments from Moroberekan in the region where the chromosome 3 oxalate oxidase is located.
- We also evaluated the Vandana/Moroberekan materials and three IR64 loss-of-resistance mutants for sheath blight resistance (Slide 2). Relative lesion height (RLH - average vertical height of the uppermost lesion on leaf or sheath expressed as a percentage of the average plant height) of lines with 3-6 candidate gene alleles from Moroberekan started at <20% at all time points. These lines were also resistant to seedling blast in Almora, India and Cavinti, Philippines. However, Vandana and Moroberekan and intermated lines with 0, 1 or 2 candidate gene alleles from Moroberekan have >20% RLH. There was also a steady increase in the RLH at each time point, reaching up ~35 % RLH at the end of 5 weeks after inoculation. In contrast, for the 24 drought tolerant lines that were not selected for resistance to blast, we observed that these were all exhibiting 20–40% RLH in the five-week time points. The loss-of-resistance mutants, IR64 DEB 6766 and IR64 DEB 6340 showed reduction in sheath blight progression over time.
- In a collaborative effort with Indonesian and IRRI scientists, a set of 186 BC2F3 lines from the Way Rarem / Oryzica Llanos 5 cross were analysed with 91 SSR, 34 STS and 8 defence-related markers. Polymorphism levels were relatively low; 22% of the SSRs and 3% of the STS markers were useful in this population. A BC_2F_4 population, now under development, will be tested with additional isolates of the rice blast pathogen.

Breeding for improved resistance to biotic stress

- Diallel crosses involving the seven inbred lines were initiated in January and the seed will be harvested and planted in June. These will be evaluated in GLS hotspots to determine whether there is synergy between the different GLS resistant sources in conferring the resistance.
- Several sets of populations consisting of crosses of *O. rufipogon* accessions with indica and japonica rice as recipients of sheath blight resistance are currently being evaluated in the screenhouse to determine introgression of sheath blight resistance from these wild rice accessions. These populations include PSBRc18 x *O. rufipogon* acc. 80671 (BC_1F_2 & F_3), PSBRc80 x *O. rufipogon* acc. 105757 (F_3), IR73885-1-4-3-2-1-6 x *O. rufipogon* acc. 80671 (F_3), B3835 x *O. rufipogon* acc. 80671 (F_3), B3847 x *O. rufipogon* acc. 80671 (F_3), B3848 x *O. rufipogon* acc. 105757 (F_3) were inoculated on May 4th and due for assessment in 14 and 21 days. Three F_2 populations of Ilpumbyeo x *O. rufipogon* acc. 80671, Ilpumbyeo x *O. rufipogon* acc. 105757, Jinmibyeo x *O. rufipogon* acc. 80671 are also currently grown and will be inoculated in 2 weeks.

Collaboration and capacity building

- An Msc student was identified in Kenya to evaluate part of the material in Western Kenya and Muguga.

- Collaborators J. Ininda (KARI) and C.M. Vera Cruz (IRRI) had a chance to meet and exchange informationa about their work under the project, during the course on "Enhancing Negotiations Skills for Women" held at IRRI at the end of March 2006.
- KARI collaborators J. Gethi and J. Danson visited Cornell in Dec., 2005. Dr. Gethi's trip was supported by a GCP travel grant, while Dr. Danson's trip was supported by the Rockefeller Foundation. Both researchers came to do shuttle DNA analysis, to meet with CU researchers, and to meet with former CIMMYT Pathologist D. Jeffers, whose visit coincided with that of the KARI team.
- Dr. Puji Lestari completed her 3 months shuttle research from October 2005 – January 2006 at IRRI. This shuttle research programme was supported by GCP. Results of this shuttle research study are reported above.
- PhD student Ms. Gay Carrillo was identified to work on integration of molecular analysis, phenotype and *in-silico* analysis for broad-spectrum disease resistance in the rice component of the project beginning January 2006.
- Dr. Gethi worked with R. Wisser of CU to analyse seven inbred lines for SSR markers at GLS hotspots on chromosomes 1, 2, 3, 4, 5 and 8 using 30 SSR markers. The lines of interested included 5 inbred line sources (TZMi 102, TZMi 711 and TZMi 712 from IITA, CML 312 and CML 384 from CIMMYT) and two recurrent parents, CML 204 and CML 373. The data indicated the lines had varied degree of homozygosity and diversity in the loci studied.

Tangible outputs delivered:

- The project's maize disease resistance panel at Cornell now consists of 106 inbreds and populations considered to possess resistance to one or more major disease, as well as large sets of genetic stocks under development from various sources. This panel has been assembled from sources including CIMMYT, NCSU, GRIN and the GEM programme. Seed increases have been undertaken and subsets of the panel are under analysis for disease and disease-related traits at CU, NCSU and KARI.
- In the lab of Peter Balint-Kurti at NCSU, a major effort in QTL analysis has lead to the identification of grey leaf spot (GLS) and southern leaf blight (SLB, also known as Helminthosporium leaf blight) QTL in 6 different segregating populations (one manuscript published, one in press, others to follow). Major QTL are being introgressed into a B73 background for detailed analysis.
- Lines (~10) with broad spectrum blast resistance in Vandana plant type are available, which were identified from pyramided lines genotyped for blast loci with favorable recombinants. Effect of blast loci on yield has been established based on evaluation of pyramided blast lines under managed drought.
- Marker analysis (n=145) of 84 blast resistant lines and 24 drought tolerant lines in the Philippines and Almora, India using both candidate gene-specific PCR primers and polymorphic SSR markers co-localising with these genes allowed selection of lines with introgressed chromosomal segments carrying targeted CG regions. The candidate defence gene markers and SSR markers within the defence genes are currently used in other mapping populations (e.g. Way Rarem x Oryzica Llanos 5) to validate functional alleles for quantitative blast resistance and identify broad-spectrum resistance to blast and other diseases.
- The replicated phenotypic analysis of 1,500 germplasm accessions for resistance to 12 *Xoo* strains has been completed. We identified new differential varieties for strains that previously belong to the same race groups. We expect presence of race-specific or qualitative type of resistance from these varieties, although some type of quantitative resistance may exist against several *Xoo* races.

Deviations from the work plan:

No significant deviations were made.

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-CNPMF
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

The use of wild relatives in regular breeding programmes is complicated by the long reproductive breeding cycle of cassava, high genetic load that is released on backcrossing, and linkage drag associated with the use of wild relatives in crop improvement. A project was initiated at CIAT to accelerate the process of introgression of useful genes from wild relatives into cassava via a modified Advance Back Cross QTL (ABC-QTL) breeding scheme. We describe here advances in the last 6 months in the introgression of resistance to delayed post harvest physiological deterioration (PPD), whiteflies, cassava green mites, and horn worm from wild *Manihot* species into cassava. We also describe new collections of wild *Manihot* species carried out by project partners in Brazil as well as progress in molecular breeding of resistance to CMD and cassava green mites resistance by partners in Africa.

1. PPD:

A source of delayed PPD had earlier been identified in an inter-specific hybrid CW429-1 obtained by crossing *Manihot walkerae* with cassava. This line was crossed extensively to the elite cassava genotypes MTAI8, CM523-7, and SM909-25 to create 3 BC_1 families (BC_1 only in the sense of crosses to cassava) or selfed to generate an S_1 family. The F_1 hybrid and 8 elite cassava genotypes namely: MCOL 1505, MPER 183, MTAI 8, CM523-7, HMC-1, MBRA 337, MCOL 2279 and CM 2772-3, having the widest variation for delayed PHD in the cultivated gene pool, were re-evaluated to obtain 3rd year phenotypic data for the novel source of delayed PPD and a wide variation of the trait in the cultivated gene pool. Results of mean PPD values at 5 days after harvest (DAH) ranged from 0% in CW 429-1 and MBRA 337 to 44.85% in CM523-7 (Table 1). At 10 DAH, mean values ranged from 0% in CW429-1 to 58% in CM523-7, respectively. The same trend was observed 15 DAH with CW 429-1 still displaying no visible sign of deterioration (Annex 1). Previous evaluation of delayed PPD shows nothing higher than 7 DAH has been found. Evaluation of inter-specific hybrid CW429-1 reveals that genes for delayed PPD have been transferred into cassava from a wild relative. BC_1 and S_1 mapping populations, a total of 3,400 plants representing 550 genotypes organised into 4 families, for delayed PPD were hardened in the screen house and transferred to the field March this year. They will be evaluated in December for delayed PPD.

2. Hornworm and Whiteflies:

The only known source of resistance to the cassava hornworm was identified in 4th backcross derivatives of *M. glaziovii* – MNGA11 (60444). High levels of resistance to white flies were also identified in inter-specific hybrids of *M. esculenta* sub spp flabellifolia (CIAT, unpublished data). The variety MNG11 was to several cassava varieties and selfed to produce BC_1 and S_1 families respectively. Several genotypes of the inter-specific hybrid CW67, a progeny of M.esculenta sub

spp flabellifolia, showing a high level of resistance to white flies was crossed to MTAI 8 or selfed to produce BC_1 and S_1 families respectively. After embryo rescue of seeds from the above crosses and screen house hardening, the number of seeds obtained was below that required for mapping of resistance genes. The low number of recovered plants was due to an unusual level of precipitation and flooding at CIAT recently. Additional crosses were made last year with MNGA11 and several inter-specific hybrids with *M. esculenta* sub spp *flabellifolia*, the family CW67, that flowers more profusely and many more seeds obtained (Annex 2). The seeds are currently being put *in vitro* from embryo axes and will be micro propagated and established in the field by August this year for 2 cycles of evaluation for the appropriate pests.

3. Cassava Green mites (CGM)

Good resistance to cassava green mites (CGM) was identified in 4 inter-specific hybrid families, CW68, CW65, CW67, and CW66, derived from a cross between cassava and *Manihot esculenta* sub spp *flabellifolia* accession. In order to identify markers associated with resistance to CGM the BC_1 derivates of the inter-specific crosses were evaluated over three growing cycles to identify resistant and susceptible individuals. Genotypes with symptom damage of less than 3, resistant, and those with symptom score of 4-6, susceptible, were bulked respectively. Molecular markers for bulked segregant analysis (BSA) were 530 SSR markers with broad coverage of the cassava genome. Till date 131 SSR markers have been evaluated in the resistant and susceptible bulks and 44 markers were polymorphic between the bulks and parental lines (Annex 3). After completion of the bulk and parental survey with all markers, individuals of the bulks and eventually the entire BC1 family would be evaluated with the polymorphic markers.

4. Molecular breeding of resistance to CMD and CGM

A total of 2,400 plantlets representing 11 F_1 families were shipped from CIAT, Colombia, in March 2006 (Annex 4) to each NARs of Ghana, Uganda, and Nigeria. This is in addition to BC_2 for CMD and CGM resistance shipped last year to Brazil and the African NARs. The materials are being hardened for field establishment during this season. The materials would be evaluated in the field for CMD and CGM resistance and put in a crossing block together with local farmer preferred and improved varieties for making crosses later in the year. The F_1 seedlings will be selected for resistance to CMD and CGM using molecular markers and phenotypic evaluations next year as well as for other traits of agronomic importance for example high starch

5. Collection of new accessions of *Manihot* species

Collection of sexual seeds of additional wild species accessions were carried out in three places:

i) From the CNPMF field collection

916 sexual seeds were collected from different accessions of 4 species: *M. anomala* (527); *M. dichotoma* (228); *M. flabellifolia* (78); *M. peruviana* (83). Most of these seeds were sowed under greenhouse condition and the seedlings will be planted at the target sites for field evaluation later in the season.

ii) Germplasm collections in the semi-arid region of Bahia State where the following cassava wild species were found and collected:

1) *M. caerulescens*
2) *M. diamantinensis*
3) *M. jacobinensis*
4) *M. glaziovii*
5) *M. dichotoma*
6) *M. maracasensis*

iii) Another collection was carried out in the "cerrado" region of Brasília and surrounds, in which 28 populations of cassava wild species were found. A total of

16 species were found and collected. The probable species found in this expedition were:

1) *M. pentaphylla*
2) *M. irwinii*
3) *M. violacea*
4) *M. falcata*
5) *M. salicifolia*
6) *M. fruticulosa*
7) *M. cecropiaefolia*
8) *M. tripartita*
9) *M. stipularis*
10) *M. triphylla*
11) *M. tristis*
12) *M. anomala*
13) *M. mossamedensis*
14) *M. mana*
15) *M. gracilis*
16) *M. tomentosa*

A part of the seeds of these collections have been established in seedling nurseries for eventual transfer to the field at CNMPF and to the evaluation sites at Petrolina, São Miguel das Matas and Tancredo Neves for field evaluation

6. Evaluation of *Manihot* species for additional genes of interest

i) Field establishment of the cassava wild species collection at Embrapa/CNPMF.

The first species/accessions established in this collection came from: a) Sexual seeds of 7 wild species sent from Embrapa/CENARGEN; b) Stakes of 3 wild species previously collected at Bahia's semi-arid region; c) Stakes of "Maniçobas" and "Pornúncias" accessions collected at Embrapa/CPATSA and at Bahia Federal University. Up to now, accessions of the following genotypes were field established:

1) *M. anomala*
2) *M. caerulescens*
3) *M. dichotoma*
4) *M. flabellifolia*
5) *M. glaziovii*
6) *M. peruviana*
7) M. tomentosa
8) Maniçoba - probably is *M. glaziovii* or *M. pseudoglaziovii*
9) Pornúncia – probably is a natural hybrid between *M. esculenta* and *M. glaziovii*

ii) Inter specific hybrids introduced from CIAT at Embrapa/CNPMF.

1098 F_1 seeds of 34 inter specific hybrids from crosses between *M. esculenta* and wild species, produced at CIAT, were sowed at CNPMF under greenhouse condition. A total of 450 seedlings from 32 hybrids were obtained. The number of seedlings per hybrid varied from 1 to 74. Some seedlings of 21 hybrids were planted in Petrolina, São Miguel das Matas and Tancredo Neves for field evaluation

iii) Inter specific hybrids introduced from CIAT to NARs of Nigeria, Ghana, and Uganda.

These were nursed in the screen house and evaluated as part of the effort to introgress Latin America material for use in the development of improved varieties of cassava. A selection would be made upon evaluation for the crossing block. The seeds were planted in September 2005 and some transplanted in November. These were crosses between *Manihot esculenta*, between

Manihot esculenta and wild relatives; and between the wild relatives. The germination rates were moderate to low. The seedlings were evaluated for disease pressure reaction

7. Training

Following the initiation of the GCP project, CIAT appointed a visiting scientist, Dr. Emmanuel Okogbenin, a breeder with experience in molecular genetics to technically backstop project activities in the NARs. The visiting scientist is placed at the NRCRI, NARs of Nigeria but oversees project activities at the African NARs as well as serving as liaison officer between CIAT headquarters (PIs) and NARs partners (collaborating scientists). The visiting scientist has successfully assisted the NRCRI, Nigeria to establish a MAS laboratory in addition to backstopping the institute in its cassava breeding activities. The NRCRI Molecular Biology Laboratory which was established last year started operations early this year. Test running of the laboratory started in February and it started functioning since March. PCR reactions and polyacrylamide gels are being run at the moment. This laboratory is now been used for diversity studies as part of preliminary activities towards applying the laboratory in molecular breeding activities. CIAT has recently provided the laboratory with primer kit (of 36 primers) for the diversity studies. The laboratory will be used for marker assisted selection activities later in the GCP project. Similar molecular marker activities have also been initiated at CRI Ghana and NAARI, Uganda.

Tangible outputs delivered:

1. Confirmation of delayed post-harvest physiological deterioration (PPD) in inter-specific hybrid CW429-1 transferred into cassava from a *M.walkerae*
2. Initiation of bulked segregant analysis (BSA) of resistance to cassava green mites in BC_1 derivatives and the identification of 44 polymorphic markers for further evaluation
3. Introduction of CMD and CGM resistant genotypes from CIAT to African NARs and establishment of a crossing block for genetic crosses to local and improved farmer preferred varieties
4. Collections of accessions of 20 species of *Manihot* germplasm in the semi-arid region of Bahia State and in the "cerrado" region of Brasília.
5. Establishment of sexual seeds of wild and inter-specific hybrids for evaluation in an attempt to identify new sources of genes for resistance to pests and diseases and other traits of agronomic interest
6. Introduction of F_1 and BC_2 molecular breeding parents with CMD resistance and tolerance to Mites to NARs partners in Nigeria, Ghana, Uganda, and Brazil for molecular breeding; setting up of crossing blocks in Nigeria, Ghana, and Uganda

Deviations from the work plan:

1. A delay in the field establishment of resistance to white flies and hornworm mapping populations: the need to increase the mapping population sizes for resistance to white flies and hornworm has meant that these mapping populations were not moved to the field at the same time with the delayed PPD populations, they will be transferred in August to the field.
2. A delay in the shipment of F_1 populations with resistance to CMD and cassava green mites to NARs partners in Nigeria, Ghana, and Uganda for molecular breeding, the plants were eventually shipped in April this year as against the original date of October last year, 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.
3. On project activities in Brazil, the main delays were: 1) bureaucratic bottle necks (plant quarantine); on the release of the plant materials (imported from CIAT) by the Ministry of Agriculture to CENARGEN 2) Delay in the transfer of the materials from CENARGEN to CNPMF; and 3) damage to the *in vitro* plants; the in vitro BC_2 plants of 28 genotypes, sent from CIAT on Dec 20, 2005, were delivered to CENARGEN on Jan 24, 2006. All the plants inside the bottles were lost due to the damages that occurred during transportation 4) delay in

the release of more than 3000 sexual seeds of five wild species shipped from CIAT on Sep 01, 2005 to CENARGEN (for quarantine), they were released to CNPMF after 5 months. These seeds were sowed at CNPMF and none sprouted. The second batch of F_1 seeds of inter-specific hybrids shipped from CIAT on Oct 18, 2005 remained 3 months in São Paulo before being releasing to CENARGEN. These constraints have caused significant delay of the original work plan and can be minimised by no cost extension of the project's activities.

4. New introductions were received from CIAT in April in Nigeria. The materials were inspected by PQS and about 1100 in vitro plants were transplanted for the hardening stage. However, due to fungal infections experienced during the hardening stages substantial losses were experienced, and efforts are underway to re-introduce more materials by June.

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

Development of Chromosome Segment Substitution Lines populations
Completion of two *O. sativa* x *O. glaberrima* populations
A1. IR64 x TOG 5681 cross

New BC1F1, BC2F1 and BC1F2 populations of this cross have been developed at CIAT in order to try to introgress some missing fragment of the *O. glaberrima* genome that were lost in the former population (see last year report).

A set of 357 SSRs was optimised in the frame of the core map project. From this set of SSRs, 120 SSRs that show polymorphism were chosen for the genotyping of the new BC1F1s. To this date, 35 SSRs have been tested on chromosomes 3 and 6. Those chromosomes were targeted because they bear two important interspecific sterility genes.

For the BC3F3s/BC2F4s (first population), a set of 40 new SSRs has been selected to complete the genotyping in order to have evenly distributed SSRs.

As a result of the genotyping of the BC1F1s, we found that the plants that were showing a high fertility don't bear the *O. glaberrima* allele at the location of the sterility genes. Crosses between those lines and the two *O. sativa* and *O. glaberima* parent are on going to verify the hypothesis that they could serve as interspecific bridges. Also, we found lines with introgressions for most of the genome segments from the *O. glaberrima* genome that were lost in the former BC3F3/BC2F4 population, especially on chromosomes 4 and 10.

In the next months, we are going to continue with the evaluation of the SSRs that we select for each one of the populations.

A2. Caiapo x IRGC103544 cross

A cross between Caiapo (an elite tropical *japonica* from Brazil) and *Oryza glaberrima* (IRGC103544) was made at CIAT HQs (C.P. Martinez), using IRGC103544 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 at Dr. Z. Lentini's lab, CIAT.

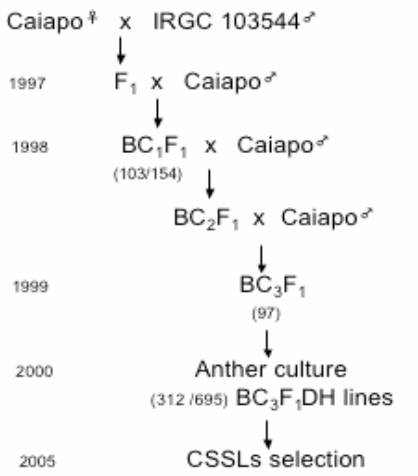


Figure 1. General scheme that shows the development of the population derived from *Oryza sativa* (Caiapo) x *Oryza glaberrima* (IRGC 103544) interspecific cross.

From this population, as a first step 312 lines were selected and genotyped using 98 polymorphic SSRs markers located in the 12 chromosomes at an average distance of 10 cM.

Each line was phenotypically evaluated for yield traits and yield components traits (panicle length, plant height, yield, sterility, 1000-seed weight, and tillering) at CIAT HQs during the 2003 dry season.

Microsatellite markers were selected according to their unique position in the genome (according to the TIGR v. 2 data obtained from the Gramene database, http://www.gramene.org/). Following this criteria, 80 SSRs markers were added to the previous data set, leading to a total of 177 markers. All the SSRs has been amplified following the same standard conditions for PCR and gel migration (PAGE/silver staining and 4% agarose gels stained with ethidium bromide when polymorphism was >10 pb).
Using the programme CSSL Finder, the lines were selected according (1) to the presence of contiguous chromosome segments that maximised the genome coverage of O. glaberrima, and (2) to the percentage of recurrent genetic background. As a result, in each line, one or few different chromosomal segments of O. glaberrima, was substituted in the genetic background of the cultivar Caiapo. The substituted chromosome segments in the 72 pre-CSSLs represent the complete genome of *Oryza glaberrima*, except for small regions of chromosomes 2, 4. Moreover, for some introgressions it was unclear if they form overlaps with other segments, due to the small size of the introgressed segments. These areas are understood among chromosome 2 (RM71-RM300), chromosome 4 (RM261-RM241), chromosome 8 (RM308-RM284), chromosome 9 (RM316-RM296) and chromosome 10 (RM474-RM239). It will be necessary to screen these regions with additional markers.

The QTL analysis by means of the programme MapDisto v.1.5 (Lorieux 2005) and QGene v.3.07 (Nelson 1997) allowed to detect 12 QTLs for 5 quantitative traits in chromosomes 1, 3, 4, 5, 6 and 11, and are being compared to QTLs found in the literature and to those obtained with the IR64 x TOG5681 population.

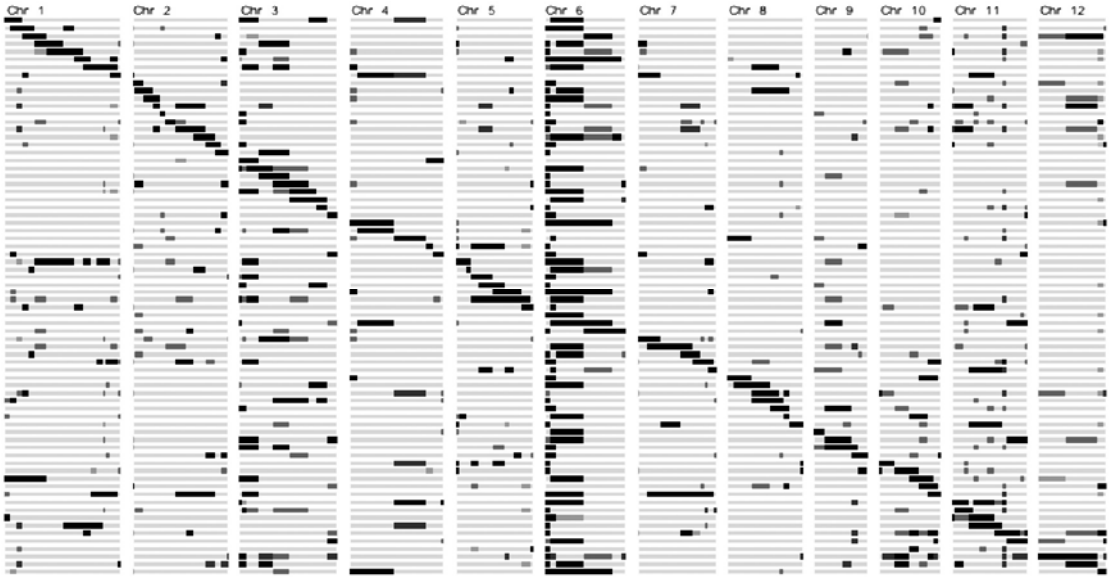


Figure 2. Graphical genotypes of the 72 CSSLs selected by the programme CSSL Finder v.0.8a11. The columns correspond to the 12 chromosomes of the rice with 114 SSRs, rows correspond to the CSSLs. The black regions indicate homozygous fragments for the *Oryza glaberrima* allele; light yellow regions indicate homozygous fragments for the Caiapó allele, red and yellow regions indicate alien alleles, orange regions indicate heterozygous and gray regions indicate missing data.

Table 1. Microsatellite markers significantly associated to 5 quantitative traits in the BC_3F_1DH population derived from the interspecific cross Caiapó x IRGC103544.

Trait	QTL	Chromosome	Interval	Positión	LOD>3.0	R^2	*F*>15.00
Height	*HT-4*	4	RM127-RM280	27.2	3.38	5.4	17.34
	TILL-3	3	RM60-RM175	10.1	5.58	7.8	23.48
Tillering	*TILL-5*	5		17.8	3.0	4.9	15.41
	TILL-6	6	RM169-RM146	146.9	4.59	6.0	18.68
	TILL-11	11	RM276-RM162	19.2	2.73	5.3	16.41
	YLD-1	1	RM209-RM229 RM259-RM292	33.5	3.0	5.1	16.60
Yield	*YLD-3*	3		56.5	4.20	6.6	20.08
	YLD-*4*	4	RM338-RM135	6.5	3.62	4.8	15.40
Sterility	*ST-1*	1	RM551-RM273 RM84-RM151	15.0	3.94	5.6	15.99
	ST-3	3	RM60-RM175	6.2	6.48	9.9	31.14
1000-seed weight	*TGRWT-4*	4	RM551-RM273	6.5	6.62	9.7	32.69
	TGRWT-6	6	RM253-RM276	146.9	4.15	5.1	15.67

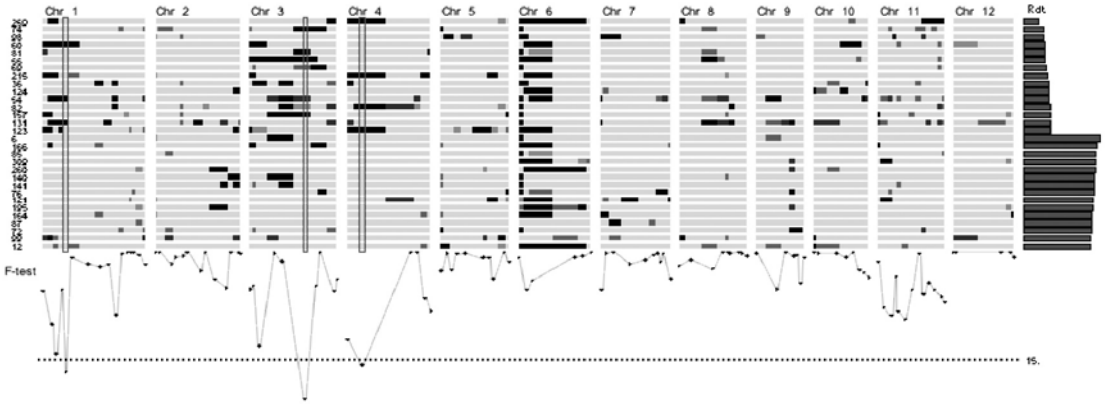


Figure 3. Correlation of the pre-CSSLs with the yield quantitative trait. The blue bars to the right of the figure indicate the lowest and higher values of the trait. The blue vertical bars show

genome regions with introgressions of *Oryza glaberrima* that contribute an additive effect on the trait. The tips located under the graph show the *F*-test value for each marker.
Ongoing activities

- To carry out an additional backcross for the selected CSSLs to clean up the genetic background,
- To fill the gaps in with additional microsatellite markers and to form bridges (overlapping fragments) between each line.

Generation of new CSSL populations
O. sativa x O. rufipogon and O. sativa x O. meridionalis:

Regarding the generation of the new CSSL population, the cultivar *Curinga* (*O. sativa* tropical japonica) was chosen as the recurrent parent for the new four CSSL populations to be developed, for its due to its better fulfill during drought assays, over other four varieties, carried out at EMBRAPA (Brazil). The donor parentals was chosen at CIAT according to their phenotypic similarities to the species described in the guide "Wild Rice Taxonomy" (*Bao-Rong Lu and Michael Jackson, IRRI*) and also according to their behavior during the F1 crossing phase. With these parameters, two new CSSL population are ready to be developed, from the crosses between Curinga x IRGC 105491 (*O. rufipogon*) and Curinga x OR 44 (*O. meridionalis*).

We are now finalising the production of BC1F1 seeds. A first backcross (BC) was made for each accession, and so far, 366 BC1F1 seeds from the cross Curinga x IRGC 105491 and 56 BC1F1 seeds from the cross Curinga x OR 44 have been obtained. 407 primers pairs that cover 141 anchors distributed every 2-3 Mb across the twelve chromosomes of the rice genome, are being tested with the parental genotypes, in order to choose one polymorphic marker per anchor. By now, 392 markers out of the 408 markers currently in use at our lab, have been optimised at CIAT. Of these, 71 markers (one per anchor) have been selected to be polymorphic for the cross Curinga x IRGC 105491, and 65 for the cross Curinga x OR 44. Primer testing is still carried out to obtain at least 120 markers between the parentals of each cross.

For the cross Curinga x IRGC 105491, one hundred BC1F1 seeds have been sowed in order to choose in 50 plants that bear introgressions that cover the entire donor genome. Sowing of BC1F1 seeds obtained from the cross Curinga x OR44 will be done soon. DNA obtained from harvest of these two populations will be genotyped at Cornell University in June.

O. sativa x O. barthii:
Three accessions from *O. barthi* IRGC 100934 (3232), IRGC 101937 (3239) and IRGC 103582 (3240) where selected to produce F1 populations using *Curinga (O. sativa tropical japonica)* as the recurrent parent. 80, 25 and 19 seeds where recovered respectively from the F1 populations. A first BC was made with the three *O. barthi* accessions and so far, 41 and 70 seeds, where recovered for BC1F1 populations from Curinga x *O. barthi* IRGC 100934 and Curinga x *O. barthi* IRGC 101937 crosses, respectively.

While originally planned to be carried out at WARDA, this work was done at CIAT. The BC1F1 seeds will be sowed and the DNA will be analysed at Cornell University by a WARDA's student.

O. sativa x O. glumaepatula
Claudio Brondani, Priscila Rangel, Embrapa-CNPAF
In the first semester of 2006 we obtained more than 300 BC1F1 seeds of the cross x *O. sativa* (cv. Curinga) x *Oryza glumaepatula* Gen1233. Fifty seeds were sowed, in order to obtain 600 BC2F1 plants, to follow the development of the CSSLs. We are collecting leaves of the 50 BC1 plants for DNA extraction, to analyse the 50 plants with 120 SSR markers at Cornell University, starting in

June 2006. This laboratory work will count with the participation of the Brazilian Ph.D. student of this project, Priscila Rangel.

Drought response screenings
The objectives of the trial were:

- To determine the drought tolerance of a segregating population of CSSLs
- To determine the contribution of various traits to drought tolerance of rice
- To identify QTLs associated with drought tolerance of rice

Materials and Methods
Plant material
A population of 93 CSSLs developed at CIAT from the cross Caiapo x *O. glaberrima* (IRGC103544) is being subjected to drought screening at the WARDA trial fields at IITA-Benin headquarters, in Cotonou. The population comprises BC3F1 doubled haploid plants and their parents. NERICA1 and 2, as well as WAB56-104, CG14 and IR64 have been included in the trial as checks.

Soil physical and chemical properties
The soil at the trial site is an Alfisol with a sandy texture (82-89%) from 0-50cm depth with a low water holding capacity. The details of the soil chemical and physical properties at the trial site are presented in Table 2.

Table 2. Soil physical and chemical properties at the trial site at Togoudo, Benin.

Properties	Depth (cm)		
	0-20	20-50	50-200
Clay (%)	7	15	41
Sand (%)	89	82	54
Silt (%)	4	3	5
PH(water)	5.3	4.2	4.2
Org. carbon (%)	1.38	0.83	0.75
Total N (%)	0.106	0.06	0.06
Avail. P (ppm)	3.4	3.2	3.3
Ca (meq/100g)	2.4	1.8	3.2S
Mg (meq/100g)	2	1.4	2.6
K (meq/100g)	0.14	2.6	1.6
Na (meq/100g)	0.14	0.19	0.15

Experimental design
In this trial a split-plot design with irrigation regime as the main plot factor and genotype as the sub-plot factor was used. Within each sub-plot the genotypes were randomised using an alpha lattice design. Two irrigation levels were used – full irrigation up to maturity and imposing 20 days drought stress from 48 days after sowing (DAS) till maturity.

Cultural practices
Due to seed limitations a non-replicated design was used. Three seeds were sown per hill at a spacing of 25cm x 25cm and 5 hills per line. Compound fertilizer as NPK 15:15:15 was applied at the rate of 200 kg/ha at 21 DAS. Nitrogen fertilizer was applied in two equal split applications as urea (46%N) at 21 DAS and 42 DAS. The plot was clean weeded throughout the crop cycle.

Data collection
Data was collected following the Standard Evaluation System (SES) of IRRI where applicable. These data include plant height, tiller number, leaf length, width and number per tiller, visual scores of leaf drying, leaf rolling and recovery, flowering delay, canopy temperature, leaf

greenness rating taken with a SPAD meter, biomass, yield and yield components (thousand grain weight, panicles per plant, filled grains per panicle, sterility).

Concerning plant height and tiller number, the mean increase in the stressed plot compared to the irrigated plot over the stress period was computed as:
RD= ((XD2-XD1)/(XW2-XW1)*100
where,
RD = relative difference
XD1=measurement before drought stress in drought plot
XD2= measurement after drought stress in drought plot
XW1=measurement before drought stress in irrigated plot
XW2= measurement after drought stress in irrigated plot
For canopy temperature, the percentage increase or decrease in temperature of rice leaves in the stressed plot relative to the irrigated plot on the seventeenth day of drought was computed as:
TD=((T2-T1)/T1))*100
where,
TD=Temperature difference
T1=Canopy temperature in the irrigated plot
T2=Canopy temperature in the non-irrigated plot

Gravimetric moisture content was determined at three different depths just before irrigation resumed on the stressed main plot. Soil samples were collected with a soil auger at depths of 0-20 cm, 20-40 cm and 40-60 cm. Fresh weights of soil cores were determined after which the cores were dried in an oven for 48 hours at 105 °C. The dry weights were then measured and soil moisture content was computed as a percentage of total soil volume.

Results and discussions

1. Gravimetric moisture content

The 20 days of drought stress imposed on the trial was sufficient to reduce soil moisture content by more than half in the upper 20 cm of soil where most rice roots are concentrated. However, soil moisture content at 20-40cm in the stressed plot was almost the same as in the top 20cm of the irrigated plot. Thus rice genotypes that are able to produce roots deeper than 20cm in this soil will be able to access moisture levels comparable to that found in the irrigated top soil layer under drought stress (Table 3.). Unfortunately in this research activity only root biomass is being measured but not root length.

Table 3. Soil moisture content (cm) at different depths

Irrigation status	**Soil moisture content (%) at different depths**		
	0-20cm	**20-40cm**	**40-60cm**
Irrigated	6.25	10.18	14.60
Non-irrigated	3.11	6.13	10.98

2. Effect of drought stress on measured traits

The first noticeable effect of drought stress in the trial was leaf rolling of plants in the drought-stressed plot. When this trait was measured after 17 days of drought stress at 65 DAS, most genotypes were found to exhibit medium to severe symptoms of leaf rolling (score of 5-9) as seen in Fig. 4. The two parents of the CSSL population, both had severe leaf rolling with Caiapo at a score of 7 and IRGC103544 at 9. Among the check varieties also, IR64, CG14 also had scores of 9 and WAB56-104 a score of 5. NERICA1 and NERICA2 both had mild leaf rolling with scores of 1.

Drought stress, on average, reduced plant height, leaf number, leaf length and tiller number but increased canopy temperature. There was an increase in the mean leaf greenness rating of the

stressed plot over the non-stressed plot after 2 weeks of drought stress. However, when the same trait was measured at 89 DAS following the culmination of 20 days drought stress at 68 DAS, the mean of the stressed plot was found to be lower than that of the irrigated plot (Table 4). This increased leaf greenness rating could be caused by increased levels of stress-induced proteins in the leaf under drought stress. However, once stress was relieved by resuming irrigation then the levels of stress-induced proteins in the leaves will reduce.

Figure 4. Frequency distribution of leaf rolling scores of 90 CSSLs, their two parents and 5 check varieties.

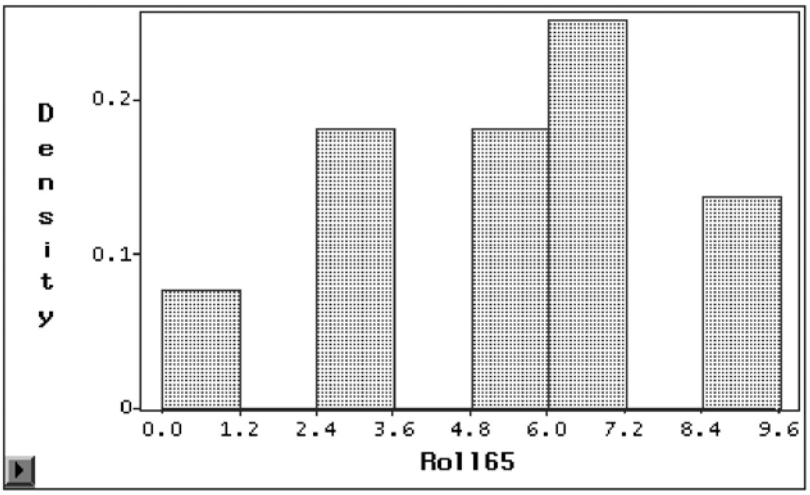


Figure 5. Relative increment (%) in plant height after 14 days of drought stress for stressed plants compared to non-stressed plants (Note: Genotypes 1-93 – CSSLs; 94 – Caiapo; 95 – O. glaberrima (IRGC103544); 96 – NERICA1; 97 – NERICA2; 98 – IR64; 99 – CG14; 100 – WAB56-104.)

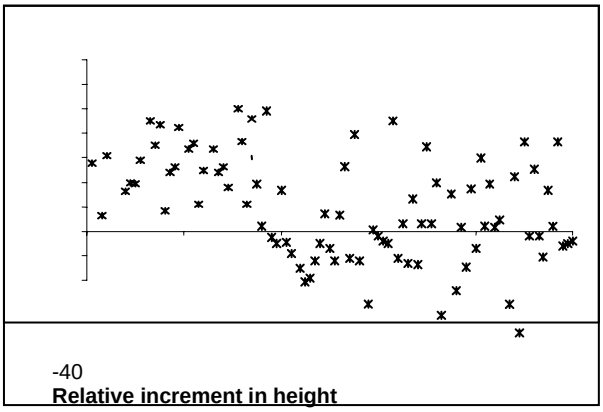


Table 4. Means of traits measured during and after 20 days drought stress on 96 genotypes of rice under irrigated and drought stressed conditions.

Trait	Irrigated	Non-irrigated	Difference
Height48* (cm)	55	67	-12
Height62* (cm)	80	80	0
Height88 (cm)	98	85	13
Leaf Temp.65	32	33	-1
Leaf no.79	5	4	1
Leaf length79 (cm)	45	40	5
Leaf no.88	5	4	1
SPAD50	43	44	1
SPAD63	45	47	-2
SPAD89	46	44	2
Tiller49	6	8	-2
Tiller65	9	8	1
Height difference (cm)	25	13	12
Tiller difference	2	1	1

* - Numbers following trait names indicate the DAS on which the trait was measured.

Thus on average leaf greenness rating in the stressed plot reduced to levels observed before stress was imposed. In the non-stressed plot, however, leaf greenness rating continued to increase over the same period.

With regards to the relative increase in plant height, it can be seen in Fig. 5 that the heights of most genotypes increased faster without drought stress than under drought stress. However, certain genotypes attained similar or higher height increments under drought conditions than without drought stress. Drought stress reduced plant heights of two genotypes and for another two there was no increment in plant height during the drought period.

Similarly for the relative increment in tiller number, most genotypes produced more tillers without drought stress than under drought stress. Furthermore, drought stress led to tiller deaths of 22 genotypes which was manifested in lower tiller numbers of genotypes during the drought period compared to before the onset of drought (Fig. 6). Nonetheless, certain genotypes manifested similar or higher increments in tiller number during drought stress than without drought stress. Both Caiapo and WAB56-104 produced no new tillers during drought stress while

O. glaberrima (IRGC103544) and CG14 under drought produced 15% and 17% respectively, of new tillers produced without drought stress. NERICA1 and NERICA2 under drought stress produced 50% and 100% of the tillers produced without drought stress.

Canopy temperature varied from 31-34 °C during drought in the stressed plots while it was 29-34 °C in the non-stressed plot during the same period. Fourteen CSSLs had lower canopy temperatures during drought stress relative to the non-stressed condition (Fig. 7). With regards to the two parents of the CSSLs, both Caiapo and *O. glaberrima* (IRGC103544) had higher canopy temperatures under drought compared to the non-drought stressed condition. However, the temperature difference was greater for Caiapo (10%) than for *O. glaberrima* (4%).

Relationship between grain yield and secondary traits

Presently work is underway to process the plants harvested from the trial in order to obtain data on yield and yield components. Further analyses will be carried out on the data collected to determine the relationships between grain yield of rice and secondary traits both under drought stressed and continuously irrigated conditions.

Figure 6. Relative increment (%) in tiller numbers during 16 days of drought stress for stressed plants compared to non-stressed plants (Note: Genotypes 1-93 – CSSLs; 94 – Caiapo; 95 – O. glaberrima (IRGC103544); 96 – NERICA1; 97 – NERICA2; 98 – IR64; 99 – CG14; 100 – WAB56-104.)

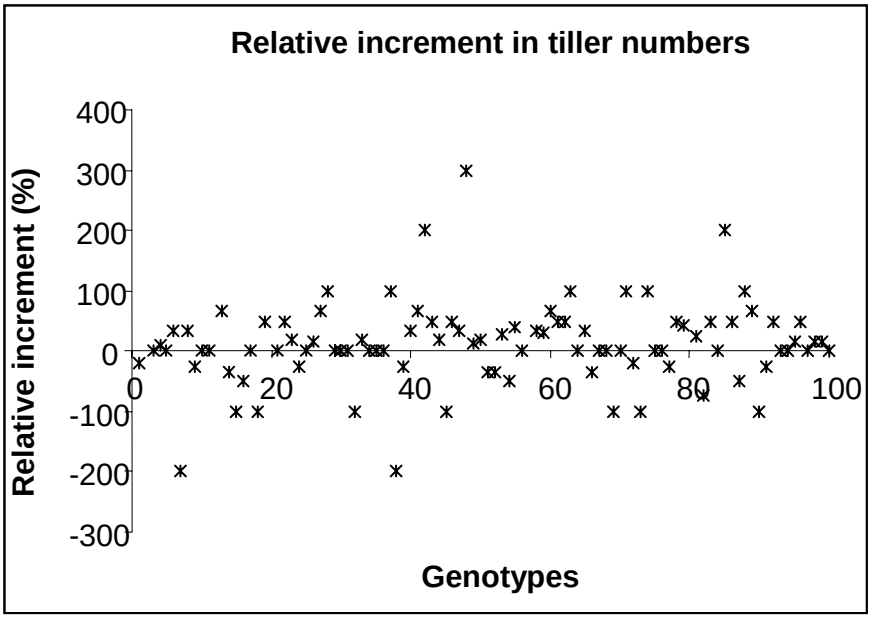
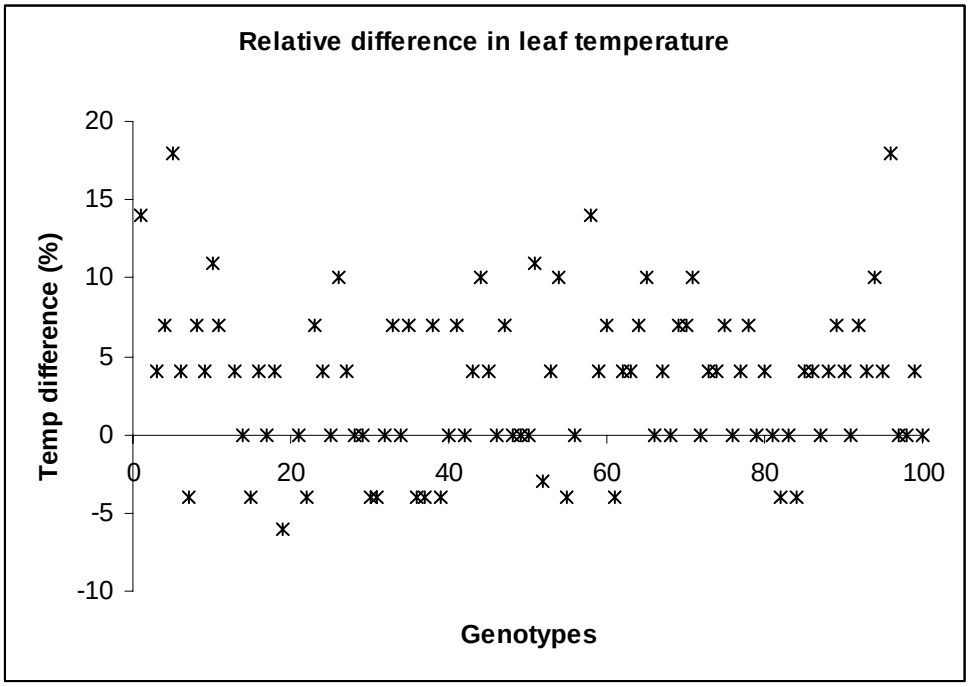


Figure 7. Temperature difference between stressed and non-stressed genotypes of rice after 14 days of drought stress (Note: Genotypes 1-93 – CSSLs; 94 – Caiapo; 95 – O. glaberrima (IRGC103544); 96 – NERICA1; 97 – NERICA2; 98 – IR64; 99 – CG14; 100 – WAB56-104.)



SNP markers development

A. Standardisation of SNP detection in rice varieties using the single base extension method.

- Setting up and validation of a multiple SNP detection system in rice genotypes:
- A multiplex strategy was assayed in order to speed up SNP genotyping in rice genes using a Luminex-100 platform.
- Thirty-two sequences containing T/C polymorphism were chosen for multiplex primer design for both, PCR and single base extension (SBE). Single base extension primers were designed using SBEprimer software developed by Zaik et al. (2003) and PCR primer pairs were designed using FastPCR software (Kalendar, 2005).
- PCR products were obtained in single amplification reactions (one primer pair), in 2 sets of 16 SNPs per reaction and one set of 32 SNPs per reaction. Single base extension was assayed in groups of 8, 16 and 32 SBE primers.
- SNP validation in eight rice genotypes including *O. sativa* (indica and japonica), *O. glaberrima*, and *O. barthii* was carried out.
- The best results were obtained doing two PCRs, each containing 16-primer pairs, and one SBE reaction. We found that 66% of the assayed SNPs were polymorphic and genotyped correctly between Nipponbare and 93-11, but when 32 PCR primer pairs were used together, percentage of polymorphic markers decayed to 44%.
- We conclude that a fast SNP validation can be achieved using a multiplex strategy that allows a significant reduction in PCR number from 32 (individual amplification) to two PCR of 16 amplicons, and one SBE reaction. Hopefully we could increase SNP genotyping system efficiency to 50 SNPs/SBE reaction/hour.

B. Development of a bioinformatic tool for automated SNP discovery and annotation in rice.

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 chromosomic 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 organised in a relational database implemented in MySQL.

We have chosen a group of 60 *Oryza sativa* drought tolerance candidate genes as subjects of SNP discovery. This set was defined using as selection criterion their orthology with *Arabidopsis thaliana* genes involved in drought response. By means of literature review we checked for experimental evidence suggesting the role of 258 plant genes in biological processes related to abiotic stress response such as heat shock, water and ion transport, oxidative stress, or either whose expression was induced by abiotic stress but were classified in more general functions such as membrane cell signaling, transcription factors etc. From these initial 258, 60 *Arabidopsis thaliana* genes were well defined as rice orthologous genes likely involved in drought tolerance. The orthology relationship was established using BLAST best bidirectional hit criterion between the chosen *Arabidopsis* genes and the *Oryza sativa* coding sequences (TIGR Rice Genome Annotation OSA1 ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/). The software tool SNPspipe v0.9 designed by our group at CIAT was then used for SNP discovery and PCR primer design in the *Oryza sativa* drought tolerance candidate genes through the comparison of Nipponbare (*indica* TIGR Build 3.0) and 9311 (*japonica* RISE WGS sequence) rice genomes. This enabled us to identify 827 SNPs in the corresponding genomic sequence and 5'-3' vecinity of the chosen candidate genes, whereas 615 SNP resulted as non redundant compared to the NCBI dbSNP *Oryza sativa* database. This results are comparable to our previous SNP discovery on iron homeostasis related genes in rice (527 SNPs dectected, 410 non redundant), where between 75-80% of which could be experimentally validated in other rice parentals (Constanza Quintero, pers. comm.). Further functional annotation and the use of higher quality sequence databases, such as the Kasalath (*indica* BAC clone genome sequence) and BAC end resequencing (IRGP 2005) would allow us to choose a representative set of SNPs for mapping the most likely candidates of the set identified by our current work.

Development of genetic tools

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.

With the purpose of making more efficient the creation of new genomic maps with evenly distributed markers and to improve localised mapping, SSR markers were selected based on the *O. sativa* genomic sequence using the Institute for Genomic Research TIGR 2004 data obtained

from Gramene database and using the bioinformatic tool CHARM from Cornell University (G. Wilson, comm. pers.). 141 anchors were formed, distributed evenly every 2 to 3.5 Mb along the rice genome and at least 3 SSRs were selected per each anchor.

357 SSRs from 124 anchors were optimised for 7 accessions of the genus *Oryza* as follows: 3 accessions from *O. sativa* (IR64, caiapo and curinga); 2 from *O. glaberrima* (TOG5681 and IRGC 103544); 1 from *O. rufipogon* (IRGC 105491); and 1 from *O. meridionalis* (OR44).

65 rice accessions with AA genotype genome were selected to evaluate their level of polymorphism using the selected markers as follows: 29 accessions from *O. sativa* type indica, 6 from *O. sativa* type template japonica, 12 from tropical japonica, 14 wild rice biotypes from *O. meridionalis*; *O. rufipogon*, *O. nivara*, *O. barthii*, *O. glaberrima*, *O. glumaepatula* y *O. longistaminata* and 3 red rice biotypes. The tetraploid species *O. latifolia* was used as an outgroup.

117 microsatellites out of 408 were evaluated corresponding to 39 anchors; 33 SSR from chromosome 6, 24 for chromosome 9 and 60 distributed at the edges of the rest of the chromosomes. These markers were evaluated using PAGE 6% and LICOR sequencer. All the results are summarised in a database that will be shared will the scientific community.

The 117 markers revealed 1255 alleles the 65 biotypes. The number per locus varied widely among these markers, ranging from 3 (RM507, RM19623 y RM172) to 30 (RM2136). As a measure of the informativeness of microsatellites, the average PIC value was 0,690 with the range of 0,25073 (RM19983) to 0,950 (RM19218).

We found good percentages of polymorphism between species crosses with an average of 89,9% and 70,4% within the *O. sativa* species.

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
D. Tharreau, AGROPOLIS
S. Kikuchi, NIAS
F. Dedryver, INRA
P. Sourdille, INRA
L. Boyd, JIC
S. Brammer, EMBRAPA
AS. Prabhu, EMBRAPA
M.C. Chaves, EMBRAPA
E. Guiderdoni, AGROPOLIS

Mid-Year Report
General Comments
- Overall the timing of the research plan was respected.
- Funds were transferred to: JIC, CYMMIT, UCD, NIAS, AGROPOLIS
- Funds were not transferred to: EMBRAPA, INRA RENNES

(contract with EMBRAPA was only recently signed and funds will be allocated shortly; following a meeting in Montpellier in February 2006 it was agreed that INRA Rennes resignes from the Project Activities – the Research activities allocated to INRA Rennes were transferred to INRA Clermont-Ferrand starting from January 1st 2006 under the supervision of Pierre Sourdille, Gilles Boutet an d Catherine Feuillet– the funds allocated to INRA Rennes for 2005 will be transferred to INRA Clermont-Ferrand in 2006 – Dr. Dedryver will continue to contribute to the Project as consultant of the INRA Clermont-Ferrand Unit).
- The CerealImmunity Annual Meeting was organised in Rome on October 2nd as a satellite event of the GCP Annual Meeting – four of the 7 institutions took part to the meeting (AGROPOLIS, JIC, UCD, NIAS). The Report of the Meeting was sent to GCP Head Office and SP2 Leader
- The planned visit of a scientist from AGROPOLIS to EMBRAPA was postponed to Year 2 given that activities were not engaged at EMBRAPA. It will probably coincide with the GCP Annual Meeting Scheduled in Brasilia in Autumn 2006.
The 2nd CerealImmunity Meeting is scheduled in Montpellier on October 8th 2006 as a satellite meeting of the rice Functional Genomics Meeting
- approximately 30% of the transcriptome data generated at UCD and at NIAS were transferred at Agropolis and analysed by Dr Lucie Michè – a number of candidates to be tested by Q-PCR and functional analysis and develop functional markers at Agropolis, UCD.

Deviations from original Plan:
- EMBRAPA has only engaged some of the planned activities for Year 1 in relation to the Cytological Characterisation of non-host interactin in rice and wheat to Puccinia and Magnaporthe strains due to delay in signing the contract and lack of additional funding.
- Dr. Lucie Michè at Agropolis hosted Miss Hale Tufan from JIC or a period of 3 weeks to carry out controlled inoculation of non-host strains of *M. grisea* in selected wheat genotypes for Affymetrix microarray experiments to be carried out at the JIC Genomics Platform.

Mitigation

As AGROPOLIS and JIC have carried out relevant experiments to ensure that the transcriptome analysis is not delayed, there is no major impact on the general research plan.
The visit of an Agropolis scientist to EMBRAPA is postponed to Year 2

Annual Report from Agropolis

Research activities at Agropolis were focused on two main areas:

1. CYTOLOGYCAL ANALYSIS of non-host interaction rice-*Magnaporthe grisea* and rice-*Puccinia triticina*. Microscopic analyses were carried out using DAB staining procedure to visualise cell death-related events and counterstaining with aniline blue to visualise the fungal structures (e.g. appressorium, infection hyphae). Cytological characterisation focused for *Magnaporthe grisea* non-host interaction with three isolates (BR32, BR29, TH3) and three rice varieties (NIPPONBARE, IR64 and Sari Celtic). Cytological characterisation for *Puccinia triticina* focused on rice viarieties IR64, Nipponbare and wheat varieties Renan, Recital and Camp Remy.

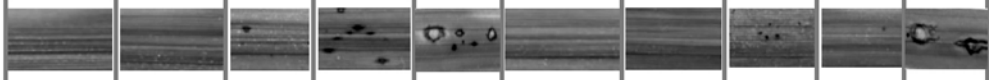
This work enabled to pinpoint clear differential responses to non-host strains of M. grisea classified as type I (absence of cell-related HR-like events) and as type II (associated to HR-like events).

The rice-Puccinia non-host interaction was found to be often associated to HR-like localised events on mesophyll cells subtending the stomata.

2. PREPARATION of RICE RNA SAMPLES for transcriptome analysis – kinetic experiments were set up according to the following experimental set-up:
- 18 day-old rice (Nb & IR64) inoculated with *Magnaporthe* & *Puccinia* (see table)
- 15 leaves cut for each condition at 8hpi, 15hpi & 24hpi (inoculation / mock) & frozen in liquid nitrogen (not sure you need so much details ...)
- RNA extraction using Trizol method, & purification on Qiagen columns
- total RNA sent to the transcriptome platforms for hybridisation (with 2 repetitions per condition)

RNA samples were sent to UCD and NIAS for hybridisation experiments in January 2006

Cerealimmunity Rice Transcriptome Experimental Design

Rice cv.	Nipponbare (*japonica*)					IR64 (*indica*)				
Fungal strain	*Pucci* BOOM 057-A	*M.g* BR29	*M.g* BR32	*M.g* CL3.6.7	*M.g* CL26	*Pucci* BOOM 057-A	*M.g* BR29	*M.g* BR32	*M.g* FR13	*M.g* CL26
Fungal host	wheat	digitaria	wheat	rice	rice	wheat	digitaria	wheat	rice	rice
Symptoms										
(Non) host type	NH-II	NH-I	NH-II	H (avr)	H (vir)	NH-II	NH-I	NH-II	H (avr)	H (vir)

Report from JIC

General comments: JIC engaged in additional microscopic activities not planned in the initial Research plan to support and complement the experimental approach (cytological analysis) undertaken at Agropolis

JIC was granted permission from GCP Head Office to hire a PhD student from 2006 to 2008

Research Activities:

Expt (1): Macroscopic and microscopic examination of yellow rust (*Puccinia striiformis*) development on rice, wheat and barley genotypes.
(This was an additional experiment, not present in the original proposal)

Four wheat yellow rust isolates tested: WYR85/22
WYR93/24
WYR90/505
WYR96/31

Four barley yellow rust isolates tested: BYR60/7
BWR80/1
BYR84/3
BYR61/31

(First figure represents the year of collection)

Species/Varieties/Genotypes tested:

Barley – variety Berac

Wheat – varieties	Lemh
	Chinese 166
	Renan
	Recital
Rice genotypes	Maratelli
	Kasalath
	NIP
	IR36
	IR64
	CO39

Macroscopic plant phenotypes were recorded 15 and 20 days after inoculation (dai). Samples were taken for microscopic examination of yellow rust development at 24 and 48 hours after inoculation (hai) and at 15 dai.

Results: Macroscopic: No visible phenotype was seen on any rice genotype at 15 or 20 dai.

Wheat Yr on non-host barley produced small, necrotic lesions. Barley Yr on non-host wheat produced lesions which were far darker in appearance (dark brown) than lesions seen in a host-avirulent interaction.

Microscopic: Yr spores were able to germinate on all rice genotypes, but on no rice genotype were Yr spores able to find and enter stomata. On wheat, both host avirulent and virulent isolates and barley non-host isolates were able to enter stomata, the development of the pathogen in the incompatible interactions being arrested at later pathogen growth stages. The barley Yr isolates were less able to find and enter stomata on the non-host wheat, than were the wheat Yr isolates on the non-host barley.

Expt (2): Macroscopic and microscopic examination of leaf rust (*Puccinia triticina*) development on rice, wheat and barley genotypes.
(This was an additional experiment, not present in the original proposal)

Two wheat leaf rust isolates tested:	WBR94/1
	WBR98/20
One barley leaf rust isolates tested:	BBR83/3

Species/Varieties/Genotypes tested:

Barley – variety	Berac
Wheat – variety	Lemhi
	Chinese 166
	Armada
Rice genotypes	Maratelli
	IR64
	CO39

Macroscopic plant phenotypes were recorded at 15 dai. Samples were taken for microscopic examination of leaf rust development at 24 and 48 hai and at 15 dai.

Results: Macroscopic: Small flecks were seen on Maratelli and IR64 at 15 dai.
Wheat Lr on non-host barley produced dark brown, necrotic lesions, with occasional small pustles. Barley Lr on non-host wheat produced small pustles.

Microscopic: Wheat Lr isolate WBR98/20 and barley Lr isolate BBR83/3 (but not WBR94/1) were able to enter stomata on all rice genotype. More Lr spores were able to infect Maratelli < IR64 < CO39. Cell death was seen in Maratelli and IR64 only.

Conclusion: Yr is unable to enter rice tissue. Lr however, can attempt to infect rice mesophyll cells and initiates an HR cell death. Lr can also invade non-host wheat and barley to a greater extent than Yr. Lr is therefore selected as cereal rust pathogen for this study.

Wheat Micro Array Analysis: At the GCP Meeting in Rome 2005 an experimental design was agreed for the wheat micro array experiments. Renan was selected as wheat genotype. We are currently running a wheat and barley Lr screen to select suitable wheat – avirulent and virulent isolates, and an interesting non-host phenotype with a barley Lr isolate.

Project Management at JIC: The NGC Programme has agreed the conversion of the Cereal Immunity project at JIC into a 4 year PhD for Ms. Hale Tufan. Ms Tufan will start this PhD in February 2006, which will run until December 2009. This, however will not jeopardise the achievement of JIC's milestones within the timescale of the project

Report from EMBRAPA

Activities at EMBARAPA focused on ongoing work on the phenotypic characterisation of wheat and rice *M. grisea* isolates at the Goiania EMBRAPA Experimental Station.

Phenotypic characterisation of non-host interaction of rice and wheat with *Magnaporthe grisea*

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 (Table1) were utilised for inoculation tests. Inoculations were made with aqueous spore suspension ($3x10^5$ 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 characterisation 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 cultivrs including IR64 with *P. recondita* races. The work is underway.

Table 1. Disease severity on wheat cultivars in inoculation tests with18 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 utilised in the inoculation tests, their orign, 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 / *Digitaria horizontalis*	Goiás	1989
Py 195	Grass / *Eluesine indica*	Goiás	1989
Py 5990	Barley /-	Rio Verde/Goiás	-
Py 3970	Rice / Bonança	Goiás	2002

Annual Report from NIAS

We have collected and analysed the complete nucleotide sequence of 32,127 full-length cDNA clone from japonica rice as reported in July 2003. Based on the sequences of the full-length cDNA clones, 22K 60mer-oligomicroarray system was established in November 2003. This microarray system is one of the globally used microarray systems. Recently based on the partially sequenced whole sequence information of 380,000 full-length cDNA clones, 22K array system will be updated to the 44K format array. Now the 44K array is final process of update. We have also contributed to the establishment of crop gene expression database in SP-4-32 project as a PI. Data obtained in this project will also be repositted in the database.

Originally project planning meeting should be held in Cancun in July 2005, but because of the attack of the Hurricane “Emily”, the meeting was postponed to October meeting in Rome. In the Rome meeting, we have discussed the condition of microarray experiments; how many times biological repeat should be made, how the experiments should be designed. At this moment, 44K

array system is not available but data should be obtained as soon as possible, so we have decided to start the hybridisation experiments with 22K oligoarray.

In January 2006 Agropolis (Lucie Miché) sent 48 samples of RNAs from IR64 infected leaves with Puccinia and Magnaporthe (5 infection processes, 3 time points and two biological replicates). Quality and quantity check of the RNA samples were made, hybridisation experiments were carried out using the 22K array with a subset of RNA samples. The data were sent to Agropolis in March 2006 for analysis.

Report from UCD

Rice transcriptome analysis by UCD (www.ricearray.org) will be carried out using RNAs isolated from rice plants inoculated with selected non-host and host-specific strains of *M. grisea* and *P. triticina* to define the overlap between host and non-host interactions and identify a set of non-host specific genes (NHG). This analysis will reveal if a common set of genes exhibits altered expression in response to non-host vs. host strains of *Magnaporthe* and *Puccinia* in rice. Inoculations were carried out at Agropolis (*Magnaporthe* inoculations) and EMBRAPA (*Puccinia* inoculations). To minimise effects of environmental variation, all samples for transcriptome analysis were grown under the same controlled environmental conditions.

We have been sent samples from the following 6 treatments: rice leaves inoculated with Puccinia (non-host), M. grisea BR29 (non-host), M. grisea BR32 (non-host), M. grisea CL26 (host, virulent), M. grisea CL3.6.7 (host, avirulent), and MOCK sampled after 8 hrs inoculation. Other samples prepared after 15hrs and 24hrs inoculation will be shipped in the end of Feb, 2006.

We will perform the first array experiments for 8hr samples with 20k chip to check whether current microarray scheme including sampling methods and selecting treatments will work well. All differentially expressed genes will be isolated from comparisons between each fungi treated samples and MOCK treatments (Puccinia vs MOCK, M. grisea BR29 vs MOCK, M. grisea BR32 vs MOCK, M. grisea CL26 vs MOCK, and M. grisea CL3.6.7 vs MOCK). Samples containing at least 300 ug total RNAs were already sent from Agropolis. Enough mRNAs (at least 2 ug) from these total RNAs was isolated at UCD. After synthesising cDNA from 0.75ug mRNAs, cys3 and cys5 were incorporated for each hybridisation. The quality of RNA will be checked in every step with nanodrop and bioanalyzer. Before hybridisation, we will check the amount of labeled probes with nanodrop and balance the cys3 and cys5 labeled probes. We have previously established that slides pre-treated with Sodium Borohydride canreduce dye effects. For hybridisation, we are using an automated hybridisation station (Tecan4800), which will reduce variation between slides. The hybridised slides will be scanned with the GenePix4000B scanner. All images will be processed using GenePix software for element identification; raw data (gpr files) will be generated. Each treatment for all time points has 4 replicates (2 biological and 2 technical replicates). R-based LMGene method will generate differentially expressed genes.

Oligonucleotides for the UCD 20k array were based on 20,230 TIGR version 2.0 gene models. All oligos were designed by the whole genome microarray design software (Picky 1) developed by HH Chou (http://www.complex.iastate.edu/download/Picky/Picky1_oligos/RiceSetData.html). The 20K array also contains 6 transgenes; hygromycin, *GUS, BAR, GFP, Cry1Ac, Luciferase.* Of them, 217 hygromycin *oligos were randomly spotted* overall 20k slide as non rice spiking controls. We have confirmed that these oligos worked well as a positive control for comparison between transgenic and non transgenic samples and as a negative control for non transgenic samples from previous array experiments.

We already set up several methods to validate microarray data. First, we will calculate ratios of hygromycin spots showing at least two fold changes. They will inform what the error rate for

oligos at low expression level is because non transgenic samples don't show expression of hygromycin genes.

Second, image data of dye-swapped slides easily indicates reproducibility of genes at high expression level and overall reproducibility will be calculated through correlation coefficients between dye-swapped slides and visualised by scatter plots.

Third, to find sources of variation (slide, treatment, dye, and other errors) between treatment and control, we will do ANOVA test encoded in the LMGene package that was developed to analyse microarray data by David Rocke at UCD. We can therefore determine if the major source of variation in the experiments using 20k chips is from the treatment as expected.

Fourth, for checking reproducibility, we make scatter plots for dye-swapped samples and calculate correlation coefficients between dye-swapped samples.

Fifth, we will do test for sensitivity through comparison between microarray data and expression profiling data based on the number of ESTs in leaf tissue. This type of analysis will inform us as to how many differentially expressed genes exist in microarray data according to expression level based on the number of ESTs in leaf. All EST information will be included in data file.

After validation, we will isolate statistically significant genes by considering both p-values and $\log_2$ ratios generated by LMGene method. To do this, all experiments will be biologically replicated two times at three time points. Dye swaps will be performed for each experiment. Thus 60 arrays will be generated for the analysis in Japonica (3 time points x 5 treatments vs control x 2 replicates x 2 dye swaps).

To find biological meaning of candidate genes, we will carry out GO (gene ontology) analysis and as original gene ontology is too complicated to apply for microarray, TIGR developed plant GOslim (http://www.tigr.org/tdb/e2k1/osa1/GO.retrieval.shtml) which is cut-down version of original GO. Through this analysis, we can get from one to twenty five GO terms for each gene.

Data analysis, deposition and public access at TIGR for consortium project data will be available. As part of NSF funded rice olignucleotide project, R. Buell at TIGR has constructed a microarray database to house data generated from the oligo arrays (www.ricearray.org). All bioinformatics resources and expertise in the oligo arrays project will be generated to this project including raw data analysis, data deposition, and web-based interface tools (GOhan, http://gohan.tigr.org/cgi-bin/gohan_runserv) including LMGene method for data-mining the expression data.

New 45K chips are now being printed in TIGR and will be available at the end of Feb, 2006. This chip contained 43,312 version 3.0 TIGR gene models and targets 44,974 rice genes. All oligos were designed by the whole genome microarray design software (Picky 2) developed by HH Chou (Bioinformatics, 2004. 20:2893). Of them, there are 40,098 unique oligo probes designed for 40,098 sequences and another 3,214 shared oligo probes designed for 4,876 additional sequences. Similar to the 20k chip, 456 hygromycin oligos were randomly spotted as internal controls. TIGR will soon upload information about the rice t-DNA insertional mutants and these will also be included in the microarray results. We can then precede efficiently to carry out functional validation for target genes by using rice t-DNA or Tos17 insertion mutants. If we get the valid information from 20k chip experiments for 8hr samples, we will repeat all experiments with new 45k chip. In the second set of experiments we will use 3 biological replicates and 2 dye swaps.

In January 2006 Agropolis (Lucie Miché) sent 48 samples of RNAs from NIPPONBARE infected leaves with Puccinia and Magnaporthe (5 infection processes, 3 time points and two

biological replicates). Quality and quantity check of the RNA samples were made, hybridisation experiments are underway.

Report from CIMMYT

CIMMYT was not directly involved in the Research activities of the Cerealimmunity Project Year 1. Manilal Williams and Ravi Singh were actively involved in the discussions and gave their availability to test a selection of rice varieties with Puccinia strains t the phenotypic level. Agropolis decided not to send the rice varieties given that no macroscopic phenotypes were observed with the 9 varieties under examination.

As part of the Year 2 activities mapping of identified NHGs by transcriptome analysis will be carried out in three wheat mapping populations, previously saturated with SSRs and AFLP markers. These three populations are:

- ITIMI mapping population – 112 RILs derived through SSD.
- Fukokumughi x Oligoculm – 107 DH lines
- Frontana x Inia 66 – 120 SSD derived RILs.

Report from INRA RENNES

INRA Rennes was not directly involved in the Research activities of the Cerealimmunity Project Year 1. Francoise Dedryver was actively involved in the discussions and contributed on the decision of using the Renan x Recital mapping, previously phenotyped for a number of disease-related characters with particular reference to Puccinia resistance. These data will be made available to the Cerealimmunity project during Year 2. Following a meeting in February 2006 in Montpellier – France, it was decided that from March 2006 the planned wheat mapping activities will be transferred from INRA Rennes to the INRA Unit at Clermont-Ferrand under the supervision of Dr. Pierre Sourdille and Dr. Gilles Boutet given the existence of a high-throughput wheat mapping facility. Dr. Devryver will continue to act ad Consultant for the INRA Unit at Clermont-Ferrand for the wheat mapping strategies and to be associated to the Cerealimunity Project.

12. Drought Tolerant Rice Cultivars for North China and South/Southeast Asia by Highly Efficient Pyramiding of QTLs from Diverse Origins

Principal Investigator:
Zhi-Kang Li, CAAS
Co-Principal Investigators:
Gary Atlin, IRRI
Ze-Tian Hua, Rice Research Institute, Liaoning Academy of Agricultural Sciences, China

Mid-Year Report

1. Research activities and progresses at ICS/CAAS and Chinese collaborators:

1.1 Replicated progeny testing of drought tolerant (DT) introgression lines (ILs) under stress and non-stress conditions

A total of 630 selected drought tolerant (DT) C418 ILs from 74 BC populations plus 42 promising pyramiding lines (PLs) with high yield and good DT derived from first round pyramiding crosses were progeny tested in replicated experiment under both stress (no irrigation) and normal irrigated conditions in the dry season of 2005-06 in Hainan. These lines were evaluated for grain yield, yield components and traits related with drought tolerance. Data have been collected and will be analysed in the summer of 2006.

1.2 Genotyping of the selected DT lines

Parental lines of 12 BC populations have been surveyed for SSR (single sequence repeat) marker polymorphism with more than 400 anchor SSR markers and an average 137 polymorphic SSR markers were identified for each BC population (Table 1). Genotyping of 307 selected DT ILs from the 12 BC populations is in progress, and has been largely completed for at least 3 populations.

Table 1. Screening and Genotyping of Polymorphic SSR Markers
Please contact PI for table

According to the genotypic data, the average introgression in the 10 DT ILs selected from C418/C71 populations was 0.088±0.154, significantly lower than the expected 0.125 and introgression in the 6 DT ILs from C418/Zaoxian14 population fit well to the expectation (Table 2). However, the variation among different loci across the genome was very large in the DT ILs from both populations. The observed numbers of loci showing excess introgression significant at 0.001 and 0.0001 were 57 and 123 times, 470 and 850 times more than expected by chance (Table 2).

Table 2. Frequencies and standard deviations of donor introgression detected by SSR markers in the 16 DT ILs selected from two BC2F2 populations
Please contact PI for table

Table 3. The frequencies of loci showing significant allelic and genotypic frequency deviation (at four significance levels) in the 16 DT ILs selected from two BC2F2 populations
Please contact PI for table

Table 4. Marker loci showing significant excess introgression in 31 DT ILs selected from C418/Teqing and C418/Manauthukha BC2F2 populations
Please contact PI for table

Figure 1. Twelve genomic regions where highly significant ($P < 0.0001$) excess introgression of donor alleles were detected in the 16 introgression lines selected for drought tolerance in the C418/C71 and C418/Zaoxian14 BC2F2 populations.
Please contact PI for figure

1.3 Development of DT japonica rice cultivars by pyramiding DT QTLs

To develop DT C418 and Liaojing454 cultivars by QTL pyramiding, we made 74 crosses between DT ILs in the C418 background and 169 crosses between DT ILs in the Liaojing 454 background. The true hybrid F1 plants from these crosses were confirmed with SSR markers and advanced all crosses to the F2 generation. The 243 F2 populations were subjected to very severe lowland drought during the dry season 2005-06 in Hainan. From 57 of these pyramiding F2 populations, we were able to select a total of 1207 DT plants, including 583 DT plants in the C418 background and 624 DT plants in the Liaojing454 background (Table 5). Seeds from these selected plants will be grown in the summer for both genotyping and preliminary phenotyping experiments.

Table 5. Pyramiding F2 populations derived from crosses between DT introgression lines from different BC populations and the numbers of selected DT from each of the F2 populations.
Please contact PI for table

2. Research activities at IRRI

2.1 Second round crosses for pyramiding for DT QTLs

A total of 37 crosses were made between 17 promising IR64 PLs each with multiple QTLs from two donors obtained from 1st round of pyramiding populations and advanced to the F2 generation (Table 6). These F2 populations were subjected to severe drought under the lowland conditions during the dry season of 2005-2006 in Hainan. From these F2 populations, we were able to select a total of 690 DT lines, which are expected to contain DT QTL combinations from 3 different donors. These lines have been progeny tested under the normal irrigated conditions in Hainan during the 2006 wet season to identify promising PLs with both superior yield potential and quality. Promising PLs from the experiment will be sent to IRRI to be tested in replicated experiments under both stress and non-stress conditions for development of DT and high yielding lines for the rainfed areas of South/Southeast Asia. Selected plants will also be genotyped with differentiating SSR markers to verify QTLs from different donors and QTL networks underlying DT in rice.

Table 6. Selection experiments on 37 F2 populations derived from crosses between 17 pyramiding IR64 lines each with multiple DT QTLs from 2 different donors conducted in Hainan, China

Please contact PI for table

2.2 Combining DT and high yield potential by QTL pyramiding

A total of 21 crosses were made between the 7 selected DT IR64 lines each with multiple QTLs from two donors. All F1 plants from these crosses were examined with at least 5 polymorphic SSR markers and 14 of the 21 crosses had true F1 plants, which were advanced to the F2 generation. Seeds from the true F1 plants of each cross were divided into two parts, grown separately under the normal irrigated and lowland drought conditions respectively in the experimental farm of IRRI during the dry season of 2005-2006 (Table 7). Two types of selection were conducted. One was conducted under the normal condition and a total of 523 plants with high yield potential were selected. The other was performed under the lowland drought and a total of 666 superior plants with best yield under drought were selected (Table 7). The selected lines will be progeny tested in the replicated experiments under stress (upland) and non-stress conditions during the 2006 wet season to identify promising high yielding and DT IR64 lines pyramided with QTLs from 3 donors and for development of DT and high yielding lines for the rainfed areas of South/Southeast Asia. These selected lines will also be genotyped with differentiating SSR markers in the summer of 2006 to verify QTLs and QTL networks underlying both DT and yield potential in rice.

Table 7. Selection experiments on 14 F2 populations derived from crosses between 17 pyramiding IR64 lines each with multiple DT QTLs from 2 different donors conducted in IRRI

Please contact PI for table

3. Evaluation of first-cycle IR 64 drought-tolerant introgression lines in drought-prone Indian environments

3.1 Overview

The IR64 drought-tolerant introgression lines (DT-IL) were evaluated in 2005 by the IRRI-India Drought Breeding Network, a network of Indian breeding centres (described below) supported by the Rockefeller Foundation, the Generation Challenge Programme, and BMZ. This network evaluates materials contributed by IRRI and other centres under naturally-occurring and artificially imposed water stress conditions. The present report on the DT-IL materials is abstracted from a larger data set involving 233 breeding lines, released cultivars, and traditional

varieties. In this report, means for the DT-IL lines and a representative sample of materials from other programmes are presented. These lines fell mainly in the 100-120 day duration range.

In 2005, all centres except CRRI, conducted trials in the wet season under naturallyoccurring stress; at CRRI, screening is conducted under artificially imposed stress in the dry season, and was initiated in January 2006. CRRI trials are therefore not reported herein. Stress trials were conducted under direct-sown upland (CRURRS, TNAU, UAS) or lowland (NDUAT, IGAU) management. Non-stress control trials were also planted at each site. Stress was enhanced by draining paddies after transplanting, or, in the case of upland trials, by planting on raised beds (CRURRS). Planting was also delayed at IGAU to intensify stress. Grain yield, harvest index, plant height, and flowering date are the common variables collected at each site. Rainfall and mean water or water-table depth was used to characterise trial site hydrology.

3.2 Test site descriptions

CRURRS, Hazaribag

Hazaribag is located 120 km north of Ranchi in Jarkhand, on the Chottanagpur plateau, in a severely drought-prone area where extremely short-duration aus-type upland rices are grown in unbunded upper fields, and short-duration semi-dwarf lowland rices are produced in shallow bunded fields that retain little water. It received 950 mm of rainfall during the growing season of 2005, slightly less than the long-term average. Upland trials suffered severe stress from panicle initiation onward. Stress was particularly severe in the managed stress irrigation regime, wherein plants under test were direct-sown on raised beds that dried quickly after rainfall.

CRRI, Cuttack

CRRI is located in an area of high monsoon rainfall and stagnant flooding. It is used by the network as a dry season screening site. Vegetative and reproductive-stage screening trials are being conducted during the 2006 dry season.

IGAU, Raipur

Raipur is located in central Chhattisgarh, where soils are relatively heavy and rainfed lowland rice is the predominant crop. The monsoon is brief, with little rain falling after September. Droughts are frequent in the region. Total seasonal rainfall was less than 1000 mm in 2005. Moderately-stressed trials were irrigated until 30 days after transplanting. Severely-stressed trials were rainfed from transplanting onward. Stress was severe around flowering.

NDUAT, Faizabad

Faizabad is located in eastern Uttar Pradesh, in a low-rainfall area where lowland rice is grown under either rainfed conditions or with supplementary irrigation. 816 mm of rainfall were received at this site in 2005, but the stress trials, which were all conducted under transplanted lowland conditions, were irrigated until 30 days after transplanting. Trials drained slowly, and no stress occurred.

UAS, Bangalore

Bangalore is located in Karnataka, in southern India. The target system in the region is dry direct-sown rice in upper fields that do not retain standing water. Varieties used by farmers are approximately 120 days in duration. Long-term annual rainfall at the research site is 767 mm. Rainfall was adequate in the 2005 field trials, and little stress was experienced.

TNAU, Coimbatore

Tamil Nadu trial sites were located in Coimbatore and at the southern locations Ramnad and Paramakudi. Rainfed rice in the region is grown via dry direct seeding in shallow bunded field. Rainfall is extremely variable at these sites, with long-term averages of about 700 mm at

Coimbatore and 500 mm at the southern sites. Little stress was experienced in 2005; seasonal rainfall was 698 and 605 mm at Paramakudi and Ramnad, respectively.

3.3 Entries

A total of 46 IRRI DT-IL entries from the first cycle of selection were screened in 2006. Performance of these lines compared to relevant checks and advanced materials from other programmes is presented in Tables 10 and 13. All DT-IL entries fell into the 100-120 day maturity class. Two trials were conducted: an advanced yield trial (AYT) conducted in 4-row plots, and an observational yield trial (OYT) conducted in 2-row plots. Entries were allocated to the OYT or AYT depending on seed availability.

3.4 Statistical analysis

Trials were grouped into non-stress, moderately stressed, and severely stressed groups based on trial mean yield, hydrology, and the occurrence of stress periods. At least one ton ha-1 in mean yield separated the stress level groupings. Means were estimated across trials within stress levels using a model that considered trials random and genotypes fixed.

3.5 Results

Observational yield trials of varieties of 100-120 days duration (OYT100-120)

Thirty-six DT-IL were included in the OYT100-120. Trial means of grain yield and agronomic traits are presented in Table 8 and trials included in each stress class, as well as the grain yield mean of each class, are presented in Table 9. Stress during flowering and filling was severe at IGAU, the only trial included in the severe-stress class. Means of grain yield, days to flower, and harvest index of lines and varieties within stress levels are presented in Table 10. On average, DT-IL lines yielded nearly 400 kg, or 80%, more than IR64 under severe stress. No improvement was observed under moderate stress or control conditions. The improvement under severe stress was associated with increased harvest index, slightly reduced maturity.

Nine of the 35 DT-IL significantly outyielded the check IR64 under severe stress. None yielded significantly less. The highest-yielding varieties under severe stress were the DT-IL DGI 75 and DSL 104-1, and the IRRI aerobic rice variety IR74371-78-1-1. These varieties were also among the highest-yielding under non-stress conditions. The UAS varieties ARB 4 and ARB 5 also performed well across stress levels.

Table 8. Trial means of grain yield and agronomic traits of genotypes maturing in 100-120
Please contact PI for table

Table 9. Observational Yield Trial of genotypes maturing in 100-120 days
Please contact PI for table

Table 10. Observational Yield Trial of genotypes maturing in 100-120 days: line means within stress levels
Please contact PI for table

Advanced yield trials of varieties of 100-120 days duration (AYT100-120)

Ten IRRI DT-IL and varieties were included in the AYT100-120. Trial means of grain yield, days to flower, and harvest index are presented in Table 11 and trials included in each stress class, as well as the mean of each class, are presented in Table 12. Stress during flowering and filling was severe at IGAU, the only trial included in the severe-stress class.

Means for lines and varieties within stress levels are presented in Table 13. Again, on average, the DT-IL lines outyielded the recurrent parent IR64 under severe stress (by 300 kg ha-1, or about

60%. However, they were not, on average, superior under non-stress or moderately stressed conditions, and were not among the highest-yielding lines under severe stress.

The traditional variety Kallurundkar was the highest-yielder under severe stress but was below average under moderate and non-stress conditions. Baranideep, an NDUAT variety, gave high yields under all stress levels. Other improved varieties combining high yield potential with high yield under stress included: CB00-15-24, IR74371-46-1-1, IR74371-3-1-1, RR272-21, and Tripuradhan. The widely-grown short-duration variety MTU 1010 had high yield potential but relatively low yield under stress. Widely-grown short-duration IRRI varieties IR64 and IR36 were moderately and highly susceptible to severe stress, respectively.

Table 11. Advanced yield trial of varieties with duration of 100-120 days.
Trial means for agronomic traits and heritability of grain yield.
Please contact PI for table

Table 12. Advanced yield trial of lines of 100-120 days duration:. Trial groupings into stress levels
Please contact PI for table

Table 13. Advanced yield trial for lines of 100-120 days duration: Line means over sites within stress levels
Please contact PI for table

3.6 Testing in India: Conclusions and Next Steps

The initial year of the IRRI-India Drought Breeding Network successfully evaluated 46 IRRI DT-IL breeding lines from the first cycle of selection in this programme. These selections showed consistent improvement over the recurrent parent IR64 under severe lowland stress at Raipur, with no change under moderate stress or fully irrigated conditions. In one trial, a DT-IL was the highest-yielding entry under severe lowland stress. The improvement in drought tolerance was associated with a slight reduction in days to flower, and an increase in harvest index under stress. These results demonstrate that phenotypic selection for yield under lowland stress can translate into gains in the target environment. Promising lines will be retested in 2006, and farmer-participatory evaluation will be initiated. A new set of lines from this programme, derived from the second cycle "pyramiding" phase of selection, will be increased in India and evaluated in 2007.

Tangible outputs delivered:

The elite C418 ILs with good drought tolerance and restoring ability identified last year was used as restorer to produce F1 hybrid with an elite CMS line. This hybrid will be recommended to be tested in multilocational yield trials under stress and non-stress conditions of North China this year. Forty-six promising initial selections and ten promising 1st round IR64 PLs with good DT and improved yield potential have been identified and sent to NARES to be tested in the multilocation yield trials in the rainfed areas of South/Southeast Asia. Hopefully, some of these lines will be released to farmers soon.

Deviations from the work plan:

Because of the personnel change in IRRI, some of the IRRI activities have been conducted in China.

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

Principal Investigator:
Marilyn Warburton, CIMMYT
Co-Principal Investigators:
Edward Buckler, Cornell University
Alain Charcosset, INRA
James Gethi, KARI
Grudloyma Pichet, NSFCRC, Thailand
Luke Mehlo, SIRDC
Tim Setter, Cornell University
Wanchen Li, Sichuan Agriculture University
Collaborating Scientists:
Marianne Bänziger, CIMMYT
Javier Betran, Texas A&M
Jose Crossa, CIMMYT
Philippe Monneveux, CIMMYT

Mid-Year Report

I. Trials conducted at CIMMYT, Tlaltizapan station, Mexico

Dec 2005: Based on the flowering data collected from the previous evaluation we reassigned genotypes into different precocity groups. These groups have been planted in two trials (well watered and stress) with three replicates. Leaf samples, silks and ear tips were harvested at mild and severe stress stages for metabolite analysis. This is the second year of field evaluation and analysis of tissue metabolites on this material.

The number of accessions was reduced from 460 to 400 (for genotyping issues). Reduction in numbers was achieved by looking at adaptability in Mexico (flowering and seed set) and extremes of flowering (MFLW under stress). Reclassifications was made to facilitate the field design (maintenance of different flowering times), with number of accessions in group 1 (n=110), group 2 (n=170) and group 3 (n=120).

Field treatment has been as follows:

- Water stress at flowering time (first planting date)
- Well-watered conditions (second planting date)
- Field design: Alpha lattice, 2 Reps, for each individual group
- Ear shoots bagged before silk emergence
- 2.5 meter row/plot 2 seeds per hole
- Measurement: Harvest target tissue at 0 and 7D at flowering

About 4800 samples were harvested from field under stress in March and April 2006.

- Three tissues – ear tips, silks and leaves
- Two sampling time points – 0 days and 7 days: means harvest at pollen shedding and 7 days after pollen shedding
- Two stress replicates
- Samples were placed in chilled 80% methanol in the field and stored for 1 month in cold storage (-20 C) for exodiffusion.

Phenotyping: male flowering (MFLW), female flowering (FFLW), plant height, chlorophyll content, SEN (leaf senescence rating during the stress period) and root conductivity (measured using an electrical capacitance meter, evaluated at the end of the stress) were collected for all entries.

Seed distribution: We are currently preparing seeds for distribution to collaborators in four countries (May 2006)

II. Metabolite analysis

Samples of leaves, silk and ear tip have been stored at -20 for 4 weeks to let the metabolites from the tissues diffuse into the medium. Samples are being processed and dry aliquots of the methanol extracts will be sent to Cornell for analysis in early June.

Metabolite analysis will be done at Cornell for sucrose, glucose, ABA, ABA glucose ester (ABA-GE). If the assay proves satisfactory for large scale throughput, phaseic acid (PA) will also be done.

Residual dry matter from methanol extractions of tissue will be dried, weighed, and ground to a powder. Dried samples will be put into labelled containers and sent to Cornell for starch determination.

III DNA

DNA has been extracted from the 460 genotypes (which was then reduced to 384, the final drought association mapping panel size) and is ready to be used to determine the population structure using SNP markers.

IV Identification of candidate genes

A first list of 20 candidate genes were developed following the meeting in Kenya based on: relevance of the gene pathway, 2) potential regulatory role of the candidate genes on the metabolites quantified in this study. A new list of 346 potential candidate genes was then generated by Tim Setter and Marilyn Warburton based on knowledge of the ABA and carbohydrate synthesis pathways and drought tolerance in maize and other plant species. This list was reduced to 58 after genes already in the Buckler SNP discovery (62 in total) were removed; genes with no sequence similarities (or conversely, genes with too many homologs, ie., gene families) in maize were removed; and genes with a lower probability of success were removed. Primers were developed for one or two contigs per candidate gene and are currently being tested for amplification; successful primers will be used to sequence a panel of 10 – 12 diverse inbred maize lines for SNP discovery. This is expected to be completed by September, 2006. The total of 112 candidate genes is considerably higher than what was originally proposed, and greatly increases the probability of finding genes that significantly affect the performance of maize plants under drought conditions.

V Association Analysis

A meeting was held on February 15, 2006 at Cornell University to discuss the project. In attendance were Tim Setter, Jean Marcel Ribaut, Marilyn Warburton, Ed Buckler, Mike Gore, and Jonathan Crouch. A phone meeting with Alain Charcosset the previous day had brought his concerns to the table as well. Aspects of the project were discussed, in particular: genotyping/SNP discovery and staffing.

It was agreed that a postdoctoral researcher would be hired as soon as possible to work on the project for two years. Part of that time would be based at Cornell's IGD and part at CIMMYT, Mexico (see attached work plan). A no cost extension will be sought, as the postdoc was hired later than anticipated. Confirmation of associations, when they are determined, is proposed using: 1) NILs to further characterise the association; 2) Re-evaluation in an independent population, including the SP1 maize composite subsets.

	2005	2006	2007
1. Select a diverse set of maize inbred lines*	Done in 2006 n = 384 (from 600)		
2. Select set of candidate genes	Done in 2006 n = 50 done (450 others)		
3. Characterise plant material (morphological and physiological)		Different irrigation regimes (WS and WW). Blocked by early, intermediate, late (384 selected plus up to 72 from Ed's quality trait association mapping) at five locations	
Genotype material	Structure the population (to be done in 2006)	Screen with candidate genes	Screen with candidate genes
4. Association analysis			
5. Development of markers			
Validate markers		Using consensus map and expressional analysis	Using genetic and breeding populations
6. Disseminate results			MB CoP

Deviations from the work plan:
There are two main issues that need to be dealt with here. First the population genotyping. Initially we were going to contract out a company called Genaissance to undertake the SNP genotyping. This company is no longer able to do this work and therefore we are trying to find an alternative genotyping company that can do this work. We are now working with Ed. Bucker at Cornell to develop a complete set of SNP markers and Illumina platform has been chosen for genotyping.

Concerns raised in the last report (November 2005) on the progress of the SNP discovery work has been addressed and this area is now progressing very smoothly.

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

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborating Scientists:
S. Taba, CIMMYT
Sarah Hearne, 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

Mid-Year Report

Notes regarding collaborators:

Luz George has left CIMMYT, and therefore the project. V. Mahalakshmi has left IITA, and therefore the project; she has been replaced by Sarah Hearne. The Philippines has not been able to participate to date, and permission is being requested to remove them from the project.

Selection of the Germplasm:

The final list of germplasm, approved by all project members, was chosen at the Beijing workshop in December, 2005. It can be found in the appendix. DNA has been extracted by So far DNA was distributed by China, CIMMYT, India, Indonesia, Thailand, Vietnam IITA and KARI (partial list) and sent to all the labs who will be participating in the genotyping. DNA amplification tests indicate that all the DNA is of good quality. Seeds were requested from the Genetic Resources Information Network in the US to compliment the germplasm that was supposed to come from KARI, but they were unable to get enough landraces, especially from neighboring countries. Seeds have been shipped to CIMMYT, who will extract the DNA and send to all participating labs by June 1. The Philippines were unable to complete the work and send any DNA. We request that they be removed from the project, as the deadline for completing the genotyping has already passed.

In total, 248 landraces and 20 populations of teosinte have been exchanged, with an additional 27 pending. This number and the distribution is well covering most of the southern (maize growing) half of the African continent and all the major maize cultivation areas of Asia. When compared to the landraces already genotyped in previous studies, we will have a complete picture of maize distribution and cultivation around the world.

Passport on any new collections done by KARI and the Vietnamese NMRI is being requested, but to date, no answer has been given.

Second organisational and training workshop:

The workshop was held on December 5 – 9 in Beijing, China, and hosted by CAAS. The programme consisted of reporting sessions, where all participants reported progress to date and the status of maize cultivation in their respective countries, as well as known historical data about the introduction and practices of cultivation. The next steps of the project were summarised, and action points documented (see Appendix 2). Laboratory demonstrations were held reviewing DNA extraction for the bulk method, PCR protocols, gel loading and running, and extraction of the bulked data from standard software (Genotyper and Genescan, ABI). Discussion of allele frequencies to characterise population diversity and results of past studies using this methodology were reported. Demonstrations of the new stand-alone analysis software written in this project were given, and there were computers available for practice and hands-on work of this, and several other, analysis packages. Finally, tours of the laboratory and surrounding cultural sites were greatly enjoyed by the participants, all organised by our very efficient hosts.

Hiring of the postdoctoral assistant:

It was decided that the project postdoc would be housed at INRA, France, and hired as soon as possible. The position has been approved by the authorities at INRA, and a suitable candidate is being sought. The duties of the postdoc will be to generate the bulk data that INRA committed to for this project, and to collate all diversity data from this, and from two past experiments, of maize population diversity using the bulk method. The postdoc will then analyse all data in this data set, which will comprise nearly 900 landrace populations, the largest ever analysis of maize population diversity.

The Following Expected Outputs from the proposal have now been met:

- A stand-alone programme for the analysis of population diversity data. User's manual for the programme is attached as Appendix 3.

- Up to three training workshops and up to 60 NARs scientists trained in population diversity lab and data analysis. (TWO have been completed now; final meeting will be held in Hyderabad, India.
- A more complete collection of global maize OPVs and landraces held in local, regional and CIMMYT genebanks.

Timeline and Milestones:
Activities highlighted in yellow have already been completed. Activities in BOLD are currently underway and progressing within schedule.

Activity	Partners Involved	Year 1	Year 2	Year 3
		Q1 Q2 Q3 Q4	Q1 Q2 Q3 Q4	Q1 Q2 Q3 Q4
Planning meeting held	All	___		
Populations collected	All except INRA	_____		
Passport data placed in Central Repository	All except INRA	_____		
DNA isolated from populations and sent to partners	All except INRA and KARI	_____		
First bulk analysis workshop held	CIMMYT and host	_____		
SSR data gathered and placed in Central Repository	All except KARI, Phil. DOA and NMRI		_____	
Write new data analysis package for bulk fingerprinting (stand alone)	CIMMYT and INRA	_____		
SSR data analysed and the most unique subset of the populations chosen for Functional Char.	All		_____	
Functional char. of drought candidate genes completed and data placed in Central Repository	CIMMYT and INRA		_____	
Last workshop held				_____

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

Principal Investigators:
Jonathan Crouch, CIMMYT
Francois Tardieu, INRA

Collaborating Scientists:
Peter Stamp, ETH, Switzerland
Andreas Hund, ETH, Switzerland
Matthew Reynolds, CIMMYT
Peter Langridge, ACPFG, Australia
John Bennett, IRRI
Claude Welcker, INRA
Yunbi Xu, CIMMYT
Jose Luis Araus, CIMMYT
Gary Atlin, CIMMYT
Yann Manes, CIMMYT
Rachid Serraj, IRRI
Jill Cairns, IRRI
Guy Davenport, CIMMYT

Richard Bruskiewich, IRRI
Philippe Lessard, Biogemma

Mid-Year Report

The overall objective of this research proposal is to elucidate and integrate common genetic mechanisms (candidate genes) underlying the maintenance of tissue growth in plants subjected to water deficit; and develop suitable strategies for comparative genomics: (i) across three species (maize, rice and wheat) for the response of leaf expansion to water deficit and (ii) across three organs (leaves, silks and roots) for one species, maize.

A. Maize activities

Leaf growth and Anthesis-Silking Interval (ASI) are the main determinants of source and sink strengths of maize, while root growth is a main determinant of the plant capacity to take up water. Because the growth of these three organs have partly common physiological determinisms (plant hydraulic characteristics, cell division and its response to water deficit, cell wall mechanical properties and their responses to water deficit), we examine the hypothesis that they may have a partly common genetic determinism. This possibility was tested in a mapping population which segregates for ASI (P1xP2 developed by CIMMYT) which has been tested for leaf, silk and root growth by project partners INRA, ETH and CIMMYT. A major component of the project was to carry out experiments in field, greenhouse and growth chamber conditions, with sufficient characterisation of the environmental conditions to allow joint analysis of all experiments.

(i) Common genetic determinisms of the responses of leaves and silks (INRA - CIMMYT)

The responses of leaf elongation rate to evaporative demand and soil water deficit have been studied in a series of 6 greenhouse and growth chamber experiments (INRA) and analysed together with 8 field experiments (CIMMYT) in which ASI was measured either in well watered conditions (3 experiments) or in water deficit (5 experiments). The elongation rate of leaf 6 was monitored continuously in 200 recombinant inbred lines (RIL) during 8 to 18 days depending on the RIL. Plants were subjected to changes in temperature, evaporative demand and soil water status in two series of experiments in the greenhouse and the growth chamber, over two years (two series of 120 RILs with 33 RILs in common both years). Each RIL was characterised by its maximum elongation rate per unit thermal time, which was similar in well-watered plants over different nights in all experiments. This maximum rate was accounted for by 5 QTLs, among which three co-localised with QTLs of ASI in well-watered conditions. In all cases, the allele conferring high leaf elongation rate conferred a low ASI, indicating a high silk elongation rate. The growth maintenance of leaves under varying evaporative demand and soil water status was assessed, for each RIL, by the slopes of the response of leaf elongation rate to vapour pressure deficit and to predawn leaf water potential. These parameters had partly common QTLs, among which 3 co-localised with ASI in water deficit. The allele for leaf growth maintenance in all cases caused a shorter ASI. In contrast, other regions influencing ASI had no influence on leaf growth. *This work therefore raises the possibility that the determinisms of the responses to water deficit of leaf and silk growths may be partly common, with a joint set of candidate genes*. Two papers reporting these results have been submitted, one in *Plant Cell and Environment* and one in *Journal of Experimental Botany* (see publication list below).

(ii) Comparison of leaf growth maintenance in field and greenhouse, consequences for modelling the GxE interaction (CIMMYT-INRA in collaboration with U. Queensland and CSIRO under the WPM project)

Two field experiments have been carried out in Mexico with the same mapping population, with or without water deficit, in order to check whether the QTLs identified for the parameters of the response of leaf elongation rate to water deficit co-localise with QTLs of leaf area obtained in the field. The area of every second leaf was collected in 15 plants per RIL in all RILs of the population over two growing season in well-watered and water-limited field conditions. This will

allow us to reconstruct the change with time in leaf area index, and therefore to carry out QTL studies of leaf area at several stages of the plant. ASI and yield data have been collected as well. Data are under analysis, *and will be the base for a modelling approach of the effect of the genetic variability in leaf growth response on yield, thereby giving a direct modelling of the GxE interaction* (collaboration with G. Hammer, under the framework of the WPM project of GCP).

(iii) Modelling the response of silk elongation rate to water deficit and evaporative demand (INRA)

The change with time of elongation rate of silks, and its responses to soil water deficit and evaporative demand have been analysed in a set of RILs of the same mapping population, which contrasted for the growth maintenances of leaves and silks under water deficit. The first results indicate that the time courses of silk growth are closely similar to those of leaves, with a depression during the day due to evaporative demand, and a maximum rate during the night. These results reinforce the hypothesis that silk and leaf responses to environmental conditions may have common determinisms.

(iv) Genetic analysis of root growth response to water deficit (ETH)

A screening system for root growth rate under water limited conditions induced by polyethylene glycol (PEG) was established (Trachsel 2005) and the two parents (P1) and (P2) were evaluated. The system consists of growth pouches (germination plotter paper covered with black foil. In Autumn 2005 the pouch system was adapted for the screening of large sets of genotypes for the QTL mapping. Since January 2006 the QTL mapping population of 208 RIL families of the P1 x P2 cross and the parents are being phenotyped under control (well watered) and water-limited conditions induced by 20% w/v PEG. Plants are germinated on filter paper and transferred to the growth pouches when the tap root reached 0.2 cm length and grown under control conditions until three days after germination (DAG). One plant per genotype and treatment is grown in a pouche. Roots of all RILs are scanned at 3, 6, and 8 days after germination. After the first scanning on day 3, the PEG treatment is applied to half of the plants. A separate, randomly chosen subset of 12 families is grown in both treatments as described above, but scanned daily. This set is used to determine and model the root growth over time for the two treatments. The QTL experiment is arranged as an alpha-lattice design. *Overall, we shall therefore access QTLs for the growth of the three sampled tissues (roots, silks, leaves) with the same genetic material, with environmental conditions controlled in such a way that experiments in the field, greenhouse and controlled conditions can be analysed jointly.*

(v) Tissue sampling for transcriptome analysis (CIMMYT-INRA-ETH)

Tissues of growing parts of leaves, silks and roots have been collected in a common set of maize lines (parents plus six extreme RILs, tolerant or sensitive) for transcript analysis:

- Growing zones of leaves were obtained in the field in 2005, in an experiment with either well-watered plants of with a water deficit during the vegetative stage. Predawn leaf water potential was measured on parental lines (20 leaf samples each) under both conditions. The whorls of 30 plants were dissected and the 5-cm basal parts of the youngest growing leaves (in which cells divide and/or expand) were sampled and pooled. The same tissue will be collected in the same RILs in the next greenhouse experiment at Montpellier, in well-watered and water deficit conditions (same predawn leaf water potential as in the field).
- For silks, the growing part of the silks wase collected in a previous field experiment with a water deficit during flowering. Mature leaves were also collected in the same plants. Silks of plants of the same RILs will be collected in the next Montpellier experiment with well-watered or water deficit.
- For roots, the growing part (1st cm apical parts of the primary and secondary roots in which cells divide and expand) will be collected in the same set of RILs at the two studied water potential of the solution in the pouches.

RNA extractions will be carried out at CIMMYT. All samples will be labelled and hybridised in the Universidad Nacional Autonoma de Mexico (Instituto de Fisiologia Cellular, Jorge Ramirez). RNA microarrays will be read in UNAM, and the data analysis including normalisation will be conducted at CIMMYT. We shall look for candidate genes (i) whose expression is associated in all RILs with growth in all tissues (comparison growing - non-growing tissues, well watered and stressed), and (ii) for which the difference between RILs in transcript abundancy accounts for the difference in growth maintenance under drought for each sampled tissue organ. This should provide interesting candidate genes, which will be confirmed by RT-PCR.

B. Rice activities

(i) Modelling the leaf growth in rice in response to environmental conditions and genetic variability (INRA)

A first set of experiments has been carried out in the growth chamber and in the greenhouse, in INRA, with the genotypes Azucena, IR64 and Nipponbare. The first results indicate that the maize model might apply to several genotypes of rice. (i) the response to temperature has been established, with a surprisingly high threshold temperature, (ii) a clear response to evaporative demand was observed, and a genetic variability was observed (to be confirmed). The results of a second experiment are being currently analysed. The next step will be to check whether responses curves to temperature, evaporative demand and soil water status are repeatable across experiments and for different leaves of a plant. The role of several traits (e.g. hydraulic conductance, ABA synthesis) will be analysed to test the model. The last step will be to analyse a larger number of genotypes (see below) by using the model.

(ii) Germplasm characterisation (IRRI)

Controlled conditions

Two greenhouse experiments were conducted to develop protocols for quantifying tissue growth over time and select parents suitable for mapping growth-related traits. Specific leaf area, dry matter accumulation, relative growth rate, and biomass partitioning during early vegetative growth were determined. Large phenotypic differences were observed for all traits. The Vandana x Moroberekan population was chosen as the best population for determining tissue growth under stress due to large genotypic differences and the absence of the semi-dwarf loci, *sd1*, in either parent which would confound QTL analysis.

Field conditions

To establish the genetic basis of leaf growth under field conditions, the Vandana x Moroberekan population (total 133 lines) was evaluated under well-watered conditions in the wet season 2005. QTL for leaf emergence, leaf elongation, and yield under well-watered were identified. Yield components and their QTLs are being analysed.

Using the Vandana x Moroberekan backcross population, the genetic basis of leaf growth under stress is currently being investigated in 2 upland environments under well-watered and water-limited conditions. One hundred and eight backcross lines (73 BC3 lines and 35 BC4 lines), plus parents and check varieties have been sown in two locations on the IRRI experimental farm. Twenty eight days after sowing (DAS), irrigation has been withheld in stress plots for 21 days. Plants were sampled, destructively, six times during the stress period in stress and 2 control plots and the number of leaves, leaf area, individual leaf length and width, total surface area, and dry weight have been measured.

Initial work on the project under controlled and well-watered conditions shows that it is possible to establish differences in tissue growth with time and identify QTL related to this trait. A delay in funding in year one prevented screening to be conducted under water-limited conditions.

C. Wheat activities

(i) Germplasm characterisation (CIMMYT)

Leaf expansion rate was measured on a contrasting set of lines based on differential response to drought in a field study in 2004, in the field as well as in controlled environments. In controlled environments indicators of soil and tissue moisture stress (leaf water potential and pre-dawn soil water potential) were measured at the same time. However, genetic diversity in tissue expansion rate showed marginal statistical significance. There was a reasonable consistency between results in the field and in controlled conditions for the reduction in leaf expansion rate in response to drought, nonetheless, the genetic variation was not large. Given the relatively low genetic variation, additional lines were selected to include for subsequent cycles based on more recent data collected on genetic resources (landraces etc). These lines show good contrast in water use and WUE in the field. In 2006, we are in the process of repeating what was done in 2005 including the new genetic resources and expect to identify genotypes with more diversity in tissue expansion rate.

(ii) Transcription factors isolation (ACPFG)

Drought-related yield loss becomes substantial when stress occurs during flowering or early grain development, and this loss results from spikelet abortion and /or reduced grain fill. One of the first events in the plant's stress response is alteration in the transcriptional regulation of gene expression. The presence of known consensus regulatory elements in the promoters of several drought-inducible genes indicates that whole classes of genes can be simultaneously regulated by binding of the associated transcription factors to these consensus elements. Consequently, over-expression of the transcription factors should cause simultaneous regulation of their downstream targets, leading to stress tolerance. Some of these transcription factors, when over-expressed under constitutive or stress inducible promoters, have been shown to increase tolerance of vegetative tissue and sometimes of the whole plant to drought.

Members of several families of transcription factors are known to be up- or down- regulated in plant tissue during drought stress. These include AP2/ERF type transcription factors (DREB), basic region leucine zipper proteins (bZIP), homeodomain-leucine zipper proteins (HD-Zip), Myb-like proteins, Myc-like proteins, some families of Zn-finger proteins, etc. The identification and cloning of transcription factors up-regulated in target tissues of drought-tolerant varieties of maize and wheat, and over-expression of this factors from late flowering and early grain specific promoters in less drought-tolerant varieties may increase grain yields under stress. The aim of these experiments has been to identify a series of wheat transcription factors that can be used as candidates for screening in the drought stressed wheat, rice and maize material.

Yeast expression libraries have been developed from wheat grain at 0-6 days after anthesis. As series of *cis*-elements have been introduced into the genomic DNA of the Y187 yeast strain and are ready to be tested in test-screens of the cDNA libraries. These elements are one DRE, eleven different ABRE , two GCC elements, one ARE (CORE), one HSE, one MYB and one bHLH elements. To date six different *cis*-elements involved in drought-responsive gene regulation have been used to screen the expression cDNA libraries from developing wheat grain. Using the DRE element two novel DREB factors and one HD-Zip factor have been identified which may be involved in drought response. These factors have been subcloned into constructs for constitutive expression in barley and are currently being used for transformation of plants. Physiological experiments are under way to examine the phenotype of the transgenic plants. These transcription factors are available for screening.

D. Capacity building

Three members of this project have organised from 3 to 12 July 2006 a training course on 'Phenotyping and water deficit' at Montpellier. Several of the methods used in the project were presented to 23 participants to GCP projects, by a combination of lectures, of field and greenhouse practical sessions and of computer practical sessions (see the report of the course). Two doctoral fellows are involved in characterisation of the variability for early growth response to water deficit in maize and rice.

Conclusion/deviations from initial work plan:

The first year results suggest that the initial hypothesis of the project, namely the existence of common genetic determinisms for the growth maintenance under water deficit in several organs, is probably at least partly correct for maize. This has been shown for leaves and silks, and the results of the root analysis will be available shortly. The common QTL analysis, microarray and RT-PCR carried out on several organs and conditions should provide interesting candidate genes for tissue maintenance under water deficit. Common candidate genes may therefore apply, and their consequences on yield under drought can be tested via the field data collected during this project. The modelling activity, common with the GCP WPM project, is under way. *If the results presented here are confirmed by the on-going experiments, this would have large consequences for breeding programmes (i) for the "pyramiding" of positive traits, because early vigour and ASI would have a partly common genetic determinism, (ii) for the early detection of positive alleles, which might be identified early in plant development, (iii) for modelling the genotype x environment interaction, because some alleles would have an effect at different developmental stages of the plant, (iv) for a common analysis in several species, which may be extended afterwards to other species such as sorghum of pearl millet.*

The part of the project which dealt with comparisons between species (rice and wheat) has been delayed by the departure of co-PIs. Initial results from the collaboration between ETH, CIMMYT and INRA for maize, and between INRA and IRRI for rice makes us confident that the years 2006 and 2007 will provide appreciable progress for rice - maize comparison, common modelling and candidate genes. This possibility remains to be evaluated for the comparison of these two species with wheat.

Tangible outputs delivered:

Diploma Thesis:

Trachsel, S. 2005. Rapid screening of maize seedling roots for drought tolerance. Diploma Thesis, ETH Zurich, Zurich.

Posters or oral presentations at conferences:

Hund A, Pa-In N, Trachsel S, Liedgens M, Fracheboud Y, Stamp P (2005) Root morphology of drought tolerant and drought sensitive maize genotypes. In: Interdrought-II. The 2nd international conference on integrated approaches to sustain and improve plant production under drought stress., University of Rome "La Sapienza", Rome, Italy, September 24-28.

Pa-In N, S. Trachsel, M. Liedgens, A. Hund and P. Stamp (2005) Quantitative trait loci controlling root and shoot traits of maize (Zea mays L.) under drought stress. Phd symposium ETH Zurich, Switzerland, October 20-21.

Pa-In, N. (2005). Quantitative Trait Loci Controlling Root and Shoot Traits of Maize. 25th Anniversary of cooperation between Kasetsart University (KU) and Swiss Federal Institute of Technology (ETH)* - International Conference on Maize Adaptation to Marginal Environments; Kasetsart University, Thailand, March 6-9.

Sadok W, B. Boussuge, C. Welcker, F. Tardieu (2006). A modelling approach to genotype x environment interaction : genetic analysis of the response of maize growth to environmental conditions. Gene-Phenotype Conference, Wageningen April 2006

Tardieu F, Chapman S (2005) Assessing crop simulation models as research tool for analysing crop responses to water deficit *Oral presentation*, In: Interdrought-II. The 2nd international conference on integrated approaches to sustain and improve plant production under drought stress., University of Rome "La Sapienza", Rome, Italy, September 24-28.

Tardieu F, W. Sadok, O. Bou Chabke, AS Voisin, B. Parent, B. Boussuge, C. Welcker, B. Muller, Th Simonneau (2005). Control of maize leaf growth under water deficit and evaporative demand : genetic and physiological dissections of involved mechanisms. Society for Experimental Biology, Congress 2005 Barcelona, oral presentation.

Torres RO, J.E. Cairns, O. Namuco and D.E. Johnson (2005) *The genetic basis of seedling vigour* Poster paper presented at Rice Genetics 5 Symposium, Manila, Philippines. 19-23rd November 2005.

Welcker C, Ribaut JM, Boussuge B, Maton C, Muller B, Tardieu F. (2005) Genetic analysis of the response of maize leaf growth to water deficit in a population segregating for Anthesis Silking Interval. *Oral presentation*, In: Interdrought-II. The 2nd international conference on integrated approaches to sustain and improve plant production under drought stress., University of Rome "La Sapienza", Rome, Italy, September 24-28.

Welcker C, Ribaut JM, Boussuge B, Maton C, Muller B, Tardieu F. (2005) Genetic analysis of the response of maize leaf growth to water deficit in a population segregating for Anthesis Silking Interval. Poster SEB conference in Barcelona

Papers submitted

Sadok W, Naudin P, Boussuge B, Muller B, Welcker C, Tardieu F. Analysis of daily time courses of leaf elongation rate in hundreds of maize lines reveals robust QTL-dependent patterns under naturally fluctuating conditions. Submitted to *Plant Cell and Environment*.

Welcker C., Boussuge B, Benciveni C., Ribaut J.M., Tardieu F. Are source and sinks strengths genetically linked in maize plants subjected to water deficit ? A QTL study of the responses of leaf growth and of Anthesis-Silking Interval to water deficit. Submitted to *Journal of Experimental Botany.*

Sadok W, B. Boussuge, C. Welcker, F. Tardieu. A modelling approach to genotype x environment interaction: genetic analysis of the response of maize growth to environmental conditions. To be published in a book out of the Wageningen conference.

16. Isolation and Characterisation of Aluminium 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
1) Sorghum

In the first year of the grant (2005), we cloned a major aluminium (Al) tolerance gene from sorghum, *Alt$_{SB}$* via high resolution mapping of a 2300 F2 population generated from a cross between SC283, a standard line for sorghum Al tolerance, and BR007, which is a very Al sensitive sorghum line. The gene is a member of the MATE (Multidrug and Toxin Efflux) family of membrane transporters and appears to encode a novel citrate efflux transporter that is activated by Al (we also have shown that root Al-activated citrate efflux is the mechanism of sorghum Al tolerance). In 2006, we have made significant progress on verifying that this gene underlies the *Alt$_{SB}$* locus, and have used further high resolution mapping and allelic variation at the locus to elaborate a list of the sequence polymorphisms that could be associated with the Al tolerance phenotype.

a) Verification that the MATE gene is the sorghum Al tolerance gene Alt$_{SB}$

We are in the process of expressing the sorghum MATE gene in stably transformed Al sensitive lines of sorghum, maize and wheat, using constructs in which the MATE gene is driven by the ubiquitin promoter. The maize and sorghum transformation is being done at Embrapa Maize and Sorghum, and the wheat transformation work is being conducted in collaboration with Dr. Harold Trick of Kansas State University. Because cereal transformation requires some time, we expressed *Alt$_{SB}$* in Arabidopsis behind the cauliflower mosaic virus CaMV 35S promoter and have scored homozygous T3 lines for Al tolerance, in order to more quickly assess the functionality of *Alt$_{SB}$*. This involves *Alt$_{SB}$* expression in the Columbia background, both in a T-DNA knockout of the wheat *ALMT1* gene - an extremely Al sensitive line that shows essentially no Al-activated organic acid release - and in the wild type Columbia. We have recently shown that this *AtALMT* gene is responsible for the Al-activated root malate exudation underlying *Arabidopsis* Al tolerance (Hoekenga et al, 2006). We have obtained a number of homozygous T3 lines in both backgrounds where expression of *Alt$_{SB}$* has conferred a significant increase in Al tolerance. Slide #1 of the Powerpoint file in the Appendix shows the best performing lines in both the wild type and *AtALMT1* knockout backgrounds. As can be seen in slide 1, root growth of wild type Col *Arabidopsis* is inhibited nearly 50% by the Al treatment, while the expression of our MATE gene in the wildtype background totally restored root growth (actually led to a stimulation of root growth with a relative root growth of 113%). The *AtALMT1* knockout is very Al sensitive, with the same root treatment inhibiting root growth by nearly 90%, while root growth in the T3 line expressing the *MATE* gene in this background was only inhibited 60% by the same Al treatment.

We also verified that the sorghum *MATE* gene indeed is *Alt$_{SB}$*, by quantifying its expression in root tips (the site of Al tolerance and toxicity) of tolerant *versus* sensitive near isogenic lines (NILs) of sorghum. Quantitative real-time PCR analysis of candidate gene expression indicates that *Alt$_{SB}$* is expressed strongly in the root tip of the Al tolerant NIL, while its expression is not detectable in the root tip of the sensitive NIL (Slide #2). When a similar expression analysis was conducted for the other two positional candidates, it was found that both are expressed exclusively in shoots (data not shown). Furthermore, expression of the *MATE* is Al-inducible, and the degree of inducibility increases strongly with longer periods of root exposure to Al (Slide #2). This correlates nicely with our recent findings regarding the physiology of sorghum Al tolerance. As seen in slide #3, sorghum Al tolerance is induced by in the tolerant genotypes, in that after 1 day in Al, root growth is inhibited by approx. 55%, while by day three, Al tolerance increases moderately (40% inhibition). By 6 days in Al there is a strong increase in tolerance, as no inhibition of root growth by Al is observed. Also, we have determined that the likely mechanism of Al tolerance is Al-activated root citrate exudation, and this citrate exudation is also induced by Al. That is, as seen in slide #3, the rate of Al-activated citrate release is low after 1 day in Al, and

increases 4- and 10-fold, respectively, after 3 and 6 days of Al exposure (Magalhaes et al, in preparation). These findings strongly suggest that our candidate *MATE* gene for the Alt_{SB} locus encodes a root citrate efflux transporter that is Al-inducible at the level of gene transcription, and is also Al-activated at the level of protein function (transport).

b) Analysis of how Alt_{SB} (MATE gene) confers differential Al tolerance

The data in slide #2 suggests that differences in Alt_{SB} expression play a key role in controlling differential sorghum Al tolerance. One of our approaches is to mine genetic diversity for Al tolerance using a diverse panel of sorghum lines to look for variations in phenotype (Al tolerance, Alt_{SB} expression, root citrate exudation) and how this variation correlates with sequence differences in and around the Alt_{SB} locus. This started by the assembly of a small (12 sorghum line) diversity panel shown in Slide #4. As seen in this slide, a detailed analysis of Al tolerance (relative root growth under +/-Al conditions) suggests that a number of alleles for Alt_{SB} exist even in this small panel. Also, one of the panel members, BR0012, was used as a common tester line and was crossed with the other 11 members of this panel. For each cross, F2 populations were generated, and the linkage between Al tolerance and our candidate Alt_{SB} (*MATE*) gene was analysed with STS markers flanking Alt_{SB} at 0.6 cM. As seen in Slide #4, for most of the lines, variation in Al tolerance is explained primarily by the Alt_{SB} locus. However, in 2 lines, Al tolerance is not linked to Alt_{SB} and this finding indicates that at least one more Al tolerance gene exists in sorghum (Caniato et al, 2006); one of our future goals will be to identify this gene or genes.

This diversity panel composed of 12 lines was scored not only for Al tolerance, but also for Alt_{SB} expression and Al-activated root citrate exudation. Data for these 3 Al tolerance-related parameters for 5 of the members of this panel is shown in Slide #5. As seen in this slide, there is a very good correlation between Al tolerance, Alt_{SB} expression, and root citrate exudation. This correlation holds up for the entire 12 member panel, and this analysis is the focus of a manuscript in preparation (Lana et al., 2006).

To further this analysis, a somewhat larger, 42 accession diversity panel that captures a wide range of diversity with regards to Al tolerance was assembled, genotyped with SSR markers and phenotyped for Al tolerance. With regards to sequence analysis, we first sequenced the cDNA clone for our candidate gene from both parents of our mapping population as well as from another Al sensitive sorghum line (the source of our BAC library), and found no sequence differences in the coding region. This fits with our findings that indicate that differences in gene expression and not protein function are responsible for a significant portion of the differential Al tolerance conferred by Alt_{SB}. Next, based upon the rationale that functional Alt_{SB} sequences are expected to be polymorphic between the alleles from the tolerant and sensitive parents and are also necessarily located between our closest flanking markers, we sequenced fragments including flanking 5' and 3' sequence, from the tolerant and sensitive parents. Several polymorphisms were found in the second intron of the gene as well as in the 3' UTR; however none of the polymorphisms found so far correlated with differences in Al tolerance in the diversity panel. Upon review of the genomic sequence for the BAC sequence, we observed a MITE-type transposable element (miniature inverted-repeat transposable element) inserted ~2 kb upstream from the translation start site for Alt_{SB}. We designed primers flanking this repeated region, and found that its size was highly variable across members of the diversity panel. This region varies in size from about 400 bp in the most sensitive lines to around 2 kb in the most tolerant one. We now have sequenced through the entire region between the translation start site and the closest flanking marker, and have found for the promoter region we have sequenced, that the highly variable MITE insertion is the only polymorphic sequence between the two parents. As seen in Slide #6, there is a significant correlation between the size of this MITE region and Al tolerance. This suggests that repeated sequences in this region may be acting as an enhancer of Alt_{SB} gene

expression. There are precedents for repeated sequences such as transposons, including MITE's, for serving as gene promoters or enhancers of plant gene expression (Bercury et al. 2001; Selinger and Chandler, 1999; Yang et al., 2001; 2005). However, it should also be noted that although the size variation of this region explains a significant fraction of the variation observed in Al tolerance ($r^2 = 0.57$), there is variation in tolerance within the size classes, particularly for the two larger allele sizes (see Slide #6). Hence other regions of the *Alt$_{SB}$* promoter, or even sequence variation within the coding region of some genotypes, may also contribute to the wide variation of observed Al tolerance in a diverse collection of sorghum lines. In addition, a more conclusive view of the role of the MITE on gene expression depends on the level of linkage disequilibrium (LD) throughout the *Alt$_{SB}$* region. Such a complete LD study is the focus of another grant pre-proposal that was submitted to the GCP 2006 call for proposals.

To further analyse the structure of this MITE, and see if the larger MITE insertions could involve larger numbers of repeats that could act as enhancers of *Alt$_{SB}$* gene expression, we selected one line as representative of each of the four MITE insertion size classes depicted in Slide #6, and sequenced the MITE insertion region. As shown in Slide #7, as one goes from the smallest MITE insertion to the largest, which correlates with increasing Al tolerance and *Alt$_{SB}$* gene expression, there is a an expansion of the ABC cluster that forms the MITE. The basic repeat consists of a 243 bp MITE with 100 bp of sorghum sequence 5' of the MITE (in relation to the *Alt$_{SB}$* transcription start site) and 21 bp of sorghum sequence 3' of the MITE, towards the gene stop codon. This region is not repeated in the most Al sensitive line, and then is repeated 3, 4, and 5 times in representatives from the 3 larger MITE insertion size classes. This analysis has identified one region we will focus on for functional analysis of the *Alt$_{SB}$* promoter, to determine whether this MITE region plays a role in increased *Alt$_{SB}$* gene expression as well as Al-inducible expression. In the future, we also will sequence through this region as well as sequence 3' of the *Alt$_{SB}$* gene in one of several possible sorghum diversity panels, to look for other sequences that may be involved in variation in either *Alt$_{SB}$* gene expression or Alt$_{SB}$ protein function (citrate transport).

2) Maize

We have two lines of investigation to identify genes and processes important for Al tolerance. First, we are using association analysis, a statistical genetic approach, to correlate particular nucleotide polymorphisms with significant differences in Al tolerance between inbred lines. Second, we are using quantitative trait locus analysis, together with an integrated genomic strategy, to identify important genes in two recombinant inbred populations.

Association analysis requires three datasets in order to be employed: 1) trait data; 2) genotype data; 3) sufficient marker data from across the genome of the study population in order to evaluate true positive from false positive results (those due to population structure or kinship). We have phenotyped the 288 inbred line association panel assembled by co-PI Buckler using our standard hydroponic methodologies for quantifying root growth under +/-Al conditions. After five repeated experiments, our estimate for the heritability of net seminal root growth (the amount of root growth that occurs during a 2-day stress treatment) is 0.65. We are using the latest form of mixed model ANOVA developed by co-PI Buckler and colleagues, which was recently published in *Nature Genetics* 38:203-208. We are employing a set of 500 SNP from across the maize genome to empirically calculate the significance threshold for Al tolerance. Our attention has principally been upon describing the maize gene families related to *ALMT1*, the Al-activated malate transporter found to be the primary Al tolerance gene in wheat, and *Alt$_{SB}$*, the Al-activated citrate transporter from sorghum described above. While maize does not yet have a completely sequenced genome, rice does and thus provides the opportunity to estimate the size of *ALMT1* and *Alt$_{SB}$*-related families in maize. Rice contains 13 *ALMT1*-like genes and 7 MATE transporters similar to *Alt$_{SB}$*. Using genome survey sequences (GSS), we can estimate that maize contains at

least 15 *ALMT1*-like genes, representing 9 of the 13 rice homologs, and at least 10 *Alt$_{SB}$*-like genes, representing 4 of the 7 rice loci. Based on the GSS contigs, we have cloned, sequenced and genetically mapped 9 *ZmALMT* and 1 *ZmASL* genes and analysed these results using association analysis. In addition, we have identified 4 additional genes from microarray data and established their possible relevance to Al tolerance via association analysis. These results are summarised below:

Gene identifier	Chromosome	Genetic Position	Significant?
ZmALMT-65133	9	212.61	No
ZmALMT-160590	10	323.11	No
ZmALMT-7313	10	234.21	No
ZmALMT-51459	5	421.81	No
ZmALMT-158506	5	427.81	No
ZmALMT-61408	5	470.51	No
ZmALMT-84410	2	273.71	No
ZmALMT-124219	10	211.51	Yes
ZmASL-49968	**?**	**?**	**Yes**
Malate dehydrogenase*	6	243.91	Yes
Isocitrate lyase	7	363.71	Yes
Pectin methylesterase	*1*	*295.71*	*Yes*
SAH hydrolase*	***4***	***225.71***	***Yes***

The genes marked with * were mentioned in passing in the last progress report.

Nineteen genes have been assessed to date, with six containing polymorphisms putatively associated with Al tolerance. The genes in standard text presumably represent components of organic acid-dependent Al tolerance pathways. The two genes in italics represent cell wall composition-dependent Al tolerance pathways, which provide novel evidence for this hypothesized tolerance pathway. The importance of two genes (**bold)** has been confirmed with linkage populations, selecting F_2 crosses between parents known to differ in the alleles detected. The superior allele of *ZmASL-49968* carried by the B73 variety is completely dominant to the sensitive allele carried by CML247. The superior allele of the S-adenosylhomocysteine (SAH) hydrolase gene (a primary metabolic gene) carried by B73 is completely recessive to the inferior allele carried by CML247. The validity of the other four genes is presently being assessed. We expect to test additional genes via association analysis in the coming year. These six genes can account for ~3% of the variance observed, such that we expect there are many genes yet to discover.

We have also made progress towards the construction of NIL that represent each of the QTL detected in the B73 x Mo17 population (Cornell) and the L53 x Al237 population (Embrapa). We anticipate that gene expression profiling via microarray hybridisation of these NIL, together with advances in genome sequencing of maize will greatly facilitate candidate gene selection. Given the present size of the QTL intervals, these will require further refinement before we can pick candidate genes based on their genomic locations.

Tangible outputs delivered:

- Elite Al tolerant sorghum hybrids developed from the breeding programme
- Identification of an Al tolerance candidate gene in sorghum
- Identification of preliminary candidate genes for maize Al tolerance
- Manuscripts in Preparation:
    - Magalhaes JM, Liu J, Guimarães CT, Alves VM, Lana U, Hoekenga OA, Wang Y-H, Coelho C, Schaffert RE, Klein PE, and LV Kochian. Positional cloning and

characterisation of *Alt$_{SB}$*: A novel membrane transporter and major aluminum tolerance gene in *Sorghum bicolor*. (In preparation)

- Caniato FF, Guimarães CT, Schaffert RE, Alves VMC, Kochian LV, Borém A, Klein PE, Magalhaes JV. Genetic diversity for aluminum tolerance in sorghum (in preparation).
- Lana UGP, Alves VMC, Guimaraes CT, Liu J, Kochian LV, Schaffert RE, Shaff, JE;, Magalhaes JV. Aluminum tolerance in sorghum is Al inducible and is due to root citrate exudation that also is Al-inducible (in preparation).
- Cançado GMA, Piñeros MA, Maron LG, Alves VM, Camargo SR, Menossi M, and LV Kochian. The ALMT1 homolog in maize is an anion efflux transporter but is not involved in maize Al tolerance (In preparation).

Deviations from the work plan:
As the GCP is aware, the process of setting up the subcontracts between Cornell and the other partners suffered a significant delay. While Embrapa has not yet received any funds for this project, this is expected within the next weeks. However, there are still very serious problems with the type of subcontract that was signed between Cornell and Moi University. Dr. Samuel Gudu (Moi University) will get in touch directly with the GCP director explaining these problems and requesting the GCP support to solve them. As a result of this, while the Embrapa Maize and Sorghum part of this project has been undertaken with some remaining funds from other sources, this is not the case for Embrapa Wheat, Embrapa Rice and Beans and Moi University. Embrapa Wheat and Embrapa Rice and Bean will start their activities after receiving the funds for year one. Moi University needs to adjust details of their subcontracts before being able to start their activities.

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, ICARDA
Stefania Grando, ICARDA
Sripada Udupa, ICARDA
Maria van Korff, ICARDA
Wafaa Choumane, Tishreen University, Syria.

Mid-Year Report

1. Determination of drought stress conditions for barley
We developed a greenhouse test (pot experiment) for the evaluation of drought tolerance for barley. The test is based on the available water in the soil and some traits (chlorophyll content and fluorescence parameters, osmotic adjustment, no. of available tillers, yield etc.) of barley.

Test for available water content: Plants of the crosses and reciprocal crosses were subjected to 70% available water content (AWC) in the soil as control (well-watered), 35% AWC (moderate drought stress) and 10% of AWC (severe drought stress conditions).

Experimental protocol:
1. Weight experimental soil (1 Kg)

2. Bake the soil and get the water content (WC) in the soil.
3. The field capacity (FC) and wilting point (WP) for the soil are determined. Then the AWC (available water content)= FC-WP are calculated.
4. The water amount for different treatments = soil WC + soil weight x AWC x 0.7 or 0.35 or 0.1,respectively.
5. Weight for experiment = pot weight + water amount + soil weight.

2. Barley plants for allelic imbalance assay
The following material was planted for the allelic imbalance assay
Control: 2 parents.
75% water content, 3 pots, 3 F_1 plants/pot, total F1 plants planted: 9
Drought stress
10% water content, 3 pots, 3 F_1 plants/pot, total F1 plants planted: 9
35% water content, 3 pots, 3 F_1 plants/pot, total F1 plants planted 9

We have planted 2 sets for each treatment (vegetative, generative).

2. Analysis with SNP primers:
SNPs for 12 genes were identified in the Oregon Wolfe barley parental lines (Appendix I). Primers detecting these SNPs were developed and used for SNP detection in the other barley germplasm lines. DNA from genotypes used as parents in 3 different crosses were tested with these 12 SNP primers. Six primers were able to detect SNPs in the genotypes used for the crosses. The other six primers (SNP 2901- SNP 86 – SNP 1293 –SNP824 – SNP2297 – SNP 957) were not able to detect SNPs in the analysed genotypes.

3. Sequencing of candidate genes and SNP primer development
50 barley genes were sequenced in 8 barley lines (Arta, H. spontaneum, Tadmor, Oregon Wolfe barley dom and rec,WI3408, Sloop, Alexis) (See attachment). Selection was based on data available at http://germinate.scri.ac.uk/barley_snpdb/dbStats_contig.html. Genes showed between 4 and 26 SNPs. Genes include nitrate stress genes and others. Primers will be designed and tested for the imbalance assay.

4. Plant material used for allelic imbalance assay:

Crosses	**Reciprocal Crosses**
Tadmor x *H.spontaneum*	*H.spontaneum* x Tadmor
Tadmor x Arta	Arta x Tadmor
Tadmor x Sloop	Sloop x Tadmor
Tadmor x WI3408	WI 3408x Tadmor
H.Spontaneum x Sloop	Sloop x *H.Spontaneum*
Sloop x WI3408	WI3408x Sloop

The crosses were planted under controlled environmental conditions and subjected to drought stress during the vegetative and generative stages. Leave tissues were collected as follows:
The vegetative stage: the second leaf from the top was collected (after 45 days) from the control and the stressed plants.

The generative stage: The second extended leaf from the top was collected from the control and the stressed plants.
For root tissues: The root of the vegetative and the generative stages were collected from the cross Tadmor x Sloop and Sloop x Tadmor cross.

For leaves: For each treatment, nine plants were planted, 3 plants/pot.

For RNA extraction: leaves from three individual plants were collected separately but mixed for RNA extraction.

For DNA extraction: Leaves from individual plants were collected. Tissues for RNA extraction were collected and conserved in liquid nitrogen. Total RNA was extracted with Promega SV total RNA isolation system. mRNA isolation was performed using Qligotex mRNA Mini kit. cDNA was prepared using Promega reverse Transcription System.

DNA was extracted from individual plants (Parents and hybrids) with CTAB protocol. Allelic imbalance assay was performed after confirming the absence of contamination by DNA. Based on the sequence of dehydrin 12 gene, 2 different SNP primers were assigned (Exon I and Exon II) and used in the analysis of the crosses Tadmor x Sloop and its reciprocal. For this cross the SNP was G/C in Tadmor and C/G in Sloop with Exon I, and C/G for Tadmor and A/T in Sloop for Exon II. The results suggest allelic imbalance between genotypes, in favour of Tadmor, which was not caused by imprinting but by some regulatory variation in cis. However, it also shows that this imbalance is not influenced by water stress. The analysis for dehydrin 12 need to be repeated to confirm allelic imbalance. Other genes will be tested with the assay. Such information should be available for the annual research meeting.

5. QTL analysis of chlorophyll and chlorophyll fluorescence parameters in Arta/Hspontaneum mapping populations

We will use the Arta/h.sponatenum mapping population to map genes that demonstrated allelic imbalance. For the RIL mapping populations we have already evaluated and determined QTLs for a number of agronomic traits. Here we add a number of photosynthesis related traits.

QTL analysis was performed on the basis of the marker linkage map constructed the Arta x *H. spontaneum* 41-1 RIL population, which contained 158 AFLP markers, 30 SSR markers and one morphological marker, covering a total map length of 890 cM. QTLs for each of the five traits in both well-watered and drought stress conditions were identified through MQM analysis using the MapQTL 5 programme.

Under well-watered conditions, genetic analysis detected fourteen significant QTLs at eight loci for the five traits. Three QTLs with high LOD scores and R^2 values were located on chromosome 2H at about 1 cM for F_m, F_v and F_v/F_m, at about 30 cM for F_o, F_v/F_m and chlorophyll content, and at about 115 cM for chlorophyll content. Beside these three loci, additional QTLs were identified on chromosomes 1H at 76 cM for F_o, 4H at 33 cM for F_m and 82 cM for chlorophyll content, 5H at 14 cM for F_m, F_v and F_v/F_m, and 6H at 98 cM for F_v.

The QTL analysis of the parameters collected on plants grown under drought conditions detected eleven significant QTLs located at five loci. A major locus was identified on chromosome 2H, accounting for a high proportion of phenotypic variance for F_o, F_m, F_v and F_v/F_m. This locus was also identified for chlorophyll content, with an $R^2 = 9.9\%$. Another locus involved in the regulation of F_m, F_v and F_v/F_m was also detected on chromosome 2H (135cM). Additional QTLs were detected on chromosomes 4H, 6H and 7H for F_o. In contrast to the situation under well-watered conditions where the contribution of the two parents to the expression of five traits was similar, under drought stress conditions Arta contributed the alleles for the traits F_m, F_v, F_v/F_m and chlorophyll content, while *H. spontaneum* 41-1 contributed the alleles for F_o thus confirming the observations at the phenotypic level. Furthermore, the major locus on chromosome 2H was also detected for chlorophyll content in plants grown under well-watered conditions, suggesting a constitutive expression for this trait.

Outputs achieved so far:

1. Experimental protocol for allelic imbalance in maize adapted to barley available now in Udine, NIAB and ICARDA.
2. Primer detecting SNPs in 12 genes were developed
3. Parental lines (8) were sequenced for 50 number of genes
4. Leave tissue from drought stressed barley plants grown under controlled environmental conditions was harvested and preserved and is available for allelic imbalance assay at ICARDA, Udine and NIAB.
5. Drought related traits mapped in Arta/H.spontaneum

Next training session on allelic imbalance is planned as part of a training course 28 of May – 8 of June 2006 at ICARDA.

SP1 COMMISSIONED GRANTS

2004—SP1CL2: SSR Analysis of Musa Germplasm

Principal Investigator:
Sarah Hearne, IITA
Collaborating Scientists:
Chris Town, TIGR
Morag Ferguson, IITA

Mid-Year Report

RNA was extracted from pools of root, leaf and stem meristem tissue using Invitrogen's concert RNA reagent following manufacturer's protocol. The pools of tissue comprised equal weights of tissue from each of: drought stressed and non-stressed musa lines, Sukari Ndizi = local name for 'apple banana' = AB, Pisang Awak = locally called 'Kayinja' = ABB, Mpologoma = Matooke banana from the Musakala clone set = AAA-EA. Sukari Ndizi is a cultivar which is reported to have a good physiological adaptation to drought. Pisang Awak is generally considered to be a cultivar that can withstand abiotic stresses fairly well (drought, poor soil). Mpologoma has a high yield potential and harvest index, but is considered (by farmers) to be sensitive to abiotic stresses (drought, fertility) and is probably the cultivar that is considered most drought sensitive within the highland banana group.

All RNAs extracted were evaluated for quantity and quality in Nairobi before being sent to Evrogen for cDNA syntheses and normalisation. Two RNA samples were sent to Evrogen; root RNA and a 1:1 quantity pool of leaf and stem meristem RNA. Samples were sent to Evrogen as per their instructions (samples precipitated with DEPC treated NaOAc and ethanol, shipped at ambient temperature). RNA quality was checked by Evrogen before cDNA synthesis (Fig 1 a and b).

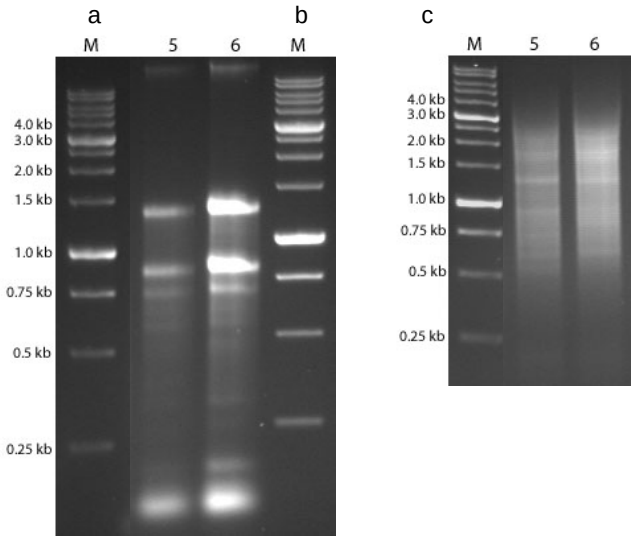


Figure 1. a) Evrogen's analysis of RNA from 5 musa leaf/meristem and 6, root; b) Evrogen's analysis of cDNAs synthesised from RNA from 5 musa leaf/meristem and 6, root.

Evrogen have completed the normalisation of both the musa leaf/stem meristem and root libraries. Ligation mixes of the normalised cDNAs will be will be shipped from Evrogen to TIGR within the next 10 days. TIGR is awaiting the cDNAs and will initiate the sequencing effort upon receipt of the cDNAs.

Tangible outputs delivered:
Two normalised cDNA pools from; drought stressed and non-stressed musa leaf and shoot meristematic tissue from drought tolerant and susceptible musa lines and drought stressed and non-stressed musa root tissue from drought tolerant and susceptible cowpea lines.

Deviations from the work plan:
The process of plant growth, stress and RNA isolation took longer than anticipated for various reasons (detailed previously). However, all issues leading to this have now been addressed.

Due to the time delay in obtaining RNA we request a NCE to November to allow TIGR to receive and complete sequencing of the libraries.

2005-01a: Completing Genotyping of Composite Germplasm Set of Barley

Principal Investigator:
J. Valkoun, ICARDA
S. Grando, ICARDA
M. Baum, ICARDA
Collaborating Scientists:
Zhang Jing,, CAAS
Joanne Russell, Scottish Crop Research Institute

Mid-Year Report
The seeds of 2692 Accessions, representing a barley collection collected from 84 countries, were provided by the International Centre for Agricultural Research in the Dry Areas (ICARDA) at Aleppo in Syria and used in this analysis.

We have received data from CAAS on the 35 SSRs loci recently, but were not able to run the analysis prior to this report. The data from CAAS needs to be completed. 500 accessions genotyped by CAAS in the first year, need to be adapted to the data format and included in the analysis.

Using Structure2.1 programme (Pritchard et al. 2000), we have analysed 15 SSR loci data on 2692 of the 3000 accessions of the barley composite collection. The analysis on 15 SSR primers detected a total of 310 alleles, ranging from 7 to 43 alleles with an average of 26.6 alleles per locus and mean PIC (polymorphic information content) value of 0.672362 (ranging from

0.372814 to 0.914002). SSR locus Bmac316, detected the greatest number of alleles (43 alleles) while HVHVAI detected only 7 alleles (see table 1). Detailed analysis for each locus is available and will be incorporated into the database (see SCSSR3907as example table 2).

Rare alleles are the one, which are present in ≤ 5% in the population. When 310 alleles detected in the composite collection were grouped into rare or common allele category, there were 252 rare and 58 common alleles at 5% and 172 rare and 138 common alleles at 1%, respectively.

Tangible outputs delivered:

- 2692 Accessions were genotyped with 15 SSR markers
- Single plants from each accession were planted in a greenhouse and seed harvested for about 2000 accessions. In case where no seed was harvested plants were replanted.

Deviations from the work plan:

Nil

Table 1. Number of alleles and the gene diversity at 15 SSR loci determined

Loci	Chromosome	No. of alleles	Gene diversity
SCSSR15864	7H	21	0.579258834
SCSSR10148	5H	29	0.816028619
SCSSR2306	5H	11	0.671354834
SCSSR25691	3H	16	0.533274004
SCSSR2748	1H	9	0.668369822
SCSSR3907	7H	32	0.914002121
SCSSR5939	5H	17	0.372814448
SCSSR08447	2H	14	0.640349137
SCSSR7970	7H	43	0.665933793
BMAC18	6H	13	0.747358836
BMAC316	6H	43	0.858131925
BMAG382	2H	23	0.683359187
BMAG211	1H	20	0.818979644
HVHVAI	1H	7	0.552136722
HVLTPPB	3H	16	0.557650141

Table 2. Fequence of alleles on locus SCSSR3907

Alleles of Scssr03907	allele freq	%freq	freq2	Classification at 5%	Classification at 1%
3907-1	0.00235	0.2350176	0.0000055	Rare	Rare
3907-2	0.04172	4.1715629	0.0017402	Rare	Common
3907-3	0.03878	3.8777908	0.0015037	Rare	Common
3907-4	0.14689	14.688602	0.0215755	Common	Rare
3907-5	0.00646	0.6462985	0.0000418	Rare	Rare
3907-6	0.06052	6.0517039	0.0036623	Common	Common
3907-7	0.00999	0.9988249	0.0000998	Rare	Rare
3907-8	0.00823	0.8225617	0.0000677	Rare	Rare
3907-9	0.12338	12.338425	0.0152237	Common	Common
3907-10	0.14571	14.571093	0.0212317	Common	Common
3907-11	0.08284	8.2843713	0.0068631	Common	Common
3907-12	0.04113	4.1128085	0.0016915	Rare	Common
3907-13	0.07638	7.6380729	0.0058340	Common	Common
3907-14	0.05523	5.5229142	0.0030503	Common	Common
3907-15	0.03173	3.1727380	0.0010066	Rare	Common
3907-16	0.02879	2.8789659	0.0008288	Rare	Common
3907-17	0.01469	1.4688602	0.0002158	Rare	Common
3907-18	0.01116	1.1163337	0.0001246	Rare	Common
3907-19	0.02174	2.1739130	0.0004726	Rare	Common
3907-20	0.01116	1.1163337	0.0001246	Rare	Common
3907-21	0.01351	1.3513514	0.0001826	Rare	Common
3907-22	0.01528	1.5276146	0.0002334	Rare	Common
3907-23	0.00588	0.5875441	0.0000345	Rare	Rare
3907-24	0.00940	0.9400705	0.0000884	Rare	Rare
3907-25	0.00588	0.5875441	0.0000345	Rare	Rare
3907-26	0.00529	0.5287897	0.0000280	Rare	Rare
3907-27	0.00353	0.3525264	0.0000124	Rare	Rare
3907-28	0.00235	0.2350176	0.0000055	Rare	Rare
3907-29	0.00235	0.2350176	0.0000055	Rare	Rare
3907-30	0.00294	0.2937720	0.0000086	Rare	Rare
3907-31	0.00059	0.0587544	0.0000003	Rare	Rare
3907-32	0.00059	0.0587544	0.0000003	Rare	Rare
Gene diversity					0.9140020

2005-01b: Completing Genotyping of Composite Germplasm Set of Wheat

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborating Scientist:
Nachit Miloudi, ICARDA

Mid-Year Report

Deliverables

1. Progress (provide specific results from funded activities during funding period; max 300 words):

1. At CIMMYT 400 putatively Mexican landraces, 2300 Iranian landraces and 400 new synthetics were grown under drought conditions in trials in spring, 2004 and 2005 (two separate trials). The drought nursery is planted each year in Ciudad Obregon, Mexico, where winter rainfall is usually exceedingly low. Unfortunately, the dry season of 2004 – 2005 was a year of record rainfall, and no useful data was gathered from this drought trial, as essentially every entry was well watered by excess rainfall.
2. The 200 most drought resistant landraces that had been analysed in 2004 were chosen for the genotyping at CIMMYT. 250 lines were identified by Nachit Miloudi of

ICARDA for genotyping as well. Please see attached table 1 for a list of the final 450 accessions to be genotyped.

3. The DNA from the 450 drought resistant wheat populations has been exchanged for genotyping

4. Genotyping with 50 SSR loci has been started at CIMMYT and ICARDA, and gel analysis is ongoing. M. Nachit and M. Warburton have agreed that the best course of action would be to send an ICARDA technician to CIMMYT for a week's worth of training, since it would be very difficult to teach these things via email or over the phone, but there are no funds left to pay for this trip. In addition to finishing the data generation, this trip would have the added benefit of capacity building, as well as building a firmer alliance between the wheat genotyping labs of both centres. A capacity building grant has been submitted to the GCP for Mr. Ahmed Alsalah, ICARDA technician, to spend two weeks at CIMMYT learning the technique.

2005-01c: Completing Genotyping of Composite Germplasm Set of Sorghum

Principal Investigator:
C Tom Hash, ICRISAT
Collaboratoring Scientists:
Hari D Upadhyaya, ICRISAT
Punna Ramu, ICRISAT
Claire Billot, Agropolis
Monica Deu, Agropolis
Jean-François Rami, Agropolis
Jacques Chantereau, Agropolis
Yu Li, CAAS

Mid-Year Report

SSR marker data generation for approximately 2970 sorghum accessions was completed at Cirad and ICRISAT. Fresh DNA samples of the initial 670 accessions genotyped in 2004 were prepared at ICRISAT and dispatched to CAAS, so that the data set previously generated at CAAS could be redone with the agreed panel of standard entries included in each run. Due to problems with the sequencer at CAAS, it has not yet been possible to produce an acceptable data set—a preliminary revised data set provided in early May 2006 still has an unacceptably large number of missing data points, an unreasonably large proportion of heterozygotes at several loci, and allele calls well outside the expected range of allele sizes for the markers used. Although CAAS has received approximately 33% of the funds allocated for this project, it has yet to contribute sorghum SSR marker data that are considered sufficiently reliable by the Cirad and ICRISAT team members.

Raw data sets generated at Cirad and ICRISAT during 2005 were passed through the AlleloBin programme for allele calling, which was developed at ICRISAT during 2005. After this procedure, a very small number of markers exhibited unexpected allele distribution patterns. These were reviewed and in a few instances manual corrections made where it was apparent that the sequencer was mis-calling allele sizes at the extreme of the expected allele size distribution. One of the 43 SSR markers for which complete data sets are currently available exhibits an unexpectedly large number of putative heterozygotes. This indicates that this marker data is likely from duplicate loci that are not ideally suited for use in genetic diversity analyses, so this marker has been dropped from the diversity analysis.

Claire Billot from Cirad visited ICRISAT-Patancheru for a week in April 2006 for finalising the analysis of the sorghum composite collection marker data sets generated to date by Cirad and ICRISAT. A preliminary analysis was completed, the DARwin software package was demonstrated to the ICRISAT team, and a list of planned publications prepared. Based on the

preliminary analysis, the landrace accessions included in the sorghum composite collection were grouped into several clusters. Many of these clusters are comprised largely of a single race and its associated hybrid races. The uniqueness of the *margaritiferum* sub-group within the guinea race continues to be supported. We now expect to contribute a final report with information from 3391 accessions x 42 markers by the end of June 2006 (including an additional 421 accessions and an updated diversity analysis).

Tangible outputs delivered:

- Generation of markers data for 2970 accessions x 42 markers was completed.
- Preliminary diversity analysis was done at ICRISAT during the second week of April 2006. This gives a clear distribution of clusters for each of the five sorghum races.
- DNA from an additional 421 accessions was isolated and prepared for export to Cirad. Marker data for these accessions will be ready by early June and the complete data analysis will be available by the end of June 2006.

Deviations from the work plan:

Portions of the sorghum marker data set that were to have been generated at CAAS have not been completed to the required standard. In the first year of this project, it was found that the marker data set generated at CAAS could not reliably be merged with data sets generated at Cirad and ICRISAT. For 2005, CAAS was requested to repeat its data generation for the 670 sorghum accessions genotyped in 2004 and to include the agreed panel of standard entries in each run. For the second year of this study, CAAS was also assigned a small portion of the total marker data generation (7 of 50 loci)—but agreed to initiate this only after demonstrating that it could produce a satisfactory data set for the initial subset of 670 accessions. As the latter has not yet been completed, CAAS has not yet initiated genotyping of the entire sorghum composite germplasm set for the 7 SSR markers assigned to it. The markers assigned to CAAS are in bins covered by markers generated by Cirad or ICRISAT, so failure to complete this will not seriously affect genome coverage offered by the reduce number of markers (42, excluding data for one SSR primer pair that detected duplicate loci) that we will ultimately be able to use for diversity analysis of the sorghum composite collection.

As there has been no breakthrough in the impasse regarding 250 entries contributed by CAAS (permission to export seed samples to Cirad and ICRISAT has been pending for over a year), and the marker data generated from DNA samples of these accessions indicates that they represent a very narrow range of diversity that is included among the samples contributed by ICRISAT and Cirad, it was decided in April 2006 to replace these entries with an additional 421 genotypes identified by the ICRISAT gene bank curator. Seed of these additional entries was sown and DNA extraction has just been completed at ICRISAT-Patancheru. DNA sample distribution will be completed this next week, with DNA marker data generation for these additional entries to be completed in June 2006.

2005-01d: Completing Genotyping of Composite Germplasm Set of Chickpea

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaboratoring Scientists:
SL Dwivedi, ICRISAT
DA Hoisington, ICRISAT
PM Gaur, ICRISAT
CLL Gowda, ICRISAT
RK Varshney, ICRISAT
S Chandra, ICRISAT

SM Udupa, ICARDA
BJ Furman, ICARDA
M Baum, ICARDA

Mid-Year Report

All allelic data for 35 SSR primers has been produced by ICRISAT. Missing data (i.e., marker x genotype) is less than 5%. The dataset has been analysed using the allele binning algorithm of Idury and Cardon (1997) in the C programme "Allelobin" written at ICRISAT. Except for TA21, TA28, and TAA58, all markers produced an allele size expected on the basis of SSR repeat motif. Final data from ICARDA on 15 SSRs loci is pending.

Using DARwin 5.0 Structure programme (Perrier et al. 2003), we have analysed the 35 SSR loci data for 2904 of the 3000 accessions of the chickpea composite collection (Upadhyaya et al. 2006). The analysis detected a total of 1116 alleles, ranging from 14 to 58 alleles with an average of 31.8 alleles per locus and mean PIC value of 0.84 (ranging from 0.47 to 0.96). SSR loci TA28, TR43, and TA22 detected greater number of alleles (50–58 alleles) than NCPGR4, NCPGR7, TR31, and TS84 (14–16 alleles). Although the average gene diversity was the same for desi and kabuli types, kabulis were more genetically diverse than the former. Similarly, accessions from the Mediterranean region revealed a high gene diversity (as detected by the range), while those from West Asia revealed lowest gene diversity (Table 1). Except for the marker TR2 that had high heterozygosity (0.80), the frequency of heterozygotes detected with the other markers ranged from 0.001 to 0.03.

A total of 602 rare (<=1% frequency in the population) and 514 common alleles were found. There were 121 alleles that were detected in all the four biological groups (desi, kabuli, pea-shaped and wild relatives), with 167 alleles specific to desi, 66 to kabuli, 52 to wild relatives and only 4 to pea-shaped chickpeas. When the composite collection was classified based on geographical information (based on country of origin), we detected 302 alleles that were common across the six regions (Africa, America, Euro-Russia, Mediterranean, South and South-East Asia and West Asia. Interestingly, accessions from the Mediterranean region had the largest number of region-specific alleles (90). This region is reported as one of the two primary centres of diversity in chickpea. The other primary centre of diversity is West Asia and accessions from this region had 57 region-specific alleles.

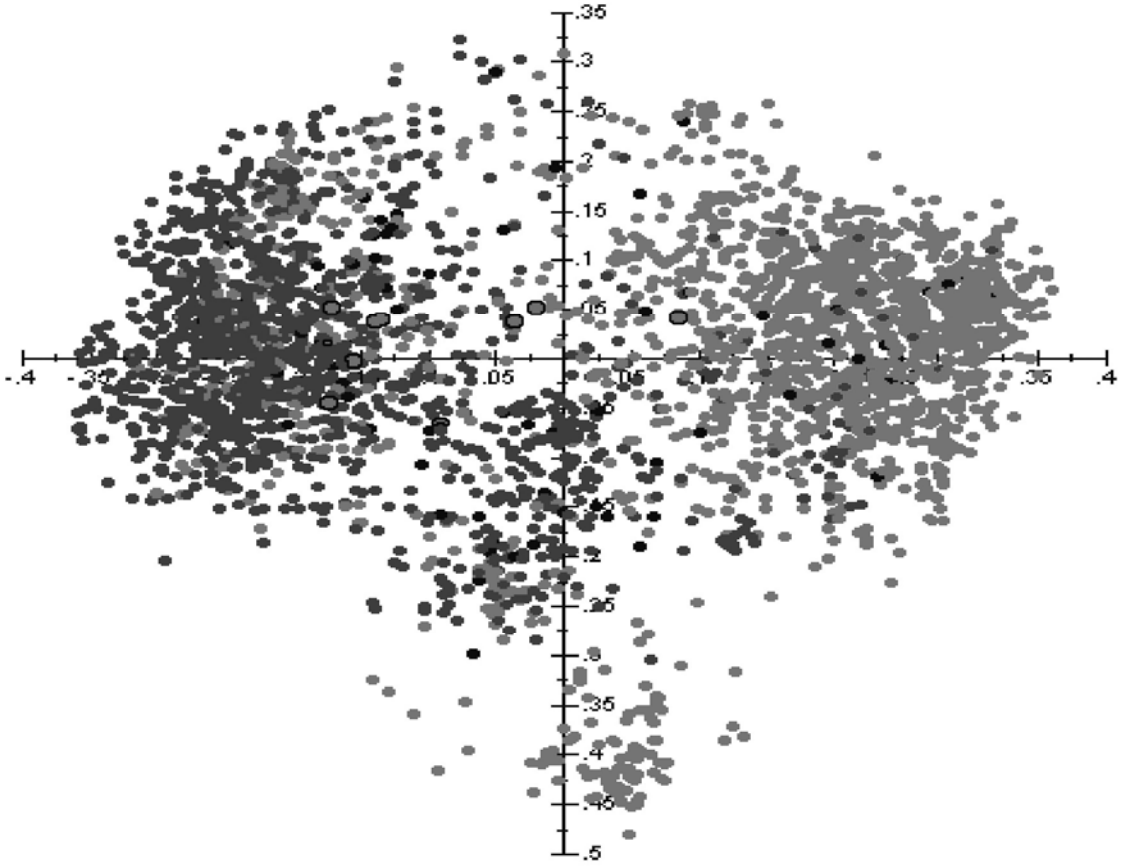
Principal coordinate analysis (PCoA) using DARwin 5.0 STRUCTURE revealed that both desi and kabuli chickpeas formed two distinct clusters; however, a number of desi chickpeas also grouped into the kabuli cluster indicating probable progressive evolution of kabuli traits from the desi chickpeas (see Figure 1).

Further data analyses are in progress to more fully understand the genetic structure of the chickpea accessions in this set. These analyses will be used, along with phenotypic information, to determine the reference set for future use in projects such as the recently funded ADOC project.

Finally, DNA allele sets are being developed to provide the major reference alleles for each SSR marker.

Table 1. Range and average gene diversity in composite collection of chickpea (number in parenthesis is the difference in the range).

Category	# Accessions	PIC Value	
		Average	**Range**
Composite collection	2904	0.837	0.468 – 0.962 (0.494)
Biological classification			
Desi chickpea	1660	0.814	0.379 - 0.954 (0.575)
Kabuli chickpea	1165	0.814	0.242 - 0.963 (0.721)
Pea-shaped chickpea	69	0.805	0.447 - 0.952 (0.505)
Wild species	10	0.758	0.346 – 0.880 (0.534)
Geographical classification			
South and South East Asia	1132	0.786	0.321 – 0.946 (0.622)
West Asia	719	0.835	0.449 – 0.959 (0.510)
Mediterranean	617	0.801	0.249 – 0.964 (0.715)
America	143	0.799	0.354 - 0.939 (0.585)
Africa	150	0.760	0.290 – 0.947 (0.657)
Euro-Russia	109	0.786	0.250 – 0.934 (0.684)
Unknown	34	0.815	0.475 – 0.935 (0.460)



Desi (1660 accessions) • Kabuli (1165 accessions)
Pea-shaped (69 accessions) • Wild species (10 accessions)

Figure 1. Genetic structure of the composite collection (35 SSR loci and 2904 chickpea accessions) of chickpea as revealed by the Principal Coordinate Analysis (PCoA) using DARwin STRUCTURE programme.

Tangible outputs delivered:
1. High quality allelic data for 35 SSR primers across 3000 chickpea genotypes.
2. Genetic structure of the composite collection.
3. Desi- and kabuli-specific alleles.
4. Region-specific alleles.

Deviations from the work plan:
Nil

2005-01e: Genotyping of Composite Germplasm Set, Tier 1, Maize

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborating Scientists:
Alain Charcosset, INRA
Mahalakshmi Visvanathan, IITA
Yu Li, CAAS
Susanne Dreisigacker, CIMMYT

Mid-Year Report

As part of it's participation in the Genetic Resources Challenge Programme "Unlocking Genetic Diversity for the Resource Poor", CIMMYT, IITA, CAAS, and INRA has committed to completing the structural characterisation of diversity of 987 inbred lines (587 from CIMMYT, 200 from CAAS, and 200 from IITA) and 467 maize landrace populations (216 from CIMMYT, 150 from IITA, and 101 from INRA) using 50 SSR markers. The germplasm was to represent the range of diversity of maize (particularly germplasm used in the developing world). A list of the entries and pertinent information can be found in Appendix Tables 1 (inbreds) and 2 (populations). From the final 1454 entries in the analysis, a core subset of 384 entries would be selected, to act as a mini-composite subset (core) for future studies needing a wide range of diversity. Because this core was selected based on neutral markers, it is not biased for any single trait, geographical area, or pedigree, and may be used to search for new variation or in association tests for any trait of interest. Although in other species, a target of 3000 entries was targeted for the core, half that number was agreed upon in maize because of the difficulties in characterising heterogeneous populations (discussed more later).

A discussion was held at the Plant and Animal Genome meeting in January, 2004, where suggestions were given as to what type of markers should be used in the structural characterisation of the composite genotype set; SSR markers were chosen as being the most advanced and advantageous marker type already optimised for use. It was determined that laboratories participating in the genotyping (CAAS, INRA and CIMMYT) would each run a subset of the markers on all of the germplasm, rather than try to run subsets of the germplasm in each lab with all of the markers; this ensured that problems of repeatability between labs was avoided. In an ensuing email conversation, representatives from IITA and CAAS determined the list of SSR markers that each lab would run, drawing on the past expertise of each of the laboratories in SSR fingerprinting. SSR markers for inbred lines and populations are listed in Tables 1 and 2. A total of 55 SSRs were run on the inbreds, in case there were many cases of missing data, in order to get the full 50 markers; however, more than expected gave poor results and fewer than 50 total were entered into the Central Repository (Table 3, below). Only 44 SSR markers could be optimised for the bulk analysis of populations, of which only 30 proved stable enough to be included in the Central Repository (Table 4, below).

In another email discussion, CAAS, IITA and CIMMYT determined the identities of the inbred lines that will be characterised in 2004; these lines can be found in Appendix Table 1. In an additional meeting held in INRA in France, the numbers of landrace populations of maize and the regions to be covered were determined, and CIMMYT is determining the identity of the landraces

at the moment. Breeder's populations were identified by CIMMYT breeders and are listed in Appendix Table 2.

Table 1. SSR primers to be run on the core subset of inbred lines. Primers in red overlap with the primers chosen for the maize populations as well.

SSRs to be used on the Inbred lines:

Primer	Lab	Repeat	Bin	Primer	Lab	Repeat	Bin
nc130	CIMMYT	Tri	5.00	phi233376	CIMMYT	Tri	8.09
nc133	CIMMYT	Penta	2.05	phi299852	CIMMYT	Tri	6.07
phi002	CIMMYT	Tetra	1.08	phi328175	CIMMYT	Tri	7.04
phi015	CIMMYT	Tetra	8.09	phi420701	CIMMYT	Tri	8.00
phi032	CIMMYT	Tetra	9.04	phi423796	CIMMYT	Penta	6.01
phi046	CIMMYT	Tetra	3.08	phi448880	CIMMYT	Tri	9.06
phi050	CIMMYT	Tetra	10.03	phi453121	CIMMYT	Tri	3.01
phi059	CIMMYT	Tri	10.02	phi96342	CIMMYT	Tetra	10.02
phi063	CIMMYT	Tetra	10.02	umc1109	CIMMYT	Tri	4.10
phi064	CIMMYT	Tetra	1.11	umc1122	CIMMYT	Tri	1.06
phi069	CIMMYT	Tri	7.05	umc1143	CIMMYT	Penta	10.02
phi072	CIMMYT	Tetra	4.01	umc1152	CIMMYT	Tetra	10.01
phi073	CIMMYT	Tri	3.05	umc1153	CIMMYT	Tri	5.09
phi076	CIMMYT	Hexa	4.11	umc1169	CIMMYT	Tri	1.04
phi078	CIMMYT	Tetra	6.05	umc1196	CIMMYT	Hexa	10.07
phi079	CIMMYT	Penta	4.05	umc1277	CIMMYT	Tetra	9.08
phi084	CIMMYT	Tri	10.04	umc1279	CIMMYT	Tri	9.00
phi087	CIMMYT	Tri	5.06	umc1304	CIMMYT	Tetra	8.02
phi089	CIMMYT	Tetra	6.08	umc1399	CIMMYT	Tetra	3.07
phi100175	CIMMYT	Tetra	8.03	umc1545	CIMMYT	Tetra	7.00
phi101049	CIMMYT	Tetra	2.10	phi034	CAAS		
phi112	CIMMYT	Di	7.01	phi011	CAAS	Tri	1.09
phi114	CIMMYT	Tetra	7.02	phi062	CAAS	Tri	10.04
phi121	CIMMYT	Tri	8.04	phi108411	CAAS	Tetra	9.05
phi123	CIMMYT	Tetra	6.07	umc1774	CAAS		
phi213984	CIMMYT	Tri	4.01	phi374118	CAAS	Tri	3.02
				umc1813	CAAS		

Table 2. SSR primers to be run on the core subset of populations. Primers in red overlap with the primers chosen for the maize inbred lines as well.

SSRs to be used on the Populations:

Primer	Lab	Repeat	Bin
nc130	INRA	Tri	5.00
phi014	INRA	Tri	8.04
phi029	INRA	Compound	3.04
phi031	INRA	Tetra	6.04
phi041	INRA	Tetra	10.00
phi046	INRA	Tetra	3.08
phi056	INRA	Tri	1.01
phi059	INRA	Tri	10.02
phi062	INRA	Tri	10.04
phi069	INRA	Tri	7.05
phi072	INRA	Tetra	4.01
phi083	INRA	Tetra	2.04
phi084	INRA	Tri	10.04
phi085	INRA	Penta	5.07
phi093	INRA	Tetra	4.08
phi102228	INRA	Tetra	3.04-05
phi108411	INRA	Tetra	9.06
phi112	INRA	Di	7.01
phi115	INRA	Compound (di/tetra)	8.03
phi127	INRA	Tetra	2.08
phi227562	INRA	Tri	1.12
phi308707	INRA	Tri	1.10
phi331888	INRA	Tri	5.04
phi96100	INRA	Tetra	2.00-01
nc133	CIMMYT	Penta	2.05
phi034	CIMMYT	Tri	7.02
phi063	CIMMYT	Tetra	10.02
phi075	CIMMYT	Di	6.00
phi076	CIMMYT	Hexa	4.11
phi079	CIMMYT	Penta	4.05
phi090	CIMMYT	Penta	2.08
phi114	CIMMYT	Tetra	7.02
phi109118	CIMMYT		
phi109125	CIMMYT		
phi299852	CIMMYT	Tri	6.07
umc1196	CIMMYT	Hexa	10.07
umc1266	CIMMYT	Tri	3.06
umc1332	CIMMYT	Tri	5.04
umc2047	CIMMYT	Tetra	1.09

It was decided to analyse the inbreds lines and the populations in different studies. The populations are heterogeneous and highly diverse, and past studies have shown that at least 15 individuals, and up to 30 individuals, must be fingerprinted in order to estimate the average allele frequencies for each molecular marker.

A study completed by Pierre Dubreuil, a postdoctoral scientist in the Applied Biotechnology Centre, has been published in Maydica (Dubreuil et al., 2006) and provides the methodology to

fingerprint populations using bulks of 15 individuals, where SSR markers are amplified from the bulks and run on an ABI automatic DNA sequencer. The heights of the peaks recorded by the sequencer are directly correlated to the frequencies of the alleles within the bulks, thus eliminating the need to run the individuals separately. However, the bulking method is not supported by any current analysis software, and thus, a SAS programme was written to take raw data from the software packages provided by Applied Biosystems (makers of the DNA sequencers) and analyse it directly.

The bulking method was found to work well in the absence of severe stutter of the markers and in the absence of preferential amplification of one allele over another from the same SSR locus. Tests were first run on bulks created of individuals with known alleles in different proportions to simulate the unknown bulks that would later be characterised. The known bulks were analysed using the new techniques and the observed frequencies calculated from the bulk was compared to the expected frequencies calculated using the data collected on the individuals prior to bulking. Only those SSR loci that had correlations of expected to observed allele frequencies above a rigorous threshold value were accepted; generally, those with low stutter and no preferential amplification, as well as lack of amplification of unexpected products in the bulk were chosen. Unfortunately, only 24 SSRs were optimised by the start of this work, of which only 15 could be run on all populations by INRA. Twenty additional SSR markers were optimised by CIMMYT. Only 15 could be run in the course of the study, however.

Genetic resources are sampled for regeneration, characterisation, evaluation, and to study phenotypic and genetic diversity, and for the formation of core subsets to access the diversity in breeding and association studies. When sampling occurs, it is very important to maintain as much of the original diversity in the subset as possible. Because the total number of entries are reduced in the core, methodologies to reduce the numbers but not the total diversity have been suggested (Brown, 1989; Franco et al, 2006). A suitable selection strategy was developed for this project in order to select the mini-composite subset of maize. Both inbreds and landraces were included in the core, but the landraces should be converted to immortalised genotypes (inbred lines) in order to preserve the same genotypes for different studies, and for association mapping (which has never been done on heterogenous, annual maize plants before).

I. Scientific activities (including tables and figures)

Markers used in the study:

Table 1 shows the SSRs chosen for the genotyping the inbred lines in this study and the lab responsible for running them on all materials to be genotyped in 2004. SSRs were run on the inbred lines at CAAS and CIMMYT only, as IITA lines and populations will be sent to CIMMYT for analysis. The number of SSRs run by each lab was the proportion of the funding received by each lab for this activity. For the populations, 24 SSRs were identified which can characterise populations in bulks, and are listed in Table 2. These are the SSRs that will be run at INRA on all populations in the study. Twenty additional primers were optimised in CIMMYT for work on the populations, but only 15 were successfully optimised in the time and funding allowed (Table 2), and these were run at CIMMYT on all populations in the study.

Germplasm analysed in the study:

DNA was extracted from lines and populations listed in Appendix Tables 1 (lines) and 2 (populations) at CIMMYT, IITA, and CAAS, and sent to all other labs participating in the analysis. DNA of two control lines, the inbred lines CML51 and CML292, were sent to all participating labs in April to be included in every gel as a quality control measure.

Inbred lines were included from the CIMMYT, IITA, and CAAS breeding programmes, three of the biggest in the developing world. Landraces were chosen to attempt to cover the entire world where maize has been traditionally grown, and major emphasis was given to geographical areas

that have not been well covered in past studies. However, Asia and Africa were still considered not well covered in this study, so a new proposal was put forth to better cover these areas, and to develop the first world wide study of traditional maize population cultivation and dispersion around the world.

Laboratory analyses run in the study:
Standard laboratory protocols in DNA extraction, PCR amplification of markers, and gel electrophoresis on an automatic DNA sequencer were identified and all laboratories agreed to follow these protocols. All inbred lines were run according to protocols listed in the CIMMYT AMG lab protocols manual (CIMMYT, 2005) and analysed according to the methodologies described in Warburton and Crossa, 2002. All populations were run and analysed according to the methods listed in Dubreuil et al, 2006.

Computer programmes and databases created in the study:
Because many of our collaborators do not have access to the expensive SAS software, the SAS programme written by Dubreuil et al. was rewritten by Franco et al as an R programme, which does the same thing (analysed bulked data output from the ABI DNA sequencer and converts them to allele frequencies, taking into account stutter and marker repeat unit). The R programme is freely distributable, along with its user's manual, and R is freely available software.

The data have been shared with all collaborators as an excel file, and have all been submitted to Tom Hazekamp at the GCP Central Repository and can be found on the GCP virtual work space. In addition, an Access database, containing many useful data manipulation tools, contains all the GCP maize data, an in addition, many past studies from CIMMYT and the Asian Maize Biotechnology Network. It is a very useful collection of tools and marker data, and can be downloaded free (with an operation's manual) at:
http://www.cimmyt.org/english/docs/manual/dbases/fingerprint_Instructions_manual.htm

Marker reference kit created:
This was finished as an SP5 commissioned project with two National Programme scientists from Thailand and China. The work was finished in December, 2005, and final reports and data submitted to the GCP.

Results (at present):

Diversity in the inbred lines:
Summary statistics for the genetic diversity in 987 inbred maize lines for 41 SSRs run at CIMMYT is shown in Table 3 below. Nine SSRs had to be removed because of high missing data; data from CAAS were turned in late and only a preliminarly analysis is reflected in the table below; they were corrected recently, and the corrected data were turned into the Central Repository. Very high allele numbers were shown for some of the CAAS SSR markers (not reasonable, thus the rescoring). The data from the missing data can still be requested from CIMMYT, but were not put into the Central Repository because it was felt that they would not a fair representation of the data. They were not used to calculate the core subsets (below).

Table 3. Summary statistics for the SSR markers run on the inbred lines including the frequency of the allele found most often, the total number of alleles, the gene diversity, heterozygosity, Polymorphic Information Content, and inbreeding coefficient. Lowest and highest values for number of alleles are highlighted.

Marker	Major Allele Frequency	Allele No.	Gene Diversity	Hetero-zygosity	PIC	f
phi062	0.7078	22.0000	0.4655	0.0000	0.4318	1.0000
phi011	0.2617	30.0000	0.8257	0.0000	0.8043	1.0000
phi108411	0.4660	15.0000	0.7313	0.0000	0.7072	1.0000
umc1813	0.2565	37.0000	0.8555	0.0000	0.8413	1.0000
phi034	0.2607	32.0000	0.8600	0.0000	0.8476	1.0000
phi374118	0.4371	40.0000	0.7594	0.0000	0.7393	1.0000
umc1774	0.3824	27.0000	0.7931	0.0000	0.7731	1.0000
nc130	0.5790	6.0000	0.5487	0.1145	0.4701	0.7915
nc133	0.6861	8.0000	0.4390	0.1561	0.3541	0.6447
phi032	0.5348	6.0000	0.5666	0.0630	0.4802	0.8889
phi046	0.5902	3.0000	0.4877	0.1098	0.3734	0.7751
phi050	0.8019	9.0000	0.3424	0.0137	0.3235	0.9601
phi064	0.2571	14.0000	0.8345	0.2084	0.8139	0.7506
phi069	0.4219	9.0000	0.6893	0.1319	0.6354	0.8089
phi076	0.5105	4.0000	0.6147	0.0672	0.5432	0.8908
phi079	0.7145	7.0000	0.4547	0.0619	0.4188	0.8639
phi084	0.6054	7.0000	0.5539	0.0517	0.4940	0.9067
phi087	0.3949	14.0000	0.7181	0.1173	0.6733	0.8368
phi089	0.6014	7.0000	0.4832	0.0472	0.3709	0.9025
phi112	0.8319	8.0000	0.3016	0.0507	0.2928	0.8322
phi114	0.4046	9.0000	0.7624	0.1922	0.7345	0.7483
phi121	0.8154	4.0000	0.3066	0.0382	0.2686	0.8754
phi123	0.5494	6.0000	0.5039	0.0558	0.3866	0.8894
phi96342	0.7589	7.0000	0.3976	0.0873	0.3669	0.7806
phi100175	0.6135	8.0000	0.5367	0.0867	0.4706	0.8387
phi213984	0.8136	10.0000	0.3265	0.1045	0.3112	0.6802
phi233376	0.4828	10.0000	0.7077	0.0839	0.6777	0.8816
phi299852	0.3982	17.0000	0.7426	0.1614	0.7056	0.7829
phi328175	0.4013	11.0000	0.7108	0.0792	0.6625	0.8887
phi448880	0.7400	6.0000	0.4199	0.0348	0.3841	0.9173
phi453121	0.3295	9.0000	0.7994	0.1359	0.7736	0.8303
umc1109	0.4883	7.0000	0.6439	0.1104	0.5816	0.8288
umc1143	0.4478	10.0000	0.6479	0.1325	0.5798	0.7958
umc1152	0.2769	21.0000	0.8168	0.3463	0.7945	0.5764
umc1153	0.3685	12.0000	0.7088	0.1487	0.6555	0.7904
umc1196	0.3949	6.0000	0.7354	0.1771	0.6946	0.7594
umc1277	0.5743	6.0000	0.5490	0.0779	0.4679	0.8582
umc1279	0.7931	7.0000	0.3576	0.0622	0.3407	0.8262
umc1304	0.8219	5.0000	0.3108	0.0470	0.2923	0.8489
umc1399	0.3137	7.0000	0.7745	0.0929	0.7387	0.8802
umc1545	0.7464	7.0000	0.4158	0.1066	0.3853	0.7439
Mean	0.5326	11.9512	0.5975	0.0867	0.5527	0.8551

Diversity in the populations:

Summary statistics for the genetic diversity in 98 Central and South American maize populations based on 31 SSRs (preliminary data) is shown in Table 4 below.

Table 4. Summary statistics of Latin American maize populations run with 31 SSR markers, including average total number of alleles, average number of unique alleles (those found in only one population), average number of rare alleles (those present in only 5% or fewer of the populations), and average and range Modified Roger's Distances (MRD).

Total no. of alleles	*254 (8.2)*
No. of unique alleles	*31 (11.6%)*
No. of rare alleles (<5%)	*77 (27.0%)*
Average and range of MRD	*0.417 (0.201 - 0.699)*

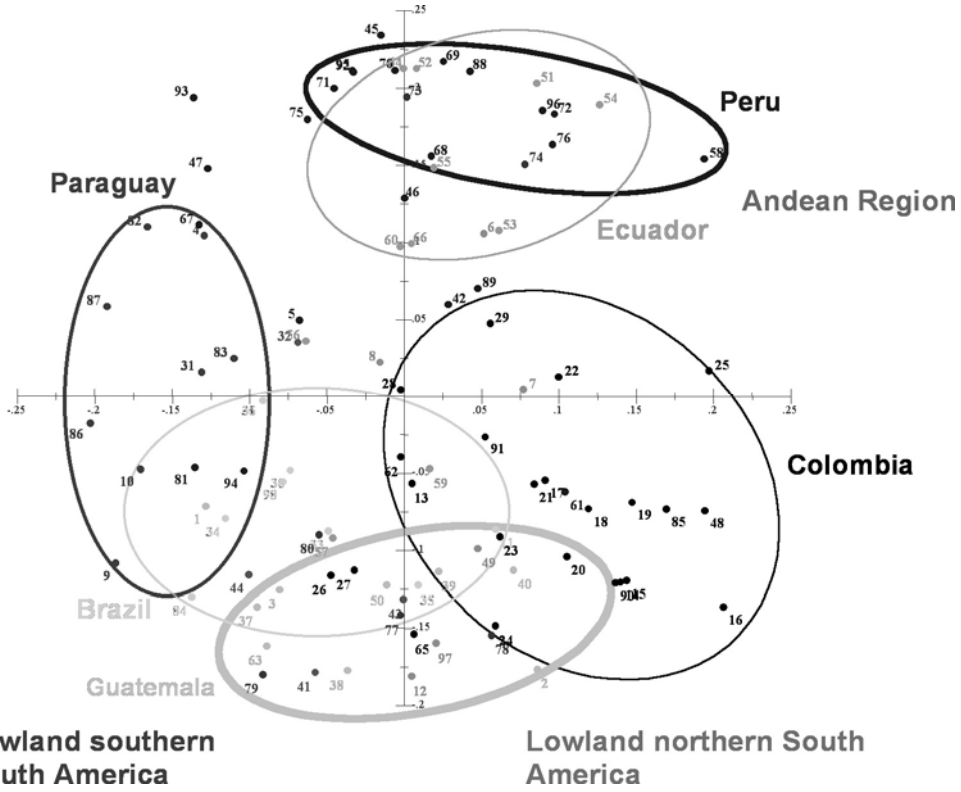
Diversity per locus for the same American landraces is shown in the Table 5. In comparison with previous studies, a high number of total alleles have been found in the 98 South American populations. In comparison with more than 400 CML lines, populations show higher allelic richness and a higher number of unique alleles (data not shown).

Table 5. Diversity per locus of the Latin American landraces, including the allelic richness and Nei's diversity index. Lowest and highest values are highlighted.

Locus	*Allelic Richness*	*Nei's diversity*	*Locus*	*Allelic Richness*	*Nei's diversity*
phi014	*6*	*0.526*	*phi102228*	*3*	*0.320*
phi029	*10*	*0.564*	*phi108411*	*6*	*0.501*
phi031	*8*	*0.668*	*phi109118*	*11*	*0.720*
phi034	*6*	*0.577*	*phi112*	*7*	*0.361*
phi046	*2*	*0.214*	*phi114*	*7*	*0.699*
phi059	*6*	*0.626*	*phi115*	*3*	*0.495*
phi062	*4*	*0.251*	*phi127*	*9*	*0.354*
phi063	*14*	*0.679*	*phi227562*	*13*	*0.810*
phi069	*8*	*0.614*	*phi299852*	*17*	*0.821*
phi072	*9*	*0.671*	*phi308707*	*8*	*0.661*
phi075	*14*	*0.791*	*phi331888*	*7*	*0.571*
phi076	*6*	*0.691*	*umc1196*	*16*	*0.791*
phi083	*9*	*0.721*	*umc1266*	*6*	*0.375*
phi084	*7*	*0.610*	*umc1332*	*6*	*0.636*
phi085	*10*	*0.741*	*umc2047*	*7*	*0.583*
Average	*8.2*	*0.588*			

A principal coordinates analysis of the American landraces based on the Modified Roger's distances between all pairs of landraces in the study (Figure 1) indicates that landraces tend to group based on geographical and altitudinal distribution, with some exceptions that may prove to be interesting examples of migration between geographical regions.

Figure 1. Principal Coordinates analysis of Latin American maize populations based on Modified Roger's distances.



Selection of core subset:

The maize GCP mini composite core were selected from the 925 inbred lines analysed (and which did not have a high percentage of missing data) and for the 317 maize populations completed by the GCP and the 276 maize populations completed in a previous study with many of the same markers. A core of 384 entries will be created as follows:

Inbreds:
27 ancestral inbreds that represent the majority of the pedigrees of modern hybrid maize will be included. These inbreds have been studied extensively, and 26 of these inbreds have been crossed to B73 to develop 25 linkage mapping populations from carefully chosen lines that together capture roughly 80% of the SNP diversity in maize. From these crosses, 200 to 400 recombinant inbred lines have been developed, and all 5000 RILs will be genotyped and made publicly available through the US Maize Stock Centre. These populations should provide a resource that unifies QTL mapping across the maize community, as well as a resource for which almost every trait will show substantial variation in multiple populations, and thus it is very important to include them in the GCP mini-composite. Furtheremore, the 27 lines were originally chosen by the Cornell Institute for Genomics Diversity lab to cover extensive maize diversity, which is also the purpose of the GCP core.

These inbreds are:

B73	CML333	Ky21	NC358
B97	CML52	M162W	Oh43
CML103	CML69	M37W	Oh7B
CML228	Hp301	Mo17	P39
CML247	Il14H	Mo18W	Tx303
CML277	Ki11	MS71	Tzi8
CML322	Ki3	NC350	

In addition, 36 inbreds will be included which are also in the association mapping panel of Ed Buckler, bringing the total overlap between the two association sets to 63 inbred lines. The 36 overlapping inbreds will be:

CML005	CML103	CML247	CML314
CML009	CML108	CML254	CML321
CML014	CML154	CML258	CML322
CML038	CML157	CML261	CML323
CML045	CML158	CML264	CML328
CML052	CML218	CML277	CML331
CML069	CML220	CML281	CML333
CML091	CML228	CML287	CML341
CML092	CML238	CML311	CML349

CML051 and CML292 will be included because they are run as controls in every fingerprinting study conducted by CIMMYT and AMBIONET.

A core selection strategy was run on the remaining set of inbred lines in order to choose a total of 234 inbred lines (169 new lines plus the lines listed above). Three selection strategies were used to choose core subsets of 160 inbreds from the initial set of 925 lines. The first method calculates Modified Roger's distances between all pairs of lines in the study, create a UPGMA dendrogram using the distances, and then to use the Gower's MRD distance between each line within each cluster to determine which individuals were chosen (using a random process) from each cluster to reach the 160 lines. The second method will be the MSTRAT algorithm developed by Guesnard (2001); and the third method will be the Simulated Annealing algorithm developed by Lui in his PowerMarker programme (V 3.23, 2006). Statistics were run to determine which maintains the most diversity and maximises distances in the core subset compared to the original sample. The Roger's method was chosen and used to determine the identity of the 160 lines listed in Appendix Table 2.

Seeds from all 225 inbreds have been requested from the corresponding institutes (unless they are already at CIMMYT) for seed increase, maintenance, and distribution. Most has been sent to CIMMYT; however, in most cases, we have very little seed. We would like to grow out all the lines in the core subset for distribution and use. We have already planned to plant out the yellow maize inbreds to use in an association mapping panel in a carotenoid study financed by the Harvest plus CP. We will find associations to genes in the carotenoid pathway (besides Y1) that increase beta carotene and precursors; this cannot be selected for phenotypically since Y1 masks all other genes affecting yellow color in maize kernals but large differences in carotenoid levels can occur. The yellow maize will thus undergo a seed increase (being planted this month) but the white maize will not; this is also made a bit difficult because some of the lines are temperate and won't flower in the field at CIMMYT without help! We could plant them in the greenhouse or further north (they should do fine in our fields in Northern Mexico, for example). Some funding

would be needed, since this was not part of the original budget, but having seeds for distribution for all lines in the core would mean that people may actually use it! We are also seeking alternate funding to develop fixed lines from the populations included in the core subset, to make them more useful for phenotypic and association studies; otherwise, they will be included as heterogenous populations and people can use them as best they can.

In addition, and most unfortunately, CAAS has been unable to send seed of the CAAS lines chosen to form part of the core subset because the Chinese Ministry of Agriculture will not give permission to remove the lines from China. I have been unable to determine when, or even if, this is likely to happen. We therefore request help from the GCP administration to deal with this policy problem. Removing the CAAS lines from the study would greatly decrease the total amount of diversity in the core minisubset, which is against the point of the core.

Populations:

30 populations were chosen using the same strategy as for choosing the 234 inbred lines and are included in Appendix Table 2 (yellow highlighted entries). This can be considered a preliminary core, as entries may change due to seed availability or suggestions from other researchers, as long as total diversity within the core does not decrease. Seeds from the populations that are not already at CIMMYT have been requested, and these populations will be handled in one of two ways: Either five inbred lines per population (on average) will be developed to come up with 150 fixed lines to include in the GCP mini composite set and phenotyped with the inbreds (selected above); or five random (non-inbred) individuals will be included per population, and DNA extracted from these individuals only once. Phenotypes of the 30 populations will be done at the population level in this case, and means of values used for association studies. Which method is chosen depends on if inbreds can be created from these 30 populations via selfing or doubled haploid technology or not. This, in turn, depends on the funding. If no additional funding is provided for inbreeding, none will be done.

II. Conclusions

The core minicomposite subset can be used as an association mapping panel for any trait of interest to the breeders. It is currently being used to find genes associated with drought and beta carotene content of yellow maize, and we hope to begin to use it for acid soils and diseases resistance genes as well. In addition, the core subset can be used to screen phenotypically only for new traits of interest, rather than screening all entries in the CIMMYT and IITA maize genebanks; because they cover a very wide range of diversity, it may be used to screen for any trait, and if this trait is found in some of the individuals, this can be used as a guide for where to look for more examples of the desired trait (ie., individuals related by pedigree, geopgraphic origin, breeding programme, selection history, etc). The utility of the core subset will be greatly expanded if funding is provided for seed increase and distribution, and for phenotyping by as many groups and for as many traits as possible.

Although the work commissioned by this project has been completed to the best of the abilities of the collaborating laboratories, work continues in the area of data analysis, and we propose the following articles may be written as joint GCP publications in the next year (discussions currently underway to identify lead and co-authors):

- A study of the maize landraces characterised in this study; between and within landrace data, relationships between landraces from similar origins vs. distant origins.
- A study of the modern inbreds from CAAS and CIMMYT; relationships and levels of diversity.
- A paper on the two IITA Reciprocal Recurrent Selection inbreds, of which two generations were characterised in this study and shifts in allele frequencies can be studied.

- All inbred and landrace data analysed together with discussion of sampling for core formation.

2005-01f: Genotyping of Composite Germplasm Set, Tier 1, Wheat

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborating Scientists:
Francois Balfourier, INRA
Nachit Miloudi, ICARDA
Xueyong Zhang, CAAS

Mid-Year Report
Deliverables
Progress in each of the activities listed in the tables below are expressed as percent of the activity finished by each partner. Activities are describes as follows:

Standard Protocols Identified: Standard laboratory protocols in DNA extraction, PCR amplification of markers, and gel electrophoresis on an automatic DNA sequencer were identified and all laboratories agreed to follow these protocols. To avoid between laboratory errors and differences in data interpretation (gel reading), it was agreed that all DNA be extracted by each lab and sent to all other participating labs; each lab then ran a subset of the SSR marker on ALL DNA so each marker is done in only one lab.

Markers Chosen: Markers were chosen by each laboratory based on past experience (those easiest and most reliable) and to ensure as complete coverage of the genome as possible. Each lab (except IITA) is running a subset of the markers in their own laboratories, the number depending on their capacity.

Work plan developed: the work plan was developed in conjunction with other SP1 members in two workshops in 2004.

Control DNA delivered: Control DNA (two lines each) for maize and wheat was extracted at CIMMYT and sent to all other labs to include in every gel for further control of error. Unfortunately, not all labs chose to use (or report) the controls.

Marker reference kit created: This was completed in conjunction with SP5 and CAAS in China. The final report has been submitted with all pertinent data.

Composite Genotype Set ID: The composite genotype set was chosen at a meeting at PAG, San Diego in 2004 and in subsequent email dialogs between all partner countries. Partners tried to compliment each other's choices to make up the set of 1868 maize and 2506 wheat genotypes. Information on each entry has been gathered and will be placed on the Central Repository as soon as the head PI of this project sees that the repository is ready.

DNA extracted and DNA exchanged: following standard protocols, DNA from each sample was extracted in the labs of the institution that proposed the lines and sent to all other labs in sufficient quantities to finish analysis.

Marker data generated: SSRs amplified on each sample and run on an automatic DNA sequencer, then gels analysed to get allele sizes for each SSR/sample combination. All data have been run, but CAAS and ICARDA are finishing the rescoring of their data, following feedback from CIMMYT at the beginning of this year. M. Nachit and M. Warburton have agreed that the best course of action would be to send an ICARDA technician to CIMMYT for a week's worth of

training, since it would be very difficult to teach these things via email or over the phone, but there are no funds left to pay for this trip. In addition to finishing the data generation, this trip would have the added benefit of capacity building, as well as building a firmer alliance between the wheat genotyping labs of both centres. If this cannot be done, CIMMYT as coordinator of the project will try to rescore the gels, but this will take longer and has not been budgeted, either. No reply was given from the GCP as to the availability of additional training funds.

Data stored and exchanged: Marker data has been placed in the CIMMYT Maize and Wheat Fingerprinting Database (http://www.cimmyt.org/english/docs/manual/dbases/fingerprint_Instructions_manual.htm) and when the data set is completed will be placed in the GCP Central Repository. Once the data are all entered into the computer, Francois Balfourier has expressed interest in generating the core subset for wheat and will try to have that ready by autumn, 2006. Seed from each entry in the core subset will be requested by CIMMYT, and again, if funds are available, multiplied for distribution. This would be cheaper than having each interested party request seeds from multiple institutions and multiplying seed every time, and should facilitate use of the core; however, past experience has suggested that seeds from CAAS will not be allowed by the Chinese Ministry of Agriculture to leave the country. CIMMYT hopes to begin using the core to seek drought associated genes via genome scanning using DARTs (we are running a pilot study now to test logisitics and feasibility) and to seek new genes, alleles, and markers for disease resistance by studying Resistance Gene Analogs in the core.

I have also initiated a conversation with my genotyping collaborators to see what articles could be written with the data, and who will take the lead on each project.

Although I commend everyone in the project for their hard work, inexperience made the project move slower than we had hoped! I am very pleased (and quite relieved) it is nearly completed and properly publicised.

In summary, all data exist but not in entirely useable format in some cases; all is expected to be finished by summer, 2006. We hope to request some small, special funds from SP1 and/or SP5 to improve the utility and quality of the data as follows:

Item	**Details**	**Amount**
Training of ICARDA staff in scoring automatic DNA sequencer gels for fingerprinting	Plane ticket, visa, one week lodging	$3,000
Seed multiplication of wheat core subset (all)	Field growout of ~3000 entries for seeds	$9,000
Totals		$12,000

We expect all data to be analysed and the core subset to be completed by Autumn, 2006, unless no funding is available to train ICARDA staff in scoring data, in which case it may be delayed a few months.

Wheat Activity	AGROPOLIS	CAAS	CIMMYT	ICARDA
Standard Protocols Identified	100%	100%	100%	100%
Markers chosen	100%	100%	100%	100%
Work plan developed	100%	100%	100%	100%
Control DNA delivered	na	na	100%	na
Marker reference kit created	na	na	80%	na
Composite Genotype Set ID	100%	100%	100%	100%
DNA extracted	100%	100%	100%	100%
DNA exchanged	100%	100%	100%	100%
Marker data generated	100%	100%	100%	100%
Data stored and exchanged	100%	50%	100%	50%

2005-01g: Genotyping of Composite Germplasm Set, Tier 1, Rice

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

Mid-Year Report

1) Genotyping of composite rice collection (2757 accessions) with panel of 51 SSR markers
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 with a split of 4 to 3, respectively, and WARDA is genotyping at Cornell.

IRRI: Genotyping on 18 SSRs is complete, and 2 weeks required to finish the remaining 8 SSRs for the 2757 samples on the LiCor 4300 genotypers. Data is available on these 8 SSRs for 1536 samples from the MJR BaseStation platform. Positive controls have been included for IR 64 and Azucena.
CIRAD: Genotyping of 7 SSRs has been finished on all samples using LiCor genotypers with positive controls. Final data was sent to IRRI on March 28, 2006.
CIAT/EMBRAPA: The 7 SSR markers were split between CIAT (4) and EMBRAPA (3). Both CIAT and EMBRAPA have finished the task. Data was obtained from EMBRAPA on March 17, 2006 and from CIAT on April 13, 2006.
CAAS: "Reanalyzed data" for samples 1 to 1536 was obtained on March 21, 2006; this data set did not include peak height, area or quality values as wanted and appeared to be identical to that obtained on October 20, 2005. Furthermore, this partial set was only 68% complete overall (7276

data points out of 10752 expected). At this time they indicated that additional DNAs would be needed to complete the data. Data for samples 1537 to 2757 was obtained on May 9, 2006. While this data set did include peak height, area, it was even less complete with only 25.5% data points present (2183 of the expected 8547 data points).
WARDA/Cornell: Genotyping on 2757 accessions was finished for 6 SSRs including positive controls on an ABI 3730. The final data set was sent to IRRI on January 15, 2006.

Tangible outputs delivered:
Genotyping data has been obtained for 36 SSRs on the 2757 accessions from IRRI, CIRAD, CIAT, EMBRAPA, and WARDA/Cornell. Partial data has been obtained for the remaining 15 SSRs. Since CAAS does not appear able to deliver the remaining data, IRRI will genotype 6 of the 7 SSRs (RM162, RM215, RM237, RM271, RM447, RM455 but not RM489) assigned to CAAS by June 15, 2006. Of the 51 SSRs in the panel, Cornell now recommends 30 of these as more robust. RM489, the other marker assigned to CAAS, is not in this set. CIRAD and WARDA/Cornell recommended that RM178 and RM19 be dropped from the data since they were found to be very problematic (these two are also not in the 30 robust markers).

Deviations from the work plan:
Genotyping is finished at CIRAD, CIAT, EMBRAPA, and WARDA/Cornell and nearly complete at IRRI. Delays at IRRI were a result of the need to order primers directly labeled with either IRDye 700 or 800 since M13 tail labeling could not be optimised for 10 of the 24 SSRs assigned to IRRI. Approximately 3 weeks is required to finish the IRRI SSRs, and an additional 2 to 3 weeks for those that were assigned to CAAS (primers for these are available with 5'-IRDye labels to preclude any difficulty in optimisation of M13 tail reactions).

Dr. Susan R. McCouch will visit IRRI on June 15-16, 2006. Analysis of the consolidated data will begin in earnest prior to her visit. Dr. McCouch has indicated that she has access to new methods more appropriate to the analysis of SSR data on inbred species and is willing to share these and her expertise for assisting in the data analysis.

A no-cost extension is requested until August 15, 2006 for submission of the final report.

2005-01j: Genotyping of Composite Germplasm Set, Tier 1, Common Bean

Principal Investigator:
Matthew Blair, CIAT
Co-Principal Investigators:
Maria Jose Peloso, EMBRAPA
Rosana Brondani, EMBRAPA
Shumin Wang, CAAS
Collaborating Scientists:
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 begun 2004 and continued into this year:
1) 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. Data is still being analysed for Brazil collection and core set.

2) 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. An article was prepared and submitted for the Mesoamerican races and one will be prepared for the Andean races.
3) Analysis of selected primary centre 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.
4) Analysis of national germplasm sets: Bolivia (171 genotypes), Brasil (560 genotypes); Caribbean (310 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 centres of diversity in Bolivia and Colombia as well as a secondary centre 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. The data is being processed for these national collections.

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 centres 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 programmes and for tracing lineages. The uncovering of introgression between the gene pools will prove invaluable for association mapping studies.

Tangible outputs delivered:
Publication of first two articles on diversity assessment of common beans with microsatellites
1. Blair MW, Giraldo MC, Buendia HF, Tovar E, Duque MC, Beebe SE (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.) Theor Appl Genet (in press)
2. Díaz LM, Díaz JM, Blair MW (2005) Diversidad genética de fríjol común (Phaseolus vulgaris L.) en Colombia. Fitotecnia Colombiana 5: 28-36

Article submitted on Mesoamerican race collection.

Marker kit prepared as a result of genotyping examples and submitted for associated project.

2005-01k: Cowpea Genotyping

Principal Investigator:
Sarah Hearne, IITA
Collaborating Scientist:
Morag Ferguson, IITA

Mid-Year Report
From January to mid-April we have been completing two tasks. Firstly DNAs from those accessions that were absent in our initial first pass genotyping were genotyped using the 16 SSRs described in appendix 1. Secondly we have re-genotyped specific samples and SSRs to fill in missing data within the datasets. All data has been extracted from GeneMapper and is being

assembled for upload into the GCP repository, this should be completed within the next 2 weeks. Data analysis is progressing on the full data sets. Preliminary data is shown in Appendix 2. Data analysis and core-set identification is being conducted with the assistance of Dr Jorge Franco Duran of the Universidad de la Republica, Uruguay. Jorge has in collaboration with CIMMYT colleagues developed tools for core set identification and statistical validation that we propose to utilise for this project.

Tangible outputs delivered:
All data for the intra and inter-accession genotyping of cowpea is available from Sarah Hearne and will be available on the GCP repository shortly.

Deviations from the work plan:
A number of platform and technical issues have resulted in significant delays to delivery of outputs (these have been described in previous reports). In addition to delivery of outputs strategies have been implemented to prevent a re-run of the issues faced during the project.

2005-01l: Genotyping of Composite Germplasm Set, Tier 1, Cassava

Principal Investigator*:*
Martin Fregene, CIAT
Collaborating Scientists:
Paula Hurtado, CIAT
Morag Ferguson, IITA
Sarah Hearne, IITA
Carmen de Vicente, IPGRI

Mid-Year Report
During the period November 2005 to May 2006 the following activities were carried out:

CIAT as lead institute has been collating and analysing the molecular data generated from 30 markers, 22 analysed at CIAT and 8 analysed at IITA. Three different molecular marker data files have been generated for statistical analysis (Molecular weight information / locus / genotype, Allele / locus / genotype, Binary data / genotype) using the relevant software like NTSYS, SAS and Popgene.

Data analysis was based on genotypes with more than 80% of complete data per set of SSR markers. The analysis included assessment of genetic structure using principal coordinate analysis (PCoA) and multidimensional scaling (MDS) based on individuals under Jaccard's similarity matrix and cluster analysis based on countries under Nei's (1972) genetic identity and genetic distance.

Preliminary results and statistical analysis are based on data set 1, which corespond to 212 alleles from 2494 genotypes evaluated at CIAT using 22 SSR markers (genotypes with more than 80% of complete data per set of SSR markers). There is a second data set corresponding to 107 alleles from 2575 genotypes evaluated at IITA using 8 SSR markers (it has between 50-100% of missing data per genotype). To complete the analysis using 36 SSR markers, 6 markers are still remaining from IITA genotyping, so the whole analysis with the complete set of markers has not been done.

A cluster analysis based on country of origin was done using both data sets to establish if a combination of data sets (1 and 2) could be done for analysing the 30 SSR loci together (22 SSR analysed at CIAT and 8 analysed at IITA). Even data set 1 is composed by the evaluation of 60% of loci and data set 2 by 20%, the analysis shows similar results: 3 main groups composed by African, American and a mix of countries (Figure 2). Set 1 shows a better differentiation between countries leaving the regions with less number of accessions in the last group (mix of countries).

Looking for enhancing the level of differentiation between groups, both data sets were put together in one binary matrix and the cluster analysis was done (data not shown). Any relevant cluster was observed and the level of differentiation was lower.

50% of the accessions have to be eliminated from the analysis since the number of missing value increased when sets 1 and 2 were put together.

Figure 2. UPGMA based on 50 countries using Nei's (1972) genetic distance. A. Set 1 (2494 genotypes evaluated at CIAT using 22 SSR markers). B. Set 2 (2575 genotypes evaluated at IITA using 8 SSR markers).

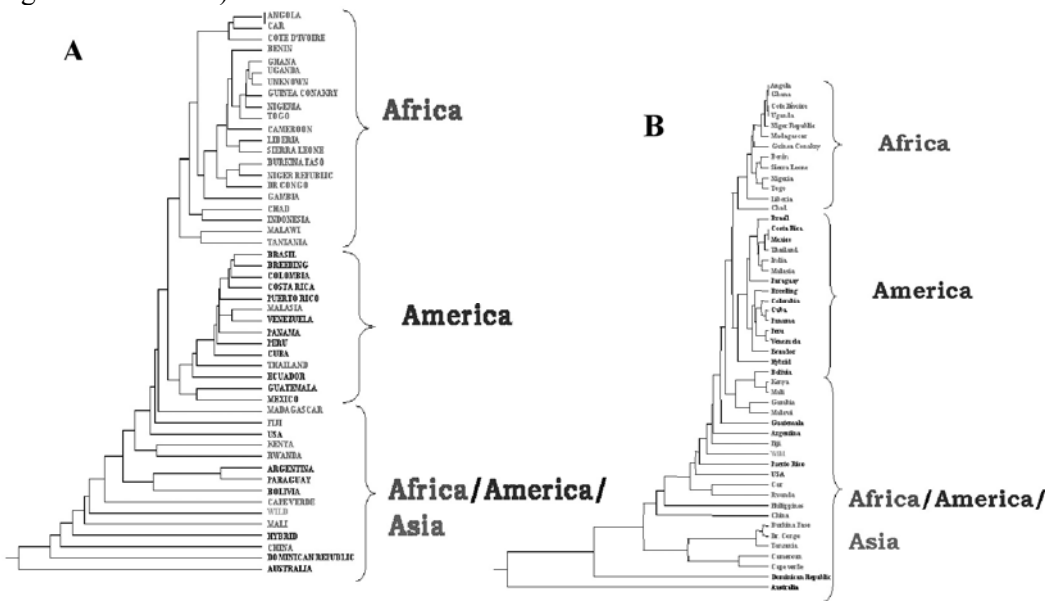


Figure 3. Multidimensional scaling plot from 2494 genotypes analysed with 22 SSR markers (set 1). Diversity estructure explaining 70% of the genetic variance.

In order to differentiate the accessions composing the 3 main groups identified by cluster analysis, a linear multidimensional scaling analysis (MDS) was done using the data set 1. MDS analysis shows how the data set 1 is divided in 8 groups explaining 70% of the genetic variance (Figure 3 and 4).

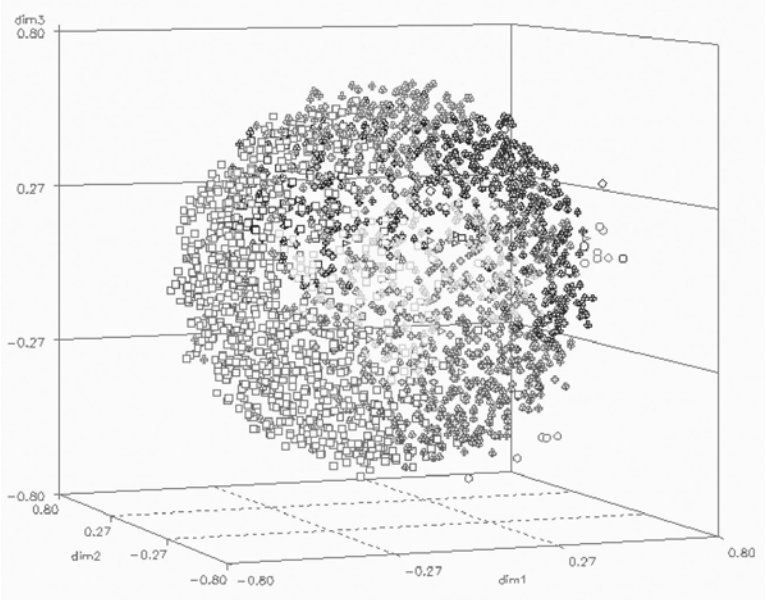
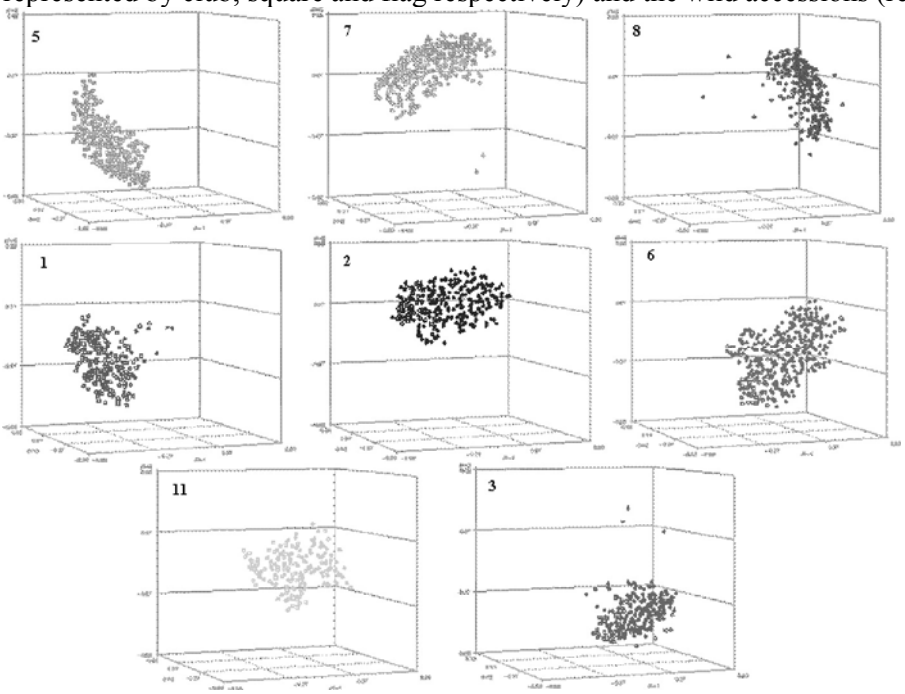


Figure 4. 8 groups generated by multidimensional scaling using data set 1. Each color reresents a group and the different shapes correspond to the germplams collecion (CIAT, IITA, EMBRAPA represented by club, square and flag respectively) and the wild accessions (represented by circle).



8 groups were generated by MDS and some countries have significant frequency inside the groups (Table 1).

These preliminary results from cluster analysis, as well as, a multidimensional scaling analysis based on individuals from data set 1 reveal a clear separation between accessions from Africa and the rest of the world confirming findings from previous studies that shows how global cassava germplasm diversity is structured by region. Sources of this genetic differentiation could be selection for adaptation to agro-ecologies, particularly diseases, found in Africa. Accessions from Asian countries have shown high differentiation that could be explained by the new introduction of the crop in this continent as well as the small effective sample sizes.

Table 1. Most frequent country per group defined by MDS

Group	Country	Frequency (%)
1	Nigeria/Ghana	20/23
2	Brasil	54
3	Wild/Mexico/Gutemala	2/4/6
5	Nigeria/Ghana	31/27
6	Brasil	63
7	Ghana/Brasil	21/23
8	Wild/Peru/Brasil	4/7/22
11	Colombia/Ghana	15/20

Up to date, the statistical results defining diversity structure in Cassava germplasm are based on genotyping of 2494 accessions using 61% of the 36 SSR markers (set 1).

It is possible to differentiate some groups defined by cluster analysis and MDS. African accessions are showing a clear sub-estructure separating Ghanaian and Nigerian accessions from the rest of the continent as previous diversity sudies have indicated. Thinking that Cassava is a crop of recient introduction in Africa, it is possible that the separation between african accessions

could be explained by the selection that farmers have done in order to preserve tolerant accessions to biotic and abiotic stresses affecting the yield.

Sub-estructure of American accessions is showing a differentiation between Brasilian and Central American ones. The source of the observed structure could be introgression from wild relatives, as well as, independent domestication events. Wild accessions are distributed in all the American groups, especially in group 3 and 8 where the separation is clear between South and Central American accessions that could be presenting introgression of some wild genes.

Small sample size could be affecting Asian accessions that are distributed in all the groups without a clear estrcture. Also the fact of the new introduction of the crop can affect their genetic diversity which is still not so clear.

Tangible outputs delivered:

1) Genotyping of 2500 cassava accesions from 3 germplam banks (CIAT, IITA, EMBRAPA) using 22 SSR markers at CIAT and 8 SSR markers at IITA
2) Three different molecular marker data files were generated:
    - Molecular weight information / locus / genotype
    - Allele / locus / genotype
    - Binary data / genotype
3) Data analysis using NTSYS, SAS and Popgene
4) Structural characterisation of diversity in global cassava genetic resources using SSR markers
5) Preliminary definition of cassava diversity using the global genetic resources in the crop.

Deviations from the work plan:

1. Last year it was informed by IITA that due to technical problems with the genotyping facility at IITA-Nairobi, there was a delay in completion of SSR analysis at IITA for the remaining 6 SSR markers. Genotyping using the 6 SSR markers is still missing.
2. The data set sent by IITA to CIAT was analysed separately from CIAT data set, since the number of alleles was higher than the usual number gotten with the six SSR markers evaluated by them. The number of missing data was higher than 50% for some accessions, so more than 40% of the whole germplasm had to be eliminated to integrate CIAT and IITA data in one analysis.

Perspectives

Based on the large amount of information generated by cassava in the first two years of GCP diversity activities (SP1 cluster 1 and 2), we would like to provide important inferences with applicability to understanding global genetic resources not only in cassava but also in other crop species. During June, we would like to extend and expand our population genetic analyses to better understand the genetics of cassava diversity and domestication. It is possible to elucidate forces behind the observed genetic diversity structure in cassava (mutations, founder effects, selection, migration, independent domestication events or small effective sample size). This crop can be used as case-study to perform analyses to understand the process of evolution in a vegetatively propagated crop, including the genetics of adaptive diversification and the forces affecting genome-wide patterns of genetic structure.

2005-01n: Genotyping of Composite Germplasm Set, Tier 1, *Musa*

Principal Investigators:
Nicolas Roux, IPGRI-INIBAP
Elizabeth Arnaud, IPGRI-INIBAP
Collaborating Scientists:
Isabelle Hippolyte, CIRAD
C.Billot, CIRAD

V. Pomies, CIRAD
P. Cubry, CIRAD
A. Weber, CIRAD
L. Gardes, CIRAD
K. Tomekpe, CIRAD
Maria Kolesnikova-Allen, IITA

Mid-Year Report

Activities performed at CIRAD:

In the past months, genotyping has been performed on 548 accessions with 12 SSR markers (see list in Table 1). Of these accessions, 192 came from IITA (1_2 to 367_2), 237 came from CIRAD Guadeloupe (1_1to 48_1 and 400 to 734), and 120 came from the Confederation of Rural Associations in Buenos Aires and La Pampa (CARBAP) (771 to 898). Because of incomplete results, the cavendish accession – giant cavendish (ITC0346- Number 805) – has been removed. Accessions from CARBAP and CIRAD Guadeloupe were extracted at CIRAD.

Gel migrations were analysed with AFLP Quantar Pro software. A standard for analysis is very useful to compare results between gels made at different times, on different machines and by different researchers. Nevertheless, since some SSR markers provided a broad range of allele sizes, the use of sterile accessions of plantain and cavendish were also very helpful when checking results between different gel migrations.

Diversity analysis will be performed with all accessions and markers using Darwin software.

Activities performed at IITA:

A total of 515 accessions (provided by CIRAD and IITA) were genotyped with 24 microsatellite markers, four of which were common to the CIRAD set. The missing value for the raw data is close to 20% and the process of filling gaps is in progress. The data that were obtained have been analysed for allelic content based on ploidy groups and final data has been compiled in the summary set; analysis is still being performed.

Tangible outputs delivered:

- 548 accessions genotyped with 12 SSR markers on Licor at CIRAD
- 515 accessions genotyped with 24 SSR markers on ABI3100 (with 20% missing values for the entire set) at IITA

Deviations from the work plan:

Some delays in work were caused by the change of personnel on the project at IITA and CIRAD.

2005-02: Supporting Distribution of Reference Germplasm (CIP)

Principal Investigator:

Enrique Chujoy, CIP

Collaborating Scientist:

William Roca, CIP

Mid-Year Report

The Potato Composite Genotype Set consists of 1084 clonal accessions (landraces, advanced cultivars, breeding lines and mapping populations) in the form of in vitro plantlets. When this project was initiated, only 469 of these accessions conformed to the CIP Phytosanitary requirements for international distribution. Now 558 accessions more conform to these requirements, making a total of 1027.

Pathogen testing followed CIP's protocols that use serology (ELISA) and nucleic acid hybridisation (NASH) detection methods. Plants with negative reactions to the pathogens were further evaluated by host range testing, using indicator plants during the winter when cool temperatures are adequate for the test. Plants with positive reactions to any detected pathogen were subjected to the pathogen elimination protocol. Here plants underwent heat treatment and meristems were excised and cultured as many times as needed to obtain healthy plants. A remaining 57 accessions of the Potato Composite Genotype Set are still to be made available as pathogen free in vitro plants.

Plans for the distribution of the Potato Composite Genotype Set to collaborators are ongoing. 272 clones were distributed and phenotyped. 716 clones were distributed for genotyping with SSR markers. From these, a tentative most genetically-diverse set of 200 accessions has been identified on SSR data. Completion of the selection of the reference germplasm set is pending further advance in field phenotyping, as well as classification of the Composite Genotype Set for maturity. A preliminary set of 42 advanced clones were distributed to China in 2005 to establish collaboration on phenotyping for drought tolerance in an important target environment during the additional year (2006–2007 growing season) that will be required to complete selection of the reference set. A new set of nearly 300 clones is being prepared for distribution and phenotyping.

Tangible outputs delivered:

1. Out of 1084 potato accessions of the Composite Genotype Set, 1027 are now virus free and available for international distribution
2. Deviations from the work plan

In vitro and field propagation, phenotyping and, in some cases, the need for virus-elimination from potato germplasm accessions, delayed the identification, thus the distribution of the Potato Composite Germplasm Set.

2005-02: Support and Distribution of Reference Germplasm, *Musa* (IPGRI)

Principal Investigators:

Nicolas Roux, IPGRI-INIBAP
Elizabeth Arnaud, IPGRI-INIBAP

Collaborating Scientists:

Isabelle Hippolyte, CIRAD
Jaroslav Dolezel, Institute of Experimental Botony (IEB)

Mid-Year Report

DNA extracts of banana reference accessions will be distributed to the *Musa* research community by the Musa Genome Resource Centre based at IEB, Olomouc, Czech Republic. The mini core of 48 accessions was defined by the GCP criteria regarding diversity representation, duplication at INIBAP Transit Centre (ITC), their use in breeding programmes and the FAO designation. (More information is included in the previous GCP report on Global genetic diversity: Marker analysis.) To increase the representation of *Musa* diversity, three accessions were added (cf table 1: N°49-50-51).

For each accession listed below, 10 extractions of 3 g of frozen leaves have been performed using the Matab method (Risterucci et al., 2000); the amount of DNA still has to be quantified. M. Rodier-Goud from CIRAD, Montpellier will bring these extracts to the Musa Genome Resource Centre (MGRC) based at IEB, Olomouc, Czech Republic. It is expected that 1 mg will be obtained per accession. Upon request, 1 µg maximum of DNA per accession will be distributed for research purposes.

Full information (morphological and molecular) on this set will be made available on the web site of the Global *Musa* Genomic Consortium (www.musagenomics.org).

Tangible outputs delivered:
See table below

Deviations from the work plan:
The project is proceeding as planned.

Extracted accessions for DNA distribution

Sample number	Reference	Species	Sub-species	Name	Country of origin	ITC code	Record number in MGIS
1	Eumusa	AAB	Nadan	Lady Finger ou AA Sucrier	India ?	ITC.0582	NEU0297
2	Eumusa	AAB	Pome / Prata	Foconah	Cameroon	ITC.0649	NEU0298
3	Eumusa	AAB	Pome / Prata	Prata Ana	Brazil	ITC.0962	NEU0310
4	Eumusa	*balbisiana*	type 4	P. Klutuk Wulung, IDN 056	Indonesia	ITC.1063	NEU0054
5	Eumusa	*balbisiana*	type 4	P. Batu, IDN 080	Indonesia	ITC.1156	NEU0055
6	Eumusa	acuminata	Banksii	Banksii 623	Papua new guinea		
7	Eumusa	*acuminata*	*microcarpa*	Borneo	Malaysia, S/E Borneo	ITC.0253	NEU0028
8	Eumusa	*acuminata*	*burmannicoides*	Calcutta 4	India, Calcutta	ITC.0249	NEU0017
9	Eumusa	ABB	Ind ABB	K. Tiparot=?Te (e)parod, THA 020	Thailand	ITC.0652	NEU0383
10	Eumusa	AAB	Plantain	Orishele	Nigeria	ITC.1325	NEU0256
11	Eumusa	ABB	Pelipita	Pelipita	Philippines	ITC472	NEU0360
12	Eumusa	ABB	Bluggoe	Dole		ITC.0767	NEU0334
13	Eumusa	AAA	Cavendish	Grande Naine	Guadeloupe	material to be sent to ITC	NEU0172
14	Eumusa	AAA	Orotava	Pisang Kayu, (IDN098)	Indonesia	ITC0420	NEU0208
15	Eumusa	*acuminata*	*errans*	Agutay	Philippines	ITC.1028	NEU0033
16	Eumusa	*acuminata*	*siamea*	Khae (Phrae), THA 015	Thailand	ITC.0660	NEU0025
17	Eumusa	AAB	Figue Pomme / Silk	Figue Pomme Géante	Guadeloupe	ITC.0769	NEU0285
18	Eumusa	ABB	Saba	Saba	Philippines	ITC.1138	NEU0361
19	Eumusa	AAA	Ambon	Pisang bakar, IDN106	Indonesia,	ITC.1064	NEU0229
20	Eumusa	ABB	Monthan	Monthan	India ?	ITC0046	NEU0350
21	Eumusa	balbisiana		Tani			
22	Eumusa	*acuminata*	*burmannica*	Long Tavoy pied		ITC.0283	NEU0016
23	Eumusa	AB cv		Safet Velchi	India	ITC.0245	NEU0152

24	Eumusa	AAA	Cavendish	Petite Naine		ITC.0654	NEU0174
25*	Eumusa	*acuminata*	*banksii*	Paliama, PNG067	Papua New Guinea, East Sepik	ITC.0766	NEU0079
26	Eumusa	AAA	Cavendish	Poyo	Nigeria	material to be sent to ITC	NEU0165
27*	Eumusa	AAB	Popoulou/Maia Maoli	Popoulou	Cameroon	ITC.0335	NEU0277
28	Eumusa	ABB	Peyan	Simili Radjah	From india through Zaire	material to be sent to ITC	NEU0357
29	Eumusa	AAA	Gros Michel	Gros Michel	Guadeloupe	material to be sent to ITC	NEU056
30	Eumusa	AS		Wompa, PNG063	Papua New Guinea	ITC.1152	NEU0020
31	Eumusa	AB cv		Kunnan	India, Kerala	ITC.1034	NEU0155
32	Eumusa	AAcv (18)	type P.jari buaya	P. Jari Buaya/BS312	Malaysia, Kelatan, Thai border	ITC.0312	NEU0117
33	Eumusa	AAcv (2)	type P.mas	P. mas / Figue Sucrée	Malaysia	ITC.0653	NEU0108
34	Eumusa	AAB	Pisang rajah	P. Raja Bulu, IDN 093	Indonesia	ITC.0843	NEU0276
35	Eumusa	AAA	Rio	Leite		ITC.0277	NEU0226
36	Eumusa	ABB	Ney mannan	Ice Cream	?	ITC020	NEU0353
37	Eumusa	*acuminata*	*zebrina*	Zebrina	Indonesia	ITC.1177	NEU0029
38	Eumusa	AAcv	Cooking AA	Tomolo, (PNG023)	Papua New Guinea, East New Britain	ITC.1187	NEU0082
39	Eumusa	*balbisiana*	type 1	Honduras	seeds from Honduras	ITC.0247	NEU0049
40	Eumusa	*balbisiana*	type3	Lal Velchi	India		NEU0051
41	Eumusa	ABB	Pisang awak	Namwa Khom, THA011	Thailand	ITC0659	NEU0347
42	Eumusa	AAA	Lujugira/Mutika	Mbwazirume	Burundi	ITC.0084	NEU0222
43	Eumusa	AAA	Lujugira/Mutika	Intokatoke	Burundi	ITC.0082	record to be sent
44	Eumusa	AAA	Ibota	Yangambi KM5	Cameroon	ITC.1123	NEU0212
45	Eumusa	AAB	Plantain	Red Yade		ITC.1140	NEU0244
46	Eumusa	AAB	Nendra padaththi	P. Rajah	Brazil	ITC.0243	NEU0282
47	Eumusa	ABBB		Yawa 2, PNG 072	Papua New Guinea, East New Britain	ITC1238	NEU0384
48	Eumusa	AAB	Mysore	Pisang Ceylan	Thailand	ITC1441	NEU0284
49*	Eumusa			Schizocarpa			
50	Austalimusa			*Musa textilis*		Neu 0001	neu001
51	Rhodochlamys			*Musa ornata*		ITC 0637	

* Not currently available in the field collection of CIRAD Neufchateau, Guadeloupe.

2005-03c: Application of Molecular Markers for Gene Pool Division and Heterosis Estimation under Drought Stress Conditions in Sweetpotato

Principal Investigator
Wolfgang Gruneberg, CIP
Collaborating Scientists:
Marc Ghislain, CIP
Jorge Benavides, CIP
Robert Mwanga, NARI

Mid-Year Report

SSR markers produce easily scorable unique alleles and/or allele combinations, which makes them an ideal system for cultivar identification. CIP produced its own SSR library and has tested these makers together with those developed at the University of Louisiana. Thirty new SSR markers were developed in this project and today a total of 50 SSR markers are available for *Ipomoea batatas.* These markers have been validated for discriminatory capacity and at least 40 of them have good discriminatory capacity.

Activity 2: SSR fingerprints for diversity assessment

All the available germplasm collection at CIP (1300 clones) and 198 parental clones from the breeding population Jewel have been propagated in greenhouse. At present, 450 accessions from the sweet potato germplasm collection at CIP and 100 clones from the breeding population "Jewel" (a total of 550 genotypes) have been screened using 20 SSR markers. This germplasm includes a balanced number of clones from different regions of the world (Central America, South America, the Pacific Ocean, China, Indonesia, South Asia and Sub-Saharan Africa. Moreover, ten accessions from *I. trifida*, which is one of the species most closely related to sweetpotato, has been included in this study. The results from these 550 clones characterised by 20 SSR markers will be recorded as data points and analysed by principal component analysis and cluster analysis. These results will be available around June 15, 2006. After that is completed, these 550 clones will be characterised using 40 SSR markers. These results of greater characterisation will be available at end of August.

Activity 3: Intra- and inter gene pool and crossings (funded by core budget or other projects)

This work will start after all the above mentioned 550 clones have been characterised using 40 SSR markers and after 100 genotypes from the breeding population "ZapSPK" have been characterised using 40 SSR markers. All crossings will be conducted in factorial designs. In each factorial design 16 parents will be used. A total of three of these crossing designs will be carried out: One for inter-gene pool crossings and two for intra-gene pool crossings.

Activity 4: Field evaluations, Heterosis and Genetic Distances (funded by core budget or other projects).

Each gene pool crossing (three gene pools) comprising 16 parents and their corresponding 64 progenies will be planted in one block, two treatments (non-stress, drought stress) and two replications (3 x 320 = 960 plots). This experiment will be conducted at three locations (La Molina, Peru; San Ramon, Peru; and Soroti, Uganda). Heterosis will be determined on the basis of the difference of the family mean and variance to the parental means.

Tangible outputs delivered:

Thirty new SSR markers were developed in this project and today a total of 50 SSR markers are available for *Ipomoea batatas.*

450 accessions from the sweet potato germplasm collection at CIP and 100 clones from the breeding population "Jewel" (a total of 550 genotypes) have been screened using 20 SSR markers.

Deviations from the work plan:
An Update Report presenting the data of the component analysis and cluster analysis of the 550 sweetpotato clones using 20 SSR markers will be presented in June 2006 and the Final Report with data using the 40 SSR markers will be presented in September 2006.

2005-03d: Molecular Characterisation of Tier 2 (orphan) Crops - Yam

Principal Investigator:
R. Asiedu, IITA
Collaborating Scientists:
M. Kolesnikova-Allen, IITA
H. Chair, CIRAD

Mid-Year Report
Most seed tubers of the core set of 393 accessions obtained from last season.s harvest remained dormant until March 2006. They were therefore planted in pots in a screen house during that month. Genomic DNA was extracted from young leaves following plant establishment. DNA quantification using a spectrophotometer is in progress. PCR working conditions for 25 microsatellite markers developed by IRD and CIRAD, France are also being optimised.

Tangible output:
Genomic DNA extracted.

2005-03e: Molecular Characterisation of Tier 2 (orphan) Crops – Lentil

Principal Investigator:
Bonnie Furman, ICARDA
Michael Baum, ICARDA
Collaborating Scientists:
Christian Jung, Universität Kiel
Aladdin Hamwieh, ICARDA

Mid-Year Report
Data extraction of all 1000 samples was completed and analysis of 30 SSR markers is underway. Primer pairs have been grouped into sets of two for more efficient analysis. A total of 10 markers have been completed on all samples. Of the remaining 20 markers, approximately 40% of the work has been completed. We expect completion of all data collection by the end of July. All data will then be aligned for the necessary formats and statistical analyses using diverse computational genomic tools will be carried out. We anticipate completion of the project, including a written project completion report and journal articles by the deadline of December 31, 2006.

Tangible outputs delivered:
- Data extraction of all 1000 samples was completed
- Analysis of 30 SSR markers is underway, 50% of the work is completed

Deviations from the work plan:
Completion of the work has been delayed due to high work volume on the available sequencer.

2005-03f: Molecular Characterisation of Groundnut (*Arachis hypogea* L.) Composite Collection

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborating Scientists:
R Bhattacharjee, ICRISAT
DA Hoisington, ICRISAT
S Chandra, ICRISAT
RK Varshney, ICRISAT
JFM Valls, EMBRAPA
MC Moretzsohn, EMBRAPA
S Leal-Bertioli, EMBRAPA
Patricia Guimarăe, EMBRAPA
David Bertioli, UCB

Mid-Year Report

A groundnut composite collection, consisting of 1000 accessions (850 from ICRISAT and 150 from EMBRAPA), was established using the available phenotypic characterisation and evaluation data. The composite collection at ICRISAT comprised of 184 mini-core accessions (Upadhyaya et al., 2002), along with other landraces, breeding lines, genetic stocks, wilds and four control accessions. The objective of this research was to fingerprint the composite collection (850 accessions at ICRISAT and 150 at EMBRAPA) using a selected set of twenty polymorphic SSR markers (10 markers contributed from ICRISAT and 10 from EMBRAPA).

At ICRISAT, DNA was extracted from all the 960 accessions (more accessions were included to account for possible missing values) following a high-throughput procedure. Each extraction plate consisted the control cultivars to optimise the extraction procedure. DNA was quantified to a working concentration of 5ng/µl. Twenty SSR primers were initially selected and screened on the 184 mini core accessions to identify ten most polymorphic markers for fingerprinting the remaining accessions of the composite collection. The selected 20 polymorphic SSR markers (10 from ICRISAT and 10 from EMBRAPA) were optimised for PCR reactions following Taguchi method. A fluorescent-based multiplex genotyping system was then used to generate four multiplexes, which were used to fingerprint the composite collection. The amplified PCR products were separated by capillary electrophoresis in an automated system using ABI 3700. SSR fragment sizes were called to two decimal places using the Genotyper v 3.7 software. Probably due to tetraploid nature of groundnut crop, more than two peaks of almost equal size or 80% of the highest peak were observed. Sometimes peaks of equal sizes were observed at two different locations/positions. All the peaks, which were of equal height or 80% of the highest peak, were recorded in a raw data file. A criteria/macro was then developed to remove the peaks that were less than 80% in peak height as compared to the highest peak and the data was reduced to two-peak situation (raw data file being retained for future reference).

The allelic data was analysed following allele binning algorithm of Idury and Cardon (1997), written in a C programme at ICRISAT and called as “Allelobin”. Less than 5% missing data (i.e. marker x genotype) was recorded in the dataset and all the markers produced allele size that was expected on the basis of repeat motif of each of the SSR markers. To consider only good quality markers in the final data analysis, we used 21 markers (10 from ICRISAT and 11 from EMBRAPA) for genotyping. Out of 21 markers, 5 showed poor quality index (Table 1), which may be due to high missing vales recorded for these markers. Data from EMBRAPA on 150 accessions is awaited.

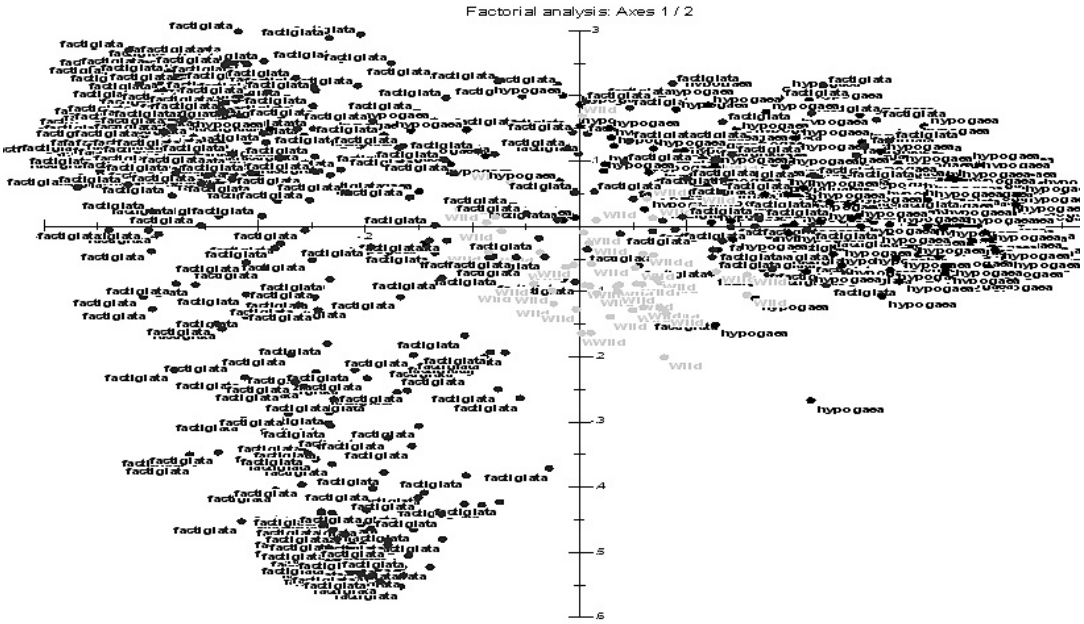
A preliminary analysis of data on 900 accessions and 21 SSR markers was carried out using DARwin 5.0 Structure programme (Perrier et al., 2003) to determine the population structure of the composite collection. The software removes all those accessions/markers that have high missing values to create the dissimilarity matrix. We chose to remove accessions having higher missing values and finally 900 accessions were considered for principal coordinate analysis considering the taxonomical classification of *Arachis*, i.e. at the level of two subspecies and six botanical varieties. The analysis detected a total of 506 alleles, ranging from 6 (7H6) to 47 (5D5) with a mean of 24.1 alleles per locus and mean PIC value of 0.797 (ranging from 0.483 to 0.923) (Table 1).

Table 1. SSR primers used in the study with information on their repeat units, quality index, number of alleles and PIC values

Primer	Repeat Unit	Quality Index	No. of Accessions Genotyped	No. of Alleles	PIC Values
1B9	(GA)19	0.2092	916	26	0.84
2D12B	(TAA)16	0.4308	901	21	0.882
7H6	(CTT)12	0.2601	928	6	0.483
8E12	(TTG)6(TAA)15	0.224	942	20	0.675
13E9	(TAA)16	0.2448	914	14	0.635
5D5	(GA)32	0.3361	910	47	0.919
15C12	(TAA)28	0.2964	868	22	0.844
17E3	(CTT)15	0.2397	919	13	0.756
18C5	(TAA)23	0.1767	901	15	0.799
19B1	2	0.254	937	29	0.688
TC1A02	(TC)35	0.4902	903	27	0.899
TC4F12	(CT)23	0.202	910	27	0.886
TC6E01	(GA)22	0.4695	923	25	0.887
TC6H03	(AG)21	0.4706	898	37	0.895
TC11H06	(AG)34	0.3032	908	27	0.923
TC1E01	(GA)29	0.1659	949	23	0.772
TC11A04	(CT)16 + (CT)33	0.2243	938	23	0.889
TC7H11	(AG)18	0.4651	912	30	0.864
TC9F10	(AG)31	0.3449	910	29	0.862
TC3E02	2	0.2935	918	22	0.615
TC6G09	2	0.2549	951	23	0.716

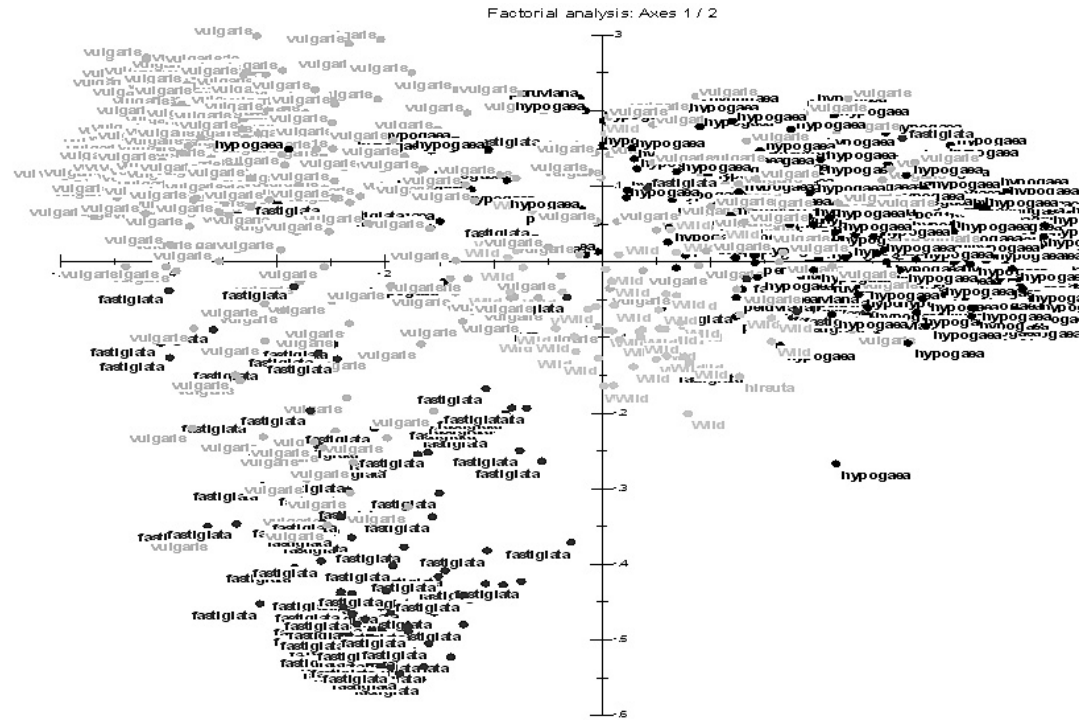
Principal coordinate analysis (PCoA) using DARwin 5.0 on subspecies revealed that both *hypogaea* and *fastigiata* formed distinct clusters however, a number of *fastigiata* accessions also grouped with *hypogaea* types (Fig. 1), which maybe attributed to the geographic origin of these accessions. Further, these accessions belong to *A. fastigiata* subsps. *fastigiata* var. *peruviana*, whose geographic origin is same as that of *hypogaea* types i.e. America. This is also confirmed when PCoA was performed considering botanical varieties (Fig. 2). The wild types formed a different cluster in both the cases and grouped with *hypogaea* types, indicating a close relation between them (Fig. 1 & 2)).

Figure 1. Factorial analysis of 900 accessions with 21 SSR markers at subspecies level.



fastigiata: Red; *hypogaea:* Black; Wild: Light Green

Figure 2. Factorial analysis at the level of botanical varieties.



fastigiata: Red; *hypogaea:* Black; *peruviana:* Blue; *vulgaris:* Orange; *hirsuta:* Pink; *aequatoriana:* Magenta; Wild: Light green

Further analysis of data is in progress to fully understand the genetic structure of groundnut composite collection. The results from genotypic data will be used to determine the reference set of 300 accessions for future use.

To ascertain the quality and position of the SSR markers, these will be checked on 15-20 plants in each of four F_2 populations, whose parents have been included in the composite collection. Only those SSR markers that showed polymorphism on the parents will be checked on the F_2 population. This will confirm the location/position of these SSR markers and would ensure appropriate peak calling. Data generated from the fingerprinting will be then subjected to statistical analysis using different computational tools.

Tangible outputs delivered:

- Genotyping data on 960 accessions with 21 (10 from ICRISAT and 11 from EMBRAPA) SSR markers.
- Information on allele size available
- Preliminary data analysis completed.

Further Plan of Work:

- Data analysis will be carried out to determine the genetic diversity and population structure.
- Mapping the SSR markers in a F2 population to ascertain the position of the markers.
- A reference collection of 300 diverse accessions will be identified.

Deviations from the work plan:
None

2005-03g: Molecular Characterisation of Tier 2 Crops: Coconut

Principal Investigator
Patricia Lebrun-Turquay, Agropolis-CIRAD
Collaborating Scientists:
L Baudouin, Agropolis-CIRAD
A. Berger, Agropolis-CIRAD

Mid-Year Report
The objective of the project was to characterise 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 (NB twelve in fact, because one has been withdrawn due to insufficient stability). The molecular characterisation consisted in adding 9 new loci and complementing the data with 456 new accessions.

The diversity was analysed and led to corroborating and refining earlier conclusions on the structure of coconut diversity and its interpretation in terms of crop history.

The analysis is detailed in the attached GCP project report.

In addition, as part of two other GCP projects (see below), ten more loci have been surveyed (for linkage disequilibrium assessment) as well as 215 trees from a breeding programme in Vanuatu (with the view to assocation mapping in the course of breeding).
The whole data set has been delivered to the GCP repository in July 2006.The final file consists in 1215 accessions x 30 SSR loci.

Tangible outputs delivered:
SSR data for 1215 coconut trees and 30 loci

The title of the dataset is : SSR genotyping data on coconut
This is a compilation of data from three projects:
A. Projet comm2005-03g consisted in genotyping 1000 accessions (individual trees) for 20 reference SSR loci distributed in the genome and analysing global structure. The PI is Patricia Lebrun-Turquay
B. Project comm2005-SP1-SPL is part of the 'discretionary' undertaking using the budget attributed to JC Glaszmann (SP1 leader) as part of the SPL arrangement; it has no formal number. It consisted in genotyping the same 1000 accessions using 10 more loci chosen to be closely linked to 10 of the reference loci; this serves for assessing the level of linkage disequilibrium in coconut germplasm. The PI is JC Glaszmann, coPIs Patricia Lebrun-Turquay and Luc Baudouin
C. Project comm2005-08, coconut part. This is part of a larger project coordinated by Carmen de Vicente which aims at assessing diversity and LD in materials from a breeding programme run in Vanuatu. It consisted in genotyping 215 trees for the same loci as above, ie 20 + 10. The PI is Carmen de Vicente, coPIs for coconut Luc Baudouin and Patricia Lebrun-Turquay

Deviations from the work plan:
No deviation.

2005-04: Validation of Diversity Arrays Technology (DArT) as a Platform for Whole-Genome Profiling in Orphan Crops

Principal Investigators:
Andrzej Kilian, DArT P/L
Carmen de Vicente, IPGRI
Jean Christophe Glaszmann, Cirad-Agropolis
Collaborating Scientists:
Eric Huttner, DArT P/L
Peter Wenzl, DArT P/L
Ange-Marie Risterucci, Agropolis-Cirad
Ken McNally, IRRI
Claire Billot, Agropolis-Cirad
Michael Baum, ICARDA
M Fregene, CIAT
Nicolas Roux, IPGRI-INIBAP
Patricia Lebrun, Agropolis-Cirad
Chandrika Perera, Coconut Research Institute, Sri Lanka
Prapit Wongtiem, Rayong Field Research Station, Thailand

Mid-Year Report

Table 1. Description of available crop data sets up to date.

Crop	SSR data	File description	Submmission date Submitted by..	DArT data	File description	Submmission date Submitted by..
Wheat	X	94 individuals x 73 SSR markers	June 13 2006[1] François Balfourier	X	94 individuals x 714 DArT markers	June 13 2006 François Balfourier
Rice				X	90 individuals x 519 DArT markers	February 15 2006 BrigitteCourtois
Sorghum	X	90 individuals x 34 SSR markers	May 5 2006 Sophie Bouchet	X	92 individuals x 520 DArT markers	May 3 2006 Sophie Bouchet
Coconut	X	192 individuals x 21 SSR markers	June 8 2006[2] Patricia Lebrun	X	223 individuals x 337 DArT markers	April 25 2006 Chandrika Perera
Cassava	X	134 individuals x 36 SSR markers	June 15 2005 Paula Hurtado	X	124 landraces x 424 DArT markers 19 wild accessions x 1710 DArT markers	May 30 2006 Andrzej Killian / Prapit Wongtiem
Musa	X	184 individuals x 12 SSR markers	June 9 2006 Isabelle Hyppolyte	X	187 individuals x 463 Pst1/BstNI DArT markers 187 individuals x 373 Pst1/TaqI DArT markers	June 9 2006 Ange-Marie Risterucci

[1] The first SSR data set corresponding to Wheat was sent by Francois Balfourier on November 16 2005. This file was corrected and the updated version was sent on June 13 2006.
[2] The first SSR data set corresponding to Coconut was sent by Patricia Lebrun on April 11 2006. This file was corrected and the updated version was sent on June 8 2006.

Regarding the Rice SSR data set (the only file not yet received), Ken McNally at IRRI reported last on June 8 2006. Some data still missing to complete the genotyping task.

The proposal to compare DArT vs SSR for all crops was as follows:

1) Analysis of SSR and DArT data (as in the original sets received) and analysis of a joint matrix of SSR and DArT data, in both cases using NTSYSpc vs 2.10 and Popgene vs 1.31:

1. Similarity matrix calculation based on Jaccard´s or Nei's coefficient using SSR or DArT data respectively
2. Principal coordinate analysis (PCoA) based on singular vectors and values for each data set
3. Cluster analysis (UPGMA) and Neighbor Joining based on Euclidian distances for each data set
4. Definition of similarity limits to differentiate the groups defined by cluster analysis
5. Estimation of genetic diversity parameters:
    a. Single / Multi-Population Descriptive Statistics
        - Chi-square test for Hardy-Weinberg equilibrium/locus (P)
        - Overall Allele frequency/locus
        - Overall Gene Frequency (allele/locus)
    b. Summary of Genic Variation Statistics for <u>all loci</u>
        - na: Observed number of alleles (mean/St.Dev)
        - ne: Effective number of alleles [Kimura and Crow (1964)]
        - I: Shannon's Information index [Lewontin (1972)]
        - h: Nei's (1973) gene diversity
    c. Summary of Heterozygosity Statistics for <u>all loci</u>
        - Observed homozygosty
        - Observed heterozygosity
        - Expected homozygosity using Levene (1949)
        - Expected heterozygosity using Levene (1949)
        - Nei's (1973) expected heterozygosity
        - Average Heterozygosity
        - Number of polymorphic loci
    d. Wright's (1978) fixation index (Fis) as a measure of heterozygote deficiency or excess/locus
    e. The Ewens-Watterson Test for Neutrality/locus
    f. Summary of F-Statistics and Gene Flow for <u>all loci</u>
        - F_{is}
        - F_{it}
        - F_{st}
        - Nm: Gene flow estimated from $F_{st} = 0.25(1 - F_{st})/F_{st}$
    g. Nei's (1972) Original Measures of Genetic Identity and Genetic distance
    h. Nei's Analysis of Gene Diversity in Subdivided Populations
        - H_t
        - H_s
        - G_{st}
        - Nm: estimate of gene flow from G_{st} or G_{cs}. E.g., $Nm = 0.5(1 - G_{st})/G_{st}$

B. Comparison of SSR and DArT in separate data sets based on NTSYS clustering:

1. Compare the clusters (trees) by means of a Chi-square test of independence (SAS version 9.1)

Hypotheses:
Null: There is no association between DArT and SSR results (SSR results are independent of DArT results)
Alternate: There is an association between DArT and SSR results

P<0.05 will reject the null hypotheses

Clarifications about individual names and numbers as well as DArT and SSR markers were sought for Coconut, Sorghum, Wheat, Musa and Cassava before starting the analysis.

Table 2. Description of data files used for statistical analysis

Crop	SSR data	DArT data	SSR+DArT data
Wheat	94 individuals x 73 SSR markers (686 alleles)	94 individuals x 714 DArT markers	94 individuals x 1400 markers
Sorghum	90 individuals x 34 SSR markers (360 alleles)	90 individuals x 520 DArT markers	90 individuals x 880 markers
Coconut	191 individuals x 21 SSR markers (243 alleles)	191 individuals x 337 DArT markers	190 individuals x 580 markers
Musa	169 individuals x 12 SSR markers (182 alleles)	169 individuals x 836 DArT markers	169 individuals x 1018 markers

Up to date, the following has been performed: PCoA, UPGMA, NJ and the two similarity limits defined for SSR, DArT and SSR+DArT in Coconut and Sorghum, as well as PCoA, UPGMA and NJ for SSR, DArT and SSR+DArT in Wheat and Musa.

Since the first two data sets available (DArT and SSR) were those corresponding to coconut, analysis for these data sets were used as a model for the comparison between DArT and SSR markers in the other crops.

Two similarity/dissimilarity limits were defined for coconut based on the results of UPGMA. With a dissimilarity of 0.66 for SSR, 0.44 for DArT and 0.44 for SSR+DArT two groups were defined in the three matrices. With a dissimilarity level of 0.33, five groups were defined in all cases. These results were used to estimate the genetic diversity parameters in the different populations identified by each marker system and also to compare both marker systems by means of a Chi-square test of independence. Comparisons were done between the groups obtained (two and five as mentioned above) with a confidence level of 95%.

The independence test rejects the null hypotheses in all the comparisons indicating that there is *association* between SSR and DArT results in coconut.

The initial proposal to do the comparison included the correlation between similarity matrices using the Pearson product-moment, which for coconut matrices resulted in 0.03. The differential nature of both marker systems makes it impossible to compare them in a single analysis.

Sorghum, Wheat and Musa data files are being processed to complete the analysis as done in Coconut. Cassava datasets will be analysed later and Rice datasets will be processed as soon as SSR data file is received.

Tangible outputs delivered:

- DArT and SSR data for Sorghum, Coconut, Wheat, Musa and Cassava
- DArT data for Rice
- PCoA, UPGMA, NJ and similarity limits definition for Coconut and Sorghum
- PCoA, UPGMA, NJ for Wheat and Musa
- Genetic diversity parameters estimation for Coconut
- Comparison between SSR and DArT (Chi-square test of independence) for coconut

Deviations from the work plan:

- Most datasets (SSR and DArT) missing in May were received in June. Only Rice SSR are missing.
- Some files needed amendments (wheat-SSR and coconut-SSR) before or during the analysis, so some steps in the individual analysis of SSR had to be repeated.
- Some analyses are in progress for sorghum, wheat, and *Musa*, since the three datasets were processed at the same time in order to have the PCoA, UPGMA and NJ that allow doing the comparison between marker systems. Those final analysis will be completed by August 1st
- Analysis for cassava datasets will be ready by August 10th
- The inferences based on the results for coconut, sorghum, wheat, *Musa* and cassava will be done in detail as soon as the analyses are completed for each crop.

A final report will be sent by August 20th.

2005-05: Assessing Ecotilling as a Methodology for Targeted Genotyping and SNP Discovery

Principal Investigators:
Kenneth McNally, IRRI
Claire Billot, Agropolis-CIRAD
Collaborating Scientists:
N. Ruaraidh Sackville Hamilton, IRRI
M. Deu, CIRAD
I. Hippolyte, CIRAD
F.-C. Baurens, CIRAD
J.-F. Rami, CIRAD

Mid-Year Report

1) The manuscript describing the modified ecotilling procedure was withdrawn from Plant Molecular Biology Reporter during September 2005 in January 2006, due to the slow response of the journal. The paper was revised to include more applications for breeding work, and this new version was submitted to Molecular Breeding on March 14, 2006.
2) For rice, ecotilling using the agarose based procedure has been accomplished on 10 candidate gene loci with contrasts to both indica (IR 64) and japonica (Nippon-bare). Most loci show frequent indica/japonica mismatches. The coding region for one of the TPP loci is conserved between indica and japonica types with only infrequent (rare) mismatches detected. Representative accessions have been chosen for sequencing during June 2006.

Primer Name	Candidate Gene	Chr	Accessions tested	
			vs. IR64	vs. Nippon-bare
DREB2	AP2 domain TF	1	1156	1156
ERF3	ethylene responsive factor AP2 domain TF	1	1536	1536
ADF_2a	actin depolymerising factor	2	1536	1536
ADF_2b			1251	1251
BZIP	bzip DNA binding protein	1	1473	1473
EXT	extensin	10	1156	1156
SUC	sucrose synthase	7	1536	1536
TPP_2a	trehalose 6-phosphate phosphatase	2	1536	1536
TPP_2b			1156	1156
MAPK_ 7a	mitogen activated protein kinase	7	1156	1156
MAPK_ 7b			1536	1536

Ecotilling has also been accomplished on a panel of 95 diverse wild species accessions representing 23 Oryza species using the LiCor-based system for 11 primer pairs at 8 candidate gene loci. For 5 candidate genes, the agarose based methods was compared to the LiCor. Within the AA genome species, the agarose based method can easily distinguish mismatches. Outside of the AA genome, banding patterns become complex and the technique is similar to fingerprinting by AFLP or RAPD. Ecotilling on a set of 190 *O. glaberrima* samples from the GCP composite collection and WARDA nominations (in collaboration with Dr. Marie-Noelle Ndjiondjop) was initiated in January 2006, using 21 primer pairs representing 12 drought candidate genes against two contrasts, Nipponbare and the *O. glaberrima* variety CG14 (IRGC 96717). Presently, 10 primer pairs have been completed on the *O. glaberrima* panel. A manuscript on the use of ecotilling outside of *Oryza sativa* is being prepared.

3) Training on agarose-based ecotilling was conducted at IRRI for 4 outside participants (3 from the Philippines during March 2006 and one from Taiwan during April 2006).
4) Specific sorghum primers corresponding to the orthologous sequences of the genes Ecotilled on rice have been designed. For each rice gene, a set of homologous sorghum EST contigs sequences (obtained from MagicDB http://fungen.org/genediscovery/) and sorghum methyl filtered genomic sequences (Orion Genomics; obtained from Genbank) was identified using BLAST. Each rice gene was aligned with the corresponding sorghum EST contigs sequences and sorghum methyl filtered genomic sequences using the sim4 software. Testing of the primer pairs on sorghum and Ecotilling using the modified procedure (T6/SP7 tailed primers and celery juice extract) are ongoing.

Tangible outputs delivered:

1) Ecotilling on agarose accomplished for 11 candidate gene loci on 1156 to 1536 *O. sativa* germplasm accessions with contrasts to both indica and japonica, a panel of 95 wild relatives has been ecotilled for 8 candidate gene loci on the LiCor system, and 10 primer pairs have been used for ecotilling on 190 *O. glaberrima* accessions.
2) Specific sorghum primer pairs corresponding to the 11 rice candidate gene loci.

Deviations from the work plan:

Sequencing has been slightly delayed, but is now on target for rice during the coming months. Drought phenotyping data from 3 seasons (DS2004, DS2005, and DS2006) on ~1500 accessions

of the 1536 O. sativa genotyped will be available for association tests with the confirmed SNP data.

2005-06: Supporting Emergence of Reference Drought Tolerance Phenotyping Centres

Principal Investigador:
Frederico Ozanan Machado Durães, Embrapa
Collaborating Scientists (EMBRAPA):
Paulo E. P. Albuquerque
João Herbert Moreira Viana
Reinaldo Lúcio Gomide
Camilo de Lelis Teixeira Andrade
Elto Eugenio Gomes e Gama
Sidney Netto Parentoni
Fredolino Giacomini dos Santos
Paulo César Magalhães
Jurandir Vieira Magalhães
Cláudia Teixeira Guimarães
Newton Portilho Carneiro
Cleber Morais Guimarães
Orlando Peixoto de Morais
Beatriz da Silveira Pinheiro
Maria José Del Peloso
Natoniel Franklin de Melo
Luiz Balbino Morgado
Edson Alves Bastos
Francisco Rodrigues Freire Filho
Walter Quadros Ribeiro Jr.
Maria da Glória Trindade
Ana Christina Sagebin Albuquerque

Mid-Year Report
A Phenotyping Platform Supporting Breeding and Genomics Programmes of Cereals and Legumes at Embrapa – Brazil. The GenerationCP, SP1-Commissioned Research Project (2005-2007): "Supporting Emergence or Reference Drought Tolerance Phenotyping Centres" - GCP Drought Phenotyping Network Project (GCP DPN Project #6)

Summary of Technical Progress (by task in the Work Plan)
This first year (*June 6th 2005 through June 30th 2006*) of GCP DPNetwork Project operations has been dedicating great attention on team organisation and improving the sites-specific qualification for the cereals and legumes phenotyping network aiming drought tolerance.

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

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

Also, during Year 1 (*June 6th 2005 through June 30th 2006*) the activities operacionalisation of GCP DPN Project, has been realised using the sites-specific experimentals with the following task technical sponsors:

Site-Specific Experimental for Drought Tolerance	Sponsor (Task co-PI)	Embrapa – National Research Unit
Sete Lagoas - MG	Reinaldo Lúcio Gomide	Embrapa Maize and Sorghum
Janaúba - MG	Paulo Emílio Pereira de Albuquerque	Embrapa Maize and Sorghum
Santo Antônio de Goiás - GO, and, Porangatu - GO	Cleber Morais Guimarães	Embrapa Rice and Bean
Teresina - PI, and, Site Parnaíba - PI	Edson Alves Bastos	Embrapa Mid-North Agriculture
Planaltina - DF, Santo Antonio de Goiás - GO, Passo Fundo – RS	Walter Quadros Ribeiro Jr.	Embrapa Wheat / CPAC-Savannah, including activities on Embrapa-Rice and Bean-CNPAF and Wheat-CNPT sites
Petrolina - PE	Luiz Balbino Morgado	Embrapa Tropical Semi-Arid

The site-especific experimental (SSE) areas for drought tolerance phenotyping in cereals and vegetables are being characterised in detail for soil properties (chemical composition, texture, structure, bulk density, soil water and porosity, nutrient content variability, biological indicators), climatic, and genotypes.

1) **Site-Especific Experimental for Drought Tolerance Phenotyping in Brazil Regions** (GCP DPN Project. **Year One Technical Report**, *from June 6th 2005 through June 30th 2006*)

Brazil - UF	**Site-Specific / Local**	**Longitude (W)**	**Latitude (S)**	**Altitude (m)**	**Crop Species**	**Genetic Material(4) / Phenotyping Strategies**
MG	**Sete Lagoas** (2)	-44.2467	-19.4658	761	maize, sorghum	Access and elite/ Preliminar
MG	**Janaúba,** North region of MG (3)	-43.3089	-15.8025	533	maize, sorghum	Access/ Preliminar
GO	**Planaltina,** DF (3)	-47.6142	-15.4528	944	wheat	Access/ Preliminar
GO	**Santo Antonio de Goiás,** GO (2)	-49.1711	-16.2811	823	rice, common bean, wheat	Access, elite, segregation/ Preliminar
GO	**Porangatu,** Nort region of GO (3)	-49.1486	-13.4408	396	rice	Access/ Preliminar
PI	**Teresina,** PI (3)	-42.8019	-5.0892	72	maize (after Aug 05) sorghum (after Jan 06)	Access/ Preliminar
PI	**Parnaíba,** PI (3)	-41.7767	-2.9047	5	maize (after Aug 05) sorghum (after Jan 06)	Access/ Preliminar
PE	**Petrolina,** PE (3)	-40.5008	-9.3986	376	sorghum, maize, cowpea	Access/Preliminar

(1) Site-Specific for Drought Phenotyping Network in Brazil Regions, (2) Sites of Excellence, (3) Sites of Reference

Precision Experimental Sites-Specific for Drought Tolerance Phenotyping in Brazil Regions:

The research programmes standard methodologies regarding drought tolerance studies do not take into account problems in the site-specific areas related to the variability of some soil properties (physical, chemical and biological attributes, mainly the soil water dynamics in the region of the effective depth of the plants root system), and climatic parameters (rainfall, air temperature, wind speed). This variability has been a challenge in the experimental selection process of tolerant genotypes to drought stresses. Although there are different layouts for field plots in the site-specific area, the mentioned effect of spatial and temporal variability normally is not identified, characterised and much less controlled. Selection, implantation, characterisation and installation of precision phenotyping experimental sites-specific for tolerance and adaptation of cereals and vegetables germplasm to abiotic stresses of water is already established with defined accurate geographical coordinates and topograph survey with a precision differential global position system (DGPS).

Some soil and environment attributes variability were determined on basis in methods and techniques of precision agriculture, DGPS, and geographical information system (GIS), which are being used in the data acquisition, storage, treatment, analysis and visualisation of the data in

each site. The spatial variability of soil physics (texture, structure, macro and micro porosity, apparent and real density, soil moisture retention curves, water infiltration rate) and chemistry properties (organic matter, fertility, some micro nutrients) was evaluated by means of topograph survey and division of the sites areas in grids of 25 m x 25 m, utilising a accurate survey laser total station Topcon Hiper and DGPS, on SAD-69 datum and UTM projection (south zone 23, 48 W a 42 W) basis. Samples of the referred soil properties were collected in the grid intersections. Soil properties contour maps were obtained by interpolation with geostatistic adjusted models (krigagem). These maps were used to divided the sites-specific areas in uniform zones for abiotic water stresses studies in cereals and vegetables genotypes. In each selected site area, the water table was deep in order to avoid soil water capilarity effect in the genotypes root systems, and it was identified high and low points in order to avoid drainage problems.

Climatic condition was characterised and hydrological water balance (Thornthwaite & Mather) was determined in each site area with a 15 to 50 years data series provenient from standard weather stations (Brazilian National Institute of Meteorology – INEMET). An automatic weather station was installed in each site area to register the following microclimatic parameters: air temperature, air relative humidity, solar radiation, wind speed and direction, and rainfall.

Protocols, Methods and Techniques of Controlling and Monitoring Plants and Soils Water Stress Levels in Precision Experimental Sites-Specific for Cereals and Vegetables Genotypes Characterisation: Drought Tolerance Phenotyping:

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

Irrigation systems and hydraulic devices were selected, designed, installed, and evaluated in all precision experimental sites-specific for accurate control and management of water application in the drought tolerance phenotyping trials. Conventional sprinkler (low to medium service pressure with sectorial sprinklers) and continuously moving straight lateral or linear-move schemes are the main two systems utilised in the sites areas. Localised drip or trickle system is been using in a few sand soil sites. In each site, the irrigation system testing and performance was evaluated and water flow rate (discharge) and applied water depths were measured (by genotypes). A standard procedure was defined for the collection of irrigation system data such as pressure, flow rate, radius of throw and spacing of sprinklers or emitters. In each genotype field plot, collectors are being placed (layout – transversally to the crop rows) for measurements of the water depths applied in the irrigations. The uniformity of distribution of the water in the irrigated plot was set to be equal or greater than 95 % (Christiansen Uniformity Coefficient or CUC = 95%). Some hydrometers are being coupled to the irrigation systems main lines. Every where, irrigation water application rate is being set to be lower than basic soil saturated water infiltration rate in order to avoid surface runoff.

The irrigation management and timing criteria (water depths- how much? and when to apply irrigation?) is being computed by means of a spread sheet (Excell) macro programme algorithm

according to soil water balance and modified Penman-Monteith equation for reference evapotranspiration (ETo) methodologies. The soil water availability and automatic weather station data are being used in the irrigation management procedures. The crop evapotranspiration (ETc) is being determined by multiplying ETo for each genotype crop coefficient (Kc). After planting, an uniform irrigation is being used in each site to assure a good germination and stand formation (with 100% replacement of the soil profile available water (AW) in the effective root system zone, complete replacement of ETc - non water stressed).

The water stress treatments or different ETc replacement level is being defined for each genotype at pre-defined crop growth phases, according to breeders and physiologists indication in order to establish the water stress with different application of water depths in the plots.

A standard procedure was established to calibrate and install the equipments and sensors of automatic weather stations in order to register automatically the microclimatic surface parameters locally for drought tolerance phenotyping purposes. The following automatic weather station configuration is being used: air temperature and relative humidity, solar radiation, rainfall or precipitation, wind speed and direction, evaporation of the water in the class A pan). The hardware and software are descripted for automatic microclimatic surface data acquisition in real time. The calibration, installation, and operation of equipments and sensors for soil water status registration in different soil layers is described for soil water dynamics registration. The gypsum blocks sensors are suitable for continuous monitoring of soil water content. The principle of its operation is that the electrical resistance of (electrodes embedded in) a porous block is proportional to its water content. Thus, the wetter a block is, the lower the resistance measured across two embedded electrodes. This type of sensor is suited to various irrigation applications mainly with soil water stress. These sensors can be left in field to automatically monitor continuously soil moisture, allowing many replicates. The time domain reflectometers (TDR) equipment combine the knowledge of the waves signal propagation velocities in the presence of water in the soil medium which affects the speed of these electromagnetic waves (slows them down sightly). The accuracy of TDR measurements depends on precise measurement of time and precise calibration with the relative volumetric content of water around the probe. The sensors to register soil water content are being installed in at least four soil depths (15, 30, 50, and 80 cm).

[4] **Drought Phenotyping Strategies and Total Genetic Material (genotypes per crop specie) per Year 1 (June 2005 – Semesters 1 and 2, through June 2006 - Semester 1), for all sites-specific in Brazil Regions.**

Genotypes [4]		**Actions Strategies of Phenotyping** [5]												**Notes** [6]
Crop Specie	**Type**	**Preliminar**				**Intermediate**				**Advanced**				
		Year 1 (Jun 2005 to Jun 2006)	**S1, 2005**	**S2, 2005**	**S1, 2006**	**Y1**	**S1**	**S2**	**S1**	**Y1**	**S1**	**S2**	**S1**	
Maize (201) **313**	Access	221	0	221										
	Selected material	36	36	0										
	Elite	6	0	6										
	CNPMS Maize Core Collection	25+25	0	50										
	GCP- Reference Sample	in definition	0	0										Partnership GCP and WPM Project
Sorghum (84) **64**	Access	64	64	64										
	Selected material	0	0	0										
	Elite	20	20	0										
	CNPMS Sorghum Core Collection	0	0											
	GCP- Reference Sample	in definition	0											Partnership GCP and WPM Project
Rice (80+144) **(81+194+ 2+51)**	Access	0	0	0										
	Selected material	81+194+2	81	194		51	39	12						
	Elite	14	14	0										
	CNPAF Rice Core Collection	0	0											
	GCP- Reference Sample	in definition	0											Partnership GCP and WPM Project
Wheat (95) **(150, 5)**	Access	200 fixed homozigotic 150	0	150										
	Selected material	(95) greenhouse (with 80, field)	0 (80)	5										
	Elite	0	0											
	CNPT Wheat Core Collection	0	0											
	GCP Reference Sample	0	0											-
Common bean (144) **177+117+ 888**	Access	76+97+4	177	0										
	Selected material	888+117	888	117										
	Elite	0	0											
	CNPAF Bean Core Collection	0	0											
	GCP- Reference Sample	0	0											-
Cowpea (80) **(114)**	Access	84 (field)	84	84										
	Selected material	30 (field)	0	30										
	Elite	0	0											
	CPAMN Cowpea Core Collection	0	0											
	GCP- 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**: characterised 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, **GCP- Reference sample:** genetic material/crop specie selected from GenerationCP Collection (negotiable and according to leader SP1 and SP3).

[5] **Phenotyping strategies** – genetic material characterisation 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 phenotipic parameter and controlled environments and with water stress, typical to conventional breeding and with emphasis to access; **Intermediate phenotyping:** characterisation using auxiliary descriptors techniques, and emphasis on selected material; **Advanced phenotyping:** specific characterisation 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 Year 1 (2005-Semester 2 and 2006-Semester 1).

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

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

In the Reference Site of Sete Lagoas, MG, it were carry out essays under greenhouse conditions, using six maize inbred lines (drought tolerant and sensitive) and different water regimes, evaluating several soil and plant traits, aiming evaluate maize performance and soil water dynamics. During Dec 2005 until May 2006, four maize, 5 sorghum and 5 rice selected genotypes were carry out during wet season, for attending GCP DPNetwork and WPModelling collaborative projects.

The initial genetic background of **maize materials** to be used in different trials for characterisations as intermediate phenotyping [object of GCP DPN Project, and to attend GCP WPModelling collaboration, during wet season (no irrigation), and dry season (differential irrigation – with and without water stress during typical critical period per each crop specie)], after November 2005, as follow:

Maize Genetic Material	**Type**	**Origin**	**Flowering Cycle**	**Adaptability**	**Characteristic under previous evaluation**	**Reference (*)** *(GCP DPN x WPM essays)*
BRS 1010	Single cross	L3 x L 228-3	Early (62 days)	Broad	Phosphorus	
L 3	Inbred line	Commercial hybrids flint type	Early (66 days)	Broad	Al^{3+} and P	**G1**
L 228-3	Inbred line	Synthetic BR 106	Early (63 days)	Broad		**G2**
BR 106	Broad base population	Tuxpeño	Early (65 days)	Broad		**G3**
BR 105	Broad base population	Caribbean	Early (65 days)	Broad		
Synthetic TS	Narrow base population	Lines Tuxpeño and Caribbean	Early (64 days)	Broad	Drought	**G4**
Synthetic PN	Narrow base population	Lines Tuxpeño	Early (66 days)	Broad	Nitrogen and Phosphorus	

Note: (*) Diverse from Embrapa Maize and Sorghum. We have already acquired enough amounts of seeds of these materials, which will be used for planting in the sites of Sete Lagoas and Janaúba-MG, Teresina or Parnaíba-PI, and Santo Antonio de Goiás-GO.

Sixty four (64) sorghum genotypes were planted in Janaúba, MG, in two areas with and without water stress. There were used two lattice of 8x8 with 3 replications each. Plots of two 5,0 m rows each, spaced 0,50 m, with 200 thousands plants.ha-1`. Dripping irrigation was used and a water stress was applied in the post-flowering time. Time to 50% to flowering, plant height, panicle harvest index (PHI), stay-green (% death leaf), 1000-grain and total grain weight and grain yield were evaluated. According grain yield, 34 genotypes had best performance (upper general average), under water stress, and were classified as tolerant, the genotypes number 64, 37,57, 43, 60, 33, 49, 27, 61, 55, 44, 35 and 13, with water stress index around de average. The genotypes 46, 17, 48, 26, 6, 36, 7, 51, 47 e 14 were very influenced by water stress imposed. The tolerant

genotypes showed a range of percent death leaf between 36 a 71% (55% of average). The sensitive genotypes showed variation of percent death leaf between 35 a 71% (53% of average). The inbred line B35, considered "stay green" standard, showed average of 25%, but with grain yield below of average for both conditions (with and without water stress).

The potential sorghum inbred lines aiming studies of drought tolerance phenotyping are: a) Tolerant sorghum inbred lines: 0025038, 0025530, 9618116, 9503086, 156-P-5-2-1, B35; b) Sensitive sorghum inbred lines: BR 008, Tx 623, 9910010, 9910272, 0025362. The sorghum inbred lines BR008, Tx623 e 156-P-5-2-1 are introduced elite materials, and the other are selected inbred lines from Embrapa sorghum breeding programme.

In both Teresina and Parnaíba, PI site-specific experimental two (02) sets (Caatinga and Cerrado) of maize materials from Embrapa- Maize "Core Collection" were planted, with and without water stress, using sprinkle irrigation. A lattice design 5 x 5 with 2 replications were used, and water stress was applied in the pre- and pos-flowering time. In Teresina, PI the four experiments were sowed on Aug 19, and, in Parnaíba, PI on Oct 03, 2005. Data are been collected for several plant traits (e.g., plant and ear height, data of male and female flowering, ear number and grain yield) and environment conditions.

In Teresina-PI, 84 cowpea materials (from Embrapa BAG-Cowpea and international access, in accordance with origin, group of maturity, and agronomic performance under water stress condition) were carry out on field. The experiment was sowed on June 2, 2005, during the terminal rain season, and the genotypes were submitted under strong hydric stress (Precipitation: June, 0.2mm; July, 4.0mm; August, 2.1mm). Another experiment with 114 cowpea materials was carry out under water stress conditions, after Nov 2005.

Porangatu-GO drought phenotyping site: Evaluation of 81 rice traditional lines from the Embrapa-Rice "Core Collection" with highest genetics diversity were evaluated. Additionally two populations F2:3 of 194 rice lines each and two tests were evaluated under drought stress, in a lattice 14 x 14 with three replicates. The experiment was planted on June 17, 2005. Plant material was subjected to adequate soil moisture - -0.035 Mpa measured at 15 cm depth until July 16, 2005 (about 23 days after emergence), when the drought treatment was applied. The evaluation was based on yield and leaf rolling under water and water stress during flowering. Also, 14 rice elite lines were evaluated under well irrigation and drought stress treatments. The elite line "Curinga", a high yielding cultivar, presented a very good adaptation under drought conditions.

Santo Antônio de Goiás-GO drought phenotyping site: The rice root diversity under drought and irrigated conditions were evaluated in a greenhouse, using PVC tube with soil column (Dark Red Latosol). 39 rice genotypes were evaluated having genetic diversity. These genotypes were well irrigated, and the soil water tension was kept lower than -0.035 MPa at 15 cm soil depth up to flowering, when two moisture treatments were applied: 1) well watered, and, 2) water stressed up to the harvesting, with the application of 50% of the water evapo-transpirated in the treatment 1. Also, the rice root diversity under drought and irrigated conditions are being evaluated in a greenhouse. The experiment was planted on Sept 9, 2005, including 12 rice genotypes having drought evaluated according plant parameters (grain yield, shoot and root dry matter, percentage of spikelet sterility, number of grain per panicle, percentage of panicle per plant, plant height and 100-seed weight).

Activities with common bean, in Santo Antônio de Goiás-GO: Evaluation of 76 breeding lines, and 97 lines from the international trial for drought and high temperature resistance plus four local test for drought resistance were planted on June 9, 2005, and, the 117 traditional lines were

planted on July 12, 2005. Also, 888 F1:2 and 367 F1:3 lines under drought were planted on June 9, 2005.

After November 2005, new rice essays were carry out during wet season, for GCP DPNetwork and WPModelling collaborative projects (in Santo Antonio de Goiás and Porangatu, GO, and, Teresina-PI sites).

The initial genetic background of rice materials to be used in different trials for characterisations as intermediate phenotyping [object of GCP DPN Project, and to attend GCP WPModelling collaboration, during wet season (no irrigation), and dry season (differential irrigation – with and without water stress during typical critical period per each crop specie)], after November 2005, as follow:

Rice cultivar	**Drought resistance level**	**Flowering (Days after emergence)**	**Reference (*)** *(GCP DPN x WPM essays*
BRS Curinga	bigger	74 (*)	**G5**
CNA 9045	bigger	74 (*)	
BRS Soberana	smaller	64 (*)	**G2**
CNA 9025	smaller	63 (*)	
IRAT 20	smaller	84 (greenhouse, in Goiânia-GO - 2004 Experiment sowed on Jan 16, 2004).	
CNA 9019			**G1**
Guarani			**G3**
Primavera			**G4**

(*) average of 106 trials from North region and Centre-Western, between 2002 and 2005;

Planaltina-DF and Santo Antônio de Goiás-GO sites for wheat field experiments: a) Outseason or "safrinha" planting period in preliminary drought phenotyping study: - 150 wheat genotypes were tested in "safrinha" (including 30 days without rain – named "veranicos", in 3 planting periods separated by about 10 days with two replications. Each planting period received water stress in different physiological stage. The main criteria for selection are grain yield and grain quality under water stress; and as the secondary criteria are been used earlyness, and initial vigour; b) Drought phenotyping using "line source system irrigation" during dry period (winter season): - a preliminary study with only five known genotypes and 13 replications were conducted using a line source irrigation system with objective of test the system efficiency. The stress was initialised during the tillering rate and the irrigation water was measured each meter from the central tubes until 15 m distance. Similar experiment is been carry out during 2006 with the genotypes selected during the "safrinha" period already harvested; c) "In vitro" drought phenotyping: - in initial development to offer selected wheat genotypes to field water essays, in near future.

Tangible outputs delivered:

Project Report (Year 1, June 2005 through June 2006) per *Objective/Activity/Outputs Expected: (See Appendix 1)*

- 1) In improvement of 02 Embrapa's Phenotyping Centre of Excellence for Drought Tolerance Studies (Sete Lagoas-MG, and Santo Antônio de Goiás-GO).

- 2) In improvement of 05 Embrapa's Phenotyping Sites of Reference (Janaúba-MG; Porangatu-GO, Teresina and Parnaíba, PI). [(2)] **Note:** The sites of Teresina, PI is been worked to carry out experiments including maize, sorghum, rice and cowpea species; and, Petrolina, PE to cowpea, sorghum and maize.
- 3) 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)]
- 4) In course, the definition and implementation of data base (climatic, soil and plant data set) and modelling (in partnership with GCP Whole Plant Modelling Project – CIRAD). During wet season (2005/2006) were carry out under WW-Well Watered (rain + supplementary irrigation) and WS (only rain) some essays with maize (04 genotypes), sorghum (05 genotypes) and rice (05 genotypes) in differents site-specific experimental (Sete Lagoas and Janaúba, MG; Goiânia, GO; and, Teresina, PI). Identical procedures is been happened during the 2006 dry season.

Deviations from the work plan:
The project was belated 2-3 months, because we had the financial resources delayed until last June 6, 2005, but technically is corrected now. Because of this our acquisitions to installation of each site, as well as acquisitions of equipment and materials to work the factor water in soil and plant, it been adjusted since July. Besides of this, the preliminary actions of our arrangement of genetic material for each crop, and some experimental schedules were made during the last first semester of 2005. We are providing to buy the equipment and material, as well to carry out our experiments planned during the wet and dry seasons.

During the 2005 semester 2nd we worked to get the planned quantity and the genetic background for each crop specie (maize, sorghum and rice; and also wheat, common bean and cowpea). In near future, we are planning to introduce from GCP reference sample aiming the partnership between GCP DPNetwork (Embrapa) and GCP WPModelling (CIRAD) Projects, which activities were initialised and carry out after Nov/2005-Fev/2006 (depending of each Site-Local). About it, is important note some legal aspects and procedures about Brazilian quarantine rules (at least with previous 3-6 months).

Summary of Personnel Commitments:
The improvement of 02 Centre of Excellence (Sete Lagoas-MG and Santo Antonio de Goiás-GO) and 05 Reference Sites (Janaúba-MG, Porangatu-GO, Teresina-PI, and Parnaíba-PI, and Petrolina-PE) to Phenotyping for Drought Tolerance has been adequately worked, according the GCP DPNetwork Project. As reference to describe these protocols for "precision site-specific experimental and farming" were prepared a technical boletin. In this first year was used the preliminary phenotyping strategy to water in different genotypes per each crop specie. In the Year 1 (June 2005 to June 2006), the GCP DPNetwork Project evaluated genotypes, as follow: - Janaúba and Sete Lagoas sites [maize: (221 access; 36 inbred lines; 50 "maize core collection"; 06 contrasting inbred elite], and, [sorghum: 64 selected material]; Teresina and Parnaíba sites – [maize: (50 + 50 "maize core collection"); cowpea: 84+30 selected material]; - Porangatu and Santo Antonio de Goiás sites [rice: 81+194+2; common bean: 177+117+888]; - Planaltina site – [wheat: 150+5].

Major Equipment Acquired:
The financial resources were available on June 6, 2005. So, we prepared an adjustment of our year 1 workplan (June 2005 to June 2006) 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 Programme); 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, June, July, August, October and November/2005.
- In: Brazilian Agrometeorological Congress, Teresina-PI, October 2005.
- In: INTERDROUGHT II – The International Conference on Integrated Approaches to Sustain and Improve Plant Production under Drought Stress, 2., 2005, Rome, Italy.
Proceedings ... Rome: InterDrought-II Committe;University of Rome, 2005. (Rome, Italy, from 24th to 28th September 2005 at University of Rome "La Sapienza").
- In: 2005 Generation Challenge Programme Competitive and Commissioned Research Project Summaries. Mexico D.F.: Generation Challenge Programme -GCP 2005 Annual Research Meeting: Mid-Year Project Reports, 2005, Rome, Italy. Proceedings ... Rome: GCP 2005 ARMeeting, 2005.
- In: 2006 Sarrah training (GenerationCP DPNetwork#6-Embrapa and WPModelling#7-Cirad Projects). Sete Lagoas, MG, Brazil, 9-12 May 2006.
- In: 2006 GCP Drought Phenotyping Network Project Workshop. Sete Lagoas, MG, Brazil, 15-17 May 2006.
- In: 2006 WUEMED training. Bologna, Italy. 5-10 June 2006. Dr. Newton Portilho Carneiro, Embrapa Milho e Sorgo.
- In: 2006 GCP DPNetwork Seminar. Porangatu, GO, Brazil, 28-29 June 2006. Dr. Reinaldo Lúcio Gomide e Dyeme Antonio Vieira Bento.
- In: 2006 GCP Drought Phenoytping Course (GCP;INRA;ENSAM). Montpellier, France, 3-12 July 2006. Dr. Alexandre Bryan Heinemann, Embrapa Rice and Bean;CIRAD.

Current Technical Status (on schedule, behind schedule, ahead of schedule):
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 programmes 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&DI, and the Embrapa's teams have accumulated good experience in plant breeding and environmental factors management. However, this GCP DPNetwork Project is creating the opportunity of get scientific and technical gains to breeding programmes 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 GCP-leadership and GCP Projects, e.g., like WPModelling Project lead our work to better possibility of scientific and technical goals.

Additional comments:
The Embrapa's experimental sites to water are been prepared, after now, also to support the GCP 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 the Generation Challenge Programme – SP1 leader to get some actions to previous choose of the genotypes per crop specie from GCP reference sample, aiming to provide seeds enough and available to quarantine period in Brazil (at least 3-6 months after come to Brazil). If necessary, the GCP 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.

2005-07: Whole Plant Physiology Modeling of Drought Tolerance in Cereals

Principal Investigator:
Delphine Luquet, Cirad
Collaborating Scientists:
Marcel De Raissac, Cirad
Michael Dingkuhn, Cirad
Jean-Claude Combres, Cirad
Scott Chapman, Csiro
Graeme Hammer, UQ
François Tardieu, Inra
Claude Welcker, Inra
Frederico Duraes, Embrapa
Elto Gama, Embrapa
Reinaldo Gomide, Embrapa
Cleber Morais, Embrapa
Edson Bastos, Embrapa
Alexandre Bryan Heinemann, Embrapa
Camilo Andrade, Embrapa
Fredolino Giacomini, Embrapa
Mark Cooper, Pioneer

Mid-Year Report

The WPM project, initiated in January 2005, aims at applying, testing, and improving modeling tools for:

- Component 1: drought environment classification (by characterising target population of environment to be accounted for within breeding strategies)
- Component 2: Plant trait characterisation and evaluation of trait impact on drought tolerance and/or yield (regarding environment typology, component 1)
- Component 3: Modeling tool development or improvement, and testing for GXE and genetics analyses.

These 3 components have obviously progressed during the last 6 months, which will be detailed below.

Component 1:

A- Alexandre Heinemann's post doc: 1st half advances (September 2005- May 2006)

The first component aims at characterising environment types by classifying within a geographic zone the kind of droughts that are encountered. This implies a geographic quantification of the variability in the degree of crop water limitations throughout the crop cycle. To achieve the goal this component was divided into 2 sub components; a) "Environmental classification of 6 different sites in Brazil based on water stress and b) "Determination of drought prone TPE for rice maize and sorghum for Brasilian savannas".

The climate characterisation for the 6 sites (Sto Antonio de Goias – lat: -16.50, long.: -49.30; Porangatu - lat: -13.30, long.: -49.11; Sete Lagoas – lat.: -19.46, long.: -44.25; Janauba – lat.: -15.00, long.: -44.00; Teresina – lat.: -5.08, long.: -42.80; Parnaiba – lat.: -3.08, long.: -41.76), considered as key breeding stations for maize, rice and sorghum, were done. This implied a huge

time investment to clean an important meteorological data set gathering pluri-annual (up to 45 years) climatic series on radiation, air temperature or humidity.

To determine the crop stress index variation among years and among planting dates for each site, a robust crop model was used (Ecotrop, or Sarra-h, Cirad). The Ecotrop model was run using a typical parameterisation for 2 rice, and one maize genotype. The behaviour of these reference genotypes was simulated for two different environments, Sto Antonio de Goias and Sete Lagoas for 22 and 45 years, respectively. Model outputs were analysed for checking the calibration. Daily crop model stress index values were averaged every 100 degrees days of the crop cycle, for each year and planting date.

Computed average stress indices associated with a given phenological stage will be now classified using statistical tools (partial least square regression and cluster analyse, assisted, by S. Chapman and Graeme Hammer, Univ Of Queensland). This will help answering questions, such as: What types of drought stress occur at a series of key testing location in Brazil? What timing of drought affects the most simulated yields?; Do the breeding station/trials represent the "TPE production" and Do "out of season" planting times hold to reliably sample some drought types?

B- Ecotrop modeling platform training (Sete Lagoas, 9-12th of May 2006)

A training session on modelling has been organised by Cirad in Sete Lagoas, at the Embrapa 'Maize and Sorghum'. The objective was to train Embrapa scientists (and particularly the ones involved in WPM, but not only) in using "Ecotrop platform", a new modular evolution of Sarra models (well known from Embrapa scientists since many years). This new platform allows characterising environment and testing genotypic parameter value impact on yield response to drought in a quite simple and flexible way, opening new opportunities by assembling "à la carte" existing (crop or environment) modules.

The training session was organised from 9 to 12th of May and gathered around 25 scientists from 4 Embrapa centres (Rice and bean, Maize and Sorghum, Cerrado, Meia Norte) and from Brazilian universities.

Most of the Embrapa scientists involved in WPM project attended the training. They have quite generally asked for an additional training (in Brazil) and assistance period by Cirad's scientists, to be organised within the 6 next months.

One or 2 embrapa scientists will come in France in the next 6 months as well, to work with D. Luquet, A. Heinemann , JC Combres on component 2 data with Ecotrop platform.

C- Short term next steps

- Alexandre Heinemann will spend 6 weeks (mid July to end August) in Australia to finalise statistical (cluster) data analyses (component 1). He will make a communication In IRRC (India, October 2006) & 1-2 publication(s) will be submitted before December 2006 on his work within the WPM.
- A training on APSIM model (Australia, CSIRO/Queensland University) is planned for some of Embrapa scientists (Camillo Andrade and others, in July 2006.

Component 2:

A- Wet season experiments (November 2005-March 2006):

Experiments for component 2 are organised both during the wet and the dry season. The relevance of studying genotype behavior variability during the wet season in Brazil relies on the

high probability of a dry spell occurrence (*Veranico*) which can be dramatic for yield, during this season. *Veranico* is variable among sites (regions), years; it is thus interesting analysing genotypic variability in tolerating this particular drought event. As defined at the beginning of the project, 4-6 contrasted genotypes of rice, maize and sorghum were chosen for experiments carried out in different Embrapa sites (Table 1).

In December 2005, D. Luquet carried out a mission for preparing wet season experiments (Visit in Goiania, Sete Lagoas, and Teresina, see component 1 for geographic positions). This mission aimed at proposing, explaining (with demonstrations) and finalising a generic experimental design and protocol with Embrapa partners, to be applied for wet and dry seasons. The final version of the experimental protocol is presented in Annex 1 (Data acquisition for field/crop management, soil, phenology and biomass monitoring; see Annex 2 for details on sites that here actually used for 05/06 wet experiments for rice, maize and sorghum).

Wet season data for maize, rice and sorghum experiments are currently gathered (M. De Raissac's mission, early May 2006). Data processing has been initiated in May 2006 by Embrapa's researchers (assisted by B. Balloy French MSc student in Brazil for 4 months with Cleber Morais). But model assisted data analyses will begin in September (D. Luquet, A. Heinemann, B. Balloy –French MSc student- and one or two Embrapa scientists who will come in France for this purpose.

NB: for this purpose as well, it was decided that Alexandre Heinemann's position in France (Cirad) should be extended for 3 months (until the end of 2006) to get involved in WPM Component 2. This will be as well for him the opportunity to present a communication on component 1 results in the IRRC (congress in October in India) and to finalise publications (2 papers are planned).

B- Dry season experiments organisation (April-September 2006)

Dry season experiments have already begun or are to be initiated in the next weeks (before June). Based on wet season experiences, some adjustments have been made (based on M. De Raissac's recommendations following his mission, May 2006). In particular, because of a lack of manpower in 2 sites: Parnaiba & Janauba (quite far from Embrapa research centres: respectively Teresina and Sete Lagoas), experiments won't be organised in these sites anymore. The same protocols (as in WS, Annex 1) will be used for the dry season; water stress treatment will consist in supplying 50% of water compared to reference treatment (see annex 2 for the experimental sites).

During the 2006 dry season, a French MSc student will spend 4 months (mid-may to mid-september) in Goiania to participate in one experiment on rice in Porangatu (co supervisors: C. Morais Embrapa, D. Luquet Cirad). The student will take advantage of his 4 month in Brazil to visit other Embrapa stations, and assist in collecting wet or dry season data that have not been gathered yet (for model assisted analysis in France).

Component 3:

The objective of the component 3 is to provide and improve modeling tools for GXE and genetics analyses.

A- Ecomeristem model:

During the last 6 months, the Ecomeristem model (Cirad, Luquet et al. 2006 & Dingkuhn et al. 2006 in Func. Plant Biol) was tested and validated. Its adaptation to account for plant response to

drought is underway and will be achieved in September. It will be tested on data external to the project first, and applied to WPM (component 2) data during the autumn of 2006.

B- APSIM combination with leaf expansion model: Karine Chenu's Post doc.

This post doc, co-supervised by INRA (France, F. Tardieu & C. Welcker) and Univ Of Queensland (G. Hammer & S. Chapman), aims at simulating the behaviour of virtual maize genotypes in climatic scenarios involving water deficit. The leaf expansion model of genetic response to temperature, evaporative demand and water deficit (Reymond et al. 2003) will be integrated in the crop model APSIM to evaluate the consequences of the genetic variability observed at 'organ-scale' on cumulated transpiration, biomass production and yield.

Karine Chenu has been recruited to work on the project in March 2006. A model has been designed to implement the leaf expansion over time and for the different leaf ranks. This model has been discussed with Graeme Hammer and Scott Chapman and should be implemented in APSIM in the next months, in Brisbane. Furthermore, a module concerning the yield sensitivity around flowering should be developed based on some recent results that pointed out some similarities in the genetic controls of leaf and silk expansion in well- and deficient-water treatments (Welcker et al., submitted).
Existing dataset from CIMMYT and INRA will be used to evaluate the model performance on the genetic variability of a RIL population, for contrasted water treatments.

Finally the behaviours of virtual genotypes characterised by their alleles at QTL positions will be simulated in chosen target environments, for different water deficit scenarios.

Tangible outputs delivered:

- Component 1 & 2: Ecotrop modeling platform training in Sete Lagoas: 25 participants, new demand for a second training (complementary) in Brazil (to use component 2 data with Ecotrop model(s)) and a huge potential fore xtending such a modeling training within the GCP (SP5).
- Component 1: Alexandre Heinemann's post doc results: model assisted climatic data analysis is finished (with a huge cleaning work on 20-40 year climatic data series) and statistical analysis is now possible (clustering, in next august in Australia CSIRO/Univ. Queensland). The results will be presented in IRCC (October 2006, India) and published before the end of the year.
- Component 2: Wet season experiments were reasonably successful, and data are now available. Data analysis and modeling will begin in September. Based on wet season experience, adjustments were decided for the beginning of dry season.
- Component 3: (1) Beginning of Karine Chenu's post doc (INRA/UQ) for one year (March 2005 March 2006). (2) *Ecomeristem* model was validated; an experiment was carried out in last April, to elaborate the module for model application to drought impact on plant growth.

2005-08: Population Structure, Phenotypic Information and Association Studies in Long-generation Crops

Principal Investigator:
M. Carmen de Vicente, IPGRI
Collaborating Scientists:
Martin Fregene, CIAT
Luc Baudoin, CIRAD
Kodjo Tomekpe, CARBAP
Merideth Bonierbale, CIP
Jean-Louis Noyer, CIRAD

Mid-Year Report
Potato

Overall Progress: Gap-filling genotyping by SSRs and phenotyping for morphological descriptors, agronomic, resistance and quality traits are in progress for the two potato populations (S. phureja germplasm collection (2x) and the Solanum tuberosum subsp. andigena advanced bred population B1C5 (4x)) selected for this study. The phenotypic data available for the respective populations at the outset of the project was partial, i.e. with considerable missing data, or with data taken in different environments and years. While this partial data showed the presence of a wide phenotypic variation for important agronomic and resistance traits, it was not considered consistent enough for an accurate QTL dissection-and effect estimations by association analysis. This project's workplan therefore included propagation of a complete set of genotypes of each population in order to improve data quality in standard trials that complement the regular evaluation of germplasm and breeding materials..

Genotyping: 13 of the 116 accessions comprising the S. phureja population and of the entire 105 bred lines comprising the B1C5 population are currently being genotyped with SSRs. To date, DNA has been prepared for all of this material and 11 of the 45 SSR markers planned in the project have been run. Complementary molecular data sets (S. phureja SP1 CGS) have been compiled for complementation with this new data, and GCP data formats downloaded for submission of the project's new data as it becomes available

Phenotyping: Both populations underwent propagation from healthy in-vitro plantlets this year, and clean tubers were obtained to initiate standard trials with the complete set of each population. At present the S. phureja population is planted in replicated trials in two locations. Filling gaps on morphological descriptors, and trait evaluations for tuber yield components, earliness, dry matter, chipping quality and vitamin C will begin in August. (Complementary resources are available for biochemical analyses). Likewise, the B1 population was be planted in a replicated trial in June and all genotypes.will be evaluated for yield components, earliness, and dry matter. Gaps in morphological data will also be completed.

Unfortunately phenotypic data for late blight resistance for all genotypes under the same environment will not be available by the end of this project due to the seasonality of late blights in Peru. We will experiment with the analysis of the pre-existing, partial information on this trait, availablel from the breeding programme, and attempt to imporve the assessment with the full data set after completion of the project.

Genetic structure and the degree of linkage disequilibrium: LD was measured by estimating the multiallelic r2 correlation coefficient on raw genotypic data of 84 out of the 116 accessions comprising the S. phureja sample using 44 SSRs. LD values (r2) were below 0.1 in 96% of locus pairs, of which 46% were non-significant. Though the linkage groups of 34 out of the 44 SSR loci tested were known, the genetic distances between them could not be established because information comes from different mapping populations. A t-test revealed that significant r2 values equal to or greater than 0.1 occur within chromosomes more oftenthan between them. LD values estimated as a function of distance for chromosome 8 , for which there was some map distance and order information, revealed that LD decreased to 0.1 within ca. 26cM (Fig 1).. It is known that LD distribution may differ among chromosomes and regions within chromosomes, so this estimate based in a small sample of SSR loci can only be taken as a preliminary estimation. However, due to vegetative propagation of potato accessions since domestication, LD might have been maintained through limited recombination. Other demographic forces such as admixture may have generated LD extending several cMs. Population structure analysis applying the Bayesian model-based clustering method in Structure 2.1 showed the highest probability for assuming the presence of 6 populations with an asymmetric proportion of the sample assigned to

each population, which is typical of highly structured populations. A neighbor joining tree was constructed with PowerMarker V3.25 (Fig 2). Since non-random mating within subpopulations affects LD patterns, familial relatedness was also analysed by calculating a relative kinship matrix using the software package SPAGeDi (Fig 3). Though 67% of the pairwise kinship estimates were close to 0, the remaining estimates were distributed from 0.05 to 1.0.which agrees with the complex familial relationship and population structure. This is a preliminary analysis to evaluate conditions for LD mapping of QTLs . Results to date suggest that S. phureja population falls into a category of a structured population that includes familial relationships within subpopulations. Taking this into account, a traditional mixed model for association studies has to include both population structure and familial relatedness matrixes as covariates to increase the power of detecting associations. We consider that our sample size as well as the number of SSR loci should be increased to obtain more accurate estimates of both population structure and familial relatedness and thus a meaningful estimate of the polygenic component.We also recognise that validation of descriptive data (eg morpjology and chromosome counts) could identify poorly classified individuals which must be removed from the population sample before analysis is complete.
All of these analyses will be carried out in the tetraploid B1 population once genotyping is completed. A great limitation is the lack of software that may deal with polyploid genetics for haplotype phase identification and association mapping analysis.

The research assistant responsible for the project received three weeks of individual; training on statistical analysis for association mapping at the Institute of Genomic Diversity during June 2006. She acquired knowledge on the principles underlying this approach, and became acquainted with the models, analysis, and output interpretation of freely available software for population structure, familial relatedness and association testing.

Cassava

In 2005, a subset of 200 cassava accessions was selected out of 800 lines generated by the breeding programme in the past 15 years. Phenotypic information on these accessions is available in different locations and years, including dry-matter content, yield, harvest index, cyanide content, commercial number of roots and traits related to farmers' interest (such as branching number, plant height and root length). To estimate LD and perform association analysis using the 200 accessions, 100 closely linked SSR markers were selected based on their distribution in the 18 linkage groups of the cassava genetic map (between 2 and 10 per linkage group).

Genotyping was completed for 138 cassava accessions with 75 SSRs and scoring is in progress. It has been proposed to complete the evaluation and scoring of the 138 with 100 SSR markers by the beginning of June 2006 so that a preliminary data set can be used as a first approximation for calculation of LD and association analysis. Genotyping and scoring of the remaining 62 accessions – to complete the 200 – will be completed between July and August 2006, and association analysis for the complete data set (200 accessions evaluated with 100 SSR markers) is planned for September 2006. Haplotypes will be defined in the 200 genotypes using the computer software Arlequin and the haplotype data will be used to calculate linkage and structural disequilibrum between alleles of the SSR loci in the selected genotypes.

Coconut

A mission to Vanuatu in June 2005 allowed the collection of leaf samples from 219 individuals representing 4 breeding generation of the Vanuatu Tall cultivar. The first three generations were represented by 22 individuals each and the most advanced by 153 individuals whose pedigrees are partly known (The grand-parents are identified). A brief visit to the DArT laboratory ensured the co-ordination between the two research teams and prepared the visit of Sophie Bouchet

(November 2005 to April 2006) who performed DArT analysis with 347 markers developed by Chandrika Perera (from CRI Sri Lanka). The same DNA samples were also analysed at Cirad with 30 microsatellite markers by Champa Kumari Bandarayanake from CRI Sri Lanka). The same DArT markers were also used in 67 individuals of a mapping population and to 91 selfed progenies. Integration of these markers to a reference linkage map will take place in the next months. Phenotypic observations (fruit yield and composition + vegetative observations) are being performed at VARTC (Vanuatu).

Besides, we developed and programmed a Bayesian method for calculating haplotype frequencies and various linkage disequilibrium parameters in Mendelian populations. This method was tested with 10 pairs of linked markers and 14 isolated loci using the four largest germplasm sections or the CIRAD reference database, namely Mozambique, Panama, Vanuatu and Brazil. The results indicate that LD is more likely to be observed and more intense in paired loci than between independent loci. There are however counterexamples i.e. LD without linkage or absence of LD in linked loci. The next steps will be:

- Splitting the populations into subsets and perform LD analyses on the subsets in order to minimise the effects of population structure.
- Adapting the Bayesian method to dominant markers in order to use it with DArT markers and to apply it to the above 219 individuals from Vanuatu.

Yam

Phenotypic characterisation:
An act of volunteer hostility which had for consequence to mix tubers of an unknown number of accessions disrupted deeply the collection of phenotypic data in 2004 and 2005. The author of this act was dismissed. Through a long but reliable work accomplished by Dr. R. Malapa, finally employed by VARTC in Octobre 2005, a collection of 200 accessions was re-assembled in which the identification of only ten clones still remain uncertain. Phenotypic data for the Vanuatu National Collection will be, at least, available only for tuber descriptors and for resistance to anthracnose for the two last years.

Validated progenies:
The approach adopted, genetic mapping, intends to place SSR loci on a genetic map and to identify zones for which LD can be analysed. To do so, we are focusing on a sexual progeny of 124 C2 individuals. This material has been obtained through clonal propagation of mixed tubers from an initial set of 88 distinct F1 clones (C1) with no individual identification. Duplicates will be eliminated once genotyping is completed. Information regarding the genitors as well as the genitors themselves is not available. Phenotypical data for the tubers will be available in august 2006. So far, this progeny is the only one available, despite the numerous controlled crosses realised successfully since 2003 but which failed to produce mature seeds because of severe anthracnose attacks on the mother plants.

It is clear that one of our major problems is due to the severity of anthracnose (Colletotrichum gloeosporioides) attacks on the island of Esperitu Santo when rain falls average 3000 mm per year favouring the rapid spread of the disease between plants. Although our controlled pollinations are efficient (average of 80% of fruit set) the development of seeds does not reach maturity because the plants are destroyed by anthracnose before the fruits reach full maturity. Embryo rescue was attempted but failed. Since the ban on Binomyl (Benlate) it is nowadays, difficult to apply a systemic fungicide on the yam plants. For the coming season (August 2006-July 2007), we are planning to spray regularly contact fungicides (i.e. Manebe or Mancozeb) in an attempt to slow down the spread of the disease, but this is an expensive endeavour.

Molecular data:
The Dr R. Malapa is presently in CIRAD, Montpellier (since April 10th, 2006). He brought with him leaf samples from the Vanuatu germplasm collection as well as those of the progeny. DNA was extracted on both series of samples. After preliminary attempts to use directly SSR markers it turned out necessary to purify these DNA extracts on a Qiagen column. This was first made on the progeny samples and germplasm collection accessions samples.

Genetic Mapping:
The first molecular results are quite unclear. D. alata is a polyploid species for which 4x, 6x and 8x ploidy levels have been described. Accessions used as genitors are usually described as 4x. For the first set of loci observed in our study, a maximum of 4 alleles is observed for each locus and for the whole progeny but each individual revealing generally from 1 to 2 alleles, more rarely 3 and never 4 or more (figure 1). This is a result which is not expected for a cross between tetraploids, dioecious and probably highly heterozygous parents. It looks like the result of a (open?) pollination between heterozygous diploids sharing some alleles. This observation needs to be confirmed on a larger number of loci but leads to questioning the ploidy levels of the cultivated yams which takes into account, so far, a chromosome basic number of 10. Recent publications concerning D. rotundata et D. trifida support the same questioning (Scarcelli et al., TAG 2005; Boussalem et al., TAG 2006). Although these results are producing new interesting knowledge on the genetics of D. alata, they also let us believe that despite the useful strength of the 88 entries we have, it will be rather difficult to map a restricted number of SSR loci on diploids with a chromosome basic number of 20, composing an heterogeneous progeny with narrow genotype groups. We are, however, pursuing molecular fingerprinting assessment of the progenies because it is necessary to clarify these results before intending to set up a new series of crosses in Vanuatu (subject of course, to their technical feasibility).

Global population structure
In spite of the difficulties related to the discrepancies of clones identities in the collection, the points 1 and 2 of the project which are respectively “to screen with molecular markers the Vanuatu collection in order to assess the global population structure and the genetic differentiation between potential parents” and “to collate phenotypic information available and attempt to correlate molecular variation with agro-morphological diversity” will be achieved in due time as planned in the project document. The objective of monitoring a minimum of 300 polymorphic loci” will be reached in a few weeks from now using both AFLP and SSR markers. The collection has been fully characterised using IPGRI type morpho-agronomic descriptors during several years (aerial and underground parts, see SPYN final report). In 2004 and 2005, the accessions maintained in the re-assembled collection have been characterised for the underground organs only (yield, number of tubers, shape of tubers, outer skin and inner skin colour of the tubers,, tuber flesh colour and the presence of nematodes). The last set of data will be available when the harvest is completed in august 2006.

Musa

Introduction

Cultivated bananas are mostly triploid and evolutionary derived from crosses within and between diverse accessions of two diploid ancestor species noted AA and BB genome for Musa acuminata Colla and Musa. balbisana Colla, respectively (Simmonds & Shepherd, 1955). Banana cultivars are usually classified into three genome groups, AAA, AAB, and ABB but the varieties grouped in the same genomic category can be very different (Simmonds, 1962; Stover and Simmonds, 1987).

CARBAP (Centre Africain de Recherches sur les Bananiers et Plantains) located in Cameroon (Central Africa) manages a banana breeding programme focusing on a predominant scheme of tetraploid x diploid crosses to develop final triploid hybrids. Several hundreds triploid plantain-derived hybrids are presently undergoing field evaluation.

This scheme usually is based on crossing 3x plantain cultivars to 2x accessions that are donors of resistance genes, selecting 4x and 2x primary hybrids from the 3x/2x progenies, and crossing 4x - 2x hybrids to produce secondary 3x hybrids. On the other hand, CARBAP is also using another method of producing triploid hybrids developed by CIRAD which involves tetraploidisation of diploid accessions using colchicine prior to crossing with another diploid.

Genetic improvement efforts actually focuse on resistance to Black Sigatoka disease, to which all plantain cultivars are highly susceptible and on the production of pure Acuminata hybrids or AAB hybrids which do not contain active sequences of Banana streak Virus.

The project objectives for Musa component are (i) to collate and complement existing agro-morphological data of natural plantain germplasm and related hybrid populations in four different agro-ecological zones, (ii) to screen with molecular markers the material mentioned above , (iii) to compile phenotypic information available and attempt to correlate molecular variation with agro-morphological data and finally (iiii) to test for association markers between markers and candidate genes controlling traits of interest.

This brief preliminary report resumes the progress in agro-morphological evaluation and in molecular characterisation and formulate an outlook for the following months.

Multi-locational agronomical and morphological evaluation

Characteristics of the three locations

Agronomic and morphological data are collected in the field at three stations located in different agro-ecological zones characterised by unimodal annual rainfall with a rainy season of eight months from mid-march to mid-november and a dry season from mid-november to mid-march

- the Nyombe station with brown ground of eutrophic type deriving from volcanic rocks. They have good physical characteristics (texture and structure) and a moderate depth; they are rich in minerals with a pH of 6 on the surface and 6.3 to 50 cm depth.

- the Mbouroukou station located on the Western slope of the Manengouba Mount with an acid ground (pH<5,5) and low level of phosphorus in spite of low aluminium level. It is a feralitic soil of muddy texture and moderately fertile deriving from an old volcanic material with contribution of volcanic ash on the surface.

- the Ekona station located on the East slope of Cameroon Mount with a relatively rich ground and high pH (6) as well as levels of phosphorus, potassium, total exchangeable bases and cationic capacity of exchange.

Experimental layout and agronomical characterisation

The experimental layout is a one block model (as for field germplasm collection) with 4-5 mats per hybrid without replication. In dry season, the plants were submitted to a light supplementary watering. The hybrids were slightly treated against nematodes and black weevil but not against the Black Sigatoka. Parental clones were also involved in the experiment. Plant spacing was 3 m

between rows and 2 m within rows to give a population density of 1,667 plants per hectare. Almost all the plants were grown for two consecutive crop cycles in the three locations and the following traits are recorded then the means and standard deviations will be calculated.

-NFL = Number of functional leaves at Flowering (NFL)
-NLH = Number of leaves at harvest (NLH)
-HMP = Height of mother plant in centimetres (HMP)
-C100 = Circumference of mother plant at 100 centimetres (C100)
-BW = Bunch weight in kilogram (BW)
-NH = Number of hands (NH)
-NF = Number of fingers (NF)
-LF = Length of finger in centimetres (LF)
-CF = Circumference of finger in millimetres (CF)

Plant material

Two types of plant material are under evaluation in three locations
- 132 cultivars of plantain (AAB) representing the whole plantain germplasm
- Four 4x/2x populations (AAAB x AA) totalising 181 triploid hybrids present in the three locations with a quality data. These populations were generated by the following crosses :

Female parent (AAAB)	**Male parent (AA)**	**Present number of triploid hybrids with quality agronomical data in three locations**
CRBP 753	DS 11	64
CRBP 776	DS 11	38
CRBP 39	DS 11	45
CRBP 956	CRBP 436	34

Molecular characterisation

Plantain (Musa subgroup AAB), a particular type of cooking banana presents a high phenotypic diversity in the Congo basin, a sub-region which is considered as its secondary diversification zone. Using 9 microsatellite loci and AFLP markers, the genetic diversity of 30 plantains landraces constituting a representative sample of the phenotypic diversity was assessed. The results confirmed a very narrow genetic base of this cultivar group.

As in banana, most of methylation pattern is transmitted through vegetative propagation and since DNA methylation is often tightly linked to epigenetic traits,
methylation sensitive amplification polymorphism (MSAP) analysis was alternatively used to survey cytosine methylation status on the same set of 30 plantain cultivars. (using 8 MSAP primer pairs representing 633 CCGG sites throughout the banana CCGG sites throughout genome). 15 loci were found to be polymorphic. A reliable degree of polymorphism was revealed allowing the classification of the samples into three clusters. MSAP seems to be a relatively useful molecular tool for highlighting differences inside the plantain subgroup but no correlation was observed between the phenotypic classification and methylation diversity.

In order to study the heredity and segregation of MSAP markers, a population of 70 triploid hybrid was characterised. This population is issued from a cross between a primary tetraploid (with AAAB genomic constitution) and an improved diploid (with AA genomic constitution). The molecular analysis of this population shows that MSAP markers are inherited. This study also indicates that the B genome do not pass systematically in the triploid progenies which could

be very useful considering that the selection of pure Acuminata triploid hybrids is one of the major objectives of the CARBAP breeding programme.

Workplan for the following months (July – October, 2006)

- Extract the DNA from the four hybrid populations established at a location at CARBAP
- Send the DNA to a laboratory in Australia for Dart analysis
- Compile the agronomical data and calculate means and standard deviations for each trait and each clone
- Comparative analysis of agronomic traits and Dart markers to examine association mapping in light of the latest compiled genetic map of Musa.

Tangible outputs delivered:

1. Compilation of phenotypic data for 200 varieties from cassava databases
2. Genotyping of 138 cassava accessions out of 200 with 75 SSR markers
3. Scoring and database development for 138 accessions evaluated with 20 SSR markers

The suitability of a hands-on data analysis workshop for LD calculation and association tests has been discussed with the project team following some requests. Communication is ongoing with Marco Bink and other colleagues at WUR as possible experts to guide the workshop. No final date has been set, but tentatively it would be held on the third week of October 2006; the venue is still to be decided.

Deviations from the work plan:

It appears that it will be very difficult to reach all the initial objectives in the period foreseen for yam. This is dramatically true for points 3 and 4. Nevertheless, a certain number of important information is being produced, in particular regarding the genetic structure of the progenies. Before the end of the project, the team is planning to focus on points 1 and 2 of the project document.

2006-01: Developing Strategies for Allele Mining within Large Collections

Principal Investigator:

N.R. Sackville Hamilton, IRRI

Collaborating Scientists:

M. Lorieux, CIAT
C. Brondani, EMBRAPA
H. Upadhyaya, ICRISAT
R. Varshney, ICRISAT
B.J. Furman, ICARDA
S. Udupa, ICARDA
M. Baum, ICARDA

Mid-Year Report

Not received

2006-02: A Dataset on Allele Diversity at Orthologous Candidate Genes in GCP Crops

Principal Investigator:

Dominique This, Agro-Montpellier, Agropolis-France

Mid-Year Report

The project "A dataset on allele diversity at orthologous candidate genes in GCP crops" (ADOC) has been accepted in February 2006 as a commissioned project from the GCP Management Team and Programme Steering Committee. The main objective of this project is to provide an initial dataset of sequence diversity for 10 to 15 orthologous candidate genes for drought tolerance, in a reference sample of around 300 accessions for seven target crops of the GCP programme (barley, rice, sorghum, common bean, chickpea, cassava and potato).

This project involves several tasks:

- **Task 1:** To elaborate a list of the best documented drought-tolerance-related candidate genes, with their location and sequence(s) in model plants
- **Task 2:** To investigate orthology relationships between crops by phylogenetic analyses
- **Task 3:** To select a priority gene list (initially 20 to 30) on this basis
- **Task 4:** To design and validate primers for amplification of whole-length gene versions from an array of GCP crops and provide first allelic sequences
- **Task 5:** To gather DNA samples from the reference samples (e.g. 200 to 300 accessions per crop) of the various crops under consideration
- **Task 6:** To produce PCR amplicons to be subjected to sequencing, produce and assemble sequence information among accessions of GCP reference samples
- **Task 7:** To construct a database and make the data available for further use

Scientific activities

Progress update by task, May15th 2006

Task 1: Joined efforts in an updated bibliographic survey and analysis of the results obtained in the different laboratories involved in the ADOC project have led to a list of drought-tolerance candidate genes, presented in appendix 1. This list involves different steps in the plant response to drought stress (signal perception and transduction, transcription control, stress response …), with a special focus on sugar metabolism and water use efficiency. A subset of seven gene families (ERECTA, DREB, NCED, LEA, Invertases, SS, ASR) have been selected during the first meeting in San Diego, analysed in detail by ortholabs and presented to the OL groups during a workshop held in Montpellier in March 2006.

Task 2: For each of the selected gene families, ortholabs have gathered sequences from general and specialised databases (mainly NCBI) and run first phylogenetic analyses, in relation with gene specialists. Activities include capacitation of CIP personnel involved with ADOC in bioinformatics and student training in CIRAD. Figure 1 presents an example of a phylogenetic tree obtained with PHYLIP software (Maximum Likelihood method, PAM distance matrix and 1000 bootstrapping) from the protein alignment for the ASR gene family (result obtained by R. Philippe, phD student, CIRAD).

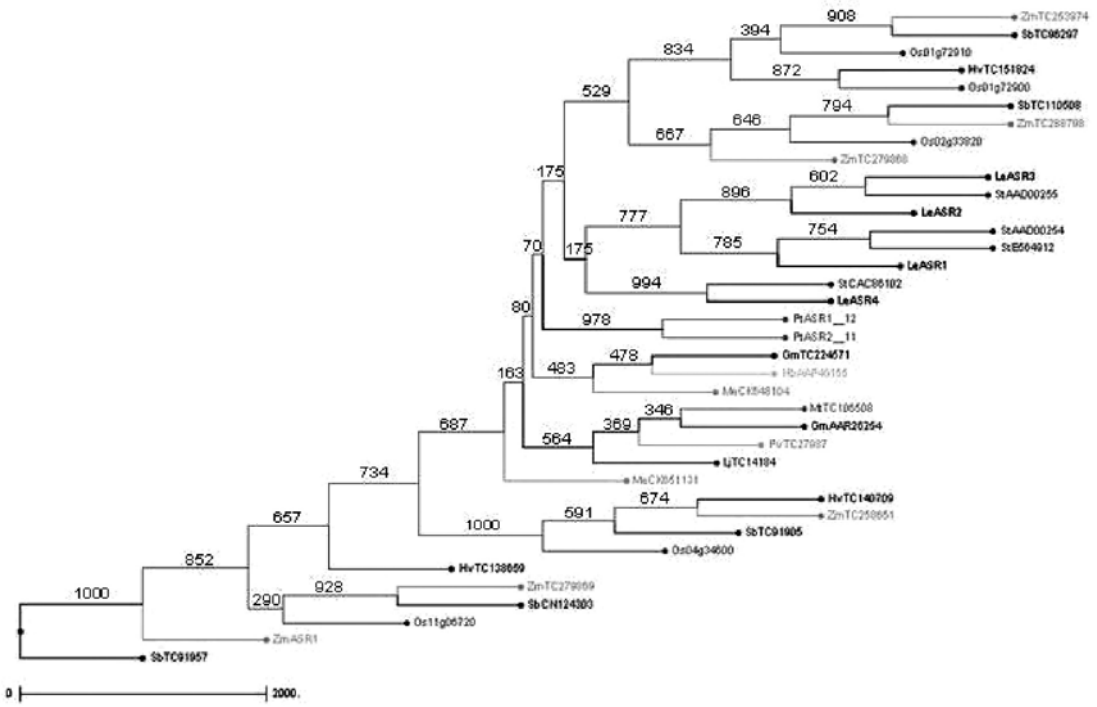


Figure 1: The ASR gene family (maximum likelihood method, PAM distance matrix and 1000 bootstrapping)

For some of the genes, like ERECTA, defining orthologs seems quite an easy task, due to the separation of ERECTA and ERECTA-like genes prior to monocots and dicots. However, for most of them, ortholog identification is not obvious from this type of analysis, because they represent multigenic families with recent duplications, and databases do not contain the complete set of genes for each of the species considered or at least their taxonomic groups. In order to complete our information on candidate gene families, it has been decided to get new sequence information from amplified genes with consensus primers for one accession of all of the ADOC species. This should allow more complete phylogenies and facilitate orthology analysis. In *Arabidopsis* and rice, for which the complete genome sequence is available, we will provide information on the different loci assigned to the physical maps as soon as possible.

Task 3: Some gene families (like dehydrins or NCED) have been discarded after this first analysis, because of the high number of potential genes and difficulty to define an easy phenotyping related to gene expression. SPS has been included in the list to complete the survey of sucrose metabolism.

Task 4: Two sets of primers will be defined from sequence alignments: Consensus primers (defined across introns) should allow us to amplify different products for each ADOC species (representing different family members) and get additional sequences. From new alignments, we will then define specific primers allowing the amplification of orthologous genes in the 7 ADOC species.

Task 5: Each crop partner has defined a first set of 7 accessions, MTA have been exchanged with ortholabs and first DNA shipments are underway. All of the 4 ortholabs will receive samples of the seven species and test primers for one or two gene families. In a parallel work, crop partners have made substantial progress in the definition of the whole set of 283 accessions to be

sequenced by CNG (3 x 94 accessions, including 1 common genotype in A1 + 1 empty well in H12 for each microplate).

Task 6: Not started yet

Task 7: Since no database can be expected very soon from GCP-SP4, one ftp site located at CIRAD has been created, from where a template Excel file (able to handle our data) will be downloadable and used by each ortholab. We have already included all documents from the project in this ftp site. Specific needs in term of databases will be transmitted to SP4 later on.

Meetings and exchanges between partners,

Since the beginning of the project, two meetings have been organised.
The initial meeting held in San Diego in January 2006 was intended to know each other, build a friendly work team and precise the work plan together. Minutes of this meeting have been already sent to the GCP group leaders.
The second workshop has been held in Montpellier in March 2006, specifically for ortholabs and gene specialists. The conclusions of the meeting are presented in appendix 2. This workshop has been very efficient in defining a common strategy among partners for the orthology work and the definition of the first set of candidate genes to be analysed. A proposed protocol for gathering orthologs in the 7 ADOC species is presented in appendix 3.

Conclusions

This initial period of the project has settled the basis of the project in defining a consensus among partners of the ADOC project in the initial choice of candidate genes and the methodology to be used. All partners have shown much involvement in this project and the two meetings held up to now were probably very helpful for that. We expect first results on orthologous relationships to come very soon and the next step will be to learn from crop partners the germplasm structuration to be expected within each crop. We hope that most partners, particularly crop specialists, will be able to meet again in Brasilia in September 2006 and this meeting should be more devoted to orthologous relationships and allelic diversity issues.

2006-03: SNP Analysis of the Genetic Diversity along the Rice Genome (HAPLORYZA)

Principal Investigator:
Kenneth L. McNally, IRRI
Collaborating Scientists:
Claire Billot, Agropolis-CIRAD
Brigitte Courtois, Agropolis-CIRAD
Dominique Brunel, INRA-CNG
Mark Lathrop, CNG

Mid-Year Report
To kickstart the process, initial meetings were held at PAG14 in San Diego during January, 2006 with further discussion in Montpellier on March 22 and 23, 2006.

Initially, SNPs from the BGI data in dbSNP, from the Feltus database, and Shen databases were filtered using perl scripts developed at CIRAD. Comparison of the masked Nipponbare and 93-11 sequences being used for the rice re-sequencing project (GCP commissioned project 35) was done at IRRI using the nucmer programme of the MUMmer package. Comparison of these analyses will result in a filtered set of SNPs for submission to CNG. CNG will design assays from this set, filtering for potential quality. From this filtered set, a final set of 1000 SNPs spaced

at ~400 kb intervals and 500 SNPs located at about ~20 kb spacing and targeting regions of interest will be chosen for the genotyping. Potential regions for the higher density include the regions around:

DREB2a, ERF3, RG171-RG157, OsCIN2 (CWI), Pi33, RYMV04, waxy, hd3a, WC (RZ247-RZ2), ERECTA, hd1, Short arm 7, Sub1, Auxin efflux carrier, ASR1, RYMV12, and ERECTA-like.

The composition of the 900 accessions will include about 400 CIRAD and IRD materials consisting of 135 lines from Madagascar, 39 lines from Guinea, 30 known intermediates, 74 *O. rufipogon,* 12 lines from Zanzibar, 75 lines from the collection of G. Second, and 30 other wild species including *O. glaberrima* and outgroups. The remaining 500 lines will be chosen from the GCP composite rice collection, specifically from those having phenotypic data as well as lines likely to be intermediates.

Tangible outputs delivered:
1) Definition of the set of 900 accessions to include in the study with 400 coming from the collections at CIRAD and 500 from the GCP composite core collection.
2) Strategy for defining the initial set of sequences CNG will use to design assays.

Deviations from the work plan:
None

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

Principal Investigator:
Jane Toll, IPGRI
Collaborating Scientists and Institutions:
A.Blum, Plantstress.com, Israel
Mahalakshmi Viswanathan
GCP consortium and collaborating institutions
Challenge Programme on Water for Food

Mid-Year Report
This project was commissioned to IPGRI/System-wide Genetic Resources Programme (SGRP) at the end of January 2006. Dr M. Viswanathan has been engaged as the principal consultant and Dr A. Blum has agreed to assist in the execution of the work.

Tangible outputs delivered:

Work plan
A work plan has been established and implementation of the project was discussed at the annual meeting of SGRP in March 2006 at CIAT in Cali, Colombia. Representatives of CG centres expressed their interest in participating in the project and provided references to specialists at their centres that should be contacted.

Collation of information
A review and collation of relevant literature and other sources of information has commenced. One initial reference source is the report of a GCP workshop entitled 'Phenotyping and water deficit', held in Montpellier in July 2004.

Questionnaire development
Work has also begun on the design of a questionnaire to validate the assembled data and obtain more precise information needed from GCP collaborators and other relevant institutions.

Deviations from the work plan:
The project is proceeding as planned.

2006-05: Development and Genotyping of a Faba Bean Composite Collection

Principal Investigators:
B.J. Furman, ICARDA
M. Baum, ICARDA
Collaborating Scientists:
G. Duc**,** INRA
M.J. Suso, IAS

Mid-Year Report
A global composite collection of 1000 accessions of faba bean is being developed in collaboration with INRA-Dijon and IAS-Cordoba. The composite collection will contain approximately 600 accessions from the ICARDA global collection and 200 accessions each from both collaborating institutions.

Development of faba bean SSR markers is underway. To identify SSR markers for faba bean we have developed enriched genomic libraries of the faba bean cultivars "91/25" and "34M" (see description below) through collaboration with a private company. We have tested motifs that were most abundant in other legumes such as chickpea and lentil. However, the most successful libraries were those that were enriched for CA, AAC, AAG and CAG. We are in the process of sequencing positive identified clones and expect to successfully identify SSR markers for faba bean over the next three months.

In addition, two recombinant inbred line (RIL) populations in faba bean (constructed from the inbred line from German cultivar Kristall (Lochow-Petkus company) "91/25" and inbred line 34Morocco originating from ICARDA-accession BPL228/ ILB141 from Morocco "34M") were received from the University of Goettingen. We are multiplying the seed in 2006. We will use the RIL populations to map the SSR markers and to identify their position in the linkage groups.

Tangible outputs delivered:
- Enriched libraries for SSR motifs developed
- CA- six out of nine sequences contain a microsatellite
- AAC- six out of nine sequences contain a microsatellite
- AAG- seven out of nine sequences contain a microsatellite
- CAG- three out of nine sequences contain a microsatellite
- A composite collection of 1000 accessions is under development

Deviations from the work plan:
Nil

2006-06: Genotyping of Composite Collection of Finger Millet

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborating Scientists:
R.K. Varshney, ICRISAT

D. Hoisington, ICRISAT
C.L.L. Gowda, ICRISAT
C.T. Hash, ICRISAT
S. Chandra, ICRISAT

Mid-Year Report

A Composite collection of finger millet consisting of 1000 accessions has been developed at ICRISAT using the available phenotypic characterisation and evaluation data, geographical origin and taxonomical information. The objective of this research is to fingerprint the composite collection with 20 SSR markers.

Presently, SSR markers are not available in the public domain. However, a set of more than 100 SSR markers, recently developed at University of Georgia, has been obtained from Dr. Katrien Devos. From this set, around 40 markers well distributed over the finger millet genome has been selected and primer pairs synthesised. These 40 SSR markers are being screened on 8 diverse accessions (included in the composite collection), one each from Burundi, Nepal, Malawi, India, Uganda, Cameroon, Kenya, and Ethiopia, to identify 20 polymorphic SSR markers. Subsequently, these markers will be used to genotype the entire composite collection.

The composite collection will be planted in the field during third week of July, and leaf material of 15 days old seedlings will be used for DNA extraction following high-throughput procedure.

Tangible outputs delivered:

- Composite collection of finger millet developed.
- Primer pairs synthesised.

Future Plan of Work:

- DNA extraction from 1000 accessions.
- Identification of 20 polymorphic SSR markers.
- DNA quantification and optimisation of PCR conditions.
- Genotyping the composite collection with 20 SSR markers.
- Allele calling and data analysis to procure information on population structure.
- Identification of a reference collection of 300 diverse accessions.

Deviations from the work plan:

None

2006-29: Preparing IITA-Cassava Reference Germplasm for Distribution and Association Mapping

Principal Investigator:

Dominique Dumet, IITA

Collaborating Scientists:

Morag Ferguson, IITA
Martin Fregene, CIAT

Mid-Year Report

The aim of this project is: 1/ To establish an *in vitro* pathogen-free reference collection of IITA cassava germplasm. This collection will be exchange with CIAT and EMBRAPPA. 2/To establish an *in vitro* pathogen-free collection of IITA cassava germplasm selected on drought tolerance. This collection will be used for germplasm drought tolerance evaluation and association mapping studies. 102 accessions have been selected to constitute the IITA reference collection. Out of these, 51 have been introduced *in vitro* and produced proper seedlings. Some of

these seedlings are now in the process of acclimatisation for virus indexing. The remaining 51 accessions of the reference collection are at different stages of the *in vitro* introduction process (stem thermo-treatment, meristem excision, subculture). 43 drought tolerant accessions have been selected; all of them are presently processed for *in vitro* introduction.

Tangible outputs delivered:
The *in vitro* collections should be ready for distribution by April 2007.

Deviations from the work plan:
As the technical work only started in April/May 2006 it is likely that germplasm will only be ready for distribution end of April 2007.

2006-30: Development and Genotyping of a Foxtail Millet Composite Collection

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborating Scientists:
R.K. Varshney, ICRISAT
C.T. Hash, ICRISAT
D. Hoisington, ICRISAT
C.L.L. Gowda, ICRISAT
S. Chandra, ICRISAT

Mid-Year Report
A composite collection of foxtail millet consisting of 500 accessions has been developed using the available phenotypic characterisation data, geographical origin and taxonomical data. The composite collection includes accessions of core collection developed at ICRISAT (155), good plant aspect scores (33), large grain size (9), high grain yield (5), early flowering (77), dwarf type (21), more number of basal tillers (25), mono tiller (59), inflorescence length (58), inflorescence width (48), and improved cultivars (10). The composite collection represents accessions from 26 countries. Landraces dominate with 451 accessions followed by 27 accessions of sp. *Glauca*, and 12 accessions of subsp. *Viridis*.

After searching the literature and contacting several research groups world wide, we realised that SSR markers are not available at present in foxtail millet and we are not aware of any laboratory that is developing the SSR markers for the species. Therefore, efforts are underway to identify appropriate cereal species, mostly the close relatives, from which SSR markers can be used for genotyping the foxtail millet composite collection. We are planning to initially screen 20 markers each of pearl millet, maize and sorghum with two genotypes of foxtail millet. Subsequently, the heterologous SSR markers that yield amplicon in foxtail millet will be used to fingerprint eight diverse accessions of foxtail millet (included in the composite collection), one each from Syrian, Arab Republic , CIS (former USSR), China, Lebanon and four from India, to select 20 polymorphic SSR markers for genotyping the entire composite collection.

The composite collection will be planted in the field during third week of July, and leaves from 15 days old seedlings will be used to extract DNA following high-throughput procedure.

Tangible outputs delivered:
- The composite collection developed.

Future Plan of Work:
- DNA extraction from 500 accessions.

- Identification of polymorphic SSR markers from pearl millet, sorghum and maize.
- Screening of these SSR markers on diverse accessions to finally identify 20 polymorphic markers.
- Genotyping composite collection with 20 SSR markers.
- Data analysis to determine the population structure and genetic diversity.
- Identification of a reference collection of diverse accessions.

Deviations from the work plan:
None

2006-31: Development and Genotyping of a Pearl Millet Composite Collection

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborating Scientists:
C.T. Hash, ICRISAT
S. Senthilvel, ICRISAT
R.K. Varshney, ICRISAT
D. Hoisington, ICRISAT
K.N. Rai, ICRISAT
R.P. Thakur, ICRISAT
S. Chandra, ICRISAT

Mid-Year Report
A composite collection of Pearl millet consisting of 1000 accessions has been developed using the available phenotypic characterisation & evaluation data, geographical origin and taxonomy. The composite collection includes 504 accessions of core developed at ICRISAT, accessions tolerant to drought (6), heat (3), salinity (20), resistant to downy mildew (42), ergot (20), rust (23), smut (15), and multiple diseases (8). The composite collection also includes accessions with high seed iron and zinc content (4), high seed protein content (20), yellow endosperm (2), high stalk sugar content (12), forage type (8), trait specific selections (197), gene pools (4), released cultivars (5), elite breeding lines (47), and 60 accessions of 7 wild *Pennisetum* species. The accessions in composite collection represent 30 countries of all the continents except Oceania. Biologically, landraces dominated with 694 accessions followed by 246 breeding lines/released cultivars, and 60 accessions of 7 wild species.

About 100 SSR markers derived from genomic DNA, BAC and EST-libraries (Allouis et al. 2001, Budak et al. 2003, Qi et al. 2004, Senthilvel et al. 2004) are available at ICRISAT. These markers will be used to initially screen on a set of eight diverse genotypes (included in the composite collection), one each from Botswana, Burkina Faso, India, and five from ICRISAT, to select highly polymorphic markers representing all the seven linkage groups.

Pearl millet being a cross-pollinated species with large intra-accession variation, the selected set of markers will also be used to screen the artificial pools comprising of different proportions of genomic DNAs of two genotypes, which are polymorphic for a given marker. These SSR markers will then be screened on the bulk DNA from 12 plants from each of 1000 accessions of the composite collection.

The composite collection will be planted in field during the third week of August, and leaf samples from 15 days old seedlings will be collected from 12 representative plants for DNA extraction.

Tangible output delivered:
- The composite collection developed.

Future plan of work:
- DNA extraction from 12 plants per accession following high-throughput procedure and pooled.
- DNA quantification and optimisation of PCR conditions.
- Screening of SSR markers on 8 diverse genotypes.
- Identification of 20 polymorphic SSR markers.
- Screening artificial pools with the identified SSR markers.
- Fingerprinting the composite collection with the set of identified markers.
- Data analysis to determine population structure and genetic diversity.
- Identification of a reference collection of 300 diverse accessions.

Deviations from the work plan:
None

2006-32: Molecular Characterisation of a Representative Pigeonpea Germplasm Sample

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborating Scientists:
R Bhattacharjee, ICRISAT
DA Hoisington, ICRISAT
Subhash Chandra, ICRISAT
RK Varshney, ICRISAT
KB Saxena, ICRISAT

Mid-Year Report

A composite collection of pigeonpea comprising of 1000 germplasm accessions was established using the available phenotypic characterisation and evaluation data. The collection comprised accessions of pigeonpea mini core (146), comparator mini core, and representative accessions of landraces, breeding lines, genetic stocks, wild species and four control cultivars. The objective of this research is to fingerprint the 1000 accessions using 20 polymorphic SSR markers at ICRISAT.

The composite collection was planted in the field and leaf material from 15-20 days old plants was used to extract DNA. Twelve plants were randomly selected per accession for DNA extraction and pooled together mainly to capture the within accession variation. DNA was quantified to a working concentration of 5ng/µl. Pigeonpea being an often cross-pollinated crop and the SSR markers showing less polymorphism, it becomes important to optimize and select the most suitable SSR markers. In this context, all the available SSR markers (~ 150) were initially screened on 15 diverse accessions (8 cultivated and 7 wild), also included in the composite collection. As a result, 33 SSR markers have been identified showing polymorphism between at least two of the tested accessions.

Since DNA is extracted from 12 plants per accession, it is also important to select the SSR markers that can detect the interpretable heterogeneity in the accessions. Therefore, a series of artificial pools having different proportions of two genotypes showing polymorphism for a given SSR marker were developed and screened with the corresponding polymorphic SSR marker. The coefficients of correlations were calculated between different proportion of alleles recorded and proportion of genomic DNA used for the corresponding accession. As a result, a total of 21 SSR

markers with highly significant correlations (r^2> 0.9) were identified.

Pigeonpea is a highly photoperiod sensitive crop and that is the reason few accessions didn't flower even after 250 days under field conditions. These include 23 accessions of wild type and 4 cultivated type. These accessions have been transplanted in glass house and care is being taken to provide appropriate conditions to these accessions to flower and produce seeds.

Tangible Outputs delivered:

- DNA isolated from 12 plants per accession and pools for all the 1000 accessions prepared contributing (approximately) equal proportion of DNA from each of the 12 plants.
- Twenty polymorphic SSR markers identified.
- PCR conditions optimised following Taguchi method.
- Fluorescent-based multiplex sets comprising of four SSR markers defined.

Future plan of work:

- Genotyping the composite collection with 20 SSR markers.
- Allele calling by considering 12 plants per accession.
- Data analysis to determine the population structure and genetic diversity.
- Identification of a reference collection consisting 300 diverse accessions.

Deviations from the work plan:
Nil

2006-33: Development and Genotyping of a Composite Germplasm Sample of Potato

Principal Investigator:
Marc Ghislain, CIP
Collaborating Scientists:
Jorge Núñez, CIP
Maria del Rosario Herrera, CIP
Guillermo Trujillo, CIP
Reinhard Simon, CIP
Edwin Rojas, CIP

Mid-Year Report

Development of a new potato genetic identification kit
On the basis of the analysis of 51 single-locus SSR markers, we have selected 24 SSR markers to constitute a new potato genetic identity (PGI) kit. In order to combine SSR databases among laboratories, we have also started to develop a ladder that should facilitate the precise identification of SSR alleles. It consists of genomic DNA of selected accessions that cover most of alleles of each SSR markers (or those with high frequency) to easily detect the size of the alleles.

Transfer of SSR data to CIP and GCP database
Because the potato crop is principally a set of polyploid species, genetic dosage information is not available for most of the genotypes. Binary matrices (0s - 1s) are stored at CIPPEX (CIP database) and will be transferred to GCP database when breeders' germplasm and mapping populations have been genotyped completely.

Development of SSR fingerprint of breeders' germplasm
The breeders' germplasm is composed of 56 advanced cultivars, 123 breeding lines and 70 landraces. Most of them have resistance or tolerance to biotic and abiotic stresses, such as late

blight and drought. Along with this material, a set of 29 landraces belonging to the core collection of cultivated potato germplasm will be genotyped. DNA has been isolated for all of the 278 potato genotypes from leaves of plants grown *in-vitro*. An additional set of approximately 60 genotypes selected from recent evaluations of breeding materials for drought tolerance will be included. All DNA will be genotyped using the 51 SSR markers following the procedure developed previously.

Tangible outputs delivered:
Molecular fingerprints of approximately 800 cultivated potato accessions.
A new 24 SSR markers PGI kit.

Deviations from the work plan:
No deviations from work plan have occurred and are expected.

SP2 COMMISSIONED GRANTS

2004—SP2CL3: Sequencing of Normalised Cassava cDNA Libraries

Principal Investigator:
Sarah Hearne, IITA
Collaborating Scientists:
Chris Town, TIGR
Morag Ferguson, IITA
Nzola Mahungu IITA

Mid-Year Report
RNA was extracted from pools of root, leaf and stem meristem tissue using Invitrogen's concert RNA reagent following manufacturer's protocol. The pools of tissue comprised equal weights of tissue from each of: drought stressed and non-stressed cassava lines, Sauti, Gomani, Mbundumali, TME 1and Mkondezi which show varying responses to drought stress from typical susceptible to rapid leaf loss and maintenance of stem meristem for rapid re-growth to stay green.

All RNAs extracted were evaluated for quantity and quality in Nairobi before being sent to Evrogen for cDNA syntheses and normalisation. Two RNA samples were sent to Evrogen; root RNA and a 1:1 quantity pool of leaf and stem meristem RNA. Samples were sent to Evrogen as per their instructions (samples precipitated with DEPC treated NaOAc and ethanol, shipped at ambient temperature). Upon receipt of the materials Evrogen concluded that the root RNA had degraded during shipment and could not be used for cDNA synthesis (Fig 1 a and b). A second samples of RNA was extracted from stored root tissue, these were quantified and quality checked again using an Agilent bioanalyzer (Fig 1 c) and were of satisfactory quality and abundance for library construction. A sample was again sent to Evrogen which was received un-degraded.

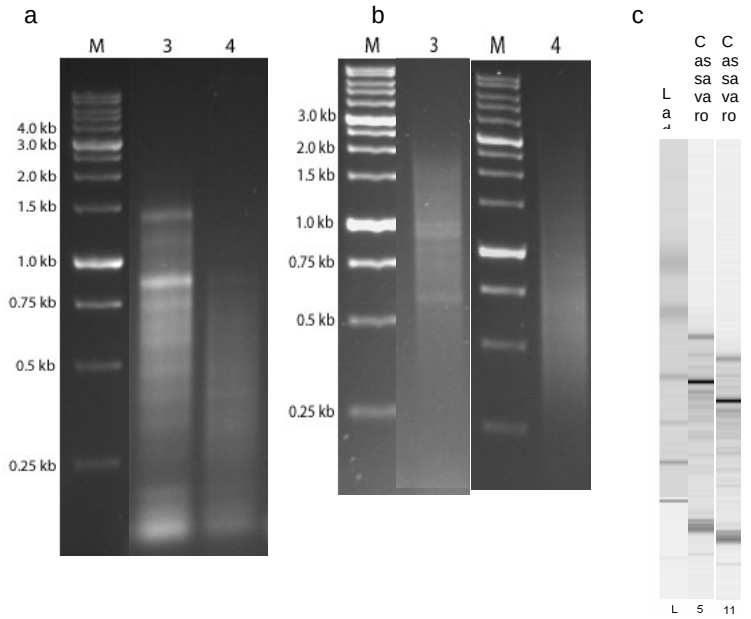


Figure 1. a) Evrogen's analysis of RNA from 3 cassava leaf/meristem and 4, root; b) Evrogen's analysis of cDNAs synthesised from RNA from 3 cassava leaf/meristem and 2, root; c) RNA from two second samples of cassava root

Evrogen have completed the normalisation of both the cassava leaf/stem meristem and root libraries. Ligation mixes of the normalised cDNAs will be will be shipped from Evrogen to TIGR within the next 10 days. TIGR is awaiting the cDNAs and will initiate the sequencing effort upon receipt of the cDNAs.

Tangible outputs delivered:
Two normalised cDNA pools from; drought stressed and non-stressed cassava leaf and shoot meristematic tissue from drought tolerant and susceptible cassava lines and drought stressed and non-stressed cassava root tissue from drought tolerant and susceptible cowpea lines.

Deviations from the work plan:
The process of plant growth, stress and RNA isolation took longer than anticipated for various reasons (detailed previously). However, all issues leading to this have now been addressed.

Due to the time delay in obtaining RNA we request a NCE to November to allow TIGR to receive and complete sequencing of the libraries.

2005-09: 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
Hei Leung, IRRI
Emmanuel Guiderdoni, AGROPOLIS
Mathias Lorieux, IRD/CIAT
Manabu Ishitani, CIAT
Tiegang Lu, CAAS
Qifa Zhang / Deming Jin, HAU, China

Collaborating Scientists:
Gyn An Pohang, Univ. Science and Technology, Korea.
Srinivasan Ramachandran Temasek Lifesciences Laboratory, Singapore
Narayana Upadhyaya, CSIRO
Venkatesan Sundaresan, UC Davis

Mid-Year Report

1. A stress associated gene SAG database of Abiotic/biotic candidate genes had been compiled for mutational analysis, identified from literature and microarray experiments.

2. From available international mutant collections, mutants were identified for these stress associated rice genes using the OryGenes DB (http://orygenesdb.cirad.fr) with about 80,000 available insertion sequences tagged by FST sequence on the rice genome. The available insertion mutants provide mutants for about 23% of the SAG candidates that are then phenotyped.

3. Gain-of-function genotypes: Overexpression constructs for about 20 SAG candidates were made and are being transformed into Nipponbare at WUR and CAAS. The T1 lines of the transformed lines are tested under greenhouse conditions for drought stress resistance. Overexpression lines of the AP2/ERF rice/Arabidopsis transcription factor genes (*CBF*, *SHN*, *HRD*) were tested in the greenhouse and display improved water use efficiency or drought stress resistance in the vegetative stage.

4. Forward Genetic screen: for drought stress of T-DNA tagged lines in 2005 revealed a large number of false positives at HAU, so efforts were shifted to reverse screens.
At CAAS a greenhouse based seedling screen of 2000 T-DNA tagged lines revealed 2 salt sensitive mutants that are analysed further.

At IRRI, 10 mutants were identified for elongated upper internode (*eui*) mutations that control elongation of the peduncle, a critical factor in sustaining fertility under reproductive stage stress. We studied their peduncle elongation rate, duration of elongation and its responsiveness to drought. The mutants were found to elongate at the rate of 10-15 cm per day when compared to 4-5 cm per day in the wild type (IR64) and the final peduncle length reaches around 50 cm (30 cm in IR64). All the mutants were found to be sensitive to drought but they recovered rapidly after re-watering and were able to show complete exsertion of the panicles (summarised in Table 1 in Appendix).

Using candidate genes that are predicted to be involved in GA metabolism, we conducted expression analysis on the panel of *eui* mutants (Figure 1). We found no knock-out mutation in the GA catabolic GA2 oxidase family genes but found four mutants (FNS4655-10, FNS6171-1, FNS4503-3-1 and FNS7120-5) that appeared to have knockout mutations at the *eui-1* locus (coding a cytochrome P450 involved in GA biosynthesis pathway, recently cloned by Zhu et al., 2006). Given the success of this approach, we are screening the mutants with more candidates involved in GA signalling.

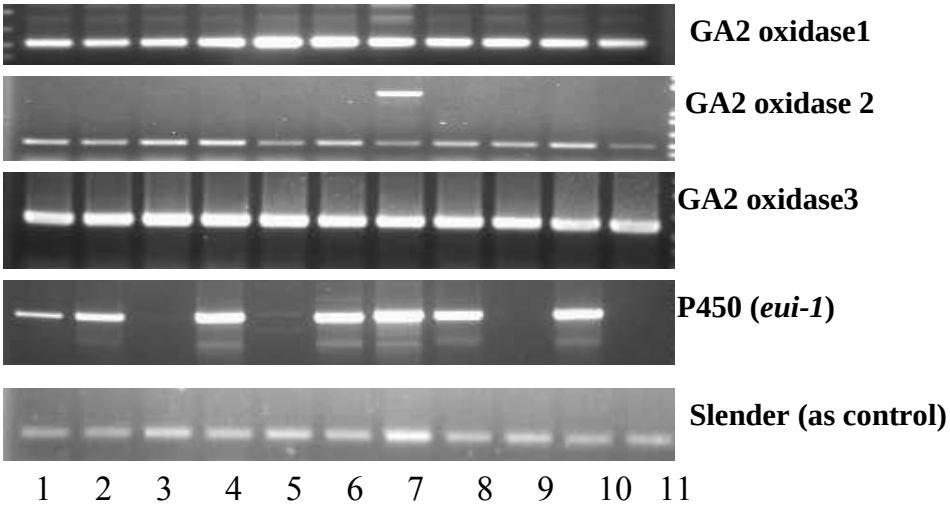


Figure 1. Analysis of transcripts in a panel of *eui* mutants, 1 = IR64 wild, 2 = D2660-1-2, 3 = FNS4655-10, 4 = D11041-8-1, 5 = FNS6171-1, 6 = N1053-3-2, 7 = D11324-2, 8 = D11031-8, 9 = FNS4503-3-1, 10 **= G7381-**

5**.** Reverse Genetic screens of SAG candidates for abiotic and biotic stress phenotypes:

Biotic screens:
At CIAT a set of 100 SAG T-DNA insertion mutants in Nipponbare were screened for sheath blight resistance/susceptibility, revealing putative mutants (two resistant and three susceptible lines) that are currently being re-evaluated prior to conduct molecular segregation analysis.
At IRD/CIRAD, from a candidate SAG list of 480 genes, 116 T-DNA mutant lines (Génoplante collection) were selected and screened for altered phenotypes to bacterial leaf blight (*Xanthomonas oryzae*/ pv. /oryzae (Xoo) /(PXO339, Phil race 9), revealing 15 candidate mutant lines.

Abiotic Drought stress screens:
At CIAT: Among the 298 SAG candidates, 62 T-DNA T1 lines of Nipponbare were tested for drought tolerance, along with some checks during the dry season starting from December to February in CIAT-Plamira campus. One month old seedlings were transplanted to the field in December, 2005 and drought started one month later after planting by withdrawing water. Under non-stress and drought conditions there were no significant differences for plant growth among the lines compared with non-transgenic plants, in terms of morphological and agronomical parameters for drought tolerance such as leaf rolling, plant height and tiller number.

At HAU in 2006 a total of 698 T-DNA mutants (for 694 identified SAGs) were selected from our mutant database (http://rmd.ncpgr.cn/) and were screened in the water-controlled field (Hainan, China). Drought stress was initiated 1 month after transplanting. Intermittent overnight irrigation was applied to avoid overstress. Visual scoring of leaf rolling and drying was carried out in field during the development of drought stress to find segregations of drought sensitivity/resistance within mutant families (20 plants/family). Eighteen mutant families showed segregation of drought sensitivity. Interestingly, most of the SAGs of the mutants showing segregations of drought sensitivity are transcription factors or protein kinases.

Co-segregation analysis of drought sensitivity using PCR to check the genotypes, revealed only 4 families (out of 16 tested) to be potentially tagged. The four SAGs for the co-segregated families encode a GT3-like transcription factor, a bZIP transcription factor, a mitogen activated protein kinase kinase kinase (MAPKKK) and a calcium dependent protein kinase (CDPK) respectively.

Since the co-segregation was done at single plant level, the results will be confirmed at family level in the next generation.

Analysis of mutations in small RNA processing genes and ERECTA:
At IRRI, we assembled a set of insertion mutations in a) genes affecting small RNA metabolism and b) the ERECTA gene that has been shown to be an 'integrator" responsible for transpiration efficiency in Arabidopsis. Insertion lines were obtained from Gyn An's T-DNA and Hirochika's Tos17 collections. Initial morphological phenotypes have been collected and seeds are being harvested for drought phenotyping and segregation analysis.

a) T-DNA insertion lines for small RNA processing genes were obtained from Pohang University of Science and Technology in Korea. These 13 different insertion lines of *japonica* rice lines (in Dongjin and Hwayoung backgrounds) at T-2 generation, carry insertions in pGA2707 (6.3 kb) or pGA2715 (8.2kb).

As of now, the mutant lines with insertions in the dicer gene, which presumably process either trans-acting siRNAs or intergenic-region-derived siRNAs, show no obvious developmental defects. The mutant lines with a disrupted gene encoding a dicer (known to process virus-derived siRNAs) showed multiple morphological characteristics. At vegetative stage, we observe dwarfism, reduced culms number, and smaller tillers. The T-DNA insertion lines in RNA-dependent RNA polymerase (important in the biogenesis of intergenic-region-derived siRNAs in plants) appear to be homozygous lethal. We are now advancing the mutants to T3 for stress phenotyping and molecular analysis.

b) Tos17 Insertion lines in ERECTA and small RNA processing genes
We obtained a set of 14 Tos17 insertion mutants in Nipponbare for ERECTA (NE 4007, ND 4028, NF 9866, and ND 6038) and small RNA processing genes (NC 0294, ND 0065, ND 4544, H 0543, ND 4559, NG 0316, NE 5023, NE 2787, NF 0589, and NE 3522). The insertion positions of Tos 17 are found in the intergenic regions, exons or introns

We have multiplied the seeds for analysis of co-segregation of molecular and agronomic phenotypes.

All the mutant lines along with Nipponbare were planted in February 2006 with an average of 18 plants/line. Vegetative stage-phenotyping was recorded 35 days after sowing and up till maturity. Based on the phenotypes observed, not all mutant lines showed segregation in Mendelian ratios. For example, ND 4028 showed lethality (7 plants), completely sterility (1 plant) and semi-dwarf (5 plants). ND 6038 showed albino (3 plants) and yellowing (2 plants). H0643 shows increased plant height, late heading (~25 days) and longer awn compared to Nipponbare. About 30% of NF 2787 segregants showed partial sterility. In the coming months, advanced progeny with insertions in ERECTA and small RNA genes will be planted in three replications and exposed them to drought conditions to observe the parameters for drought such as leaf rolling, plant height, tiller number and also stomatal conductance.

Tangible outputs delivered:
Reverse genetic screen of 694 SAG T-DNA tagged lines for drought sensitivity identify 18 mutant families, out of which 4 mutants (GT3-like and bZIP transcription factors, MAPKKK, CDPK) show co-segregation of insert and phenotype.

Reverse genetic screens of biotic SAG insertion mutants revealed 15 putative mutants susceptible to bacterial leaf blight and 5 putative mutants with an altered response to sheath blight.

Gain-of-function overexpression lines for AP2/ERF transcription factor genes (CBF, SHN, HRD) revealed improved drought resistance and water use efficiency phenotypes at vegetative stage.

Deviations from the work plan:
Forward genetic screens for drought revealed a high proportion of false positives at HAU, redirecting the project to reverse screens for SAG insertion mutants.

Drought screens in the field at CIAT were not successful due to uncontrolled environmental conditions. An alternative screen-house based large tray screen will be carried out on the SAG candidate mutants.

Due to delays brought about by difficulties in phenotyping a non-cost extension of the project is foreseen for some labs.

2005-10: 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:
Tom Payne, CIMMYT
Collaborating Scientists:
Peter Langridge, University of Adelaide, Australia
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 California, Davis
Bikram Gill and Bernd Friebe, Kansas State University
Perry Gustafson, USDA-ARS, University of Missouri
Adam Lukaszewski, University California Riverside
Mark Sorrells, Cornell University

Mid-Year Report
Not Received

2005-11: Legume Mutant Resource Development

Principal Investigator:
Matthew Blair, CIAT
Collaborating Scientists:
W. Broughton, Univ. Geneva
P. Lariguet, Univ. Geneva

Mid-Year Report
We are continuing the development of the TILLING populations for common bean with the following achievements:

1) Chemical mutagenesis: CIAT received from Univ. of Geneva a second set of 163 M1 families (Pv 1008 to Pv 1215) bringing the total to 514 families developed by 9 ethyl methane-sulfonate (EMS) mutagenesis. Univ. of Geneva and USDA is continuing further mutagenesis with an additional 4 kg. of seed that we prepared at CIAT as described above however the

EMS concentration used has been lowered to 35 mM. A total of 870 seed from these 163 families was planted in the screenhouse (we reduced the number of individuals per family to six but we still find it useful to plant several individuals per family to ensure that at least two to four plants per family produce enough seed for the seed increase in the field).

2) M1:2 Generation advance: in an enclosed screenhouse, we harvested the M2 generation from the previous set (Pv 271 to Pv 544) and produced a total of 1482 new M2 progeny. This work brings to a total of 3,152 the number of M2 progeny generated.

3) M2:3 generation advance: we are increasing 686 M2 single plant selections from the second increase of M2 progeny in the field at CIAT during the 2006A season (April to June 2006). This will bring to a total of 1,152 the number of M2:3 lines generated and field increased as 466 M2 single plant selection were increased in the 2005B season (September to December 2005).

4) Single seed descent from M2:3 to M3:4: we are increasing seed from the 463 lines produced in the field experiment in the 2005B semester (September to December 2005) by single seed descent in the greenhouse. The increase in the greenhouse was precipitated by flooding which affected a March planting of these same genotypes in the field at CIAT (note: the M2:3 generation discussed in point 3 was not affected by this flooding, thankfully).

5) Phenotypic screening: the first set of 800 M1:2 plants and the second set of 1000 M1:2 plants have or are being 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 in both the greenhouse and the field growing cycles. A database is being constructed with these photographs and characterisation data.

6) DNA extraction: a miniprep DNA extraction technique developed last year was applied to extract DNA from an 827 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 analysed in the phenotyping described above.

7) CelI production and testing: a protocol was followed to obtain celery juice extract and testing of the CelI enzyme is underway with known SNPs in PCR products from allelic variants at the SR2 (SCAR), Banyuls and DREB (candidate gene) loci. Once we have pools for the mutant population these EcoTILLING protocols will be transferred to the TILLING population and mutations will be sought in these same genes as well as other 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 analysed as a proof of concept for the value of the mutant stocks generated by this project.

8) Databasing: as mentioned last year a database is being created for the mutant collection which will contain notes and photographs of phenotypic differences (see slide) as well as information on DNA extraction and seed production.

Tangible outputs delivered:

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

Deviations from the work plan:
The project is continuing well and we are mostly on schedule. We have produced a large number of M2 progeny (almost 3000 which was the goal of the project) and have increased or are increasing these by single seed descent to the M3 and M4 generations, however many of the lines are related within a more limited number of families. We have run into the expected problem of slow seed increase for common beans (unlike many cereals, common beans given their large seed size have a low multiplication ratio) and unexpected problems such as recent flooding in our production site and for this reason we are planting several M1:2 plants per family and single seed descents with remnant seed to ensure we do not lose the line. A bottleneck is the production of M1 plants due to the high rates of deleterious mutants at EMS concentrations of 45 to 50 mM. The Univ. of Geneva has reduced the concentration of EMS to try to ameliorate this bottleneck but we will need to evaluate the frequency of mutation at the new concentrations. DNA extraction continues to be by a standard technique that gives good quality (see slide) but we are trying to work on a way to increase throughput. We have overcome the potential bottleneck of the extraction of CEL I endonuclease (McCallum et al., 2000) and we are excited about using acrylamide and agarose gel based method of TILLING and EcoTILLING (K. McNally, pers. communication).

2005-12: A Saturated Potato Mutant Population for Functional Genomics among Solanaceae and Tuber Crops

Principal Investigator:
Marc Ghislain, CIP
Collaborating Scientists:
Meredith Bonierbale, CIP
Alberto Salas, CIP
Maria Herrera, CIP
Glenn Bryan, Scottish Crop Research Institute
Robbie Waugh, Scottish Crop Research Institute
Dani Zamir, Hebrew University of Jerusalem

Mid-Year Report
At CIP, as a result of homozygosity assessment of existing *S. verrucosum* material in 2005, we identified two accessions with the lowest heterozygosity and planted five plants from each accession for selfing. All plant material was processed in Huancayo field station and berries have been recently been collected. After proper drying and germination, 50 plants will be grown from each of the five plants of the accession TRHRG 161 and analysed using 5 AFLP primer combinations. The following protocol has been developed at CIP to assess lethal dose (LD) of EMS (ethyl methane sulphonate) treated seeds. 2000 seeds will be treated as follows: 400 seeds will be treated with 0, 2.5, 5.0, 7.5, 10 % of EMS following the protocol established for tomato seeds (provided by Hebrew Univ.). Seeds will be germinated on filter paper and transferred carefully to seedbed. Observations of lethality will be after germination and on 1 month-old plants.

At SCRI seed of CPC54 population 04/04a had been sown in glasshouse late spring 2005. 346 plants were sampled for DNA extraction in September 2005. AFLP's were run on all 346 plants using three PstI/MseI primer combinations. Overall this population showed unacceptably high levels of marker polymorphism, with ~15 of AFLP fragments showing some level of polymorphism. However, 40 plants, which showed a uniform AFLP fingerprint, were selected. These were transferred to a location with additional lighting to improve formation of berries and thus increase the yield of seed. Seed from these plants will be re-sown in the glasshouse to generate more seed in 2006. A further 58 plants were isolated that, according to their AFLP fingerprint, looked completely different from the 40 homogeneous plants. Bulked seed from these was retained for future work. The remaining 218 plants were separated from the others and

bulked seed from these is to be used in mutagenesis studies to calculate doses of mutagen required.

At the Hebrew University, this February the Zamir group treated diploid seed of *Solanum verrucosum* with EMS. The germination was high and the seedlings were transplanted to the field in early March 2006. The M1 seedlings were sown in the nursery and produced small tubers and upon transplanting we detected three M1 seedlings that did not produce any tubers as well as other mutants with distinct phenotypic alterations. Just after the transplanting it rained heavily and the temperature was low and seedling lethality in the field was high. The surprising result was that the tuberless mutants were not affected by the stress and presently they are flowering and still did not produce tubers; seed derived from selfing of the tentatively dominant mutants will reveal the genetic basis of the detected phenotype. We planted the available diploid seed again in the nursery and will transfer them to the field in June for seed increase. One of the long-term objectives of this project is to compare mutant repertoire among Solanaceae species. In 2005 The Zamir group also generated an EMS population of eggplant. Approximately 600 m2 families were planted in the field (24 plants per family) and are continuously surveyed and photographed for mutant variation.

Development of a bioinformatics framework to access Solanaceae mutant populations

The *Solanaceae Genome Network* (SGN; http://sgn.cornell.edu.) is a relational database (MySQL) that hosts all published gene/EST/map information from Solanaceous species in a comparative genomics format. Our objective is to expand the "The Genes that Make Tomatoes" (http://zamir.sgn.cornell.edu/mutants/) to a site that will be called "The Genes that Make Solanaceae". For this reason the Zamir group constructed a phenotypic catalog that will include specific phenotypes that are relevant to the Solanaceae species that are being mutagenised in this project – potato and eggplant.

Tangible outputs delivered:

- Selfed seeds of highly heterozygous lines.
- First generation of EMS potato mutants of *S. verrucosum* generated.
- One tuberless mutant identified.

Deviations from the work plan:

No deviation from of the first year workplan.

2005-13: Crop Gene Expression Profiles and Stress-gene Arrays

Principal Investigator:

Tiegang Lu, CAAS

Co-Principal Investigators:

Guozhen Liu, Beijing Genomics Institute, China
Shoshi Kikuchi, NIAS
Manabu Ishitani, CIAT
Raveendran Muthurajan, IRRI
Manavalan Laxmi Praba, IRRI
Lafitte Renee, IRRI
Hei Leung, IRRI
Kouji Satoh, NIAS
Qiang Gan, BGI

Mid-Year Report

1. Drought Responsive Gene Expression Analysis in Rice during Vegetative Stage (CAAS-BGI-NIAS-IRRI)

The experiment was undertaken with an aim of identifying genes or gene sets that are common or unique to genotypes with different phenotypic response to drought. We have completed the microarray analysis of IR 64 (drought susceptible rice variety) and Apo (drought tolerant rice variety) using two chip platform, the 22K Agilent rice oligoarray and the BGI 60K oligoarray. For each of the oligoarrays, we conducted 3 biological replicates for two genotypes (Apo vs IR64) under two water regimes (drought stress vs control). Drought stress was imposed during the vegetative stage (33 days after sowing). Water was withheld until pots reached 20% field capacity (FC) or till the development of severe leaf wilting symptom. In rice, wilting started after 5 days of stress and leaf sampling was done on 6^{th} day of stress in both control and stressed plants.

Table I shows the number of genes that showed significant drought responsiveness for their expression in the leaf tissues of rice genotypes IR64 and APO. The analysis of 22K Agilent Oligo array data revealed that among the significantly expressed genes, 77 genes were commonly up-regulated and 153 were commonly down regulated in both IR64 and APO. The existence of variety specific response may be related to the divergent degree of drought tolerance between two rice genotypes.

Table1. Classification of drought inducible genes in IR64 and APO. The drought inducible genes were categorised into three groups: (1) genes induced in APO; (2) genes induced in IR64; and (3) genes induced in both. The genes with expression log ratios (drought stressed/unstressed) greater than 2 times the average of the three experimental sets (P = 0.01) were considered as drought stress-inducible genes.

Expression	APO	IR 64	Common in both APO and IR64
Number of genes Up regulated during drought (> two fold change)	467	89	350
Number of genes Down regulated during drought (< 2 fold)	543	98	420

The tolerant genotype Apo shows more number of changes. We are now examining the genesets which are specific to the tolerant genotype Apo. Apo specific changes include upregulation of about 170 genes and down regulation of about 204 genes where the difference between IR64 and Apo was found to be more than two fold.

We also investigate another alternative way to contrast the expression patterns. Instead of looking single genes with significant differential expression, we asked whether there are regions of contiguous genes that are co-expressed together (we call it region of correlated expression, RCE) as a response to the conditional stress. In this analysis, we scanned the entire genome with a 10-gene window space, and asked whether the genes within this space were correlated in their expression (either up or down regulated). Such regions were detected statistically by comparing their occurrence with that by chance. This method was tested using multiple gene expression datasets and shown to be robust.

When applied to the dataset of Apo, we detected approximately 10 RCE using a 6-gene window. Some of these RCE appeared to co-localise with previously observed QTL regions for leaf-rolling, leaf-drying, osmotic adjustment. Though the QTLs are coarsely defined, this approach appears promising in identifying regions of chromosome with potential contribution to drought tolerance. We expect complete analysis of the entire datasets from Agilent and BGI oligoarrays

should yield new insights into the relationship between gene expression and phenotypic expression of drought tolerance.

2. Stay green phenotype (CIAT)

At CIAT, we continued to establish drought phenotyping method for stay green phenotype in rice in the green house, and to evaluate the genotype with the parental lines under water limited conditions for molecular phenotyping. Experiments consists of three rice varieties, Fedearroz 50, known as a stay green genotype, its mother P5413-8-3-5-11 which is a commercial variety calls Oryzica Llanos 4, and its father PC 1274-6-8M-1-3M-1.

The first experiment comprised intermittent drought stress(starting at 100% of field capacity-FC level, drying out until 25% FC and keeping available water at that level) and control treatment (100%FC). Plants were established under well irrigated conditions during 1 month, after that treatments were imposed on three plants of each genotype. Plants were kept until grain filling stage. Chlorophyll content was measured at 21 and 110 days after treatment-day, photosynthetic rate was taken at 21 day and number of tillers and plant height at 40 day.

During the first experiment a reduction in growth was observed (Fig. 1). The number of leaves, tillers and plant height were affected by the drought stress. There were no differences between control and 75%FC, and between 50%FC and 25%FC, it means that available water under 50%FC has strong effects on the growth and development of rice, plants began to delay the flowering time and kept themselves with the low amount of water they took out from the soil each watering time.

Fedearroz 50

Oryzica Llanos 4

PC 1274- 6-8M-1-3M-1

Figure 1. Three genotypes of rice exposed to drought conditions, cylinders in each photo from the left to right has different FC of 100, 75%, 50% and 25%, respectively.

Chlorophyll contents were similar in the parents at 21 day, however Fedearroz 50 increased the chlorophyll under water stress conditions. In the physiological maturity of the grain chlorophyll contents were lower in the control plants, the father has lower chlorophyll content its senescence began early that in the other genotypes. No differences in chlorophyll content were showed under drought stress conditions at maturity stage (Fig. 2).

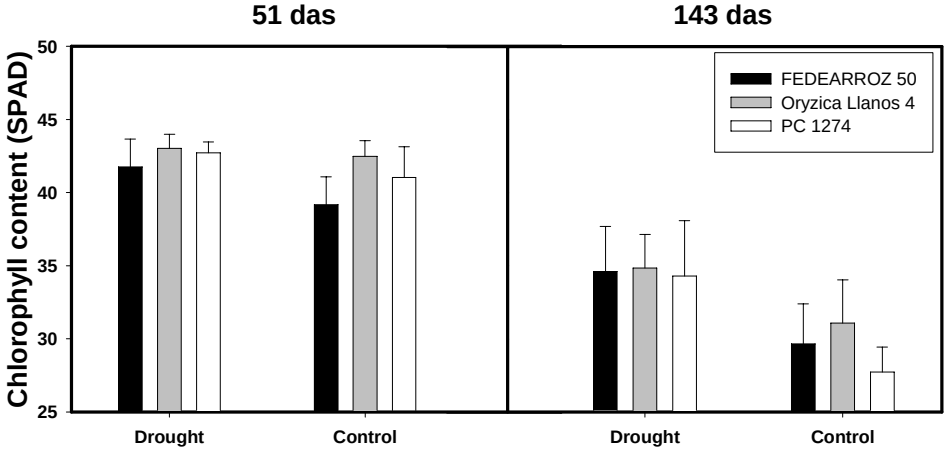


Figure 2. Chlorophyll content during the vegetative and reproductive stages of three rice genotypes under intermittent drought conditions

Second experiment comprised 2 types of drought: intermittent drought and terminal drought stress (starting 100%FC and drying out) and control treatment. Treatments were imposed at flowering time around 100 days after sowing-day. Transpiration rate, stomatal conductance and chlorophyll content were measured 80 day while plants were growing, chlorophyll was measured after treatments were imposed at 105 and 129 day and number of tillers were measured at 108 and 140 day moment of maturity stage.

While plants were growing in the second experiment, we found that Fedearroz 50 seems to have lower stomatal conductance and transpiration rate as a mechanism of low use of water keeping its normal growth (Fig. 3) under well-watered condition.

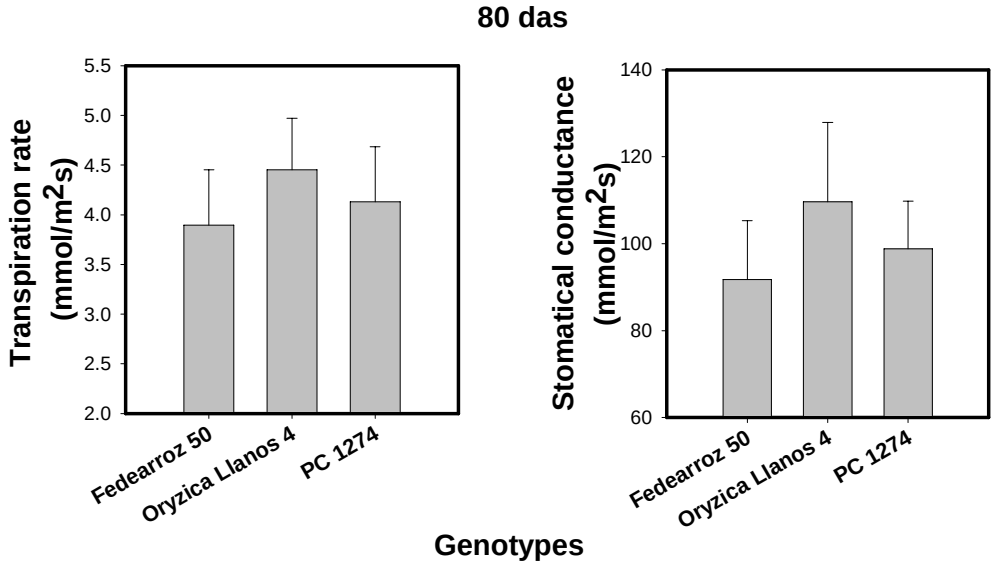


Figure 3. Physiological attributes of three genotypes of rice during the vegetative growth under control conditions.

Chlorophyll contents were near to 40 SPAD during the vegetative stage, there were not differences between genotypes. At 129 days under the control condition Fedearroz 50 kept more chlorophyll than its parents, that response was not showed in the drought treatments (Fig. 4).

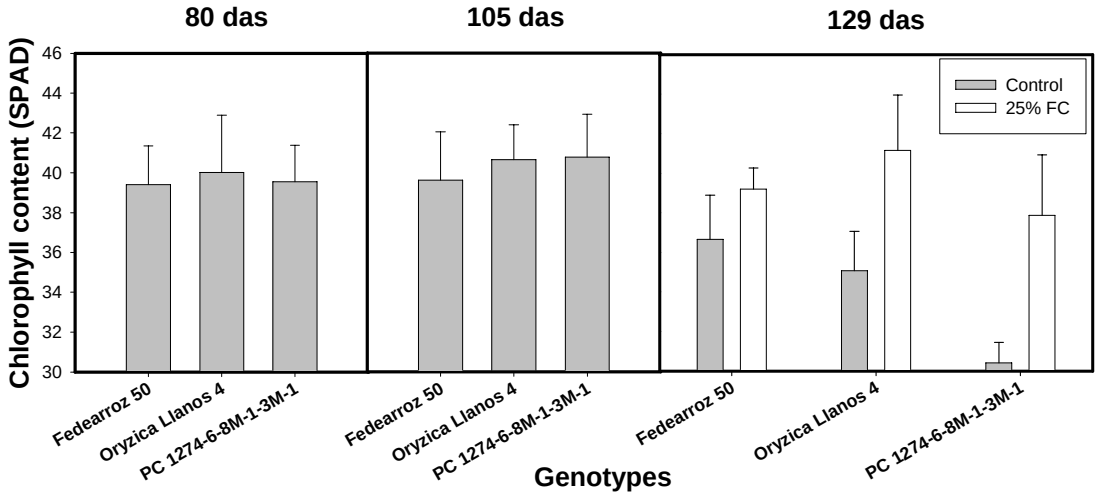


Figure 4. Chlorophyll content of three rice varieties during vegetative stage and at the reproductive stage after 15 days of intermittent drought

The third experiment had a quick water loss in the plants at the vegetative stage for terminal drought due to hot and dry conditions in the greenhouse. We could not find significant differences between the genotypes in terms of physiological aspects on the stay green genotype.

The stay green phenotype of Fedearroz 50 was found particularly under non-stress conditions when compared with PC 1274-6-8M-1-3M-1 as shown in Figure 4. Under different water limited conditions, we did not observe consistent phenotypic difference between the genotypes for the stay green phenotype. This could be due to little genetic difference between the materials. We harvested the leaf materials that indicated physiological differences under control and drought stress conditions for gene expression analysis.

Tangible outputs delivered:

- Complete gene expression datasets on two gene chip platforms for two rice genotypes (Apo vs IR64) with contrasting response to drought
- A short gene list (<100) with strong differential expression unique to drought tolerant genotype Apo.
- Analytical means to extract regions of correlated expression, potentially leading to alignment between correlated expression regions and QTL.
- Detailed phenotypes of three rice varieties, Fedearroz 50 (stay green) and it parental lines (Oryzica Llanos 4 and PC 1274-6-8M-1-3M-1) determined under water stress conditions. It sets the stage for sampling of RNA for expression analysis.

Deviation from original workplan:

The exploration of rice oligo arrays as a heterologous gene expression assay for wheat and maize was discontinued. Since both maize and wheat arrays are now available for experimentation, expression analysis will be done on “homologous platforms”. Since the rice oligoarray platform is the most accessible, the group will concentrate on applications of rice chips to generate quality

data to compare with expression data from different platforms to infer function of orthologous genes. We redesigned the experiments to better integrate with existing projects on drought.

The group will discontinue the experiments with artificially induced stress (e.g. PEG treatments). Rather the group will concentrate on experimental treatments and genetic materials that are directly relevant to on-going work in the dissection of drought resistance mechanisms.

1. Conduct gene expression on rice reproductive stage (during anther development) under drought and heat stress using two genotypes (IR64, Moroberekan). We may include a third genotype, N22, if time and resource permit. This set of experiments will be linked to the competitive grant on **"**Identifying Genes Responsible for Failure of Grain Formation in Rice and Wheat under Drought."
2. Examine gene expression at vegetative growth in key genotypes (parents used in producing mapping and advanced backcross populations) under drought stress. This set of experiments will be linked to "Determination of a common genetic basis for tissue growth rate under water-limited conditions across plant organs and genomes."

2005-14: Stress Response-enriched EST Resources for Targeted Species

Principal Investigator:
Sarah Hearne, IITA
Collaborating Scientist:
Chris Town, TIGR

Mid-Year Report

RNA was extracted from pools of root, leaf and stem meristem tissue using Invitrogen's concert RNA reagent following manufacturer's protocol. The pools of tissue comprised equal weights of tissue from each of: drought stressed and non-stressed cowpea lines, Dan Ila (type II drought tolerance), Tvu11986 (type I drought tolerance) and Vu7778 (susceptible).

All RNAs extracted were evaluated for quantity and quality in Nairobi before being sent to Evrogen for cDNA syntheses and normalisation. Two RNA samples were sent to Evrogen; root RNA and a 1:1 quantity pool of leaf and stem meristem RNA. Samples were sent to Evrogen as per their instructions (samples precipitated with DEPC treated NaOAc and ethanol, shipped at ambient temperature). Upon receipt of the materials Evrogen concluded that the root RNA had degraded during shipment and could not be used for cDNA synthesis (Fig 1 a and b). A second samples of RNA was extracted from stored root tissue, these were quantified and quality checked again using an Agilent bioanalyzer (Fig 1 c) and were of satisfactory quality and abundance for library construction. A sample was again sent to Evrogen which was received un-degraded.

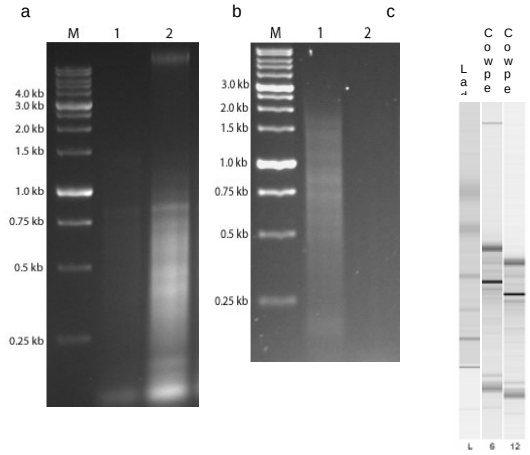


Figure 1. a) Evrogen's analysis of RNA from 1 cowpea leaf/meristem and 2, root; b) Evrogen's analysis of cDNAs synthesised from RNA from 1 cowpea leaf/meristem and 2, root; c) RNA from two second samples of cowpea root

Evrogen have completed the normalisation of both the cowpea leaf/stem meristem and root libraries. The efficiency of nomalisation is being assessed by Evrogen using the leaf /stem meristem library, the results should be available shortly. Ligation mixes of the normalised cDNAs will be will be shipped from Evrogen to TIGR within the next 10 days. TIGR is awaiting the cDNAs and will initiate the sequencing effort upon receipt of the cDNAs.

Tangible outputs delivered:
Two normalised cDNA pools from; drought stressed and non-stressed cowpea leaf and shoot meristematic tissue from drought tolerant and susceptible cowpea lines and drought stressed and non-stressed cowpea root tissue from drought tolerant and susceptible cowpea lines.

Deviations from the work plan:
The process of plant growth, stress and RNA isolation took longer than anticipated for various reasons (detailed previously). However, all issues leading to this have now been addressed. No additional funds have been required and all monies intended for sequencing are available for transfer to TIGR.

Due to the time delay in obtaining RNA we request a NCE to November to allow TIGR to receive and complete sequencing of the libraries.

2005-15: *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, UK
Jaroslav Dolezel, Institute of Experimental Botany, Czech Republic

Mid-Year Report

Objective 1: Consolidate existing publicly available Musa genomic resources and make new resources available (ESTs, BACs, sequence data)

Activity 1: *Consolidate publicly available Musa genomic resources (EST collections, BACs)*
The EST cloneset from Syngenta has been verified by sequencing of eleven key clones identified from sequence comparison with biotic and stress related genes. 90% of sequences are from the expected EST, and the availability of complete, high quality sequence from the ESTs (rather than the single-read in the database) is valuable for comparison with genomic sequence data from BACs and genomic PCR products. Given the history and long storage of the clone-set, finding only one erroneous clone is a good result. The clones now fully sequenced include some drought-responsive, cellulose synthase and biotic stress related genes.

Activity 2: *Production of DNA pools from Calcutta 4 BAC library*
The MA4 library, which is ordered in 144 x 384-well plates, has been pooled and is now available in the following format:
1) 18 tubes with super pools (each prepared from 8 x 384-well plates)

2) 9 x 96-well plates with row and column pools (each prepared from 8 plates), and individual plate pools
The pools were already sent to University Leicester and the preliminary results indicate that they are suitable for PCR screening.

Activity 3: *Continue to work to make private and nationally held data and resources available for public-good plant breeding*

- The genomic BAC library of cv. Grande Naine was sent by CIRAD to the Musa Genome Resource Centre (MGRC) in December 2005.

Objective 2: Establishing international mapping populations with the view of high precision genetic mapping

Activity 4: *Generate a new mapping population and to map it*
The mapping F1 population "Borli" (coming from the cross *Musa acuminata* 'Borneo' X *Musa acuminata* 'Pisang lilin' which was previously chosen) was generated. The *in vitro* plantlets were transferred to green house and leaves from 180 individuals were harvested between January and February 2006. Extractions of these individuals have been performed using the Matab protocol (Risterucci *et al.*, 2000). Periodically leaves from the population are harvested and frozen for further analysis if needed.

Risterucci A.M., Grivet L., N'Goran J.K.A., Pieretti I., Flament M.H., Lanaud C (2000) A high-density linkage map of Theobroma cacao L. . Theoretical and Applied Genetics, vol. 101, n. 5/6, p. 948-955).

Activity 5: *Ploidy characterisation: chromosome counting on roots tips in association with flow cytometry*
7 plants were analysed by chromosome counting. They were all diploids. Flow cytometry was performed on 56 plants. All these plants gave diploid levels by this method. Flow cytometry will be extended to all individuals. Further chromosomes counting will be conducted upon results obtained by flow cytometry and/or SSR markers analysis.

Activity 6: *Develop a back up mapping populations (if the first one cannot be obtained)*
Two other mapping populations: "Malaga" and "HDPam" mentioned in previous report are in standby but are not considered here since Borli was finally selected for its earlier availability, the number of individuals and the greater heterozygocity between parents

Activity 7: *Generate SSR markers from EST sequences*
The 143 SSR markers defined from the NIAS sequences (see objective 4 activity 14) are under analysis at the moment on the population.

Activity 8: *Map SSR markers*
Until now, 100 SSR markers have been amplified on the 180 Borli individuals. SSRs were revealed following the protocol of Roy et al. (1996) applied with the automated infrared fluorescence technology of a sequencer LICOR IR2. The images of migration are analysed using the AFLP quantar pro software. Among these first SSR markers, 50 primers pairs gave rise to polymorphic patterns in the progeny. Among the others, some did not amplify, some gave monomorphic patterns, some gave unexpected features that could be interesting to analyse and some are not yet analysed.

Roy R, Steffens DL, Gartside B, Jang GY, Brumbaugh JA (1996) Producing STR locus patterns from bloodstains and other forensic samples using an infrared fluorescent automated DNA sequencer. Journal of Forensic Sciences JFSCA 41(3):418:424.

Objective 3: Developing publicly available Musa maps anchored with genetic and EST markers with reference to gene and genome sequences from rice and other models

Activity 9: *Characterise major repeat classes in relation to genomic diversity*

614 repetitive DNA sequences, which were isolated from Calcutta4, were analysed for homology with known DNA sequences and genomic arrangement. Sequences displaying interesting properties were physically mapped by PRINS on metaphase chromosomes of M. acuminata 'Calcutta4' and M. balbisiana 'Tani'. This work significantly expands the knowledge of the repetitive fraction of the Musa genome and long-range organisation of chromosomes.

Activity 10: *Test sequences identified for polymorphisms between genomes and accessions*

Activity 11: *In silico EST detection on rice genome*

Objective 4: Selective sequencing and immediate publication of A and B genome BAC clones bearing genes potentially involved in important traits; sequencing of random and targeted ESTs, and immediate publication of sequences

Activity 12: *Defining BACs for sequencing through PCR and EST hybridisation approaches*

5 candidate clones have been sent to NIAS via the Musa Genome Resource Centre (MGRC). In order to compare A and B genome sequences, Calcutta 4 (AA) BAC library was hybridised with 4 probes of the ADH1 locus. CIRAD will send these clones of interest to NIAS as soon as fingerprint is achieved. After automatic annotations of AA genome ADH1 locus, Pisang Klutug Wulung (BB) BAC library will be hybridised with homologous probes defined on banana sequences as heterologous probes doesn't give clear signals on BB genome.

Screening by hybridisation to high density arrays of BACs on filters:
Biotic stress genes were successfully screened using this approach. Clones containing NBS-LRR and GLPL domains have been identified in clones and the corresponding BACs pulled out and verified by PCR screening.
Unfortunately the equivalent screening with ESTs from was highly equivocal. A series of candidate BACs was identified from poor hybridisation results, where the observed patterns did not correlate well with the grid-spotting patterns. It was uncertain if the problem was poor probe-target homology, multiple use of the filters, or otherwise. Therefore, the strategy was changed to BAC pooling below:

Screening by PCR amplification from BAC pools The gene DRFP (drought responsive family protein) has been identified in plate pool 32, row F, column 2. HSP (Heat shock protein, class 1 LMW) has been identified in plate pool P22 and plate pool 18 (fainter band), row pools H and J, column pools 6 and 7. These BACs are now being processed to verify the identification of the target sequences in BACs. As of May 2006, LTIP (low temperature induced protein) has been found in plate 52 pool, LTSRP (Low temperature and Salt responsive protein) is in plate 16, 70HSP (70kDa Heat shock protein) in plate 9 and STP5 (salt tolerance protein 5) gives a faint response in plate 4 (See appendix 1 below for example result).

Activity 13: *Production of cDNA libraries*

At Embrapa Genetic resources and Biotechnologies, the following cDNA libraries were produced and validated:

a) From leaves of Pisang Klutuk Wulung (PKW) under normal growth conditions
b) From roots of PKW under normal growth conditions
c) From leaves of Pisang Klutuk Wulung (PKW) under drought stress conditions
d) From roots of PKW under drought stress conditions

The two last libraries are still under way since some problems were encountered in the final steps of the established protocol. Nevertheless this libraries should be ready by the end of June 2006.

Activity 14: *EST Sequencing and annotation*
NIAS performed 5'-terminal sequencing of the 9 216 cDNA clones from two Musa species (MACVLIMFLS and MA4LINFES) was completed. From the data, low-quality sequences and redundant sequences, were removed .resulting in 8 177 sequences with 524 bases of average read length. These sequences are analysed by BlastX (six-frames) against NCBI nr protein database. In total, 4 217 ESTs are detected with 1 544 Blast hits with less than 10^{-20} as E values.

Besides abundant sequences with the similarity to proteosome subunits, protein synthesis, and photosynthesis proteins, there are some ESTs which share significant sequence homology to leucine zipper proteins, nucleotide binding proteins, plant hypersensitive reaction-related proteins, and other stress-related or disease resistance proteins. EST sequence data obtained from NIAS have been treated with "ESTtik" and "SSR Analysis tool" pipelines at CIRAD. 143 SSR loci have been defined in the ESTs.

Activity 15: *BAC sequencing*
NIAS performed shotgun-sequencing of 7 BACs which were selected according to biotic stress related gene sequences (PR, SOD, and WRKY) by CIRAD in France and EMBRAPA using 5 EST libraries. The addresses of the seven selected BACs were transmitted to the MGRC based at IEB, Olomouc in Czech Republic which sent the BAC clones to NIAS for sequencing. Each of the BAC sequences were successfully assembled to single contig (702 912 bases in total). For the initial attempt to the Musa annotation, we have analysed these sequences with RiceGAAS which was originally designed for rice genome annotation. In total 157 gene models are predicted. These include many hypothetical or unknown proteins, but further detailed analysis will reveal the novel genes in the regional sequences around the stress-related genes, and thus DNA markers for the map-based cloning of the trait genes might be constructed. All the sequence information for the BACs is submitted to INIBAP for further application within community.

EMBRAPA also did automatic annotation of these BAC clones and an initial manual annotation is under way.

Objective 5: Establishment of Musa databases and front-ends to sequence databases; public distribution of clones; training and dissemination of information about Musa genomics and the uses of this information
Activity 16: *Sequence annotation and comparison of BACs sequences from A and B genome, and if needed comparison with rice sequences*

Activity 17: *Integration to Musa genetics portal and MGIS (Musa germplasm information system) as appropriate*
A significant effort has been made to improve information sharing between partners involved in the project. In that prospect, the Global Musa Genomic website has been overhauled to gather all the relevant information. A database on sequencing status has been developed. Data were collected and integrated into it. Currently 25 BAC clones and 3325 BAC ends clones were sequenced. A genome Browser has been set up for a user-friendly visualisation of the sequence and the annotation publicly available. A database for the storage of EST is also installed and is waiting for result of analysis. As long as the data are produced and analysed in the project, they will be integrated and consultable via the GMGC website.

Tangible outputs delivered:
- Identification of drought-responsive, cellulose synthase and biotic stress related genes in the available Syngenta clone sets
- DNA pools (plates/rows/columns) from BAC library of Calcutta 4 (MA4) suitable for PCR screening
- BAC library *Musa acuminata* Cavendish (MAC) available at the MGRC

- DNA extraction from 180 individuals of the BORLI mapping population
- 143 SSR markers identified from EST sequences
- 50 SSR markers were mapped
- 614 repetitive DNA sequences from Calcutta4 characterised for homology with known DNA sequences and genomic arrangement
- 2 cDNA libraries of PKW have been constructed (from leaves and roots)
- Correlation between results obtained by chromosome counting and flow cytometric measurements
- 8177 Musa EST sequence from 5' end form 2 cDNA libraries
- Genomic sequences for 7BACs (702 912bp) selected by biotic stress markers
- The sequence data have been incorporated into the Musa genomics web site of the Global Musa genomics Consortium

Deviations from the work plan:
Activity 11: The initial strategy of EST sequencing for the first year was reading both ends of 5,000 clones, and map paired ESTs onto rice genome sequence.
As the objective of obtaining ESTs has changed to establish SSRs as much as possible, only 5'-ESTs from 10 000 clones were analysed within the budget allocated to NIAS. As 5'-ESTs are part of protein sequences, it would not be avoided to be mis-mapped because of tandemly or distantly repeated genes. Presently NIAS is sequencing 5'-end of the same clones for paired mapping.

Activity 15: Sequencing only 7 instead of 17 BACs selected. Screening of biotic stress response BACs was over and 7 BACs were sequenced. However, as it can be found in the Activity 12, it was hard to detect discrete abiotic signals in the filter hybridisation. So the strategy was changed to pooled PCR screening. So far, screening is under way at University Leicester

2005-16: Validation of Conserved Orthologous Markers

Principal Investigators:
Merideth Bonierbale, CIP
Jizeng Jia, CAAS
Michael Baum, ICARDA
Pat Heslop-Harrison, U. Leicester
Pat Heslop-Harrison, U. Leicester (subcontractor to IPGRI/INIBAP)
Collaborating Scientists:
Lifeng Gao, CAAS
Qingming Sun, CAAS
Reinhard Simon, CIP
Leticia Portal, CIP
Nicolas Roux, INIBAP
Sripada M. Udupa, ICARDA

Mid-Year Report
COS validation in potato
'COS I' primers: A total of 62 Conserved Orthologous Set (COS) I genes from Cornell University were investigated for utility in potato, before COS II markers became available. Primers were designed at CIP for 55 COS I, and 10 representative primers and 49 probes were provided by F. Wu and S. Tanksley. Three potato populations (BCT, PD and PCC-1) were used to survey polymorphism and map the COS into pre-existing frameworks, using PCR and CAPS (from 5 enzymes at random) for the primers and RFLP for the probes. Forty-nine primers gave amplification fragments, of which 26 were polymorphic by simple PCR. Of 19 polymorphic primers with clear amplicons, 11 mapped to potato locations predicted by synteny with tomato in

at least one cross, and the remaining 8 mapped to unexpected locations. CAPs (1) and RFLP (7) permitted the mapping of 8 additional COS I, of 22 that had monomorphic PCR amplicons.

110 'COS II' were selected from the SOL Genomics Network (SGN; http://www.sgn.cornell.edu), including 30 with annotations suggesting roles in disease resistance, drought tolerance or other agronomic functions, and primers were synthesised commercially. Of these, 75 have been screened for polymorphism in the three potato populations. Forty-six were prioritised for validation by mapping and re-sequencing, and of these, 34 were polymorphic with simple PCR. Twenty-six have been mapped in at least one of the three populations, coinciding with locations predicted by synteny with tomato. Five of the remaining polymorphic markers mapped to unexpected positions, and two are still pending segregation analysis in the PPC-1 population.
54 PCR products amplified from the parents of the BCT, PD or PCC-1 population have been sequenced and analysed by BLAST.

15/18 COS I products sequenced showed high similarity to the respective target sequences with up to three regions of high identity, e.g. 122bp (100%). Three showed low similarity to the original sequences. Eleven of 15 COS I markers are considered to be validated in potato because they mapped and sequenced as expected.

31/36 COSII products sequenced with high similarity to the original/targeted sequences, with e-values up to 0 to 9e-49, and up to three regions of similarity ranging from as short as 29 bp (93%) to as extensive as 392 (99%). Sequences returned for five PCR products isolated from agarose were difficult to interpret, probably due to their size or complexity, and will be purified again. Thirteen of the 31 markers are considered to be validated in potato because they mapped and sequenced as expected. Seven additional products showed high similarity regions and mapped to expected locations but with low probability; seven markers were monomorphic and are pending segregation analysis by CAPS designed on sequence information, and four mapped markers are pending sequencing. Of seven markers for which positions in tomato are not available at SGN, five have been mapped with confidence and validated by sequencing in potato at CIP. The remaining two mapped to potato positions with low probability.

To date, 24 COS markers have been validated as orthologous by synteny and sequence homology between tomato and potato (Table 1; Appendix II), 11 belonging to COS I and 13 to COSII marker sets. Twenty additional promising COS markers are pending validation by either sequencing or mapping in potato. The markers validated in potato cover 11 chromosomes, all except chr.VIII, Chr.II is the most populated with five COS markers (T1395, T0706, T0086, At5g37260, At1g67730) and chr.I, chr.VI, chr.VII and chr.XII have three COS markers each (Figures 1 and 2, Appendix I).

COS validation in other dicots

Samples of bean (M. Blair, CIAT: DOR364 and G19833) and sweet potato (CIP: Beauregard and Fenshugai) were included in the screening with 30 COSII primers, and 30% gave clear PCR products. Moreover, samples of Fava bean (91/25, 34 Morocco), Lentil (ILL 4605, ILL 5888), Chickpea (ILC 1272, ILC 3279) from ICARDA and additional sweet potato genotypes from CIP (Wogabolige, Zapallo, SPK 004 and Xushu 18) were tested with 10 COSII primers validated in potato. Six (T0989, At5g58490, At3g26060, At2g42810, At5g62390, At1g08830) with clear bands in various dicot samples were selected for validation by sequencing. The products of these six COS were confirmed to have regions of high similarity to the original target sequences in one or more species. At5g62390 amplified homologous products in all species except chickpea (Figures 3 and 4, Appendix I). The significantly similar regions ranged from 38 bp (89%) to 371 bp (80%) with e-values from 1e-49 to 4e-4. T0989, a COS marker validated for chromosome 12 of potato whose putative function is pyruvate dehydrogenase E1 alpha subunit *Arabidopsis*

thaliana showed the most extensive similarity across all species except sweetpotato, which gave non-specific bands.

COS application to potato diversity assessment
Six COS (T1065, At1g44575, T0761, T1252, T0721, T0675) were tested in genotypes of four wild (*S. chiquidenum*, *S. cajamarquense*, *S. paucissectum* and *S. piurae*) and one cultivated (*S. phureja*) potato species previously evaluated for disease resistance phenotypes (Figure 5, Appendix I). Most markers were informative for allele diversity and a putative functional polymorphism was observed. COS T1065 (basic chitinase ortholog mapped to Chr IX potato/tomato homologues) reflected relative diversity for disease resistance among three accessions of *S. paucissectum*, i.e., the accession with the greatest variation in disease resistance also showed higher allele diversity with this marker. PCR products of *S. paucissectum*, *S. cajamarquense* and *S. chiquidenum* accessions amplified with COS T1065 were sequenced and found to have high similarities with the original target sequence, even when bands of significantly different in weight were tested. The expected values ranged from e-95 to e-152 with significant similarity regions of between 281(98%) and 259 (91%) often showing two regions of similarity.

Overall, COSII markers showed more clear product amplification and better map coincidence in number and expected positions, (85 % of 43 segregating COSII markers and 41 % of 27 segregating COS I markers were mapped) and had sequences corresponding to original genes. However similarity region for COS II were generally shorter (e.g. 37 bp of 100%), as compared to COS I markers, which frequently showed high similarity regions of over 122bp (100%).

The COS markers validated in potato enrich comparative genetic maps between tomato, potato and *Arabidopsis*, and appear useful as candidate genes for functional diversity assessment. Sequences of products from amplification of about 45 COS in potato mapping parents and progeny samples are given in Table 2; Appendix II.

Bioinformatics:
A database for COS markers at CIP is available at the following link: http://www.potgenebank.org. This database presents a summarised view of progress, while more details are available in CIP's LIMS. The full database is available to collaborators in Excel format. Online access can be enabled for collaborators
Finishing documentation on COS identification strategies is scheduled for later 2006 in the context of the ADOC project (commissioned GCP). Integration of COS information into metabolic context along with gene expression data is scheduled once the BioCyc software (manages metabolic pathway, biochemical compounds, and can overlay gene expression data) currently under evaluation at CIP is established.

a. Summary ICARDA
Fifty abiotic stress (drought, cold and ABA [abscisic acid]) induced gene sequences from microarray experiments of *Arabidopsis* were used to identify putative orthologous 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. Primers were designed at conserved regions (Table 3, Appendix II). Information on amplification of these primers with faba bean, chickpea, lentil and Lathyrus are presented in Table 4, Appendix II. Meantime, we sent DNA samples of the legume species to CIP and IITA – Nairobi (D.J. Kim) to test the primers they developed on these legumes. All eight DNA samples (two samples each from a species) from *Vicia*, Lentil, *Lathyrus* and *Cicer* were tested with 16 COS primer sets,

which have been developed for phaseoloid clade (cowpea and common bean). Only two out of 16 had a good amplification because the COS primers were specifically designed for tropical legume. These COS primers are mainly developed based on cowpea/common bean/soybean sequences. We will be designing some more primers for the galegoid clade specific COS primers, primarily based on chickpea and *Medicago* sequences, for COS amplification of these cool season legumes. On the other hand, some of the *Medicago* derived COS markers work with this panel of legumes effectively. Results obtained to date are based on PCR amplification, as re-sequencing across species was not performed

I. Monocots:

b. Summary CAAS

CAAS focused on the development and application of EST-SSR and disease resistance markers across wheat, rice, maize and barley. Primers were designed on wheat sequences from public databases and full-length cDNA libraries at CAAS, for testing in wheat, barley, rice and maize (Figure 6, Appendix II). Of 243 non-repetitive and putatively homologous EST-SSR primers, 154 (63%) primers amplified products across the four species, On PAGE, all the clear bands were scored as successful amplications, as primers were designed from ESTs, but since PCR reactions based on genomic DNAs differ in length from the original sequences they were considered tentative successes. Of the total set, 216 (89%) produced amplicons in wheat, 211 (87 %) in barley, 187 (77 %) in rice, and 166 (68 %) in maize. The percentage of primers functioning decreased with the increased phylogeny distance. 14 primers produced PCR products only in wheat. On the contrary, some primers did not work in wheat, but did work in the other three species.

About 40 EST-SSR primers appeared genotype-specific, yielding amplification in only one accession of each species. Only four primers produced PCR products of the same size among the four species and, as expected because barley and wheat shared more homologous ESTs, these two species had the highest number (12) of identical amplicons. Approximately 6 % (15/243) of primers amplified at least one PCR fragment of a larger size than expected in wheat genomic DNAs because of intron insertions.

In summary, about 82% primers produced expected products in wheat. The degree of primer conservation determined the efficiency of utilisation of transferable primers across species. *In silico* analysis showed that the variation in microsatellites of wheat, rice, maize and barley came not only from the differences in the numbers of repeat units and their location shift, but also from mutations in the regions flanking microsatellite repeats.

Diversity assessment, map development and gene function

Twenty-one primers reveal polymorphism in wheat varieties and have been used for genetic diversity analysis, and 12 primers showed polymorphism between parents of three mapping populations. Ten have been mapped to Opata 85 X W 7984 (Figure 7, Appendix I). 25 ESTs contained LRR (leucine-rich repeat) or NBS-LRR (nucleotide binding site–leucine-rich repeat) domains, 22 contain PK (protein kinase) or PK-LRR domains, and four other ESTs were related to disease resistance.

Technical comments –

PCR products were sequenced for validation, but expected results were not obtained. While short PCR products (<300bp were clearly detected by PAGE, products were often weak in agarose. Difficulties encountered in isolating bands from agarose were such that few products were obtained, and subsequent linkage and transferability was poor. PCR product sequencing is still in progress and additional segregation results are pending. The key to success to developing COS markers is to identify conserved sequences between species (at least from two) and design

primers from the conserved parts. However, in fact it is not easy to find such conserved sequences conferring abiotic stress, especially to design SSR primers.

c. Summary INIBAP

Four sets of genes were targeted in the Musa section of the project:
1) Abiotic stress-related
2) Biotic stress related
3) Quality related
4) COS genes where Musa EST sequences showed high homology to rice, *Arabidopsis* and other well-characterised sequences.

Of eight COS markers (Fulton et al., 2002) used, two amplified the expected genes from Musa. COS1263 gave a strong amplification of 600bp in all Musa accessions studied. COS1006 amplified a single fragment of 220bp in all triploid AAA accessions and gave weak products in other triploids. Another COS marker gave retro-element fragments while the others gave poor products.

Refinement of primer design, ideally based on EST comparisons, was 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, For example:
A) One primer redesigned from Musa ESTs isolated the polygalacturonidase gene. Interestingly, this was absent from the Musa EST database of 30,000 sequences. SNPs were found in the successfully amplified sequences. B) Genes for dihydrofolate reductase and chalcone synthase were targeted using existing sequences 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 (dehydration-responsive element) binding factor 1 and DREPP2 protein.

Some 70 primer pairs (Table 5, Appendix II) have been designed and ordered, tested by PCR on genomic DNA from 2 to 50 different Musa accessions, and the products cloned and sequenced. This work has been done by nine researchers/ Fellows from developing countries. A major difficultly with Musa is the lack of mapping populations and measured diversity in abiotic traits; current information largely being anecdotal or from work on materials to which access is limited. Biotic stress responsive genes are better defined from model species, and a number can be studied. Quality traits, particularly those in colour and ripening characters, show obvious variation between varieties, and have economic and potential nutritional consequences, as well as being of high interest to collaborators. The primers have been applied across various accessions of Musa, including cultivars with A and B genome in various combinations, and non-domesticated species such as *Musa textilis* and *Musa ornata*. Table 3 (Appendix II) gives the sequences of primers, which have been used within the work for isolation of sequences from genomic DNA.

Disease resistance genes

Appendix III [available from phh4@leicester.ac.uk] gives a poster describing the isolation and relationships of NBS-LRR type resistance genes. 50 sequences were analysed and showed high diversity, and a general pattern of relationships, which followed the genome phylogeny. Based on plant resistance gene homology motifs, degenerate oligonucleotides primers were designed to isolate the NBS regions between P-loop and GLPL amino acid motifs in 42 Musa cultivars and species using PCR. Amplification resulted in products of 520 and 620 bp, which were cloned and sequenced, showing heterogeneity in the genomic sequences. 75% of the sequenced clones had uninterrupted open reading frames (ORFs) and were used for deduced amino acid multiple

alignments and phylogenetic tree construction. The most conserved motifs in Musa NBS region were GMGGVGKT (P-loop), VWXCTV (RNBS-1), VLDDVW (Kinase-2), SKILVTTR (Kinase-3) ASYWKKCAF (RNBS-2) and GLPLALKT (GLPL). The most frequent sequence changes in deduced amino acid occurred between the P-loop to Kinase-2 and the Kinase-2 to GLPL motifs, reflecting their specificity in signalling activation. All sequences showed homologies to resistance protein and Resistance Gene Analogues (RGAs) from species such as *Oryza, Nicotiana, Linum, Lycopersicon/Solanum, Lactuca* and *Arabidopsis*. Multiple alignments grouped the Musa NBS region as a non-TIR resistance protein, and phylogenetic analysis classified them into four distinct clusters. Our findings suggest that the NBS region is primitive in the *Musa textilis* (T genome) and *Musa balbisiana* (B-genome). The clones obtained are useful as a marker system for BAC library screening to identify properties of the resistance genes in Musa and may be valuable for mapping and diversity analysis and for testing linkage with biotic stress resistance. These sequences have the potential to determine germplasm relationships and diversity, to provide markers, and to select particular genes showing resistance to biotic stresses.

Abiotic stress

Part of the abiotic stress work was carried out with a Generation Challenge Fellow (Luis Rodriguez-Zapater, see Fellowship report, separate funding) and a trainee. The sequences have since been reamplified and primers optimised by redesign, and sequences have been obtained. Within the quality traits, we targeted genes in the anthocyanin pathway, which effect fruit color and some nutritional characteristics. Patterns of evolution among these genes in divergent plant species have shown that the structural genes are more consistent and the adaptive evolution of color may be mediated more by regulatory genes. In our first study, the sequences were amplified using newly designed primers to chalcone synthase. The neighbour joining tree showed that the sequences from the Green, Dwarf Cavendish, and Pisang Lilin comes under a single cluster while the sequence from Red banana group fell into a separate group.

Quality and morphoogy

The anthocyanidin reductase (ANR) primers amplified at 288 or 284 base pair fragment from the genomic DNA of all the 16 accessions tested from Assam, India (submitted to Genbank EMBL) starting in exon 2 of the gene, spanning an intron of 83 or 87 bases, and ending in the next exon. Comparison with EMBL database sequences showed significant homology at both nucleotide and predicted protein sequences to putative anthocyanidine reductase from *Vitis vinifera*, legumes and *Oryza sativa*. We did not find homology to any *Arabidopsis thaliana* gene: this enzyme seems to be absent from *Arabidopsis*. The primers were both designed within coding regions, and spanned an intron (from position 136/132 to 219), which showed either a 7 base insertion at position 156 or a 3 base insertion at position 212. The number of synonymous sequences is high, and overall, the primers are a useful polymorphic intron-spanning marker. The CHS chalcone synthase gene fragment was 500 bases long, and included 20 bases of the last gene exon, a TGA stop codon and untranslated 3' sequence. The translated region showed high homology to sequences from rice, Arabidopsis and other species The Myc primers amplified a 534 bp fragment of Myc-like anthocyanin regulatory protein gene from genomic DNA. The fragment included a short intron of 87bp and parts of two exons. All three of the quality related genes are apparently real – no stop codons or frameshifts were evident. The sequences for the three genes affecting the anthocyanin biosynthetic pathway showed high levels of polymorphism between the accessions analysed.

Conclusions:

The sequencing of orthologous genes is very important to understand the variation present between different species.

This work has set the scene for a larger scale and more systematic study of genes in comparisons between rice, Musa and other model species. Primers working from Musa (and some others) are

now being tested with DNA from other Challenge programme members (rice, Takuji Sasaki; cassava, Martin Fregene).

However, it is important to state 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 across all species using straightforward methods. While genes are conserved between species, and this can be detected by bioinformatic approaches at the DNA level (>60% homology over more than 100 bp would be a typical threshold) or protein level (>30% similarity over 20 amino acids), the identification of conserved regions of the genes for primer design, and then amplification of specific regions of the DNA proved to be much more difficult than expected.

Previous studies in Musa have taken advantage of information about the individual genes being targeted, making primer design more straightforward. However, the design of primers around leucine (with six possible codons and importance in many resistance genes) is hardly possible because the conserved sites seem to be inappropriate separations for genomic PCR. Conservation was often low between primers and some publications based on work with model organism are perhaps overly-optimistic regarding gene conservation.

Tangible outputs delivered:

A. Markers and other kinds of information

- 100 COS ll primers (SGN) synthesised
- 46 COS ll (SGN) amplified, surveyed for polymorphism and/or mapped in potato
- 24 COS markers validated by mapping and sequencing in potato (stage 4)
- Comparative maps of tomato and potato have been enriched with gene-based markers (COS)
- 46 EST-SSR primers synthesised
- 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
- 33 sequences of Musa COS markers have been submitted to Genbank (AM259287 to AM259319) and an additional 87 sequences are being checked for read quality before submission.
- >70 potato sequences are being checked for read quality before submission to Genbank.
- Database of COS markers and DNA sequences from GCP crops

B. Training and Communications

- Two posters were presented at the GCP Annual Research Meeting in Rome: *"Conserved Genes and Universal Primers in Bananas and Plantain"* and *"Toward Diversity Assessment with Orthologous Gene Markers"*. M. Bonierbale, R. Simon, L. Portal, M. Baum, S.M. Udupa J. Jia , L. Gao , N. Roux, S. Kumar, and P. Heslop-Harrison.
- An oral presentation was delivered at the GCP Annual Research Meeting in Rome; the Powerpoint file was left with the programme management-
- *Personnel funding and consumables:* Dhairyasheel Desai, MSc (Pune, India); Ojay Benedict Oka, MSc (Nigeria); Suneetha Reddy Eluru, diploma (India); Suresh Kumar, MSc (India); Emmanuel Otwe, PhD (Ghana); Azhar Mohamad PhD (Malaysia)
- *Consumables only from GCP:* Ashalatha Nair (University Institute staff member) (Commonwealth Fellowship), Kerala, India; Pratap Handique (University Institute staff member) (Department of Biotechnology, Government of India) Guwahati, Assam, India; Zul Mohamed (University Institute staff member) (British Council) Malaysia; Emmanuel Otwe (Ghana) and Azhar Mohamad (Malaysia) (both consumables with personal support from home countries, and project-specific personal funding).

- *Invovled in aspects followed up in this SP project:* Generation Challenge Programme
  Fellowship-holder: Luis Rodriguez-Zapata (Mexico)

Deviations from the work plan:
Coordinated exchange and testing of most successful primers developed in each lab was not accomplished as planned.

2005-17: Comparative QTL Mapping for Drought Tolerance

Principal Investigator:
Mathias Lorieux, IRD/CIAT
Co-Principal Investigators:
Matthew Blair, CIAT
Stephen Beebe, CIAT
Idupulapati Rao, CIAT
Manabu Ishitani, CIAT
Nour Ahmadi, Agropolis-Cirad
B. Courtois, Agropolis-Cirad
J.-F. Rami, Agropolis-Cirad
A. Ghesquière, Agropolis-IRD

Mid-Year Report

CIAT

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

1) Genotyping – DREB marker development: Further analysis of the DREB transcription factor genes cloned from common bean last year by M. Ishitani's group found that the homologues we were working on last year were of the DREB1 subfamily that are induced by cold temperature rather than drought so this year we worked on the drought-inducible DREB2 subfamily. The status of the first DREB genes (ie. the DREB1 subfamily) was confirmed based on gene sequence analysis and expression studies and the genes were renamed appropriately (PvDREB1A, originally DREB2 pv; and PvDREB1B, originally DREB3 pv). This year we continued to look at additional DREB homologues that would be expressed under drought condition (ie. DREB2 homologues). To isolate these, data-mining was conducted with the CIAT and UNAM (Mexico) collections of ESTs (Ramirez et al., 2005) using the sequence information of DREB genes from *Brassica napus*, *Capsicum annum*, *Gosypium hirsitum*, *Helianthus annus*, *Oryza sativa*, *Glycine max*, *Arabidopsis thaliana* and *Lycopersicon esculentum*. Three additional putative DREB genes were identified by the BLAST searches (and were named PvDREB2A, originally EST clone CV535836; PvDREB2B, originally EST clone BQ481823; and PvDREB2C, originally EST clone CV529541), respectively, based on sequence similarity with DREB2 genes. BLASTp analysis showed that the five putative DREB genes from bean have high sequence similarity to the following DREB genes: CBF-like protein from *Glycine max*, dehydration responsive element binding protein 1 like protein from *Prunus avium*, DREB2 from *Glycine max* (same GmDREBb), DREB3 from *Glycine max*. Based on the sequences of the five PDREB genes we have designed specific markers for genetic mapping of this gene family. Alleles for 5 parental genotypes were cloned first for the cold-induced DREB1A and B clones. Sequence analysis has revealed a total of 9 SNPs and one indel between Andean and Mesoamerican gene pools for DREB1A. These polymorphisms allowed us to map the DREB1A locus to linkage group b04, the first DREB gene to be mapped in common bean. Meanwhile, a single SNP was identified between a drought tolerant genotype (BAT477) and a drought susceptible genotype (DOR364) and a dCAPS assay was designed for mapping of this gene but unfortunately it has not been possible to detect the polymorphism in the BAT477 x DOR364 population. Similarly, an analysis of DREB1B reveals that is highly conserved and that will make it difficult to map this gene. Meanwhile we are hoping

to detect sequence level polymorphisms for the DREB2A and DREB2C genes based on two molecular markers made for the sequences described above. We have also re-sequenced at both 5' and 3' ends the plasmid clones from CIAT that were identified as DREB-homologous ESTs to extend the sequence information for the genes and to assist in marker development and eventual cloning of a full-length gene with native promoter sequences. We are also attempting to determine the genomic arrangement of the DREB genes and the number of independent chromosomal regions containing these homologues.

2) Phenotyping: A second field trial was conducted at Palmira, Colombia (June to September). The trial included 121 RILs of MD 23-24 x SEA 5 along with 5 checks (Cowpea, Tio Canela 75, DOR 390, EAP 9510-77, SEA 15) and 2 parents (MD 23-24, SEA 5) to determine genotypic differences in tolerance to water stress conditions and to generate phenotypic data for eventual gene tagging. An 11 x 11 partially balanced lattice design with 3 replicates was used. Two levels of water supply (irrigated and rainfed) were applied. Details on planting and management of the trial were similar to those reported before. Experimental units consisted of 2 rows, 3.72 m long by 0.6 m wide. A number of plant attributes were measured at mid-pod filling in order to determine genotypic variation in drought resistance. These plant traits included leaf chlorophyll content (SPAD), photosynthetic efficiency (Fv/Fm); leaf area index; canopy dry weight per plant; shoot nutrient (P) uptake; shoot and seed ash content; and shoot and seed TNC (total nonstructural carbohydrates). At the time of harvest, grain yield and yield components (number of pods per plant, number of seeds per pod, 100 seed weight, pod wall biomass proportion relative to pod biomass) were determined. Seed P, ash content and TNC (total nonstructural carbohydrates) were measured. Pod harvest index (dry wt of pods/dry wt of shoot biomass at mid-pod fill x 100) and grain filling index (100 seed weight of rainfed/100 seed weight of irrigated x 100) were also measured. Greenhouse studies with parents were continued to compare them to drought tolerant advanced lines from the breeding programme (BAT 477, BAT 881, BFB 139, DOR 364, EAP 9510-77, G 21212, MCD 2004, RAB 651, RAB 655, SEA 5, SEA 15, SEC 16, SEN 34, SER 7, SER 8 y SER 16). Plants were grown in small plastic cylinders (50 cm long and 7.5 cm diameter) inserted into PVC tubes. The trial was planted as a randomised block in split-plot arrangement with two levels of water supply: 80% field capacity (well-watered) and 40% field capacity (drought stress) as main plots and genotypes as sub-plots. To maintain 40% field capacity, water was added on top of the cylinder 2 times a week by monitoring the amount of water loss due to evapotranspiration. Treatments of water stress were imposed after ten days of initial growth of plants that were established with seed. Water stress was maintained by weighing each cylinder every week and applying water to the soil at the top of the cylinder. After 5 weeks of growth (at 38 days after germination and 28 days after drought treatment), leaf area, shoot biomass distribution, root biomass, root diameter and root length distribution in different soil depths were determined.

Tangible outputs delivered:

1. Phenotyping protocol for field
2. Clone sequences
3. Publication on ESTs: Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs: Building a foundation for functional genomics. Plant Physiology 137: 1211-1227.

Deviations from the work plan:

No deviation from plan.

CIRAD-AGROPOLIS

This project relay on a new mapping population and an integrative map produced recently by Agropolis (projects funded by EU and French Government) for one of the most popular mapping resource used by the rice scientific community for QTL mapping, the IR64 x Azucena cross. The

new mapping population is composed of 178 RI lines genotyped with 226 SSR markers covering the genome with an average distance of 10 cM. To date, some 1700 additional markers from Previous IR64 X Azucena map and from other maps have been *in silico* mapped or interpolated in their respective intervals spanned by the framework SSR markers. The forthcoming integration of markers from the other popular rice map, Nipponbare x Kasalath will pave the way for, projection of QTLs identified with this mapping population.

Activities related to output 3.1: Synthetic map produced with positions of QTLs and candidate genes for root development and for other drought related characters from rice and sorghum.

- Integration of specifics tools into Rice-BRCdb (http://rice-brcdb.cines.fr) in order to display the integrative map: CMap to visualise the integrative map and to compare it with other maps. Specific link with Gbrowse (specially adapted to rice genome by CIRAD http://orygenesdb.cirad.fr/) to facilitate the connection between genetic studies to genomic data and annotations and vice versa.
- Inventory of the existing QTL results in the literature on rice roots and other drought-related traits. Fifty five articles have been reviewed and more that 1400 QTLs for some 90 traits involved in drought tolerance have been identified. The afferent information (QTL position, confidence interval, …) were synthesised in a database compatible with the CMap visualisation system. The next step are -1-, the projection of these QTLs on the reference map and -2- Implementing an meta-analysis to identify QTLs that are important and common across genetic backgrounds, and QTLs that are specific of a genetic background or environment.
- A single stay green QTL consensus map gathering the information of several QTL studies available in sorghum was constructed. The map data and QTL positions were uploaded into a database. After marker and locus name curation, the QTL positions were projected on the Klein2004 map (Menz et al., 2002; http://www.gramene.org) which shares the maximum number of common markers with the individual QTL maps and is well anchored to rice at least for chromosome 1. The chromosome nomenclature used for the consensus map is the one defined by Kim et al. (2005). QTL projections were performed with the MapDB software. The resulting consensus QTL map was projected on the rice pseudo molecules (TIGR V3) to identify the genomic region corresponding to each stay-green QTL.

Activities related to output 3.2: BC4F3 lines produced with narrow chromosomal segments carrying QTL identified and candidate genes in the sub-segments identified in silico

- Fifty four BC3F4 lines of the IR64 x Azucena have been fine-genotyping at 4 QTL target area on Chromosomes 1, 2, 7 and 9.
- Fifty four BC3F4 lines have been phenotyped for root development under hydroponic conditions. The QTL involved in deep root of 5 NIL have been selected, for BC4 lines development: one NIL for each of target QTLs on Chromosomes 1 and 2, and 3 NILs for Chromosome 9.
- Seventy two CSSL from Caiapo x IRGC 103544 (O. sativa x O. glaberrima) cross have been phenotyped for root development under hydroponic conditions. CSSLs from Nipponbare x Kasalath kindly provided by the Dr Yano, RGP, NIAS, Tsukuba need to be seed increased, before phenotyping works.
- Backcrosses between the selected BC3F4 lines and the recurrent parent IR64 have been realised in order to produce large recombinant populations for the 3 target QTLs (chromosomes 1, 2, and 9)
- Annotation data for the genomic area carrying the QTL of chromosomes 9 have been assembled for in silico determination of candidate genes.

Tangible outputs delivered:

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

- The position of the sorghum stay-green QTLs on a single sorghum consensus map.
- The position of the sorghum stay-green QTLs on the rice pseudo-molecules.
- Effect of QTL 9 for deep root development confirmed in BC3F4 NILs
- BC4F1 materiel produced.

Deviations from the work plan:
CSSL lines from Nipponbare x Kasalath cross kindly provided by the Dr Yano (RGP, NIAS, Tsukuba) need to be seed increased before phenotyping works.

2005-35: 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
D. Star, Perlegen Sciences, Inc
Jan Leach, Colorado State University
C. Robin Buell, TIGR

Mid-Year Report

1. Project design

This project provides partial support (15% of current funding) for undertaking genome-wide SNP discovery by re-sequencing multiple, diverse rice varieties through DNA-DNA hybridisation on high density oligonucleotide arrays and is a partnership between IRRI, Perlegen Sciences, Colorado State University, and TIGR with additional funding from IRRI and the USDA-CREES.

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 hybridised to the array.

The design of the project involves 2 phases:
First phase (supported by current funding) will generate data on 21 diverse rice varieties for 100 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 43% of the genome will be covered in this phase.

Funding for a second phase to complete the genomic coverage and extend the set of rice varieties from 25-30 is being sought through partnerships under the umbrella of the International Rice Functional Genomics Consortium. 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 accomplished by interrogating around 3,900 features.

2. Status

- IRRI has assembled a collection of 25 varieties/germplasm for seed increase and purification. A main criterion for choosing varieties of germplasm was their comparative diversity while demonstrable utility for plant breeding was an added benefit. From 5-8 individual plants were genotyped using a panel of 49 SSRs. Phenetic analysis of the SSR fingerprints by DarWin4 resulted in the dendrogram presented in the attached file. All plants from a variety grouped together, indicating that the varieties are distinct. For Rayada, Aswina, and Co 39, traditional types, higher within variety diversity was noted. We have selected a single representative plant from each variety. Seed from this plant has been sown for DNA production and for further seed increase.
- Nominated varieties from Taiwain (Tainung 67) and China (Minghui 63 and Zhenshan 97B) have been included in the set.
- Release 4 of the high-quality BAC-by-BAC japonica sequence of Nipponbare (International Rice Genome Sequencing Project) was obtained on September 1, 2005. This sequence has been masked for repetitive DNA by a pipeline developed at McGill University by Thomas Bureau and Doug Hoen and is similar to that used for the IRGSP Nature publication. The whole genome shotgun indica sequence of 93-11 (Beijing Genomics Institute) has also been masked by this pipeline. Both sequences have been aligned to one another using the mummer tool (TIGR). 91.5 Mb of unique sequence were identified for the chip design. The remaining 8.5 Mb will consist of filtered, unanchored BACs, sequences from Kasalatha and FR13 A, mintochondrial and chloroplast genome, and sequence from the low repetitive fraction (present in 2-10 copies) where unique stretches are sparse.
- Cs-banded DNA has been prepared for the additional 17 varieties now in Phase 1: Aswina, Rayada, FR13A, N22, Swaran, Sadu-cho, Pokklai, Minghui 63, Zhenshan 97B, Tainung 67, Azucena, Moroberekan, Gerdeh, Dom sufid, Cypress, M202, and Nipponbare.

Tangible outputs delivered:

- Nipponbare and 93-11 genomic sequences have been masked for repetitive sequences by the pipeline developed at McGill and implemented at IRRI. The Nipponbare and 93-11 sequences have been aligned to one another.
- DNAs from the 17 additional varieties have been prepared and will be shipped to Perlegen within the coming week.
- A publication on the project appeared in the May 1, 2006 issue of Plant Physiology (141:26-31):

McNally KL, Bruskiewich R, Mackill D, Leach J, Buell CR, Leung H. "Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes.

Deviations from the work plan:

Growth of plant materials to obtain sufficient material (over 100 g fresh weight) for DNA preparations required longer than expected. Masking of the genome and defining the regions to be tiled also took longer than predicted.

SP3 COMMISSIONED GRANTS

2005-18: Development of Low-Cost Gene-Based Trait Assay Technologies in Cereals

Principal Investigator:
Casiana M. Vera Cruz, IRRI
Co-Principal Investigator:
Manilal Williams, CIMMYT

Collaborating Scientists and Institutions:
Jianli Wu, China National Rice Research Institute
Eduardo Redoña, Philippine Rice Research Institute
Masdiar Bustamam, Indonesian Agricultural Biotechnology and Genetic Resources Research Institute
Usha Barwale-Zehr, Mahyco Research Foundation, India
Valerie Verdier, IRD-Agropolis
E. Raman Babu, VPKAS, India
Chughtaisajjad Ur Rehman, National Agricultural Research Centre, Pakistan
K.M. Karunaratne, Field Crops Research & Development Institute, Sri Lanka
Bhuiyan Safiful Alam, BARI, Bangladesh
Dil Prasad Sherchan, NARC, Nepal
SME breeding companies
Instituto de Ciencia y technologia Agricolas (ICTA), Guatemala
Instituto Nacional de Investigación Agropecuaria (INIA), Venezuela
Kenya Agriculture Research Institute (KARI), Kenya

Mid-Year Report

A. Recipient germplasm materials from NARES collaborators

Maize. 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 as source material for sequence comparison of the alleles at the opaque2 locus. The non-QPM sources have been selected based on adaptability and superior performance in respective geographical zones. Currently, there are no available sources of QPM for tropical highlands. These donor sources of QPM also contain excellent kernel modifying capacity and have been extensively used in line conversion activities at CIMMYT in the past.

The donor and recipient genotypes, selected based on their widespread use, are as follows. QPM donors: white tropical lowland CML159, CLQ-RCW Q01, CLQ RCW Q50, CLQ 4203 (251-57); yellow tropical lowland CML161, CML165; white sub-tropical CML176. Recipient lines: white tropical lowland CML343, RCW01 (309, 310), CL RCW22; yellow tropical lowland CML348 (G26 SEQC3); white highland CML 244, CML349; Ethiopian white F-7287, A-7018.

Rice. All DNA and seeds of recipients (in the case of India) have been collected and maintained at IRRI. These DNA are used in all technologies as test materials for developing the methods. In the NARES breeding programme, these recipients are being used in the crossing work to introgress their target Xa-genes.

In China, for example, Dr. Jianli Wu is developing three breeding populations of rice for this project. One of them is from Hui593 where he introgressed xa5, Xa7 and Xa21 genes. The size of the population developed is around 1,800 lines. They are currently growing this population (F_2) this summer in Hangzhou and will be advanced by single seed descent until the technologies being developed can be applied for MAS.

B. Project team linkages and training activities

Shuttle researcher and Visiting Collaborator:

1. Ms. Jie Chen who arrived at IRRI in September 2005 has made excellent progress in developing the FRET-based detection assay and other techniques. She left in May 2006 to rejoin her group at CNRRI, China.
2. Dr. Debra Skinner, a consultant from CIMMYT with experience in high throughput genotyping techniques joined IRRI on 27 April 2006 and stayed until 16 May 2006 to assess and become familiar with high throughput genotyping technologies under development at

IRRI and has been exploring the feasibility for technology transfer to CIMMYT with the goal of adopting a system for maize systems. This visit allowed close interaction between IRRI and CIMMYT researchers involved in the project.

C. Sequence comparisons for Opaque2 alleles

Initially, attempts were made to identify useful polymorphisms between the chosen QPM donors and recipients using the microsatellites located within the Opaque2 sequence region. The co-dominant microsatellite markers phi057 and umc1066 were used to amplify these regions from select donor and recipient lines. For umc1066, 8 recurrent genotypes and 1 donor genotype were sequenced. The umc1066 marker gave the expected 140bp amplification product which aligned with the Opaque2 sequence in the database. However, sequence comparisons among the donor and recurrent parents demonstrated that there were no conserved polymorphisms between the donors and recipients. Due to incorrect priming and cloning difficulties with the SSR marker phi057, only a couple of recurrent genotypes were used for sequence comparisons.

D. Primer design for Opaque2

Due to the lack of useful sequence polymorphisms based on sequencing the microsatellite amplification products, the reference sequence of Opaque 2 in Genbank and maizegdb was used to design primers to amplify larger regions for sequence comparisons. The positions of introns and exons was determined by aligning the genomic and cDNA sequences, and all Opaque 2 sequences in databases were used to identify regions, in exons, with high degree of conservation. Five sets of primers were designed in these regions, to increase the likelihood of amplification across diverse lines. Primer sets #1 and #3 have been used to amplify and clone a total of 1 kb of the Opaque 2 locus from the 16 genotypes listed above. Amplification products from primer set #5 have been cloned and will be sequenced if necessary, and the remaining two primer sets could also be used to search for more polymorphisms.

Sequence comparison revealed that all QPM donor sources had the same sequence, indicating a possible common source for the opaque 2 mutation in CIMMYT donor lines. Several potentially useful SNPs and indels were found that were polymorphic between the donors and some or all of the recipients. Following observations were made from the results of sequence alignment comparisons.

- All donor sequences are essentially identical.
- There are several polymorphisms among the tropical and sub-tropical white recipients, including 2-3bp indels which could be useful, but assays would have to be tailored directly to the populations being tested.
- Ethiopian White recipients (F-7287, A-7018) are quite diverged from other recipient lines, with several conserved SNPs.
- Three polymorphisms have been selected for further study; they are:
    - SNP1 (315) C/T
        - perfectly cosegregates with o2 phenotype
        - C found in 7/7 donors and T in 9/9 recipients
        - Tm shift primers designed and being optimised
        - T allele primer works well, but C allele primer had to be redesigned.
    - SNP2 (612/613) AT/TC/AC
        - AT cosegregates with 7/7 donors and 1/9 recipients.
        - Tm shift primers designed and being optimised for maximal allele discrimination.
        - The primers work quite well, and result in amplification products with 5^{o}C melting temperature differences which can easily be identified with the real time melting curve analysis.
    - SNP/INDEL 3
- 7 base insertion present in 6 of 8 recipients.

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

1. For detection of Xa21 polymorphisms in rice

We continued the development of dot-blot technology in rice using Xa21 as probe. Ninety six DNA samples [composed of replicates of 26 near-isogenic lines with single bacterial blight resistance (Xa) genes and Xa-gene pyramids, 42 popular and breeding varieties, and two no template controls (NTC) as negative controls] were tested using SNP-based probes specific for the resistant allele (SNPDB_513-530_Xa21R) and susceptible allele (SNPDB_513-530_Xa21S) of the Xa21 resistance gene. High and low stringency wash concentrations were 0.5X SSC + 0.1% SDS at 60°C and 2X SSC + 0.1% SDS at room temperature, respectively

Using Xa2118-mer probe, there was distinct discrimination between R and S alleles. Spots on the S blot had a more distinct/solid color development compared to the R blot. Repetition of the dot blot method produced consistent results indicating that marker genotyping was reproducible. Color development was not detected for some samples using either R or S probes presumably due to problems occurring during PCR, spotting or blotting. Stringency washes were homemade, thus, reducing the cost in detection procedure, while the washing and blocking buffers and other detection reagents were obtained from Roche Diagnostics.

Current work involves developing a high throughput method for dot-blot, from 96 to 384 spots using a replicator. This also allows cost reduction by making the technology high throughput. We are also conducting a cost analysis for this method. Since the method has been found reproducible, we are also currently validating the technology on a Basmati-derived RIL population (F_9) to determine the robustness of the methodology in detecting the Xa21 allele in the RILs.

2. For detection of Opaque2 polymorphisms in maize

A two-base SNP (SNP2) was used to evaluate the utility of the dot blot technique for genotyping donor and recipient lines for the o2 mutation. A protocol was developed and several parameters based on the existing methods for rice were tested to determine the optimum assay conditions for o2 polymorphism detection.

- Commercially labeled 20-mer DIG-labeled probes were used. 18-mer probes have also been used with good results for Xa21 (single base SNP). Increase in probe length gives increased specificity, but this is unnecessary with PCR products. Increase in probe length leads to smaller delta-Tm between probe-target match and mismatch hybrids, requiring more exact hybridisation conditions to distinguish genotypes. Thus, 18-20mer probes work well.
- Probe was used at concentration of 10 nM, and lower concentrations were not tested, therefore, probe concentration 10 nM was able to detect the o2 allele.
- Two lengths of PCR target were evaluated for use as the DNA to be blotted and probed, 160 bp and 530 bp (1.2 kb for xa5). Once the optimum PCR dilution was established, (1:5 for the 160 bp value, and 1:5-1:10) both probe lengths worked well. Thus, target length is flexible up to 1.2 kb.
- The dilution of PCR products (determined empirically with water) and the amount spotted were varied. PCR products spotted without dilution did not spread evenly and made larger spots. PCR products diluted 1:5 or 1:10 spotted evenly, made smaller spots and hybridised more efficiently. The amount spotted varied from 0.5 µl to 2 µl. The 0.5 µl of a 1:5 dilution was sufficient to see good signal, therefore use of a replicator may be possible. Therefore, PCR products should be diluted, and can go as low as 0.5 µl dots.
- Hybridisations were carried out in DIG EasyHyb (Roche) and two temperatures were tested, 42°C and 45°C. Both temperatures worked well, although the second hybridisation at 42°C did show light background cross-hybridisation, therefore 45°C may be better. Thus, hybridisation temperature must be determined empirically, 45°C for these 20-mer two-base SNP probes.
- Length of hybridisation was not varied – overnight (12 hours) was used.

- Antibody concentration was not varied – used 1:5000 dilution from stock.
- Concentration of the substrate was varied between 1:50 and 1:75 dilution, with the lower concentration working as effectively as the higher one. Thus, substrate concentration should be reduced as much as possible without compromising signal to save on reagents.
- Time of color development was determined empirically, by monitoring the reaction every 5 min over time. The color development took 25-30 min, and longer development resulted in slight background. The use of control genotypes on the blot would allow evaluation of the results as the development proceeds, to stop the reaction before background accumulates, usually 25-30 min.

An attempt was made to hybridise with genomic DNA. DNA was treated by denaturation at 95^{o}C in 0.5M NaOH, and 2 µl of 250 ng/ul was spotted 4 times over to accumulate ~2 µg of genomic DNA. There was no hybridisation detected. The binding capacity of the Roche nylon membrane is not described but maximum signal is likely obtained with 1 µg/ µl concentrations, due to the surface chemistry of the membrane. In this case, it may not be possible to bind enough maize genomic DNA to give detectable hybridisation, as the genome size is large (2500 Mb). However, it is worth trying to optimise conditions further to determine whether detection is possible. The use of rice genomic DNA may be more feasible as the genome size is less than 1/5 in size.

F. Progress on the development of the PCR ELISA-based detection assay

New PCR primers were tested for xa5-allele specificity using 96 priority lines. PCR reaction conditions (especially annealing temperature) were optimised until PCR primers discriminated between xa5 resistant and susceptible lines. In preparation for the ELISA step, PCR products were labeled with digoxigenin-11-dUTP during PCR (dig-11-dUTP was obtained from Roche Diagnostics).

PCR-ELISA (using a Roche Diagnostics kit) was performed using rice varieties IR24 and IRBB55 as susceptible samples, IRBB5, IRBB54 and IRBB56 as resistant samples, and no template control (NTC) PCR product as negative control. Phosphate-buffered saline (PBS), used to dilute the PCR products, was also used as a negative control. Preliminary results indicate that there was a good correlation between the ELISA absorbance readings and gel electrophoresis of PCR products. Threshold value could not be computed, however, because of very high absorbance reading value from the NTC sample; no or very low absorbance values were obtained from the PBS (negative) sample. NTC sample contained all the PCR cocktail (including DIG-dUTP) except the DNA template. It was hypothesised that free and/or excess DIG-dUTP in the wells might not have been thoroughly washed away and might be reacting with the anti-DIG antibody.

To determine the optimum concentration of DIG-label to reduce non-specific binding, DIG-labelling was done during PCR using two kits containing different ratios (1:3 and 1:20) of dNTPs:DIG-dUTP. Three setups per primer pair (xa5 F2-Res + xa5 R3-Bio and xa5 F2-Sus + xa5 R3-Bio) were done: no DIG, 1:20 DIG and 1:3 DIG. PCR products (2uL) were resolved in 1% agarose gel, stained with SYBR safe and viewed under UV light. The remaining PCR products were diluted 1:100 using PBS (phosphate buffered saline) prior to ELISA. IRBB5 and IR24 DNAs were used as template for the xa5 allele-specific PCR. A no-template-control (NTC) was included as negative control.

Detection was done using a Roche PCR-ELISA (DIG Detection) Kit (11 636 111 910). Background readings were still present in the “no-DIG” set-up which might suggest that the anti-DIG-POD conjugate binds to the wells at some extent. Increasing the number of washes to six or seven might decrease color development in these wells. The ratio of DIG concentrations (1:3 vs. 1:20) gave no significant difference in their optical density values. Lower DIG-dUTP concentrations of 1:50 will be used in future experiments.

G. Progress on the development of the FRET detection assay

Control test using synthetic oligonucleotides. In order to first establish the FRET detection assay, a control test was performed using synthetic oligonucleotides (using R and S Xa21 alleles separately and then as a mixed template) as single-stranded DNA template for the single base extension (SBE). This control was used to ensure that SBE and FRET could successfully occur using single-stranded template, in the absence of contamination caused by the dNTPs and primers which were carried over from PCR. SBE and FRET were successful and because a higher fluorescence compared to the background was observed at 610 nm and 670 nm when Rox-ddATP and Cy5-ddGTP were used, respectively. As expected, an intermediate fluorescence was detected for the mixture of templates and dyes.

Generation of single stranded DNA (ssDNA) template. One of the most important steps in FRET is to generate ssDNA template for SBE. The PCR products were treated with different concentrations (0.25 to 10 U) of Lambda Exonuclease to degrade double-stranded DNA (dsDNA) and generate ssDNA, Results indicated that 10 U of Lambda Exonuclease was necessary to generate ssDNA.

Validation of FRET assay using Xa21 SNP. Oligonucleotide primers for PCR and SBE based on the sequence of the Xa21 resistant lines (Hui333, Hui593 and Hui811) and Xa21 susceptible lines (IR24 and Hui161) were synthesised. PCR products were purified and SBE was performed according to the previous protocol (Takatsu, et al, 2004)[1] with modifications. Gradient test were made to modify annealing temperature (Tm) so that single copy PCR products were generated. The optimum Tm was 66 ^{o}C. Fluorescence was detected using the recently-acquired FLUOstar OPTIMA microplate fluorometer (BMG LabTechnologies Ltd.). The results indicated that FRET occurred and while there were moderate levels of background fluorescence, resistant and susceptible lines could be distinguished. This experiment will be repeated with improved PCR clean up procedures aimed at reducing background.

H. Progress on the development of the modified TAM ('Microarray-based genotyping') detection assay

A standard set of 384 DNA samples of rice accessions was prepared for validation using xa5. This set included numerous control accessions with and without xa5 genes, popular rice varieties and a segregating recombinant inbred population for four bacterial blight genes and quality rice. Control genotypes such as near-isogenic line for xa5 were replicated and also prepared using crude DNA preparations which are commonly-used in MAS. Samples were used for PCR using gene-specific xa5 primers. PCR conditions were optimised such that a single 300 bp product was produced. After PCR, products were purified by using a sodium acetate/ethanol precipitation method. Prior to array printing, samples were mixed with an equal volume of DMSO which is a recommended printing buffer that also denatures DNA. Quality control checks of printed slides indicated successful printing of PCR products.

Detector probes (12-mer) that were specific to the R and S xa5 alleles were mixed with hybridisation buffer. Hybridisation was performed using a GeneTAC HybStation (Genomic Solutions) which permits automated hybridisation using commercial buffers. Two hybridisation temperatures were used: 35° and 40° C. Scanning analysis indicated that hybridisation did not occur when 40° C was used; a partial hybridisation occurred when using a hybridisation temperature of 35° C, however the signal was extremely weak and effects of uneven hybridisation were observed.

Subsequent hybridisation was performed using low stringency conditions (35° C) using a manual hybridisation chamber. Low and medium stringency post-hybridisation washes were prepared using freshly-prepared wash buffers. Hybridisation using only the R xa5 detector probe occurred although differences between R and S genotypes could not be discriminated. The manual

chamber permitted more uniform hybridisation and used less probe compared to using the automated system, which will reduce future costs for detector probes.

The recent results with 37° C manual hybridisation for 14 hrs with 12-mer R probe showed good discrimination between the R and S complementary oligos spotted on the slide; no hybridisation was observed on the S oligos. However, cross hybridisation was observed with the PCR products from S genotypes, probably due to length of the probe in relation to the template DNA and DNA purity.

We therefore used 26-mer probes in two concentrations in the succeeding experiments. From the 384-well plate previously described, the microarray slides were printed using a GeneTac G3 Arrayer. Synthetic oligo nucleotides (single-stranded DNA) that are perfect complementary matches for the xa5 detector probes were spotted on slides as positive controls for hybridisation.Commercially-labeled 26-mer detector probes (AlexaFluor 647) was initially prepared by resuspending it in filtered-sterilised nanopure water. Prior to hybridisation, the probe was mixed with the hybridisation buffer (Unihyb, 2x concentration). Using a GeneMachines manual hybridisation chamber, 35uL of the probe (50uM and 100uM) was hybridised at a temperature of 58°C (submerged in a water bath) for 14 hours. The microarray slides were scanned using a Perkin-Elmer ScanArray GX.

Slight cross-hybridisation was observed but was less as compared to previous experiments using a 12-mer detector probe (hybridisation temperature at 37°C). No detectable difference was observed in hybridisation strength when using 50uM and 100uM probe concentrations. 50uM probe concentration will be used in subsequent experiments. Hybridisation experiments using probes of longer length (50-mer) will also be done.

Based on Shirasawa et al. (2006)[2], cross-hybridisation can be reduced by including in the hybridisation solution, aside from the exisiting labeled oligonucleotide (complementary to the resistant allele), an unlabeled oligonucleotide (complementary to the susceptible allele). This unlabeled oligonucleotide will specifically bind to the S allele, lowering the incidence of hybridisation between the labeled oligonulceotide and R allele.

I. Survey

A survey questionnaire has been developed which is being used to assess the current status of MAS in plant breeding and identify high priority traits for the development of new DNA markers for diverse crop species. The survey has also been designed to assess current obstacles for MAS in plant breeding by requesting respondents to evaluate the importance of possible constraints and also to identify the most important factors successful MAS application. A web-based version can be accessed at http://irriwww/irrihome/survey/marker/questions.asp.

Tangible outputs delivered:

1. Successful application of the dot-blot technique for screening rice germplasm for bacterial blight using Xa21 SNP probe.
2. Protocol evaluated for dot-blot technique for screening maize germplasm for QPM using o2 SNP probe.
3. Preliminary establishment of methods for validation of PCR-ELISA, FRET and modified-TAM techniques.
4. Sequence comparisons and oligonucleotide probes of Opaque2 alleles.
5. A web-based survey questionnaire has been developed; rice breeders were the first respondents.

Deviations from the work plan:

- E. Raman Babu of VPKAS, India who is also the major partner for the maize component was not able to join the project in early 2006 for 3-4 months at IRRI. Instead, Dr. Debra Skinner joined to work on the project from April 27-May 16, 2006.
- For the 2005 budget, a no-cost extension was approved from January to October 15, 2006.

2005-19: Evaluation and Deployment of Transgenic Drought-Tolerant Varieties--Potato

Principal Investigator:
Marc Ghislain, CIP
Collaborating Scientists:
Amélie Gaudin, CIP
Roland Schafleitner, CIP
Kazuo Watanabe, University of Tsukuba, Japan

Mid-Year Report

The proposed study focuses on stress-related regulatory genes *AP2/EREBP*-type transcription factors, known as *DREB/CBF*. We aimed at characterising the expression of *Arabidopsis thaliana DREB1A* gene driven by the stress inducible promoter *rd29* and selected target genes in transgenic potato under water shortage. Prior to the initiation of this project, the collaborators at the University of Tsukuba in Japan (Celebi-Toprak et al., 2005) have genetically engineered potato plants. A detailed work plan has been submitted to Dr. K. Watanabe for review and improvement. However, the Peruvian regulations on the exchange of genetically engineered organisms are still lacking the final set of norms and procedures to implement existing laws and regulations on biosafety. Therefore, the plant materials have not yet been transferred to the CIP biosafety greenhouse, where the initial experiments were planned to be conducted.

Thus, to enhance our knowledge on potato DREB endogenous gene expression, our activities are currently concentrating on the characterisation of endogenous DREB transcription factors under drought in a set of Andean clones from the CIP germplasm collection. Our research objective is to understand how much of expression differences of orthologous *DREB/CBF* genes are due to gene regulation and to genetic backgrounds. We are addressing this question by the initial expression characterisation of endogenous DREB1A and selected target genes expression levels using Real time PCR and microarrays in a set of Andean clones showing different responses to water shortage (contribution from other GCP activities).

Four Andean landraces — Cceccorani, Puca Pishgush (of the Stenotomum Group), SA2563 and Sullu (of the Andigenum Group) — and one modern potato variety (Perricholi), representing a wide spectrum of sensitivity to dehydration in the potato germplasm, were submitted to water stress at the CIP La Victoria field station in Huancayo (Peru) at 3200 masl. Two biological repeats of drought exposed and control plants were sampled for molecular biological analysis at two time points during water shortage (-0.3Mpa and -0.6Mpa). We measured an array of physiological parameters (plant water status, stomatal resistance, leaf osmotic potential and photosynthesis efficiency), and morphological, yield and growth evaluation parameters to have a complete overview of the specific clone response to drought. Tuber yield, harvest Index, biomass partitioning and sugar contents were evaluated. We first choose to examine changes in DREB and target genes transcript levels under drought in SA2563 and Sullu, both belonging to the same landrace group Andigenum, with similar phenology and tuberisation period. Both clones displayed drought tolerance, but exhibit different response to water shortage. Sullu retained a high harvest index and apparently allocated more resources to tuber formation, while SA2563 produced a high amount of total biomass.

Under the drought conditions applied, the yield loss of the two resistant clones amounted only to 25 – 27%, while sensitive clones had a yield drop of 50 – 70%.

Gene expression profiling performed in collaboration with The Institute for Genomic Research (TIGR) revealed that some DREB endogenous genes are up regulated in both clones at moderate water shortage; expression of selected target genes is under analysis. Corroboration of chip expression dataset will be done by real time PCR using SYBR green I assay. Real Time PCR primers were designed based on tentative consensus sequences of the potato gene index database of TIGR. Standard curves and efficiency for all primers have been established. Relative quantification of transcript abundance in challenged and control plants is being done according to Pfaffl et al. (2002) using the potato cytochrome B oxidase gene (TIGR StGI TC116542) as internal standard to correct for different amounts of RNA input for cDNA synthesis. Significance of differences in transcript abundance between drought-exposed and control plants will be determined according to t-test.

Gene expression studies combined with physiological analysis are key to understanding the mechanisms that control adaptation to water stress in Andean clones. Changes in DREB endogenous genes expression levels monitored by microarrays and Real Time PCR assays of DREB1A genes and their selected target genes will mirror the multiple environmental stimuli resulting in a unique pattern of gene expression.

Tangible outputs delivered:
Characterised landraces for drought tolerance in order to relate to gene expression studies.

Deviations from the work plan:
Awaiting the biosafety authorisation to import transgenic potatoes in Peru.

2005-19: Evaluation and Deployment of Transgenic Drought-Tolerant Varieties—IRRI

Principal Investigator:
John Bennett, IRRI
Collaborating Scientists:
Rachid Serraj, IRRI
Idupulapati Rao, CIAT
Matthew Reynolds, CIMMYT
Vincent Vadez, ICRISAT
Fernando Ezeta, CIP
Kazuko Yamaguchi-Shinozaki, JIRCAS
Akira Kikuchi, University of Tsukuba

Mid-Year Report
Drought is an important limitation on productivity for all of the mandated crops of the CGIAR system. Although lines of several mandated crops are now available with shorter duration (drought escape) and DNA markers are being identified for deeper and more penetrating roots (drought avoidance), there is also a considerable potential for deploying drought-tolerance genes that enable plants to survive and recover from unavoidable periods of low plant water status, especially at the sensitive reproductive stage. Studies on the model plant Arabidopsis thaliana have been particularly effective in revealing the genes underlying both constitutive and inducible components of drought tolerance. The Generation Challenge Programme (GCP) now wishes to systematically and comprehensively test the hypothesis that this knowledge will be of agronomic value across a range of GCP-mandated crops, notwithstanding the existence of significant variations in gene structure and behavior between and within species. The present proposal focuses on the transgenic approach and on the single gene that has already been introduced into

the largest number/range of mandate crops the DREB1A transcription factor of Arabidopsis (AtDREB1A).

DREB/CBF genes are a small family of transcription factors that bind to the drought responsive elements (DRE) found in the promoters of many drought-responsive genes of Arabidopsis, rice and other plants. Enhanced tolerance of drought, cold and salt was observed in Arabidopsis plants in which the AtDREB1A gene was over-expressed under the control of the viral CaMV35S promoter. However, stunted growth was also observed in these plants, because of the constitutive nature of the promoter. When the promoter of the stress-responsive gene rd29A replaced the CaMV35S promoter, enhanced stress tolerance was observed without the stunting (Kasuga et al., 1999).

This project will consider how effective this gene is against drought. Does it have a yield penalty without drought stress, is it effective in different genetic backgrounds and across the range of GCP-mandated crops, and does it affect other agronomic traits, product quality or weediness of the crop and its relatives? The existing $DREB^{+}$ transformants are found in representatives of cereals, legumes and clonal crops, the three pillars of the comparative biology framework of the GCP. The constructs used to generate these transformants were kindly shared by JIRCAS with CIAT, CIMMYT, ICRISAT, IRRI and the University of Tsukuba under Material Transfer Agreements.

The Goal of the Project is to facilitate the delivery to farmers of drought-tolerant varieties of CGIAR mandated crops. The specific Objectives are:
• To evaluate the impact of DREB1A constructs and different insertional events on the growth, yield, quality and other agronomic traits of drought stressed and =unstressed plants of four CGIAR mandated crops (groundnut, potato, wheat, rice);
• To evaluate the impact of DREB1A constructs on the tolerance of these plants to different durations, intensities and timings of drought stress;
• To examine the impact of DREB1A constructs on weediness traits (seedling vigor, seed shedding/shattering, dormancy);
• To provide six graduate students or visiting scientists from developing countries with training in physiology, molecular biology, breeding, weed science and biosafety protocols relevant to the evaluation and deployment of transgenic drought-tolerant varieties.
• To hold a 3-day workshop on 'Liability Issues related to Deploying Transgenic Crops' in Rome (October 2005).
• To hold a 3-day expert consultative group meeting on 'Biosafety Testing of Abiotic Stress Tolerant Transgenics' at CIMMYT (probably April-June 2006).

Groundnut (ICRISAT)
In three early trials the two most promising events selected from the previous three trials (these events had shown consistently higher TE in RD2 than in the non-transformed parent JL24, both under water stress and water deficit).Two experiments have been carried out using DREB1A transgenic events. The first of these experiments involved RD2 and RD11, The purpose of this experiment was to confirm the superiority of these events and compare them to known germplasm for high (JUG24) and low (TAG24) TE level. Results confirm that RD2 has a higher TE than JL24 under water deficit conditions. It also showed that RD2 had higher TE than the best reported genotype (JUG24) for TE under water deficit. This experiment needs to be reconfirmed with more genotypes reported with high TE. In the second experiment, we tested the activity of the photosynthetic systems at different stages of water deficit in 4 transgenic events and non-transformed parent JL24. Results are still being analysed.

Another experiment is in progress where we are testing the effect of water deficit at different stages (vegetative, flowering and mid-pod filling) in two transgenic events (RD2, RD11), and

wild type JL24. We are using our standard dry-down technique to apply the stress at these different stages. At each stage, we do measure transpiration efficiency (TE), so that sets of plants are harvested before imposing the stress, and others (well-watered and water stress) harvested after water stressed plants have depleted all the soil moisture. Since we are also interested to measure and compare the effect of a drought spell at different stages, we are also keeping two more sets of plants at each stage of treatment imposition. Those two sets are exposed also to either well-watered or water-stressed conditions, and then water-stressed plants re-watered when their relative transpiration is between 10-20% of that of control. Plants from both treatments are then well-watered until harvest. Another experiment is now scheduled to test the effect of (vapor pressure deficit on the response to progressive exposure to water deficit in contrasting transgenic events and wild type JL24.

Rice (CIAT)

The goal of this work is to screen all the DREB transgenic lines that we have generated for lines showing drought tolerance. So far a total of 20 T1 rice lines have been tested for drought tolerance in the screenhouse: 5 lines for AtDREB1A in cultivar Palmar; 5 lines for AtDREB1A in cultivar Cica8; 5 lines for OsDREB1B in Palmar; and 5 lines for OsDREB1B in Cica8. Plants at the vegetative stage were exposed to terminal drought. Agronomic parameters such as height and tiller number and physiological measurements were conducted once the drought response was seen. We identified 3-5 putative drought-tolerant lines from this work but the experiments must be repeated to confirm the phenotype under different drought conditions and at different developmental stages.

We are in the course of selecting homozygous T2 lines from more than 40 T1 lines lines. By November we expect to have identified lines carrying single insertion events with one transgene copy. A rainout shelter (15 m x 7.5 m) is under construction at CIAT HQ campus. Initially, we will test non-transgenic lines to establish field drought screening conditions. We are expecting to obtain a biosafety permit from the Colombian National Biosafety Committee within 3 to 4 month. Once we receive the permit, we shall concentrate on field evaluation of the DREB transgenic lines for yield penalty under control conditions and yield performance under drought conditions.

Wheat (CIMMYT)

Wheat cultivar Bobwhite was transformed with the *DREB1A* gene under the control of the *rd29A* promoter. Both promoter and coding region were derived from Arabidopsis. Using RT-PCR, the expression of the *DREB1A* gene was compared with the expression of the endogenous housekeeping gene *ACTIN* before and after water stress in individual plants of T2 families. The *actin* gene was constitutively expressed in all plants before and after water stress, but transcripts of the *DREB1A* gene were detected after only 2 days of water stress.

Physiological differences between the check (Bobwhite) and transformed T2 lines were sought in pot and field experiments. While pot experiments in which moisture is withheld can give an indication of the ability of plants to survive and recover from drought stress, potential crop productivity can only be estimated under more normal field soil profiles where rooting behavior is realistic. This environment permits roots to penetrate deeper into the soil profile where soil moisture may be extracted preferentially by drought-adapted genotypes, a phenomenon which has been shown to be associated with yield under drought. Therefore, the best performing transformation events selected from small pot experiments were re-tested in normal field soil profiles for evaluation of yield and a series of drought adaptive traits.

Two experiments were grown in screen-houses where small plots were sown in field-type soil profiles. In Experiment 1, a total of 10 transformation (T) events with *rd29A*::*DREB1A* gene were grown in plots of $1m^2$ in a replicated design in two moisture environments: (i) well watered and (ii) pre-anthesis drought stress where plots were watered after anthesis. In Experiment 2, the four

best performing T events (#2, 4, 7 and 10) were grown in plots of $2m^2$ in four moisture environments: (i) well watered, (ii) pre-booting stress, (iii) pre-anthesis stress, and (iv) continual moisture stress. In both experiments the background wheat cultivar Bobwhite was used as a check. In Experiment 1, while no significant differences among genotypes were apparent in biomass measured at anthesis, the *DREB* gene appeared to provide a benefit in recovery from drought as indicated by final yield. In Experiment 2, however, no yield advantage was associated with any of the treatments. Physiological parameters such as leaf water potential did not indicate a benefit of *DREB* except perhaps for the event #10. Further replicated trials were conducted on the four leading T lines, but only T event #2, a late maturing line, showed some differences from the control in canopy temperature and soluble carbohydrate content at anthesis. However, measurements of C isotope discrimination or osmotic adjustment did not suggest that T event #2 resulted in improved drought adaptive mechanisms. A cooler plant canopy is often associated with late maturity.

Further improvements in the initial screening of transformants are now planned, with emphasis on measurement of rooting depth and water use efficiency. By focusing on these traits under more realistic growth conditions, it is hoped that small but significant differences in plant performance will be identified.

Tangible outputs delivered:
Techniques and germplasm
- Agreement on use of dry-down technique for imposing stress across crops and on several key parameters for evaluating plants, including biomass, leaf water potential, C isotope discrimination and osmotic adjustment.
- Confirmation of higher TE for groundnut line JL24 through DREB transgenic event RD2 (higher TE than that of a reported high-TE genotype)

Publication:
- First draft of overview paper on DREB studies in crop plants.
- Presentation of $DREB^+$ groundnut studies at conferences and as draft manuscripts (one submitted to PNAS).

Deviations from the work plan:
Postponement of workshop on Biosafety Testing.

2005-20: Optimising Marker-assisted Breeding Systems for Drought Tolerance in Cereals through Linkage of Physiological and Genetic Models

Principal Investigator:
Scott Chapman, CSIRO
Jiankang Wang, CIMMYT
Collaborating Scientists:
Mark Dieters, University of Queensland
Graeme Hammer, Agricultural Production Systems Research Unit (APSRU)
David Bonnett, CSIRO
Greg J. Rebetzke, CSIRO
Francois Tardieu, INRA
Claude Welcker, INRA
Mark Cooper, Pioneer

Mid-Year Report

1. Completion of a case study on efficient use of marker-based selection in wheat
QuCim/QuLine was used to determine the proportion of target genotypes 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 Australian wheat parents were used, i.e., Sunstate, Silverstar/TIN, and HM14BS. The best strategies using marker based selection in genotype building for the top cross among the three parents have been studied.

An extension of this case study has been undertaken to incorporate selection for a quantitative drought-adaptive trait (coleoptile length). This study will be written up as a paper in the next 6 months and aims at demonstrating effective methods to combine selection for 'perfect' markers for single-gene traits with selection for QTL.

2. Maintain and update of the simulation module QuCim/QuLine

QuCim/QuLine has been updated based on the QU-GENE library version 2.2.04. The QuCim/QuLine User's Manual was revised accordingly. QuCim/QuLine version 1.3 was returned to The University of Queensland and will be distributed to those licence holders upon request. QuCim/QuLine in the current status can simulate almost all breeding activities in our wheat breeding programme, including male master selection, female master selection, parental selection, single cross, backcross, top cross, double cross, doubled haploid, marker-assisted selection etc. QuCim/QuLine can simulate not only CIMMYT's wheat breeding programme but also in principle any other breeding programmes for selecting inbred lines, which means all major food cereals in the world, plus basically all leguminous crops. It can also be used when inbred lines are developed that later are to be used in hybrids, such as in the case of maize. Hence its potential to increase breeding effectiveness encompasses all CGIAR mandated crops.

3. Modification of the QuCim/QuLine breeding module to realise the linkage with physiological models

A new functionality called Plug-In (available in the QUGene library for previous work in sorghum breeding), has been added to QU-GENE module QuCim/QuLine so that the physiological model can be implemented through the linkage between QuCim/QuLine and a crop growth model. We are currently modifying the simulation module QuCim/QuLine so that the simulation model and physiological model can communicate with each other.

In Oct 2005 and May 2006, interactions were established with the INRA research team (Tardieu et al) to obtain published physiological and map data from their research into QTL for a physiological model of leaf elongation rate (LER) in maize. In the next 6 months, simulation experiments will be undertaken using the LER model to demonstrate selection for QTL related to adaptation of leaf growth to drought induced by low humidity and/or by drying soil. These experiments will aim to identify effective selection methods to combine together genes associated with these types of complex traits.

Tangible outputs delivered:

1. In the completed case study on efficient use of marker-based selection in wheat, we found population size was minimised with a three stage selection strategy (in TCF1, TCF2 and DH lines). Enrichment of allelic frequencies in TCF2 reduced the total number of lines required from >3500 to <600. Eight of the genes were present at frequencies >0.97 after selection, while

the tin, reduced-tillering gene was at a frequency of around 0.77 in the final selected population due to its strong repulsion-phase linkage to Glu-A3 in this cross and the 'imperfect' nature of the tin marker. Therefore, the presence of the tin allele following marker assisted selection must be confirmed by other methods. The identified optimum strategy to combine the nine target alleles in the topcross can be divided into three steps: Step 1 - selection for Rht-B1a and Glu-B1i homozygotes, and enrichment of Rht8, Cre1, and tin in TCF1; Step 2 - select homozygotes for one target allele, e.g. Rht8, and enrich remaining target alleles in TCF2; and Step 3 - selection of the target genotype (last row in Table 3) in DHs/RILs.

2. One manuscript on the efficient use of marker-assisted selection has been prepared and submitted to Molecular Breeding for consideration of publication. The authors and title of the manuscript are: Jiankang Wang, Scott Chapman, David Bonnett, Greg Rebetzke, and Jonathan Crouch. Population genetics and simulation models on efficient use of marker-assisted selection in plant breeding. Molecular Breeding (under review).
3. A conference paper to be published by Kluwer was presented to the 'Gene-Plant-Crop' conference in Wageningen April 23-26 2006: Accounting for variability in the use of markers for simple and complex traits S.C. Chapman, J.K. Wang, G. Rebetzke, D. Bonnett.
4. A clear process has been established for the issue of licences to use QUGene and its associated breeding modules by researchers. Under the terms of a pre-existing arrangement, researchers associated with non-commercial breeding programmes in developing countries and who are collaborators with CIMMYT are requested to contact either Jiankang Wang (j.k.wang@cgiar.org or wangjk@caas.net.cn)or Peter Ninnes (p.ninnes@cgiar.org) for the issue of a licence. All other users (including from centres other than CIMMYT) should contact The University of Queensland (Mark Dieters m.dieters@uq.edu.au or Andrew Cecil a.cecil@uniquest.uq.edu.au) to arrange either one of 3 licence types: commercial (negotiated); academic (AUD$50/computer/year) or for student research use (free).

Deviations from the work plan:
The projected began in July 2005, which was six-month later than the scheduled date in the original proposal. Due to the late starting, the application for a no-cost extension has been submitted and approved by GCP director. Under the adjusted workplan, we should be able to maintain the same deliverables on an adjusted time scale.

2005-21: Planning for Effective Product Development, Delivery, and Use

Principal Investigator:
Victoria Henson-Apollonio, IPGRI

Mid-Year Report
The projects/contact that will be followed/utilised, in this study are:

- "Unlocking the genetic diversity in peanut's wild relatives with genomic and genetic tools"/EMBRAPA;
- "Revitalising Marginal Lands: Discovery of Genes for Tolerance of Saline and Phosphorous Deficient Soils to Enhance and Sustain Productivity"/CIMMYT;
- "Development of low tech gene-based trait assay technologies in rice and wheat"/ CIMMYT; and,
- "Simulation of marker-assisted selection strategies for optimising molecular breeding systems for drought tolerance in cereals"/ CAAS.

Although follow-up after our meeting with the principal investigators of all the projects commissioned by the SP3 during the GCP, Annual Research Meeting, has been slow, the synergy between this project and the inclusion of new requirements for product development and delivery planning has increased the need and relevance for this project. (The guidelines that were provided by the PI in 2005, were used in the preparation of the 2006 call/request for proposals.)

With the hiring of a consultant, Ms. Karine Malgrand, we anticipate that the data collection for the case studies will be completed by the time of the 2006-GCP ARM, with the workshop to be held in January 2007. In addition, introductory workshops for new (2007+) GCP grantees will be initiated before the end of 2006, using data from the case studies as examples.

(With the retirement of Dr. Zenete Franca from the ISNAR division of IFPRI, it is anticipated that we will consider hiring Dr. Franca as a consultant to prepare and deliver workshop materials. She has indicated her interest in working on this project workshop).

As indicated in the budget table, below, we are asking for a no-cost extension. The Time frame and milestones have also been adjusted.

Tangible outputs delivered:
Draft TORs for consultancy.

***Deviations from the work plan*:**
–Please see the adjustments to the time frame, (changes highlighted in yellow).

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. The SP5 workshop has helped us to more clearly define the project and as a result we expect out institutional meeting to be more effective.

	2005				2006				2007			
	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
Formation of institutional teams												
Start-up meeting on project strategy[1]												
Research for development and delivery maps												
Development and delivery maps drafted and evaluated												
Selection of specific products/outputs for products and distribution plans												
Initial product distribution and delivery plans drafted												
Preliminary studies of regulatory constraints prepared												
Participation in biosafety workshop		?										
Participation in workshop on 'Liability Issues related to Deploying Transgenic Crops'				?								
Mid-term review and planning colloquium[2]												
First annual report submitted												
Draft publication, based on joint summary Year 1												
Complementary skills/assets analysis												
Draft of complete product development and delivery plans for selected products												
Products plans completed												
Initiation of workshop materials production team												
Preparation of workshop materials												
Product development and delivery workshop												
Evaluation of workshop materials												
Publication of workshop materials for broader distribution												
Terminal report on task												
Publication, based on joint summary Year 2												

[1] Meeting held during the 2005, GCP Annual Research Meeting.

[2] Meeting to be held during the 2006 GCP Annual Research Meeting.

Terms of Reference for Consultant
GCP-SP3 Project Number 21:
Planning for Effective Product Development, Delivery and Use.

With assistance fro the Project PI, Victoria Henson-Apollonio, Consultant will carry out a series of interviews, via telephone and e-mail, to complete a set of case studies from the project in the SubProgramme 3 portfolio including the projects[1]:

- "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;
- "Revitalising Marginal Lands: Discovery of Genes for Tolerance of Saline and Phosphorous Deficient Soils to Enhance and Sustain Productivity"/IRRI;
- "Development of low tech gene-based trait assay technologies in rice and wheat"/IRRI and CIMMYT; and,
- "Development of low tech gene-based trait assay technologies in rice and wheat"/IRRI; and "Simulation of marker-assisted selection strategies for optimising molecular breeding systems for drought tolerance in cereals"/CSIRO.

The focus of these case studies will be to:

1) Identify the full range of products (potential public goods) produced and anticipated by each project and assist the project scientists in preparing a written description of each product.
2) Determine the proximal and end-users for each product.
3) With the project scientists, analyse the skills, assets, and services needed to:
    a. Develop the product to the end-stage, deployable level
    b. Distribute the end-stage product for effective uptake and use
4) Help project scientists carry out a self-assessment to determine the available in-house skills and assets for product development and delivery and thus to also identify gaps.
5) With the project scientists, list existing and potential partners that can provide those skills, assets, and services that are not available within the project base institutions.

The deliverables of this project will be:

1) Notes, interview materials, or any other data that is generated by the project.
2) An inventory of products from each of the listed projects
3) A list of users for each product.
4) Corresponding lists of needed skills, assets, and services for product development and delivery
5) Self-assessment data from all of the projects.
6) Lists of partners for each of the projects.

The desired date for completion of this project is 15August 2006.

[1] The particular cases that are included in this study may vary from this list, based on the progress of these projects and the input from the new SP3 Director.

SP4 COMMISSIONED GRANTS

2004—SP4CL2: Establishing Fingerprinting/Allele Data Repositories for All GCP Partners

Principal Investigator:
Rajesh Sood, IPGRI/SGRP
Collaborating Scientists:
Dag Terje Endresen, IPGRI/Nordic Genebank
Tom Hazekamp, IPGRI
Milko Skofic, IPGRI
Samy Gaiji, IPGRI
Guy Davenport, CIMMYT

Mid-Year Report

This project commenced in 2004 as one component of GCP subprogramme 4, cluster 2: Genetic Resources, Crop and Genomic Information and Analysis Systems. Cluster 2 was aimed at establishing the necessary informatics infrastructure to support the GCP activities related to acquisition, storage, retrieval and analysis of phenotypic, genetic and genomic data.

The cluster workplan cited the need for a repository or data warehouse to aggregate both internal and external data sources that could be integrated with existing information systems. Its deliverables included a prototype system to handle passport data, fingerprinting and allele data generated by the research carried out in GCP-supported activities.

Project Achievements:

At the end of 2004, a study of various database models for storing fingerprinting and allele data repositories was completed. A template using an Excel spreadsheet was developed, validated and expanded to meet GCP requirements. The template was circulated among partners and modified to accommodate their data. It was then posted in the GCP virtual workspace.

In order to continue building upon the achievements of this project, a no-cost extension was granted to facilitate additional data collection in 2005 and 2006 in conjunction with the GCP Central Registry project, in which IPGRI is the lead institution.

Achievements in 2006:

During this reporting period, the technical infrastructure for the GCP fingerprinting data repository was successfully put in place. Through the activities of the System-wide Information Network for Genetic Resources (SINGER), an initiative of the CGIAR System-wide Genetic Resources Programme (SGRP), a BioCASE wrapper was installed in almost all participating centres. Once the Extensible Markup Language (XML) schema for fingerprinting data has been embedded in all BioCASE installations, the GCP Central Registry will automatically harvest data and make it available online.

Data collection efforts:

1. In collaboration with the GCP Central Registry project, efforts were made to collect data from various partners during 2006. These efforts included requesting data by telephone and personal visits to the focal points of participating centres.
2. Data-collection efforts were further strengthened by the availability of fingerprinting-data templates online through the Central Registry.

Tangible outputs delivered in this project:

1. The fingerprinting data submission template is available online at the GCP bioinformatics website: http://www.generationcp.org/bioinformatics.php?da=0526518.
2. With this template, users can upload their data files into the Central Registry; the data-upload mechanism for the Central Registry is now up and running.

Deviations from the work plan:

Delays encountered in the course of implementation included the following:

1. Partners indicated that since their scientific work was still in progress, they were not ready to provide fingerprinting data to the project team. To date, no fingerprinting data has been obtained.
2. Automatic data harvesting scripts are being tested and are in the process of being integrated into the Central Registry. In order to harvest fingerprinting data in the Central Registry automatically, the template – already developed – must be translated into an XML schema; the construction of a fingerprinting XML schema for use with BioCASE is currently underway.

Conclusions and lessons learned:

The collection of data for the data repository will continue under the GCP Central Registry project; the Central Registry will be an integrated repository of plant genetic resource information generated by research supported by the GCP. As these data repositories were designed to support GCP activities by facilitating access to and exchange of phenotypic, genetic and genomic data, their efficacy as research tools depends on the extent to which data is contributed to them.

2005-22: Development of GenerationCP Domain Models

Principal Investigator:

Richard Bruskiewich, IRRI

Collaborating Scientists:

Reinhard Simon, CIP
Manuel Ruiz, CIRAD
Tom Hazekamp, IPGRI
Masaru Takeya, NIAS
Jane Morris, ACGT
Guy Davenport, CIMMYT

Mid-Year Report

- Production release of GCP models for GCP platform and network (MOBY) application essentially completed, archived in CropForge "Pantheon" project, in use for software development, and posted with "models-in-plain-English" narratives at http://pantheon.generationcp.org, cross-linked to GCP platform and MOBY technology documentation, with glossary and ontology inventory (indirect link to this provided via http://www.generationcp.org/model). Ontology site still available at http://ontology.generationcp.org (but may be updated by September)
- GCP models reported in an invited paper in a special edition of the "OMICS" journal:
- Richard Bruskiewich, Guy Davenport, Tom hazekamp, Thomas Metz, Manuel Ruiz, Reinhard Simon, Masaru Takeya, Jennifer Lee, Martin Senger, Graham McLaren and Theo Van Hintum. (2006) Generation Challenge Programme (GCP): Standards for Crop Data. OMICS A Journal of Integrative Biology. 10(2):215-219

Tangible outputs delivered:

- GCP domain models published at www.generationcp.org/model with glossary and ontology inventory

- Release reported in an invited paper in a special edition of the "OMICS" journal (in press)
- Production release of models for GCP platform and network application to be completed and published with "models in plain English" narratives posted to CropForge now, and to the web site by the end of May, 2006
- MOBY data type and Pantheon Java implementations in CropForge (under Pantheon/Demeter project)

Deviations from the work plan:
None except perhaps slower than anticipated (originally) delivery of production-ready models.

2005-23: Implementation of Web Services Technology in the GCP Consortium

Principal Investigator:
Milko A. Skofic, IPGRI
Collaborating Scientists:
Samy Gaiji, IPGRI
Rajesh Sood, IPGRI
Tom Hazekamp, IPGRI
Reinhard Simon, CIP
Richard Bruskiewich, IRRI
Mathieu Rouard, IPGRI
Dag Terje Endresen, IPGRI
Graham McLaren, IRRI
Martin Senger, IRRI
Javier De La Torre, University of Madrid
Donald Hobern, GBIF

Mid-Year Report

Establishment of data-abstraction layer
A data-abstraction layer has been established by deploying BioCASE at almost all data provider sites. Currently, only passport data is mapped using this technology, but as the Development of Generation Challenge Programme domain data models project progresses, other domain schemas will be extracted and made available to BioCASE data providers. In the interim, the data submission templates activity will be tapped into in order to generate schemas usable by BioCASE.

Training
BioCASE was most recently deployed at the Centre for Genetic Resources, the Netherlands (CGN), where training was conducted in mid January 2006. CIMMYT has not yet installed BioCASE because of a lack of resources, but data will be available through this technology once installation is completed. The Africa Rice Centre (WARDA) was trained on BioCASE, but bandwidth problems prevent it from deploying services on site; a solution is being devised to mirror its data in IPGRI.

Other external data providers have also been trained and their passport data is now available via BioCASE, including ILRI, the EURISCO web catalogue, the Germplasm Resources Information Network of the United States Department of Agriculture (USDA/GRIN), the Malaysian Agricultural Research and Development Institute (MARDI) and the Asian Vegetable Research and Development Centre (AVRDC). This has been achieved with the support of SINGER, and all data providers have published their passport data using the Generation CP passport domain model. SINGER also uses the Generation CP Phase I passport model to publish its data through BioCASE, which allows better integration with the Generation CP.

BioMOBY/BioCASE integration
In collaboration with the Application and development of web services project, in particular with Martin Senger, work is underway to provide access to BioCASE through BioMOBY services. The user interface for designing such services will be provided by the Dashboard application (http://biomoby.open-bio.org/CVS_CONTENT/moby-live/Java/docs/ Dashboard.html) and the tool to generate the services will be MoSeS (http://biomoby.open-bio.org/CVS_CONTENT/moby-live/Java/docs/Moses.html), a component of Dashboard. The first MOBY services to take advantage of this integration are being developed and will serve as a basis to complete the tool.

Enhancement of BioCASE protocol
In addition to bridging the gap between BioCASE and MOBY from the MOBY community, work is underway to enhance the functionality of BioCASE by supporting a new protocol called TAPIR (http://trac.pywrapper.org/pywrapper/). This protocol offers several advantages over BioCASE, such as interoperability between schemas and support for other protocols, and will offer full support to BioCASE, which means the upgrade will be transparent to users.

Besides finalising TAPIR, support for BioMOBY will be included in the protocol, providing a solution similar to that provided by Dashboard to MOBY users. This dual approach will allow better integration of the two technologies as well as sustainability, and will open new data sources to both the MOBY and GBIF communities.

Data collection
Collaboration with the Management of Generation Challenge Programme Central Registry project has taken place to provide a means of searching datasets shared within the GCP. Most notably, a tool has been developed to harvest data sets published via BioCASE and to index relevant fields. This can be used to gather data-quality information during indexing activities. Another activity that is sponsored by SINGER is the development of a desktop tool to harvest data from BioCASE/TAPIR providers; a company has been outsourced to develop an application that will allow users to search, harvest and store data locally from any BioCASE/TAPIR provider. This tool will be especially useful for users interested in collecting specific data and performing statistical analysis.

Tangible outputs delivered:
- Training to CGN was provided in mid-January.
- An initial BioMOBY web service, which takes its data from BioCASE, has been developed. This modular service already provides a basis on which to build the Dashboard and MoSeS interface. Information and instructions on how to implement the service are available at http://moby.generationcp.org/bb_services/docs/index.html.
- A plan to finalise the TAPIR protocol and integrate the MOBY protocol into TAPIR will be produced by the end of May.
- Functionality to harvest BioCASE providers has been implemented in the Generation Challenge Programme Central Registry.

Deviations from the work plan:
It was originally expected that the BioMOBY/BioCASE mapping tool and the enhancements to the BioCASE protocol would be finalised by the second quarter of 2006. The initial BioMOBY/BioCASE service provides a skeleton and the workflow to create other services of its kind, and scientists with knowledge of MOBY can already build upon that service, but more work is needed on both the Dashboard user interface and TAPIR to support BioMOBY; it is not yet a deployment-ready solution. Dashboard and MoSeS were chosen as the reference technologies to deliver a deployment-ready solution because they are the first tools to provide a user friendly

interface to MOBY development and they take advantage of the research and knowledge of the MOBY Java platform creators. By September, the project team should have the means to provide tools that will allow centralised assistance in the deployment of MOBY services.

2005-24: Application and Development of Web Services Technology

Principal Investigator:
Richard Bruskiewich, IRRI
Collaborating Scientists:
Reinhard Simon, CIP
Natalia Martins, EMBRAPA
Mathieu Rouard, INIBAP
Shoshi Kikuchi, NIAS
Masaru Takeya, NIAS
Koji Doi, NIAS

Mid-Year Report

- Task funded consultant, Mr. Martin Senger, completed the design, implementation and production release of the MOBY Services Support (MOSES) and associated MOBY Dashboard tool supporting web services provider specification, implementation and testing of web services.
- GCP version of the MOBY Central Registry commissioned at IRRI for GCP web services testing.
- MOBY project documentation consolidated and enhanced on GCP Wiki.
- Specialise MOBY scripting support created and posted to CropForge under the GCP MOBY Project including automated Unix cronjob monitoring of registered, implemented ("live") web services.
- GCP MOBY web site set up (http://moby.generationcp.org) for public documentation (outside GCP Wiki) for MOBY web services.
- Project meeting held in Pretoria, South Africa to plan development of the MOBY Rice Functional Genomics demonstration network (2006 deliverable).
- Task team translating GCP domain model into MOBY data types (posted in Pantheon CropForge project, see http://pantheon.generationcp.org/moby for details).
- Martin Senger visited Montpellier, France (May 1-5th, 2006) to continue design and pilot implementation of Rice Moby network.
- NIAS scientists visited IRRI May 8-17th, 2006 to develop microarray-based web services (among other things).
- Mathieu Rouard worked collaborative with GCP colleagues to code web services using the newly available GCP MOBY data types, in collaboration with IRRI staff.

Tangible outputs delivered:

- MOSES tool kit and MOBY Dashboard
- GCP domain model compliant MOBY data types (documentation posted on http://pantheon.generationcp.org/moby)
- Selection of MOBY web services implemented and accessible at a number of GCP partner sites (documentation at http://cropwiki.irri.org/gcp/index.php/MOBY_Rice_Network)

Deviations from the work plan:

None except definition of GCP MOBY data types slightly delayed pending GCP domain model stabilisation. The only team meeting held was in Pretoria so project travel budgets may be largely unspent to date for all partners.

2005-25: Creation and Maintenance of Data Templates

Principal Investigator:
Guy Davenport, CIMMYT
Collaborating Scientists:
Richard Bruskiewich, IRRI
Brigitte Courtois, Agropolis
Andrew Farmer, NCGR
Andy Flavell, UoD
Sarah Hearne, IITA
Tom Hazekamp, IPGRI

Mid-Year Report

Templates for Mapping, QTL and Phenoptying data are nearly complete and will be released on the GCP Bioinformatics portal in June. Work on a SNP genotyping template has yet to be started, since we are still waiting for guidelines on the storage, format and minimal information on SNP data to be delivered by SCRI/UofD by the end of May.

The original genotyping data parser was extended into a generic parser and validater which is available on CropForge. This parser was written in such a way that it can be used for all existing and planned templates and with little changes could also be used for other text data formats. It can be used either in a command line mode for use with the central registry or through an API by the RCP template editor. The first release of this editor will be made on the GCP Bioinformatics portal in June. A set of XSL scripts and other Java code has been developed to convert the XML to a number of different text formats. Specifically, the original requirements for genotyping formats from the 2004 Zaragosa workshop meeting will be fulfilled (see attached file). Plug-ins for the Germinate and ICIS databases will also be developed to allow data to be uploaded directly to these databases are envisioned before the Annual Research Meeting.

Uptake of the templates has been poor, with only a few datasets submitted to the central registry. A more proactive approach will be attempted. It is still hoped that the conversion tools provided by the RCP client will provide incentive to use the templates. However, support will also be provided in filling templates as further incentive starting in late May/Early June, which will be advertised by email and news item on the GCP web site.

Tangible outputs delivered:
Generic parser and available on CropForge.

Deviations from the work plan:
Final Deliverables for 2005 project will be available in June 2006 and not May 2006 as mentioned in the last report.

2005-26: Management of the Generation CP Central Registry

Principal Investigator:
Tom Hazekamp, IPGRI
Collaborating Scientists:
Marco Bink, WUR
Subhash Chandra, ICRISAT
Guy Davenport, CIMMYT
Samy Gaiji, IPGRI
Reinhard Simon, CIP
Milko Skofic, IPGRI
Rajesh Sood, IPGRI
Dag Terje Endresen, IPGRI/NGB

Mid-Year Report

Technical Management of Central Registry

During the reporting period, the technical management of the Central Registry has been reviewed, and a schedule of daily backups of the site and the underlying databases has been implemented. The database management system (DBMS) underlying the Central Registry was upgraded to Postgres version 8.0.1. In addition, documentation describing the technical management procedures is under development.

Building up the Central Registry resource collection

At the beginning of 2006, the Central Registry contained mainly descriptive data on GCP datasets (72 in total). At the end of January, an email was sent to all Principal Investigators who had previously provided descriptive data on GCP datasets that were produced within their respective projects. The Principal Investigators were urged to visit the Central Registry site (http://gcpcr.grinfo.net), review their entries and upload data files if available. They were provided with instructions, including individual user IDs and passwords, to enable them to login and modify text or to upload data files. Only a few responses were received from Principal Investigators, indicating that they were ready (or nearly ready) to provide data sets for upload to the Central Registry. From February onwards, a number of Principal Investigators were approached individually in email and telephone communications and, if possible, personal meetings to solicit the upload of data files. IPGRI personnel working with other networks such as the SINGER Network were also mobilised to contact potential providers within the GCP and remind them to submit data to the Central Registry.

Assistance was provided to bring the resources online, including assistance in documenting, formatting and uploading the data files. To date (30 June 2006) 75 datasets have been registered and 10 files related to datasets have been uploaded. A few files are in the process of being submitted. For most datasets, there is descriptive information indicating whether data files are available (or when they will become available) together with the email address of a contact person. Very few are currently available as direct downloads from the Central Registry, however. The preparation of the data for publishing in the Central Registry is often a painstakingly slow process. Usually the data files need some additional work such as double-checking of the content and providing additional documentation and formatting, all of which require a substantial amount of time.

Efforts will continue to upload more data files to the Central Registry. Collaboration in the collation of data with GCP Project 2005-25 (Creation and maintenance of templates for Generation CP data storage in repositories led by CIMMYT) can help to generate a greater push to publish data using GCP templates. The 2 projects jointly prepared a submission on the Central Registry and Data Templates for the GCP e-newsletter. The submission was published in the newsletter of 17 June 2006.

In addition to the e-newsletter article, a follow-up activity has been developed together with the SP4 Leader. The Leaders of SP1, 2 and 3 have been asked to individually contact the PI's within their programmes and urge them to register and upload datasets to the Central Registry using the appropriate GCP Data Templates. Assistance is available from Tom Hazekamp (IPGRI) for the Central Registry and Guy Davenport (CIMMYT) for the Data Templates. Regularly GCP Management will be provided with information on the status of the Central Registry's data collections. This information will enable them to undertake actions within their respective sub-programmes to further facilitate the flow of data into the Central Registry.

Development of the Central Registry
Further testing of the Central Registry application was performed early in the year to eliminate programming bugs and improve functionality. Scripts for the data harvesting of Web services have been developed; these activities were implemented primarily by Dag Terje Endresen.

The scripts for data harvesting of Web services are in the process of being incorporated in the Central Registry application. Once these are integrated, an index system can be implemented on data sources available as Web services. This will allow a more direct and detailed discovery at the unit level (i.e. 'accession' level). Relevant new developments on open-source datamart technology are being monitored and evaluated by CIP.

As an added-value activity, the Central Registry project is collaborating with three other GCP SP4 projects to produce written assessments of analysis tools for SP1, SP2 and SP3. Tool revision for SP2 (specifically an R-based library bioconductor for microarray analysis) has begun. A list of software packages that have already been reviewed by the French Agricultural Research Centre for International Development (CIRAD) in 2004 and 2005 was shared among the collaborating Principal Investigators (courtesy of Brigitte Courtois) in addition to the written software assessments. A section on the GCP Wiki has been created where the assessments can be published. The Wiki-type environment allows users to provide comments on the assessments and will help keep the documents current.

Content Management
A first sweep was made through the descriptive data held by the Central Registry. Where possible, consistent formatting was applied, obvious typographical errors were removed and acronyms were spelled out. Activities to further clean up the data and enhance their presentation are underway. Although some entries are very well documented, many offered very minimal documentation. Data providers were urged to make the descriptive data as complete as possible.

Help desk
Help was provided to users mainly by email. Most of the requests were related to the documentation and formatting of data files.

Tangible outputs delivered:
- A fully functional Central Registry application (http://gcpcr.grinfo.net) has been developed allowing the registration of datasets, upload of data files and registration of Web-service entry points. It features user-friendly and flexible query facilities and allows the download of centrally stored data files.
- To date (30 June 2006), 75 datasets have been registered in the Central Registry (http://gcpcr.grinfo.net) and 10 files related to these data sets are available for direct download.

Deviations from the work plan:
No deviations from the work plan have been made to date.

2005-27: Integration of the High Performance Computing (HPC)-Facilities in the GenerationCP Toolbox

Principal Investigator:
Anthony Collins, CIP
Collaborating Scientists:
Reinhard Simon, CIP
Roland Schlafeitner, CIP
Subhash Chandra, ICRISAT

Jayashree Balaji, ICRISAT
David Hoisington, ICRISAT
Rajeev Varshney, ICRISAT
Richard Bruskiewich, IRRI
Thomas Metz, IRRI
Manuel Ruiz, CIRAD
Guy Davenport, CIMMYT
Marcos Costa, EMBRAPA
Guy Davenport, CIMMYT
Marcos Costa, Embrapa

Mid-Year Report

CIP: HPC use case summary (2005/6)

Primary use to support gene annotation via BLAST of potato gene expression experiments (Roland Schafleitner, Hannele Kreuze, SP2) using the TIGR chip results.

Support of SP1 (Jorge Nunez, Marc Ghislain) to complement gene pool analysis with SSR markers.

ICRISAT: HPC use case summary (2005/6)

Computational SNP mining from DNA sequence data. A pipeline of open source tools for the identification of SNPs (Single Nucleotide Polymorphisms) from nucleotide sequence data was installed on the HPC at ICRISAT. These include a part of the TGICL for the clustering of sequences using megablast; PCAP (Huang et al., 2003) for the assembly of clustered sequence datasets; PolyBayes (Marth et al., 1999) for the discovery of SNPs in assembled datasets and SNP2CAPS (Thiel et al., 2004) for the conversion of SNP markers to CAPS markers (cleaved amplified polymorphic sequences) for cost effective genotyping of SNP markers. The entire pipeline has simple web interfaces for job submission, retrieval and viewing of output files.

Parallel implementation of the population genetics software. Implementation of the Structure software within a parallel framework was carried out to allow simultaneous execution across all nodes.

IRRI: HPC use case summary (2005/6)

Rice genome repeat masking and alignments for the Perlegen rice SNP discovery project: Various sequence analysis tools (e.g. BLAST) and repeat databases were combined into a pipeline to mask out repetitive sequences in published rice genome assemblies and the remaining unmasked sequences realigned against the genome to detect low copy sequence redundancy. The various available rice genome sequences (from different germplasm sources) were also cross-compared to one another. The ultimate objective was to identify a set of ~100 megabases of "single copy" sequences for use in oligomer design for the DNA hybridisation-based Perlegen SNP detection arrays.

Ongoing: NIAS; Japanese scientists at IRRI collaborating on design and further software integration of microarray analysis tools and data.

One target is to tie HPC analysis of such data into the GCP platform (and the Task 2005-32 microarray repository and data mining task) using SoapLab web services.

CIP and IRRI Access Grid video conferencing for Bioinformatics

Access Grid video conferencing via Internet2 is now available between CIP and IRRI, as a forward-looking step to promote North/South and Collaborator dialogue between Bioinformatics

communities. CIAT and CIMMYT have recently connected to Internet2 as a gateway to the major global Bioinformatics establishments, particularly in USA, Europe and Australasia.

Tangible outputs delivered:
CIP:
- Posters presented at Solanaceae Genomic Workshop, Italy 2005 and GCP ARM, Rome, Italy: Nunez J. et al. Gene pool structure of cultivated potatoes assessed by SSR marker analyses.
- HPC user information poster and brochure presented at ARM.

ICRISAT:
- Parallel implementation of a pipeline of software tools for SNP identification from DNA sequence data, used to mine for SNPs from public chickpea and groundnut EST data, contributing to a GCP SP3 activity and an ICRISAT-Israel collaborative project (Figure 1).
- Structure software implemented in parallel for the analysis of sorghum and chickpea datasets (SP2 activities), reporting a speedup of almost four (when comparing the speed of execution on eight processors of the HPC with the speed of execution on one desktop)

IRRI:
- Microarray data analysis:
- HPC Analysis: user/projects/experiments/datasets/hpc analysis done (Table 1)
- General analysis pipeline/Workflow diagram (Figure 2)

Deviations from the work plan:
ICRISAT:
Addition of the SNP2CAPS software to the pipeline of SNP detection tools. This software allows the deduction of CAPS marker candidates for SNPs by the identification of putative restriction sites in any primer pairs flanking the SNP sites. The cost of a CAPS assay is generally low, and manual conversion of SNPs to CAPs a difficult process, which makes this software a useful addition for the user.
IITA:
With the departure of Frank Noonan from IITA, the planned collaboration with IRRI was assumed by IRRI without overall budget variations.
General:
All HPC sites have experienced some technical support and infrastructure logistics problems in 2005/6, leading to the no-cost extension. The pending 2006 proposal places strong focus on more specific deliverables for the GCP 2006 ARM, with a broadening of the user community.

2005-31: Development of Ortholog-Function Display Tools

Principal Investigator:
Richard Bruskiewich, IRRI
Collaborating Scientists:
Kimmen Sjölander, University of California-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
- IRRI staff met at UC Berkeley with Kimmen Sjölander (December 2005) to discuss project design & development

- Task funded NARES postdoctoral scientist, Dr. Samart Wanchana (from Thailand) started work on January 2nd, 2006 and has made good progress in the initial compilation of integrated ortholog gene data sets, in collaboration with external partners
- Dr. Wanchana attended the March, 2006 ADOC project meeting in Montpellier, France to coordinate efforts
- GCP domain modeling pertinent to the gene catalog design and usage has progressed significantly further with the task 22 funded Pretoria review meeting. The revised model and associated ontology are being applied to this project. Discussions with CIRAD, ICRISAT and EMBRAPA task participants undertaken in Pretoria, with some technical objectives established for coming months, pertinent to this project.
- Work progresses on gene catalog database, data entry tool, web interface and web services design for the gene catalog, in collaboration with the Pantheon GCP platform and network software development project (http://pantheon.generationcp.org).

Tangible outputs delivered:

- Prototype gene catalog database and data entry tools using GCP model coupled with ICIS and GMOD Chado schemata established at IRRI and being loaded with project data.
- Preliminary web based user interface to the catalog will be available by end of August
- Preliminary web service data types and services specified and partly implemented by end of August

Deviations from the work plan:

Time delay: Previously reported 1 year's delay in the recruitment of the GCP postdoctoral fellow for this task, coupled with delays in pertinent domain model development, means that the progress in the project design, implementation and data curation has been seriously delayed to the end of 2005.

2005-32: Development of Crop Gene Expression Database and Data Mining Tools

Principal Investigator:
Shoshi Kikuchi, NIAS
Collaborating Scientists:
Masaru Takeya, NIAS
Kohji Doi, NIAS
Koji Suzuki, Hitachi Software Engineering Co., Ltd.
Richard Bruskiewich, IRRI

Mid-Year Report

There are two objectives of this project. One is establishment of crop gene expression data base mainly from the microarray data. The other is to prepare tools for the mining of gene expression data. Last year (2005) we have mainly focused on the construction of gene expression database, mainly by modification of already existing Rice Expression Data base http://red.dna.affrc.go.jp/RED/ . This year we have tried to make an important tool for the promoter analysis.

Many plant physiologists have problems of data mining process after obtaining massive gene expression data from the microarray analysis. Especially clustering analysis of differentially expressed genes, many plant physiologists want to know the commonness of the biological function of clustered genes and commonness of the regulatory system of transcription. For the elucidation of the former problem, accurate and trend contemporary gene annotation is required and for the elucitation of latter problem, mapping position of the gene, promoter sequence information and cis-regulatory element information are necessary. In case of rice, under the favor

of complete genome sequence and massive collection of full-length cDNA clones, those kinds of information are available. But for the plant physiologist or many users of microarray system, it is a laborious work to obtain those kind of information gene by gene.

We have established a pipeline system for the promoter analysis. In this report, we will show the scheme of the pipeline as follows. And currently, we are validating the results generated by the system.

The pipeline system is constructed to list up cis-element candidates particularly corresponding to user-defined gene lists, using information of upstream sequence of genes. Frequency of appearance of motifs in upstream sequences of listed genes is counted, and is compared with that of all genes in KOME database. The system consists of several small perl scripts. One of these call MEME, an open software for motif search (Timothy et al 1994). Another script carries out association rule analysis to evaluate its specificity of cis-element candidates listed by MEME, and filtered by values of list index. The motifs showing good value (list>1.0) are listed as the final cis-element candidates. This strategy depends on the idea that motifs specifically frequent should play specific roles on expression regulation of listed genes.

It is known that some cis-elements need to interact with other element to function. This is important phenomena to understand the gene expression regulatory mechanism of cis-elements. Therefore, the system of next version will also enable to evaluate co-existing of multiple cis-elements in a single upstream sequence, using same algorithm described above.

We are planning to open this system for the biologists using microarray system for the gene expression analysis and also planning to publish the manuscript introducing this system
We are also collaborating with Richard Bruskiewich's research group in IRRI to implement the system in BIO-MOBY system. From 8th May to 17th May, visited IRRI for the purpose.

2005-33: Development of an Integrated Decision Support System for Marker-assisted Plant Breeding

Principal Investigator:
Subhash Chandra, ICRISAT
Collaborating Scientists:
Jose Crossa, CIMMYT
Guy Davenport, CIMMYT
Graham McLaren, IRRI

Mid-Year Report

Project Goal and Structure

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: A1: Analyse potentially useful free software, A2: Select software for inclusion in iMAS, A3: Develop iMAS system, A4: Develop & incorporate online decision guidelines, A5: Test iMAS system, A6: Refine iMAS system, A7: Develop iMAS user manual/tutorial, A8: Release of and Training in iMAS, A9: Consultation and support.

A3: The testing for correct functionality of IRRISTAT, GMendel, PlabQTL, Win QTL-Cartographer, PopMin and GGT within the iMAS system has been completed and problems encountered successfully resolved. The mapping population module has been completed. Completion of association mapping module is in progress. TASSEL for association mapping, for which permission has only recently been received, will be next incorporated into iMAS and its functionality tested.

A4: Online decision guidelines for IRRISTAT, GMendel, PlabQTL, PopMin and GGT have been prepared and incorporated into the system. Those for Win QTL-Cartographer are in the process of being incorporated. Guidelines for TASSEL will be prepared after it is incorporated into iMAS and its correct functionality successfully tested.

A5-A6: The system, as it is being developed, is being continually tested, refined and problems resolved. A one-day local iMAS workshop at ICRISAT has been planned at the end of May to get feedback from molecular breeders - particularly from local NARS - the target group of this system. After that, the system will be refined, as necessary, and sent to CIMMYT and IRRI for similar feedback from (NARS) molecular breeders there.

A7: Development of an on-line tutorial is in progress and is likely to be completed by July/August.

Tangible outputs delivered:
A3: The mapping population module has been completed.

A4: Online decision guidelines for IRRISTAT, GMendel, PlabQTL, PopMin and GGT have been prepared and incorporated into the system.

Deviations from the work plan:
No deviation.

2005-34: GenerationCP Software Engineering and Collaboration Platform

Principal Investigator:
Thomas Metz, IRRI
Collaborating Scientists:
Reinhard Simon, CIP
Edwin Rojas, CIP
GCP Scientists as focal points and users of the collaboration platform

Mid-Year Report
IRRI - This project is a continuation from 2005, with increased budget and scope. The use of the collaboration systems (CropForge and GCPWiki) is increasing, not only by the early adopters. For example, currently ICRISAT is being guided through the process of establishing a project, and managing their source code (ICRISAT LIMS) on the CropForge system. A new application for the GCPWiki was its use during a recent GCP workshop, when the workshop report and documentation was collaboratively developed while the workshop was progressing. Hardware and software upgrading are on track, but slightly delayed due to budget availability. Consultancies on both collaboration systems, CropForge and GCPWiki, have been conducted, and the reports have been circulated. While increased uptake of the collaboration systems can be observed, it is gradual and requires close support. Extensive training materials are currently under development.

CIP - Funds were received in March 2006. The Wiki-concept as a means of institutional communication has been coordinated with CIP's Head of Communication. The Wiki-concept has

been applied in several of CIP's websites (where adequate), for example for DIVA-GIS, the World Potato Atlas, etc. Two staff, one consultant, and one trainee have been trained to use a Wiki, along with three scientists. It is hoped that this training will eventually lead to a more active use of GCP Wiki by CIP scientists. Experiences with an issue-tracking system (JIRA) employed at CIP give indications of improved software documentation and quality.

Tangible outputs delivered:

- After running for about 10 month under an initial prototype installation, the CropForge system has been upgraded to a newer version via a commercial support contract. The new CropForge version now supports Subversion as a source code repository, which was an outstanding request mainly from the Java developer community.
- Two dedicated servers have been purchased and installed, and the transfer and testing of the CropForge and CropWiki systems to the new servers is currently ongoing. Both servers are now housed in a dedicated server room with UPS, air-conditioning, as well as better physical access control.
- Two consultants (Michael Janich, Dave Edwards) have reviewed the use of the collaboration systems (GCPWiki, CropForge) and their reports have been circulated. Recommendations from their reports will be implemented during the rest of the project phase 2006.
- User and focal point training has been provided during two GCP workshops (Platform Development & Domain Modeling – March 2006, Pretoria, South Africa).

Deviations from the work plan:

Project is on track with only minor deviation from the project time table.

2006-08: Data Analysis Support for Existing Projects in SP2 with Emphasis on Integrating Results across Gene Expression and QTL Mapping Experiments

Principal Investigator:

Guy Davenport, CIMMYT

Co-Principal Investigators:

Richard Bruskiewich, IRRI
Shoshi Kikuchi, NIAS
Andreas Magusin, JIC

Collaborating Scientists:

K. Satoh, NIAS
Masaru Takeya, NIAS
Hei Leung, IRRI
Jose Crossa, CIMMYT
Yunbi Xu, CIMMYT
Ramil Mauleon, IRRI

Mid-Year Report

Project initiated in August 2006, no report yet available.

2006-16: Development of an Integrated GCP Information Platform

Principal Investigator:

Graham McLaren, IRRI

Collaborating Scientists:

Guy Davenport, CIMMYT
Reinhard Simon, CIP
Richard Bruskiewich, IRRI
Martin Senger, IRRI

Akinnola Akintunde, ICARDA
Manuel Ruiz, CIRAD
Jayashree Balaji, ICRISAT
Andrew Farmer, NCGR
Maseru Takeya, NIAS
Jane Morris, ACGT
Natalia Martins, EMBRAPA

Mid-Year Report

The following activities were agreed for this project:

- GCP workbench middleware: framework and components – M.Senger and all partners.
- Generalised query engine and result integrator - Richard Bruskiewich.
- Germplasm Genotype visualisation tools- Akinnola Akintunde.
- Hibernate adaptors to the ICRIS database - Jayashree Balaji
- Genomedium interface for genotype, QTL and map data analysis and visualisation - Guy Davenport.
- Genomic Sequence analysis and visualisation - Manuel Ruiz.
- Geographical data analysis and visualisation - Reinhard Simon.
- Web Service/Internet Data Source Integration - Martin Senger.
- BioMOBY: using the MOSeS Java tool kit, and the MOBY Dashboard
- BioCASE (wrapped as BioMOBY services)
- Genomic Diversity and Phenotype Connector (GDPC)
- Taverna Workflow Engine: will be connected to the GCP domain model.
- Gene expression data analysis and visualisation - Masaru Takeya.

A one-day planning meeting was held in conjunction with a domain model and platform review meeting in Pretoria in March. Activity leaders agreed to specify outputs and milestones for each activity using a form on CropWiki. The specifications are to include milestones for biological applications to be demonstrated at the ARM.
http://cropwiki.irri.org/gcp/index.php/Platform_Development_Activities_for_2006

Martin Senger, Guy Davenport and Richard Bruskiewich have been preparing the technical foundations for the workbench through the following activities:

- Updating GCP Domain model and deriving Biomoby data types.
- Defining two core "glue" interfaces - "GCP Data Sources" and "GCP Data Consumers". These interfaces are now available, together with examples and tutorials on how to use them.
- Documenting Pantheon (the platform). The official URL is http://pantheon.generationcp.org.
- Commissioning the GCP "Sibyl" Pantheon platform search engine web site at http://sibyl.generationcp.org.

Martin Senger is integrating the workbench to the GCP network through the following activities:

- Wrapping BioCASE data by Biomoby services, now running on top of SINGER at IPGRI (http://moby.generationcp.org/bb_services/docs/index.html).
- Visiting CIRAD/INIBAP and working on the Rice Moby Network using Biomoby and the Dashboard.
- Visiting Terry Casstevens regarding integration of GDPC into GCP.
- Attending the "Virtual Plant Information Network" meeting at NCGR, Santa Fe, NM, USA

Tangible outputs delivered:

Core interfaces for GCP data sources and consumers defined within the Pantheon code repository and tutorials and documentation developed for their application for different use-cases.

Deviations from the work plan:
The workplan has been retarded by continued delays in finalising the domain model and by continued work on the 2005 deliverables. Both these constraints should now be relieved and members must now specify specific use cases to be implemented by ARM.

2006-17: GenerationCP Data Quality Improvement and Assurance

Principal Investigator:
Thomas Metz, IRRI
Collaborating Scientists:
GCP Scientists and institutions

Mid-Year Report
This project was part of the 2005 platform development project. It has become an independent project under a new a new PI in 2006. The focus of the project was changed to deliver concrete and immediately usable outputs that have an effect on data quality.

The project has had a slow start, with the community building aspects (mailing list, GCPWiki use) taking longer than expected.

A mailing list for GCP partner focal points has been started and initial content has been posted on the GCPWiki.

So far 6 GCP partners have made a commitment to contribute, or at least expressed a strong interest to contribute (IRRI, ICRISAT, INIBAP, NIAS, ICARDA, CIMMYT, CIP). Direct contacts with GCP partners during the next month are expected to get this commitments and expressions of interest turned into work plans, budgets, and deliverables.

Tangible outputs delivered:
- A mailing list for GCP partner focal points has been established (gcpcomm-dataquality@lists.cropforge.org) with currently 12 subscribed GCP partner focal points.
- On the GCPWiki, a section on the data quality project has been established, and initial content has been added to the page (http://cropwiki.irri.org/gcp/index.php/DataQuality2006).
- IRRI has fully documented its implementation of linking bar-coding and electronic scales to its genebank information management system. (http://cropwiki.irri.org/gcp/index.php/IRRIbarcoding)
- ICRISAT has prepared a work plan and budget for publishing phenotyping protocols and allele calling software that it has developed. (http://cropwiki.irri.org/gcp/index.php/ICRISAT_Data_Quality_Work_Proposal)

Deviations from the work plan:
As indicated in the January-May report section, the project has had a slow start, explained at least partly due to a new mode of operation (extensive use of collaborative tools), the new focus of the project, and the need to develop GCP partner work plans at the beginning of the year. Taking into account the actual budget release, the project has only had slightly more that 2 months of operational time.

During the next few weeks the PI will have direct partner contacts (NIAS, CIMMYT, CIP, ICARDA), which are expected to speed up and intensify the process.

2006-18: Creation of Institutional Bioinformatics Capacity (CIAT)-2006

Principal Investigator:
Joe Tohme, CIAT

Co-Principal Investigators:
Fernando Rojas, CIAT
Mathias Lorieux, CIAT
Martin Fregene, CIAT
Matthew Blair, CIAT

Mid-Year Report
Not received.

2006-19: Creation of Institutional Bioinformatics Capacity (CIMMYT)

Principal Investigator:
Guy Davenport, CIMMYT

Mid-Year Report
Not received.

2006-20: Creation of Institutional Bioinformatics Capacity (CIP)

Principal Investigator:
Reinhard Simon, CIP
Collaborating Scientists:
Edwin Rojas, CIP
Sara Villanueva, CIP
Luis Avila, CIP
Magna Schmitt, CIP
Enver Tarazona, CIP

Mid-Year Report
Funds arrived in March 2006. Virginia Tech was informed of this. As it is anticipated that this activity will have co-funding from the HCP activity for 2006, transfer of funds to Virginia Tech is pending until HPC proposal for 2006 is approved.

Tangible outputs delivered:
None so far.

Deviations from the work plan:
None.

2006-21: Creation of Institutional Bioinformatics Capacity (ICARDA)

Principal Investigator:
M. Singh, ICARDA
Collaborating Scientists:
M. Baum, ICARDA
K. Chabane, ICARDA
A. Akintunde, ICARDA
K. El-Shamaa, ICARDA
H. Abed; ICARDA

Mid-Year Report
A number of software – STRUCTURE, Distruct, Strat, Tassel, and DARwin – were examined for their functionalities, self training was carried out on these and support was provided to the scientists. A PC was procured and Linux operating system was installed. Sequence assembly software suit Phred, Phrap, Consed of packages was procured and their installation was

completed. Plans are underway to train two staff member in the areas of bioinformatics and statistical genomics this year seeking the opportunity of conferences and workshops.

Tangible outputs delivered:

- Bioinformatics support on Structure 2.1 and DARwin 5 software for studying the population structure
- Sequence assembling facility using Phred, Phrap, Consed
- Training material on QTL estimation

Deviations from the work plan:
None.

2006-22: Creation of Institutional Bioinformatics Capacity (ICRISAT)

Principal Investigator:
Subhash Chandra, ICRISAT
Collaborating Scientists:
Jayashree B, ICRISAT
Dave Hoisington, ICRISAT

Mid-Year Report
Activity 1: Capacity building

Scientists and technical staff	Number of postgraduate students	Number of doctoral students	Trained in
		3	Sequence analysis
	5		Sequence analysis, Perl scripting for sequence analysis applications.
17			Use of the LIMS application

Activity 2: Support to researchers:
Support provided in the areas of compiling tools for ortholog identification, mining of public datasets, use of AGL-LIMS with respect to upload and retrieval of information, linkage mapping, QTL analysis, molecular diversity analysis, development of phenotyping protocols.

Activity 3: Maintenance/Updation of databases and information management systems:
The EST-SSR database has been updated and now includes candidate conserved orthologous sets (COS) markers along with their primer pairs. Sequence data in the Environment stress transcripts database is currently being annotated with FUNCAT categories and conserved protein motifs as identified from Pfam databases. These additional annotations is hoped to increase the usefulness of this resource for the local user community. Users of the Inventory management system are also being provided with database management support. The user accesses all databases and management systems via the intranet.

Tangible outputs delivered:
Bioinformatics support to researchers.
Updated databases and support with using Information management systems.

Deviations from the work plan:
None.

2006-23: Creation of Institutional Bioinformatics Capacity (IITA)

Principal Investigator:
Dong-Jin Kim, IITA
Collaborating Scientists:
Trushar Shah, ILRI/BECA
Andrew Farmer, NCGR

Mid-Year Report
The bioinformatics platform has been developed in IITA-Nairobi/BECA with a goal of a user interface platform, especially for the African scientists. This bioinformatics project website was just open in May 2006, and an outline and some contents of the current project will be found.

1. Home:
http://www.becabioinfo.org/bbrweb/cms/iita-bioinformatics/
2. The current legume COS-Cowpea SNPs project
http://www.becabioinfo.org/bbrweb/cms/iita-bioinformatics/CowpeaSNP/putative_cos.php
3. Comparative sequence analysis of the legume sequences from NCBI: four tropical legumes (Vu, Vr, Pc, Pv) are blasted with reference legumes(Mt, Lj, Gm) and themselves.
Cowpea: Vu
http://www.becabioinfo.org/bbrweb/cms/iita-bioinformatics/CowpeaBLASTResults/contig_singleton_list.php?querydbname=VUGI
Vigna radiata
http://www.becabioinfo.org/bbrweb/cms/iita-bioinformatics/CowpeaBLASTResults/contig_singleton_list.php?querydbname=VRGI
Phaseolus coccineus (PCGI)
http://www.becabioinfo.org/bbrweb/cms/iita-bioinformatics/CowpeaBLASTResults/contig_singleton_list.php?querydbname=PCGI
Phaseolus vulgaris (PVGI)
http://www.becabioinfo.org/bbrweb/cms/iita-bioinformatics/CowpeaBLASTResults/contig_singleton_list.php?querydbname=PVGI
4. GCP-bioinformatics wokshop
The Bioinformatics workshop will be held in the fall of 2006.
http://www.becabioinfo.org/bbrweb/cms/iita-bioinformatics/GCPBioinformaticsWorkshop/

Tangible outputs delivered:
Project website installed in BECA/IITA-Nairobi.

Deviations from the work plan:
None.

2006-24: Creation of Institutional Bioinformatics Capacity (IPGRI)

Principal Investigator:
Samy Gaiji, IPGRI
Collaborating Scientist:
Dag Terje Endresen, IPGRI/Nordic Gene Bank (NGB)
Tom Hazekamp, IPGRI
Javier de la Torre, IPGRI/Global Biodiversity Information Facility (GBIF)/Botanical Garden of Berlin (BGBM)

Mid-Year Report
As part of its contribution to the Generation Challenge Programme, IPGRI has committed to recruit a Senior Bioinformatics Scientist based at IPGRI Headquarters in Rome within its

newly created 'Biodiversity Informatics' project. In 2005, an informal consultation with Consortium partners was carried out in order to best identify the activities and qualifications required for this position. In the last quarter of 2005, a position announcement was released to journals such as Nature and New Scientist. More than 90 applications were received by the end of October 2005. A selection panel has been created with the support of AGROPOLIS, an external Institution. Interviews were held in early January 2006 and by the end of February 2006, the selection panel had identified a series of candidates suitable for the position.

While the selection process for the Senior Bioinformatics Scientist is still ongoing, it was decided to recruit additional informatics expertise in the form of short-term contracts to address specific needs, particularly in the following fields:

a. development of web services technology
b. deployment of web services at various GCP locations
c. help-desk support to GCP partners
d. deployment of geographic information system (GIS) web services technology (Open Geospatial Consortium [OGC] compliant)

The following experts were recruited:

a. Dag Terje Endresen from the Nordic Genebank (NGB)
b. Javier de la Torre from the Botanical Garden of Berlin (BGBM) in association with the Global Biodiversity Information Facility (GBIF)
c. Tom Hazekamp (consultant)

The main activities carried out by these experts focused on evaluating appropriate technologies for inter-operability and their integration into the GCP informatics platform. The following technologies were considered:

a. database interoperability using BioMOBY and GBIF-compliant protocols (e.g. BioCASE and TAPIR) with released GCP data models
b. GIS interoperability using OGC-compliant protocols

The outcomes of these activities are as follows:

a. A BioCASE wrapper was installed in GCP partner locations and at non-GCP organizations such as the International Livestock Research Institute (ILRI), the United States Department of Agriculture Germplasm Resources Information Network (USDA-GRIN), the Malaysian Agricultural Research and Development Institute (MARDI), The World Vegetable Center (AVRDC) and the Food and Agriculture Organization of the United Nations (FAO). A help desk has been provided to partners.
b. The TAPIR wrapper was tested against GCP data models for passport data with SINGER and EURISCO data sets. Modifications to the latest release have been documented and the changes have taken effect as of December 2005.
c. Integration of BioCASE data providers within the GCP Central Repository and routines for automatic replication have been developed.
d. Essential software requirements for the integration of TAPIR and BioMOBY protocols have been identified. Routines for mapping GCP passport data model from TAPIR to BioMOBY have been tested.
e. A GIS OGC-compliant server (http://gis.grinfo.net) has been developed; basic routines for GIS web-enabling using Google Map through Web Features Services (WFS) and Web Mapping Services (WMS) have been evaluated.

Tangible outputs delivered:

a. A BioCASE wrapper has been deployed at all GCP locations (where required, with the exception of CIMMYT) as of April 2006.
b. GCP BioCASE providers are activated and accessible online (providing access to passport data).
c. The various BioCASE data provider entry points are registered at the GCP Central Registry. Procedures for data-harvesting have been developed and are in the process of being integrated into the Central Registry.
d. The TAPIR protocol was finalized by GBIF/BGBM and is ready to be integrated with BioMOBY.
e. An OGC GIS-compliant server is deployed and ready to provide access to GCP geo-referenced datasets provided through BioCASE.

Deviations from the work plan:

The project is proceeding as planned.

2006-25: Creation of Institutional Bioinformatics Capacity (IRRI)

Principal Investigator:
Graham McLaren, IRRI
Collaborating Scientists:
Richard Bruskiewich, IRRI
Thomas Metz, IRRI

Mid-Year Report

One NRS staff position is funded from this project to improve critical mass in research informatics at IRRI.

One NRS staff member traveled from IRRI to CIMMYT in May 2006 to attend the ICIS workshop and work with colleagues at CIMMYT on common solutions to the management of genotyping data in cereals.

Tangible outputs delivered:

Prototype ICIS implementation of genotyping database with CIMMYT wheat genotyping data.

Deviations from the work plan:

None.

2006-34: Installation and Implementation of the ICRISAT LIMS at the Biosciences Eastern and Central Africa (BecA) Facility and IITA-Ibadan

Principal Investigator:
Dave Hoisington, ICRISAT
Collaborating Scientists:
Jayashree B, ICRISAT
Etienne de Villiers, ILRI
ME Ferguson, IITA
Sarah Hearne, IITA
Santie de Villiers, ICRISAT
Rosemary Mutegi, ICRISAT
Dan Kiambi, ICRISAT

Mid-Year Report

An initial teleconference/demo was held in February 2006 to present the basic philosophy and functionality of LIMS to ILRI, IITA and ICRISAT staff based in BecA. Subsequent to this, ICRISAT staff based in BecA (Dan Kiambi and Santie de Villiers) received further training on LIMS while in Patancheru, India for various ICRISAT meetings. Dan and Santie then presented an in-depth demo of the LIMS to ILRI and IITA staff in BecA in March 2006. During this presentation a list of required versus desired modifications/additions were developed. These have been reviewed by the bioinformatics team at ICRISAT-Patancheru and discussed during the Platform developer's workshop in Pretoria (20-24 March 2006). The list indicates: (i) functions that are not available in the LIMS but which staff at BecA would like to use; (ii) modules that need to be re-designed; and (iii) protocols and methods that need to be incorporated that are not available in the present application. Staff from IITA, ICRISAT and ILRI (Nairobi) indicated the need for redesign in some modules especially with regard to setting up PCR experiments, and some modifications in other sample tracking components.

Currently, the LIMS application is being ported to the open-source postgreSQL database, and the implementation of modifications requested is in progress. Additional and Modified modules were demonstrated to ICRISAT staff from Nairobi (Santie de Villiers, during her visit in June 2006) who, after discussions with colleagues at Nairobi has indicated the need to integrate some additional features to improve use of the application at the Nairobi facility.

Plans are in progress to complete installation of the LIMS at BecA in August and provide hands-on training in its use and support to ILRI, IITA and ICRISAT staff soon after.

Tangible outputs delivered:

- Working version of the application provided to ILRI/IITA/ICRISAT in Nairobi.
- Support for its installation and use with the free MS-SQL server Express edition.
- List of required modifications and additions.

Deviations from the work plan:

None.

2006-35: Data Analysis Support for Existing Projects in SP1 with Emphasis on Sampling Germplasm (DASSP1)

Principal Investigator:

Marco Bink, WUR

Co-Principal Investigators:

Hans Jansen, WUR
Fred van Eeuwijk, WUR
Xavier Perrier, CIRAD

Collaborating Scientists:

Paula Hurtado, CIAT
Claire Billot, CIRAD
Reinhard Simon, CIP

Mid-Year Report

PART I – Support in germplasm sampling and statistical association analysis.

Phase I (deadline: July 1, 2006)

Set up default protocol for complete analyses of SP1 project, in close collaboration with scientists of two SP1 projects. Identify complications/pittfalls and possibly provide guidelines for an

advanced protocol. These experiences will be linked to the training materials for association analysis (SP5 project).

Phase II (deadline: December 1, 2006)
Describe and finetune default protocol for complete analyses and make available through GCP website.

PART II – Decision Support System for Germplasm Sampling

1. Additional developments in connection to sampling and association analysis.
1.1 Hypervariability of SSR markers
SSR markers, used in almost all SP1 projects for association studies, are characterised by a frequent hypervariability with a large number of alleles and consequently frequent very low allelic frequencies. These low frequencies are a real problem in linkage disequilibrium estimation.

It is assumed that the allelic diversity is the result of (i) the genetic structure of the population due to demographic and breeding events, and (ii) secondary and recent mutational events following a stepwise mutation model (gain or loss of one repeat unit at each mutation). For disequilibria due to structures, the first level is the most informative and the second level can be considered in part as a noise. So the reduction of allele number will be achieved in pooling (grouping) alleles that differ by a small number of repeats, assuming that they result of a recent stepwise process. Several strategies will be explored to pool alleles. The first one is based on aggregation on statistical kernels.

For that the first work was to define the nature of these kernels and the distribution of the number of repeats under a strict stepwise mutation model has to be specified.

Distribution of SSR repeat number under a stepwise model
We consider that SSR undergo at each generation a stepwise mutational model (SMM) (Ohta and Kimura 1973) assuming equal probabilities of gain or loss of a repeat (the symmetric single-step SMM)

Then for the first generation, a locus can undergo or not a mutation, and, if a mutation occurs, it can be a positive mutation (gain) with a probability of 0.5 or a negative mutation (loss) with a probability of 0.5. Then if m is the initial number of repeats, the resulting number of repeats is m-1, m or m+1.

After a second generation the resulting number of repeats is m-2, m-1, m, m+1 or m+2, and so on for following generations.

We are looking for the distribution of the number of repeats in a population descending after t generations of a common ancestor with m repeats. We have to compute the probability P(n) that a taxon differs from the ancestral form for n repeats ($n = 0, \pm 1, \pm 2, \cdots$) after t generations.
This stochastic process is a Markov chain whose state space is given by the integers $n = 0, \pm 1, \pm 2, \cdots$. It can also be looked at as a random walk for a walker on a straight line who at each point of time either takes one step to the left (loss of a repeat), stays at the same place (no mutation) or takes one step to the right (gain of a repeat).

Assuming symmetric model, the distribution is symmetrical and P(-n) = P(n), so we will focus only on positive values of n.

The net increase of n repeats can occur by having:

- n gains that increased the length of one repeat
- n+1 gains and 1 loss, for a total of n+2 mutations
- n+2 gains and 2 losses, for a total of n+4 mutations
and more generally:
- n+k gains and k losses, for a total of n+2k mutations

Thus the probability P(n) is the sum over all values of k of the probability of having n + 2k mutations, times the probability that, out of n+2k mutations, n+k are gains and k are losses.

$$P(n) = \sum_{k=0,1,\ldots} P(n+2k \text{ mutations}) \times P(n+k \text{ gains}/\, n+2k \text{ mutations})$$

Under classical assumptions on mutational process, the probability of n+2k total mutations after t generations is given by a Poisson distribution with parameter μt (assuming a per-generation mutation rate of μ).

$$P(n+2k \text{ mutations}) = e^{-\mu t} \frac{(\mu t)^{n+2k}}{(n+2k)!}$$

The probability of n+k gains (and k losses) out of n+2k mutations is given by a Binomial distribution

$$P(n+k \text{ gains}/\, n+2k \text{ mutations}) = \binom{n+2k}{n+k}\left(\frac{1}{2}\right)^{n+k}\left(\frac{1}{2}\right)^{k} = \frac{(n+2k)!}{(n+k)!(k)!}\left(\frac{1}{2}\right)^{n+2k}$$

Then, P(n) the probability of a net change of n repeats after t generations with mutation rate μ,

$$P(n) = \sum_{k=0,1,\ldots} e^{-\mu t} \frac{(\mu t)^{n+2k}}{(n+2k)!} \frac{(n+2k)!}{(n+k)!(k)!}\left(\frac{1}{2}\right)^{n+2k}$$

$$= \sum_{k=0,1,\ldots} e^{-\mu t}\left(\frac{\mu t}{2}\right)^{n+2k} \frac{1}{(n+k)!(k)!}$$

$$= e^{-\mu t}\left(\frac{\mu t}{2}\right)^{n} \sum_{k=0,1,\ldots}\left(\frac{\mu t}{2}\right)^{2k} \frac{1}{(n+k)!(k)!}$$

The total number of mutation cannot be greater than the number of generations:

$$n + 2k \le t \Rightarrow k \le (t\text{-}n)/2$$

However if assuming that all terms in the sum for k greater than (t-n)/2 are negligible, then, an approximation of P(n) is:

$$P(n) \approx e^{-\mu t}\left(\frac{\mu t}{2}\right)^{n} \sum_{k=0}^{\infty}\left(\frac{\mu t}{2}\right)^{2k} \frac{1}{(n+k)!(k)!} = e^{-\mu t}\left(\frac{\mu t}{2}\right)^{n} I_n(\mu t)$$

where In(μt) is the n-order modified type 1 Bessel function of μt.

Validation
This result has to be confirmed and conditions of validity of the approximations have to be specified.

For that we have written a simulator that generates a stepwise process with μ and t as parameters. This simulator gives us the exact distribution. In parallel, the approximation can be computed using the Bessel function available in Mathematica or Excel.

Further developments
We plan now to derive from properties of the Bessel function an approximation of this distribution by a Gaussian.

Then the observed distribution of alleles for a locus will be looked at as a mixture of Gaussian distributions on some kernels K. Their number, their position and their range have to be estimated in maximising the likelihood of the observed data, probably using a Expectation-Minimisation

algorithm. When the kernels will be known in position and range, each allele will be assigned in probability to a kernel and it will be pooled to the kernel of maximal probability.

Tangible outputs delivered:
Part I: The output will be delivered in the next period.

Part II: The latest version of the DARwin software has been made available through the CIRAD website (ftp://consult-darwin@auvergne.cirad.fr/).

Deviations from the work plan:
Part I: The work progress is behind schedule, as priority was given to the associated SP5 project on development of training material (we were contacted by C. deVicente to participate in a workshop to be organised in summer 2006).. In the next period the arrears in deliverables will be made up.

Part II: none.

SP5 COMMISSIONED GRANTS

2005-CB01: Training Materials for Molecular Genetic Diversity Analysis

Principal Investigator:
Christian Poisson, Cirad
Collaborating Scientists:
Carmen de Vicente, IPGRI
T.Fulton, Cornell

Mid-Year Report
A learning module in the field of genetic diversity of global genetic resources is being developed and was planned to be finished at the end of 2005. Unfortunately most of the Cirad scientists implicated in this project have been strongly mobilised in other CGP projects. On the other hand, we received very late, at the end of 2005, some contributions from GCP partners to complete the training materials and only few of them were usable in the frame of the topic of the genetic diversity analysis. So a no-cost extension was requested in order to finalise this project up to 15 may 2006.

From November 2005 to may 2006 new support notes have been finished, about DNA libraries (BAC, cDNA.), supports for laboratory practice (DNA extraction, PCR amplification.), population genetics and structure and other support notes are on going, but the project is not achieved at the moment. A review of the situation is planned on 8th of June with all the scientists implicated and a new no-cost extension will be probably requested.

Tangible outputs delivered:
See excel file.

Deviations from the work plan:
The scientists in charge of the realisation of these support notes have been heavily overloaded and have very much difficulties in finding the consecutive time necessary to complete the work. They probably also underestimated the time they needed for the realisation of such a work.

We have recently associated a new partner from INRA in order to finalise as quickly as possible these support notes. Modalities will be discussed during the 8th June meeting.

2005-CB02: Development of Training Materials for a Course in Genomics and Comparative Genomics, and Design of Course Curriculum

Principal Investigator:
Theresa 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 used either as a self-tutorial or as the basis for a course of approximately 2 weeks duration; includes definitions of terms, illustrations of concepts, photographs, real-life examples, appropriate applications, lists of key references, and other items as appropriate.

Target Outputs: Training materials and curriculum, Powerpoint presentation format.

Tangible outputs delivered:

A first draft of most of the slides, not yet finalised, is available on http://www.igd.cornell.edu/Comparative%20Genomics/Comparative%20Genomics%20Proj.html (with frequent updates as I finish more slides). Currently the slides are organised into 4 sections:
1. Background, importance, history: includes a brief history of how genomics began, the potential benefits, background concepts such as DNA sequencing, definition of the terms genomics and comparative genomics
2. Basic concepts: An overview of the basic concepts needed to understand genomics, such as genome structure, EST development, methods of comparing genomes, defining synteny, etc.
3. Methods and tools: Databases, software used in sequence retrieval and alignment, use of ESTs, gene expression and microarrays, computational and statistical methods
4. Applications and findings: Use of model plants, lessons learned, examples, limitations

In addition there is a glossary and lists of Resources and tools publicly available

The attached Appendix gives the titles of all the slides at this time.

Deviations from the work plan:

There have been no major deviations from the work plan, with the exception that it has been going slower than expected. In addition, I have had some difficulty getting feedback, so while I am still working on the slides, I would like to ask that they go out for review anyways so that I can get some comments and feedback while I am finishing them.

2005-CB04: Online Bioinformatics Course

Principal Investigator:
Richard Bruskiewich, IRRI
Collaborating Scientist:
Guy Davenport, CIMMYT

Mid-Year Report

The Principal (R. Bruskiewich) is currently working with CRIL bioinformatics staff to compose the required 2 week online GCP crop bioinformatics course over the course. The course will be prototyped first on the GCP Wiki (http://cropwiki.irri.org/gcp/index.php/GCP_Online_Bioinformatics_Course) until ready for publication as a regular web site.

2005-CB05: Publication and Dissemination of Microsatellite Kits for Crops Genotyped in the First Tier

Principal Investigator:
Carmen de Vicente, IPGRI
Collaborating Scientists:
Ken McNally, IRRI
Claire Billot, Agropolis-Cirad
Martin Fregene, CIAT
Morag Ferguson, IITA
Matthew Blair, CIAT
Marc Ghislain, CIP
Marilyn Warburton, CIMMYT
Zhang Jing, CAAS
David Hoisington, ICRISAT

Mid-Year Report

The following table summarises the progress of communications, submission of progress reports and submission of data at the time of finalising this report.

REFERENCE MICROSATELLITE KITS - CURRENT STATUS
Last updated: 13/06/06

Crop	Responsible person	No-cost extensions approved	Reports May 2006	Report approved?	Last communication	Current status
Rice	Ken McNally	15-Oct-06	Report received on 5/15/2006	approved	5/15/2006	No cost extension approved
Musa	Morag Ferguson	-	Report received on 5/12/2006	not approved	7/7/2006	no data received
Cassava	Martin Fregene	-	Report received in 5/15/2006	approved	5/15/2006	Complete data received
Bean	Mathew Blair	-	Report received on 5/27/2006	approved	5/27/2006	Complete data received
Maize	Marilyn Warburton	-	Final report received in January 2006	approved	1/1/2006	Complete data received
Wheat	Marilyn Warburton	-	Final report received in January 2006	approved	1/1/2006	Complete data received
Potato	Marc Ghislain	-	Final report received in December 2005	approved	12/1/2005	Complete data received
Sorghum	Claire Billot	-	Report received on 6/1/2006	approved	6/1/2006	Complete data received
Barley	Zhang Jing	-	Report received on 4/28/2006	not approved	5/25/2006	no data received
Chickpea	David Hoisington	-	Report received on 6/9/2006	approved	6/9/2006	Complete data received

Complete data received:
Maize, wheat, potato, cassava, common bean, sorghum, chickpea

Progress reports received but data not provided in the appropriate templates:
Musa

Tangible outputs delivered:
Green highlighted cells in above table

Deviations from the work plan:
It is impossible to predict progress or give an accurate estimation of a likely delivery date for barley and *Musa*.

The next step is to begin preparing the publication of the data sets already available once there is agreement with GCP on the format and content. Communication regarding the two missing crops will continue. It is hoped that the research team that have reported the completion of activities will provide the data on their crops in a timely manner and in the appropriate templates.

As a result of the unavailability of data for some crops, which has delayed publication of the results, a no-cost extension of this project is requested until 31 December 2006.

2005-CB08: Functional Genomics to Improve African Crops

Principal Investigator:
Roeland van Ham, Wageningen University

Collaborating Scientists:
Dave Berger, University of Pretoria
Jan-Peter Nap, Wageningen University

Mid-Year Report
Due to the decision of Dr. Oscar Vorst to change his carreer path, the execution of the project was seriously delayed. In agreement with all involved, execution and budget was moved to the summer period of 2006 (July – Sept) and prof. dr. iJan-Peter Nap (PRI) will replace Dr. Vorst. In close consultation with the Pretoria group, a preliminary scheme for the workshop was agreed upon and is being finalised. See appendix.

Tangible outputs delivered:
Not applicable yet.

Deviations from the work plan:
In addition to the change in dates, also the content of the project is adjusted to current needs and desires, in combination with the expertise and interests of the visiting scientist. The workshop will cover less high-level bioinformatics, but focus more on the needs and home possibilities of the African participants. The research will focus less on microarrays and more on the possibilities of future collaboration between Wageningen and South-Africa (and possibly other workshop participants) in the 'omics' characterisation of target and/or future crops.

2005-CB09: Molecular Markers for Allele Mining: A Workshop

Principal Investigator:
M. Carmen de Vicente, IPGRI
Collaborating Scienitst:
J.C. Glaszmann, CIRAD

Mid-Year Report
Contributions to the publication of a proceedings document were requested and received. Abstracts were copy-edited by a professional English editor. The final document was adapted to the IPGRI Style Guide v 2.1 (Oct. 2005) and submitted to IPGRI Headquarters for clearance on 24 March; a request for publication has also been submitted to IPGRI Headquarters.

Tangible outputs delivered:
Proceedings of the workshop

Deviations from the work plan:
Publication of the proceedings has been delayed as a result of internal clearance procedures.

2005-CB13: The Institute for Genomic Diversity's Interactive Resource Centre

Principal Investigator:
Theresa M. Fulton, Cornell University

Mid-Year Report
New items added to the Resource Centre include 2 tutorials, one for the popular software programme Mapmaker (Lander et al.) and one for the QTL programme QGene (Clare Nelson). These tutorials explain the concepts of the programmes and demonstrate basic usage. They include screenshots so can be used as stand-alone information, but downloadable data files are included as well as links to the programmes themselves (they are both freely available) so one can

work through the tutorials while actually using the programmes if desired. I have used both these tutorials in classes and received very positive feedback.

Recently, together with a colleague at IGD, Dr. Alexandra Casa, I successfully received an additional grant from the Syngenta Foundation to add a new section on Sorghum and Millet Resources. This project just started (April, 2006) but already available are several resources that should be useful to scientists working on sorghum and millet. These include:

- Literature: a list of key references, organised by section (marker-assisted selection, comparative genomics, etc.). Articles publicly available include a downloadable pdf; more than 30 of these are available. Articles still under subscription restrictions are linked to the abstract.
- Maps: Links to the publicly available maps of sorghum.
- Data: Links to available data. In addition, Dr. Sharon Mitchell of IGD has compiled an excel file of freely available microsatellite (SSR) data, which include contains primer sequences and genetic information for more than 200 mapped *Sorghum bicolor* SSRs. This file is available for downloading.

Next, we hope to develop tutorials for sorghum-related databases and map information. Sorghum is one of the most important crops world-wide, in particular developing countries, but the data and information available to researchers is extremely disjointed and hard to navigate. We hope to make it more accessible and truly usable.

In addition, a "Queries" page has been added as a bulletin board for questions from users, to users. Databases, protocols and other links are added as I find them. Questions to the helpdesk and the number of hits to the various pages show that the most widely used information is still literature, protocols, and the Molecular Marker training modules.

Next priorities include a tutorial for "Using databases" (much requested) and information on buying supplies and reagents for molecular biology laboratory work. Regarding this last item, discussions have been made with Dr. Jedidah Danson and others at Ammanet who are most keenly interested in this resource and will be willing to advise.

As the budget for this project was cut this year, I have not been able to hire website assistance as hoped. I would very much like suggestions or constructive criticisms about the design and structure of the pages, as I feel this is my weakest point. I have attempted to make the pages very simple and easy to navigate, while using the basic color strategy of the GCP website, but without adding pictures or things that could make it difficult for my users to load (as many are still using dial-up internet services).

Tangible outputs delivered:
Outputs (the website) are available at http://irc.igd.cornell.edu
In particular, please see: "Tutorials" and "Sorghum and Millet Resources" for newest items, in particular http://irc.igd.cornell.edu/SorghumMillet/SorgMillLiterature.html for available references

Deviations from the work plan:
None.

2005-CB15: Distant Policies: A distance learning module for scientists on genetic resource policies and their implications for freedom-to-operate

Principal Investigator:
Niels Louwaars, WUR

Collaborating Scientists:
Victoria Henson-Apollonio, CAS-IPGRI
Zeze Sampaio, Embrapa

Mid-Year Report
The text for the module was finalised March 2006 and currently (Late April) the digital environment ia adapted to the needs of this particular course and the text is introduced to it in such a way as to create an interactive learning environment. Test run for the course is planned for June 2006.

A start has been made with the preparation of the additional module on Freedom to Operate measures by individual scientists. This module will use the experiences obtained from the current (4 week) course.

Tangible outputs delivered:
Presentation of the course environment as presented during the annual conference 2005.
Text of the current course will be presented during the test run.

Deviations from the work plan:
There is some delay in the preparation of the text due to a 6 months stay of the PI in Rome. Apart from the inclusion of his experiences during that stage (which should be considered positive), it has resulted in a small delay in the finalisation of the first module.istant learning

2005-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
Collaborating Scientists:
M.J. Sampaio, EMBRAPA
M. Blakeney, University of London

Mid-Year Report
1) The intellectual property rights (IPR) helpdesk, "IP Matters", is online. The webpage is hosted by the GCP sever, located at CIMMYT; it was made available in April 2006.
2) A broad range of materials is available from the site and its web links.
3) In 2006, a helpdesk expert panel will be coordinated from the CAS office at IPGRI.*
4) Links to the GCP/WUR distance learning site and the GCP asset inventory site will be functional when these sites are in operation.
5) Additional materials will be by available on a monthly basis.

Tangible outputs delivered:
A group of experts has been assembled, a webpage has been launched, and an initial list of materials has been uploaded onto the site.
The URL for the web site is: www.generationcp.org/iphelpdesk.php?da=0629604

Deviations from the work plan:
* Original plans for the helpdesk expert panel, coordinated by Maria Mendoza from IFPRI, have been revised because Maria has needed to reduce her workload for personal reasons.
Development of a full range of materials to be added to the website, and coordination of the interactive question and answer section has yet to be implemented. An additional legal specialist hired by CAS, will be working with this project starting in Oct 2006.

2005-CB17: Reporting for Product Distribution: An asset inventory system for the Generation Challenge Programme

Principal Investigator:
Victoria Henson-Apollonio, IPGRI
Collaborating Scientist:
B.H. Rao, ICRISAT

Mid-Year Report

1) In this project, a set of Internet-ready forms and a database to house the information in these forms has been created in 2006.
2) A test site was launched on the Internet in April 2006.
3) Testing should be completed by August 2006 and the site will be available for use by GCP scientists at that time.

Tangible outputs delivered:
Draft templates, an Internet webpage and a preliminary database have been created. The test site URL is: www.gcpais.icrisat.org.

Deviations from the work plan:
The project is proceeding as planned.

2005-CB23: Genotyping Support Service

Principal Investigator:
M. Carmen de Vicente, IPGRI
Collaborating Institutions:
Coconut Research Institute, Sri Lanka
Vanuatu Agriculture Research and Technical Centre (VARTC), Espiritu Santo Island, Vanuatu
Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), France
DArT P/L, Australia
International Coconut Genetic Resources Network (COGENT), IPGRI
University of Southern Mindanao, Philippines)
Centro Nacional de Pesquisa de Mandioca e Fruticultura Tropical (CNPMF, Embrapa) Brazil
Namulonge Agricultural and Animal Production Research Institute (NAARI), Uganda
National Root Crops Research Institute (NRCRI), Nigeria
Crop Research Institute (CRI), Ghana
Centro Internacional de Agricultura Tropical (CIAT), Colombia and Nigeria
Instituto Nacional de Investigación Agropecuaria (INIA), Uruguay
Instituto de Investigaciones Agropecuarias (INIA), Chile)
Centro Internacional de la Papa (CIP), Peru

Mid-Year Report

The Generation Challenge Programme (GCP) launched the Genotyping Support Service (GSS). The GSS is intended to offer National Programmes, with ongoing efficient plant breeding programmes, opportunities to access molecular markers to improve their work. Hopefully, the expected results will directly translate into successful stories for the participants.

During the first year, started in May 2006, the GSS was launched to address Cassava, Potato, Groundnut, Coconut and Musa, and it was meant to reach and address national breeding programmes in Africa, Asia and Latin America.

Contacts with the respective GCP crop representatives have been made, with the following objectives:

- Introducing the GSS,
- Informing on putative GSS National Programme partners and requesting feedback on choices,
- Inviting other suggestions of additional candidates.

Also, contact was made with GCP genotyping laboratories designated for each crop to learn about their willingness, availability and timing to participate in the service.

Candidate National Programme partners were also contacted, with a description of the GSS and their nomination to participate while requesting their expression of interest. At the moment of this report, eighteen (18) National Programmes have been contacted and seventeen (17) replied with a positive response. An application form was circulated requesting details on the type of germplasm, traits phenotyped, capacity for DNA extraction, and training needs.

As of July 26th, applications from Bangladesh (groundnut), Nigeria (cassava), India (Musa) and Uganda (cassava) have been received. More are expected within the coming days.

Tangible outputs delivered:
None.

Deviations from the work plan:
The project was delayed 5 months, but recovering time as fast as possible.

2006-09: Training Course on Phenotyping

Principal Investigator:
François Tardieu INRA
Collaborating Scientists:
Graeme Hammer, U. Queensland
Greg Mc Lean, APSRU
Delphine Luquet, CIRAD
Rachid Serraj, IRRI
Vincent Vadez, ICRISAT
Claude Welcker, INRA

Mid-Year Report
Phenotyping for complex traits such as drought tolerance involves methodological and technical choices which will eventually determine the quality of the whole phenotyping process. The training course aimed at providing participants with theoretical and practical elements for these choices. Twenty three scientists (level : PhD or master) participated to the course, 6 from Africa, 6 from south Asia, 4 from east Asia, four from Latin America and 3 from Europe. The latter were participants to ongoing GCP projects which received no financial support from the GCP.

The course lasted 8 working days (from 3 to 12 July 2006) and the time was split into three parts of similar durations (i) lectures, aimed at providing and discussing the theoretical frameworks, (ii) practical sessions in the greenhouse or in the field aimed at a direct contact with the tools involved in phenotyping, (iii) computer sessions in which each participant could work on practical examples and directly test methods of analysis and reasoning. The course consisted in two sessions :

Characterising the plant environment and the stress.
- *Lectures aimed at presenting the physical bases* necessary for measuring micrometeorological conditions and soil water characteristics, plus those for using techniques based on energy balance such as canopy temperature.

- *A practical experience of measurements* of micrometeorological conditions, soil water status and plant water status was provided during sessions in the greenhouse and in the field. In each of them, 3 groups of 8 scientists participated to the following activities : (i) measurements of stomatal conductance and photosynthesis, sampling protocols, (ii) plant temperature, leaf temperature and evaporative demand: sensors and methods, (iii) measurements of soil water balance and transpiration in the greenhouse, (iv) plant water status (field), each participant measured a leaf water potential, (v) soil water status (field), methods and spatial variability, (vi,) using thermography to identify differences in stomatal conductance between treatments and genotypes.
- *Computer sessions* allowed each participant to get familiar with (i) the calculation of thermal time for characterising the plant development in different genotypes and locations, (ii) direct analysis of meteorological data (plant temperature, air temperature, vapour pressure deficit), (iii) calculation of soil water balance and the concepts of soil available water and relative transpiration.

Assessing phenotypes : trait evaluation, assessment of the behaviour of genotypes with contrasting traits under different environmental conditions.
- *Lectures provided the framework of analysis* allowing one to place individual traits in plant development and their contribution to yield formation. This allows one to weigh their roles, according to soil characteristics and climatic scenarios. Courses specifically addressed transpiration efficiency and the assessment of plant development in the analysis of genetic differences.
- *A field practical session presented methods* (i) to evaluate and/or infer phonological stages and leaf establishment in multiple field trials (ii) to measure or infer light interception in a large number of genotypes, (iii) to indirectly assess plant status in an experiment with hundred of genotypes by using probe genotypes.
- *Practical sessions were dedicated* to (i) calculations of "hidden" traits from phenotypic measurements (ii) statistical methods for computing data with a practical approach, from descriptive statistics to curve fitting, (iii) experimental designs and protocols.
- *Two days have been dedicated to a first experience of modelling the behaviour of genotypes with contrasting traits under a large range of climatic scenario*, via a computer session using the APSIM model. The main objective was to guide the intuition about the role of the climatic scenario on the respective importances of traits, and on the ranking of genotypes.

Tangible outputs delivered:
- Booklet of the course, will be posted on the GCP website
- CD rom with all power point presentations, conclusions of the course and spreadsheets for calculating environmental data and soil water balance.
- Six A4 pages presenting techniques (measurements of plant temperature, of evaporative demand, of soil water content, of intercepted radiation, calculation of evapotranspiration). These pages will be posted on the GCP website

Deviations from the work plan:
None.

2006-11: Establishment of training materials for a course in association study/linkage disequilibrium mapping (TM_AS)

Principal Investigator:
Marco Bink, WUR
Collaborating Scientists:
Hans Jansen, WUR
Fred van Eeuiwijk, WUR
Marja Thijssen, WUR
Ed Buckler, Cornell

Ian Mackay, NIAB
Claire Billot, CIRAD

Mid-Year Report
Phase 1 (deadline 1 May 2006):
1. Collaborating scientists have been approached for available training materials. Until sofar, only one (I. Mackay, NIAB) responded by sending lecture notes. However, these lecture notes provide a good basis for the development of a comprehensive set of training materials for association/linkage disequilibrium mapping.
2. A search on the internet has been carried out to find presentations with materials on association/linkage disequilibrium mapping. Many of these presentations are related to problems in human genetics.
3. A literature search on on association/linkage disequilibrium mapping has been carried out. A list of references to relevant literature will be part of the training materials. The list will be subdivided into different topics of association/linkage disequilibrium mapping, e.g. measurement of linkage disequilibrium, haplotype reconstruction, population structure, false discovery rate and prospects of association/disequilibrium mapping in different practical settings.

Tangible outputs delivered:
The first/final output will be delivered in the next period.

Deviations from the work plan:
The work progress is slightly behind schedule. In the next period the arrears in deliverables will be made up.

2006-15: Fellowships and Travel Grants

Principal Investigator:
Carmen de Vicente, IPRGRI

Mid-Year Report
The Generation Fellowship awards were established to facilitate innovative research related to the central theme of the Generation Challenge Programme, i.e. unlocking genetic diversity of crops for the resource-poor. The Fellowship programme is aimed at scientists who want to conduct research outside of their home countries/institutions for a period of three months to one year. The Generation Fellowship places primary emphasis on research in the four thematic subprogrammes: 1) genetic diversity of global genetic resources, 2) comparative genomics for gene discovery, 3) trait capture for crop improvement, and 4) genetic resources, genomic, and crop information systems. Up to 8 fellowships per year are awarded, and the maximum award per fellow will be up to US$25,000, which is intended to cover travel, living expenses, laboratory consumables, and conference participation.

Applications are invited from crop science researchers from developing country research institutions (National Agricultural Research Systems), who hold at least a Master of Science degree (MSc), or equivalent, in a relevant subject area. Applicants should demonstrate they are engaged in a related ongoing research activity in their home country. Priority is given to scientists from National Agricultural Research Systems already involved in GCP research projects. For more information on the Fellowships programme and to access the application materials, please see the GCP Capacity Building Corner:
http://www.generationcp.org/capcorner.php?da=0531908.

The Generation Travel Grant Programme is a key component of the GCP Capacity Building Subprogramme (SP5). Sixteen Travel Grants are available per year to cover the expenses of

developing country National Programme scientists working at or in collaboration with a GCP Consortium Institution (list of GCP Consortium Institutions: http://www.generationcp.org/consortiummembers.php). The purpose of the Travel Grant Programme is to encourage and promote collaboration between the GCP and NARS institutions, foster linkages within current GCP projects, and provide training opportunities for developing country scientists. The grant may be requested to visit a GCP Consortium Institution or any other Advanced Research Institution to get training in concepts and/or techniques necessary for the advancement of the GCP research (first priority), to participate in any training event organised by the GCP (second priority), to participate in the annual GCP research meeting (limited number of spaces available), or to participate in conferences whose subject is relevant to the work of the GCP. Preference will be given to applicants with links to current GCP projects and for whom the travel grant will be used as a learning experience.

The maximum grant award will be $5,000 USD, which is intended to cover travel, accommodation, and conference participation, if applicable.

A Generation Travel Grant Committee evaluates all applications and selects the recipients. The deadline for travel grant applications is the 20th of each month. Winners are notified early the following month. Applications are evaluated as they are received until all 16 grants are awarded. So far in 2006, there have been 25 applicants for the first round of Fellowship Grants. Applications are currently being accepted for the second round.

There have been 94 applicants for Travel Grants as of the July 2006 round. Applications are currently being accepted for the August round.

2006-28: Regional PGR Courses

Principal Investigator:
Marja Thijssen, WUR
Collaborating Scientists:
Niels Louwaars, WUR
Victoria Henson-Apollonio, IPGRI
Zeze Sampaio, Embrapa

Mid-Year Report
During the latter part of 2005, opportunities were sought to create a regional course. This was negotiated with ICARDA and a regional course will be held in November 2006 in Iran in conjunction with a number of other PGR-related courses. The Netherland sGovernmnet will provide most of the funding for the participants and lecturers. In the mean time, the curriculum, developed with the assistance of the 2005 project will be used again in a global course in Wageningen with participants from Asia, Latin America and Africa, with inputs from all 4 collaborators.

Tangible outputs delivered:
In the first half of 2005: curriculum tested in a global course.

Deviations from the work plan:
None – everything according to plan.