Final Technical Report

Date Report Submitted: April 6, 2010

Name and institutional affiliation of person submitting: Mathias Lorieux, IRD/CIAT

Project information

GCP project number:
G3005.10 (Competitive grants 2005-2007 Project # 10)

Project title:
Exploring Natural Genetic Variation: Developing Genomic Resources and Introgression Lines for Four AA Genome Rice Relatives

Location:
Latin America, Colombia

Lead institute, Principal Investigator:
CIAT, Mathias Lorieux (IRD/CIAT) & Joe Tohme (CIAT)

Collaborators (institute[s], researcher[s]):
Susan R. McCouch, Cornell University, USA
Claudio Brondani, CNPAF-Embrapa, Brazil
Baboucarr Manneh & Marie Noelle Ndjiondjop, WARDA, Côte d'Ivoire
César P. Martinez, CIAT, Colombia
Miguel Diago Ramirez, Fedearroz, Colombia

Total project budget:
1,074,900 US$

Grant Period: (as in original project contract)
Start: 01/2005; End: 12/2008

Reporting period: (including no-cost extension period, where applicable)
Start: 01/2008; End: 12/2009

Report

I. Executive summary

Cereals provide the majority of calories consumed by humans. Cereal production faces growing challenges due to increasing human population, changing nutritional requirements and variable environmental conditions that require new approaches to crop production. Wild relatives of modern crop species have survived for millions of years using natural genetic defenses to endure biotic and abiotic aggressions. These wild relatives represent a valuable source of under-utilized genetic variation that is available to plant breeders and represent an invaluable source of genetic information for modern genomics research initiatives. A systematic approach is required to identify and characterize genes from wild species that can be used to enhance crop productivity in a range of environments and under diverse cultural conditions. Using rice as a model, we propose to (1) develop four libraries of interspecific lines called Chromosome Segment Substitution Lines (CSSLs), targeting chromosomal introgressions from different rice relatives, (2) develop a set of 140 molecular markers (called SNPs) identified in genes associated with tolerance to abiotic stress (drought, acid soils, mineral deficiencies or toxicities), (3) validate the utility of the SNPs by using them in the development of the CSSLs in this project and exploring their value in breeding programs for other cereals (4)

analyze a set of advanced CSSLs generated from Asian x African rice crosses for their phenotypic response to drought stress. Generating such resources and knowledge will contribute to the objectives of Subprograms 1 and 3 by (i) utilizing *natural genetic diversity* to develop whole-genome libraries of CSSLs as a permanent genetic resource for both breeding and genomics-based research (ii) producing high-throughput, cost-effective markers to *facilitate access to genetic diversity* in a range of different cereal species (iii) making the CSSLs available to breeders and geneticists so that the intersection of their efforts will continue to generate new knowledge.

II. Introduction/Background

The future of crop improvement depends on the availability of genetic variation. Most modern crop varieties have undergone a genetic bottleneck associated with the process of domestication resulting in a restriction of the genetic options that are available to plant breeders. There is a larger pool of genetic variation available in landraces and wild relatives of crops. These resources are known to contain many interesting traits for breeding, including good to strong tolerance to abiotic and biotic stresses and various nutritional traits of interest (Sun et al 2001). However, it is often difficult to utilize these natural sources of genetic diversity because of fertility barriers, linkage drag, the time and resources required to recover useful recombinants. This project is designed to take advantage of the unexploited reservoir that exists in the wild relatives of cultivated rice (*Oryza sativa* L.) through the development of introgression lines that will be of immediate use to breeders and will simultaneously serve to enhance our understanding of the "wild alleles" that contribute favorably to plant performance under drought stress.

In this project, we propose to develop a comprehensive toolkit of genetic and genomic resources that will allow breeders and geneticists to explore and more efficiently utilize wild relatives in crop improvement. Our approach involves the systematic introduction of foreign alleles from four different AA genome *O.* species into one elite, highly productive *O. sativa* varieties and to provide introgression lines (ILs) or sets of overlapping CSSLs as the basis for genetic analysis and applied plant breeding.

Complex phenotypes can be dissected genetically by evaluating and comparing CSSLs as a first pass. By linking the information about gene identity back to the Gramene database, we will be able to provide a dictionary of genes with known function that are contained in each of the CSSLs. This dictionary is also a key ingredient in enabling comparative approaches to the study of phenotype-genotype relationships. Once an association is established between a phenotype and a specific introgression line, it is possible to use advanced forward and reverse approaches to genetic analysis to mine down and identify the gene(s) that are directly involved (Yano, 2001; Jander et al., 2002), but it is also feasible to use the introgression line directly in a breeding program without knowing the identity of the gene(s) that are involved (Gur and Zamir, 2004).

In addition to the targeted introgression of traits that can be identified phenotypically in the wild material, such as biotic or abiotic stress tolerance, it has been demonstrated that alleles hidden in low yielding, agronomically undesirable ancestors can enhance the productivity of many of the world's most important crop varieties. These yield-enhancing alleles are the basis of 'transgressive variation' and may confer an advantage in both favorable (irrigated) and unfavorable conditions (drought and weed competition) (Moncada et al., 2000; Gur and Zamir, 2004). Thus, the use of wild and exotic germplasm for CSSLs construction carries with it the possibility that favorable transgressive segregants will be identified, providing the basis for studies aimed at understanding the genetic basis of transgressive variation associated with resistance or tolerance to drought and acid soils.

III. Activity/Output table (see Appendix A—modified)

IV. Scientific activities (including tables and figures)

Launching meeting

A meeting was held 21-24 February/05 in CNPAF-Embrapa headquarters in Goiania, Brazil. Protocols and work plan for the development of these populations were defined. Criteria for choosing the recurrent parents were agreed upon, mainly yield under stress conditions, tolerance to drought stress under field conditions, tolerance to main diseases, seedling vigour, performance in farmers fields, stay green or late senescence, and improved grain quality.

Universal Core Genetic Map of the rice genome

To facilitate the creation of easily comparable, low-resolution genetic maps with evenly distributed markers in rice (*Oryza sativa* L.), we conceived of and developed a Universal Core Genetic Map (UCGM). With this aim, we derived a set of 165 anchors, representing clusters of three microsatellite or simple sequence repeat (SSR) markers arranged into non-recombining groups. Each anchor consists of at least three, closely linked SSRs, located within a distance below the genetic resolution provided by common, segregating populations (<500 individuals). We chose anchors that were evenly distributed across the rice chromosomes, with spacing between 2 and 3.5 Mbp (except in the telomeric regions, where spacing was 1.5 Mbp). Anchor selection was performed using in silico tools and data: the O. sativa cv. Nipponbare rice genome sequence, the CHARM tool, information from the Gramene database and the OrygenesDB database. Sixteen AA-genome accessions of the Oryza genus were used to evaluate polymorphisms for the selected markers, including accessions from *O. sativa*, *O. glaberrima*, O. barthii, *O. rufipogon*, *O. glumaepatula* and *O. meridionalis*. High levels of polymorphism were found for the tested *O. sativa* x *O. glaberrima* or *O. sativa* x wild rice combinations. We developed Paddy Map, a simple data- base that is helpful in selecting optimal sets of polymorphic SSRs for any cross that involves the previously mentioned species. Validation of the UCGM was done by using it to develop three interspecific genetic maps and by comparing genetic SSR locations with their physical positions on the rice pseudomolecules. In this study, we demonstrate that the UCGM is a useful tool for the rice genetics and breeding community, especially in strategies based on interspecific hybridization.

The Paddy Map Database was created and is available at http://mapdisto.free.fr/-PaddyMap.

This work was published in: Orjuela et al (2010) A universal core genetic map for rice. *Theor Appl Genet* **120:** 563-572.

Development of CSSL populations

O. glaberrima (Caiapo x MG12)

A BC3F1 population was obtained at CIAT HQs from the cross between Caiapó (an elite tropical *japonica* variety from Brazil) and *O. glaberrima* MG12 (alias IRGC103544) (César P. Martinez). From these lines, anthers were collected and a population of 695 BC3DH lines was obtained through in vitro culture of anthers. A subset of 312 BC3DH lines was genotyped with polymorphic SSRs selected from the Universal Core Genetic Map (see above). The population was evaluated at the field level for various agronomic traits at CIAT HQs. Fourteen QTLs for plant height (3), yield (3), tillering (3), 1000-seed weight (2) and sterility (2) located on chromosomes 1, 3, 4 and 6 have been detected. One highly significant QTL for resistance to the Rice Stripe Necrosis Virus (RSNV) could be

located on Chr. 11 and fine mapping of this major QTL could be envisaged using BC4F2/F3 lines. All the analyses were performed considering 1) the ANOVA1 F-test value and 2) the graphical genotypes of lines showing extreme phenotypes for the considered trait. The QTLs positions were approximated regarding the IR64 x TOG5681 BC1F1 population developed at CIAT. In order to optimize and to purify the CSSL population, 59 BC3F1DH subsets of the 64 candidates were backcrossed to Caiapo and selfed to obtain 59 BC4F2 families. For each one of these BC4F2 families, a minimum of 60 individuals were planted at CIAT HQs with the aim of identifying plants with the target fragment. These materials (about 4,200 plants) were collected for seeds in the field and evaluated with microsatellite markers for their genetic background and for presence of the targeted *O. glaberrima* segments. The BC3DH population is available to the scientific community upon request and has been already distributed to eleven laboratories. The BC4F3 lines are also available.
Part of this work was published in: Gutierrez et al (2010) Identification of a *Rice stripe necrosis virus* resistance locus and yield component QTLs using *Oryza sativa* x *O. glaberrima* introgression lines. *BMC Plant Biol* 10: 6.

O. glaberrima (IR64 x TOG5681)

Two *O. sativa* x *O. glaberrima* sub-populations of BC3F2s (120 lines) and BC2F3s (233 lines) were obtained by IRD (M. Lorieux and A. Ghesquière) from the cross IR64 (an elite *indica* variety from IRRI) x TOG5681 (*O. glaberrima*). The lines were grown at CIAT HQs during the second semester of 2004 and the resulting BC3F4 and BC3F2 lines were analyzed using a set of 144 SSR markers.
The genotypic data were analyzed with the program CSSL Finder (M. Lorieux 2005, http://mapdisto.free.fr). The program uses a heuristic to determine the subset of lines that maximizes the representation of the donor genome and minimizes the undesired genetic background. It also displays graphical genotypes of the lines. The data were analyzed with the full set of 90 markers and, from the 353 lines evaluated, 64 were selected to be part of the introgression lines population. As the number of lines selected by the heuristic may depend on the number of markers selected for the analysis, we analyzed again the data with a subset of 75 markers, pulling out the ones that were redundant in terms of genome location. This way, 56 lines were selected. The size of the introgressions was measured based on the position of the SSR markers on the rice genome (TIGR release v. 2).
Although the majority of the *O. glaberrima* genome was conserved in the two sub-populations, it appeared that some parts of the genome of *O. glaberrima*, especially on chromosomes 3, 4, 6, 10 and 11 were lost. Anticipating this result, we developed new lines from the same cross. A total of 125 BC1F1 lines were produced and 62 of them could be successfully backcrossed to IR64 to advance to the third generation. From the graphical genotyping analyses, we could select candidate lines for backcrossing that allowed us to fill the few gaps encountered in the initial BC2F4/BC3F3.

O. meridionalis

We have been developing a series of introgression lines from 87 BC1F1 lines obtained from an interspecific backcross between the cultivated rice *O. sativa* BRSMG Curinga (an elite tropical *japonica* from Brasil) and its wild relative *O. meridionalis* accession OR44 (or W2112) from Queensland, Australia.
Foreground and background selection of 516 BC2F1 lines led to the selection of 60 lines that were subsequently backcrossed. Six seeds from each of the 60 lines were sown to generate a population of 360 BC3F1 plants. Foreground pre-selection was done at CIATto select candidate BC3F1 lines, using SSRs and MITEs markers. Graphical genotypes of

130 lines with introgression target fragment obtained using CSSL Finder software, are shown in Figure 1.

We then developed 848 doubled haploid lines from the pre-selected BC3F1 lines through anther culture at CIAT and were screened for the presence of the target fragment and genetic background at Cornell University. We observed a significant loss of target fragments in chromosome 1, 2, 3, 4, 5, 6, 7, 10 and 12, meaning that *O. meridionalis* may possess unfavourable genes for anther culture (Figure 2). Forward selection of additional 67 BC3DH lines is currently being done at CIAT. Lines were chosen to fill missing chromosome fragments in a last effort to complete the Curinga x OR44 CSSL library. Additionally, selected BC3F2 lines will be selfed for at least 4 more rounds to recover fixed alleles in some remaining missing target introgression fragments.

Figure 1. Graphical genotypes from forward selection of 130 BC3F1 lines showing introgression target fragment (brown); fragment reduction or background recovery (light pink); missing data (gray). Genotyping was done with 130 polymorphic markers (120 SSRs and 10 MITEs)

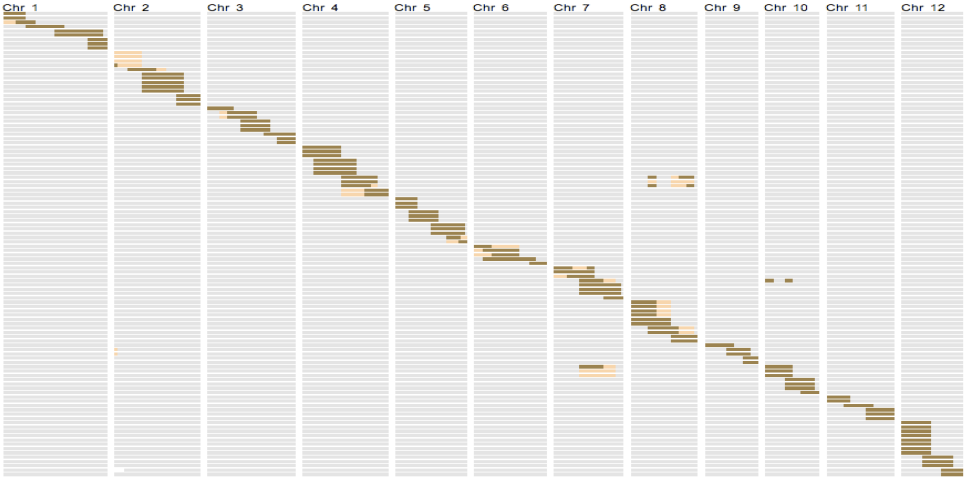

Figure 2. Graphical genotypes of 32 BC3DH lines chosen from forward and background selection of 94 BC3DH lines. Lines were selected using the CSSL Finder software based on desired introgression fragment length and maximum background recovery: introgression target fragment (red); background recovery (light pink); missing data (gray)

O. rufipogon

In a similar approach to the one used for the development interspecific lines with *O. meridionalis* as donor, we developed a series of CSSLs with the *O. rufipogon* species. The crosses and anther culture were made at CIAT, between Curinga and *O. rufipogon* acc. IRGC-105491.

In the second semester of 2009, the BC_3DH population was genotyped at for a *positive* selection with SSR markers. This type of selection allowed to check the plants that keep the desirable wild introgressed fragment. From each family, around three plants (total of 184 plants) were chosen to be negatively selected, selecting those with the smallest amount of donor genome in their background. These 184 plants were genotyped with an Illumina 384 SNP chip, designed at Cornell. 248 SNPs were polymorphic between the parents, with an average of 5 cM between them. Using CSSL Finder, 84 plants were chosen according to their fragment of interest and background recovery to be the CSSL library between Curinga and IRGC-105491. The final coverage in the BC3DH population was excellent. The population is ready for distribution to partners and we will also propose BC4F2/3 lines by the end of 2010.

O. glumaepatula

We developed a series of CSSLs with the *O. glumaepatula* species as donor. The crosses and anther culture were made at Embrapa-CNPAF, Brazil, between Curinga and *O. glumaepatula* acc. GEN1233.

153 BC3F1 plants were selected based on SSR genotyping and the BC4F2 seeds are now available. The BC4F3 families of *O. glumaepatula* x *O. sativa* cross are in the field to proceed to selfing for seed increase and further fixation.

142 BC2F2 plants were evaluated for yield-related traits at an experimental field in Porangatu, Goias, Brazil, under two conditions, one fully-irrigated and one under water stress. A QTL analysis was performed and eight QTLs were detected in both conditions, from which four were detected in the first treatment and four under water stress. A scientific paper will be submitted with the results from this experiment.

O. barthii

The cultivar Curinga, used as female parent, was crossed at CIAT with the *O. barthii* IRGC101937 accession. The F_1 plants were backcrossed to Curinga to produce the population of BC_1F_1 lines. A total of 80 BC_1F_1 seeds have been obtained and sown at two times. Among them, 64 have germinated.

A BC2F2 population was produced at CIAT and the seeds were sent to WARDA for sowing and genotyping.

214 BC2F1 lines bearing 54 *O. barthii* segments were selected among the 600 genotyped plants. The number of BC_3F_1 seeds produced per plant varied from 5 to 62. Due to germination problems with the BC3F1 seeds, we had to use alternative lines that bear the desired *O. barthii* fragment but contain a higher proportion of residual genetic background.

Genetic maps

Construction of a new IR64 x TOG5681 genetic map

The new population derived from the cross (IR64 x TOG5681) x IR64 (125 individuals) has been evaluated for 141 anchors belonging to the Universal Core Genetic Map. This allowed us to compute a new interspecific genetic map, in order to confirm and to improve the previous *O. sativa* x *O. glaberrima* interspecific genetic map (Lorieux et al 2000), to

derive graphical genotypes of the lines to monitor the derivation of BC2 lines, to validate the SSRs of the Core Map for their genetic location, to provide information about the interspecific sterility loci and to allow a direct comparison of the location of the *O. glaberrima* introgressions to the location of wild species introgressions obtained from this project.

Eight regions showing segregation distortion (SD) have been found on chromosomes 1, 2, 3, 6, 7 and 11. Of these regions, six matched with the twelve *O. sativa/O. glaberrima* sterility loci described so far: S30(t) (Li et al 2005), S29(t) (Hu et al, 2006), S19 (Taguchi et al 1999), S1 (Sano 1990), S21 (Doi et al, 1999) and S3 (Sano 1983). Two new regions with SD that have not been described for the cross were found in the long arms of chromosomes 3 and 6.

As expected, the strongest SD was found to be located at the short arm of chromosome 6, corresponding to the expected position for the locus S1.

O. meridionalis

Due to uncertainty on some of the genotypes identity in the original BC1F1 population, which could generate mistakes in the choice of the BC2 and BC3 lines, we decided to build a new *O. sativa* x *O. meridionalis* genetic map from a new BC1F1 population. This allowed us to confirm the location of a few markers and confer more robustness to the choice of the lines.

Drought stress screenings

The two interspecific *O. sativa* x *O. glaberrima* populations comprising 54 CSSLs from the cross IR64 x TOG5681 and 93 CSSLs from Caiapo x MG12 (alias IRGC103544) were subjected to drought screening at WARDA trial fields at IITA-Benin headquarters, in Cotonou, Benin (B. Manneh). The IR64 x TOG5681 population was screened in hydromorphic soil and the Caiapo x MG12 population was screened in upland soil. The trials were conducted during the dry season of 2006/2007. A split-plot design with irrigation regime as the main plot factor and genotype as the sub-plot factor was used in the trials. Within each sub-plot the genotypes were randomized using an alpha lattice design. Two irrigation levels were used for the IR64 x TOG5681 population – full irrigation up to maturity and imposing 20 days drought stress from 48 days after sowing (DAS) till maturity. For the Caiapo x MG12 population three irrigation levels were used – the two irrigation levels described above plus a third comprising intermittent drought stress imposed by irrigating every third day during the 21 days drought period. Data collected included: soil moisture content, plant height, tiller number, leaf number, length and width per tiller, leaf rolling score, leaf drying score, recovery ability, days to flower, leaf temperature, leaf greenness rating, grain yield and biomass per plant. Results from the analysis of yield data for the two populations are presented herein. Data analysis was performed using the Mixed Procedure of SAS (version 9.1). This same procedure was used to compute least squares means for grain yield of each genotype under continuous irrigation and under drought stress. Means were calculated for each genotype in both the fully irrigated and drought stressed plots. Mean separation for irrigation levels was done using the PDIFF option in the Proc Mixed model.

IR64 x TOG5681 population

Grain yield was significantly reduced by 33% on average in this population, from a mean of 2886 kg/ha to 1939 kg/ha following 21 days drought stress. Amongst the genotypes, percentage yield loss varied from 3-88% implying a wide range of drought tolerance within this CSSL population. This was also supported by the ANOVA output from SAS

which also showed a significant genotype × irrigation interaction ($p < 0.01$). However, there was significant correlation between grain yields under drought stress and without drought stress. Nine CSSLs were found to yield higher under drought stress than under continuous irrigation. For the parents, IR64 yielded higher than average under both continuous irrigation and under drought stress while TOG5681, the *O. glaberrima* parent, yielded significantly less than average under the same conditions. Under both drought stress and continuous irrigation transgressive segregation for grain yield was exhibited in this population because several CSSLs yielded higher than IR64 (the high yielding parent) under these conditions. This implies that *O. glaberrima* has contributed several genes in this cross that either alone or through epistatic effects can increase grain yield of rice in these conditions.

Caiapo x MG12 (alias IRGC103544) population

Significant genotype-by-irrigation level interaction was expressed by this population and yield (8.56 g/plant) under continuous irrigation were significantly different from those of the two stress treatments – intermittent drought stress (7.55 g/plant) and continuous drought stress (7.70 g/plant). Meanwhile yields under the two stress treatments were not significantly different from each other. In this trial, Caiapo out-yielded *O. glaberrima* (MG12) in all conditions. However, transgressive segregation for grain yield was expressed within the CSSLs in all three irrigation treatments. The *O. glaberrima* MG12 parent exhibited high sterility under all treatments and this led to the very low yields recorded for the genotype in this trial.

Laboratory setup at Cornell

We set up a small lab to train visiting scientists in G40 Emerson Hall. It will accommodate up to 2-4 scientists at a time. It was designed with the idea that scientists with different levels of background training could learn to work in a basic molecular laboratory. The lab is set up to run PCR, both horizontal and vertical agarose gels and polyacrylamide gels. Data collection/analysis mechanisms are also provided. These basic functions will allow a scientist to learn how to do marker-assisted analysis of their plants. We tried to provide equipment that was adequate to accomplish the tasks required at reasonable prices. By doing so, we are hoping that visitors will be able to see the value in purchasing certain pieces of expensive equipment and on what things they can make due with lesser quality. Ultimately we want them to be able to return to their countries able to set up a lab of their own that will fit their budgets and resources, and with the skills to make that lab produce useable results.

SNP markers development

A set of 180 SNPs were selected in rice chromosomes 1, 2, 3, 5, 6 and 7 and six PCR-multiplex experiments were designed. PCR-multiplex reactions were carried out in two sets of 30 markers, one for each chromosome. Single base extension reactions were carried out in multiplex of 13 to 19 SBE primers. No fluorescent signal was observed for 47 SNPs, eleven were monomorphic and seven were classified as no-specific since one allele could not be recognized alone and was observed always together with the alternate allele as a heterozygote. The remaining 41 SNPs (68%) were scored with high fluorescent signals and were also polymorphic, although two of them did not reproduced the polymorphism between Nipponbare and 93-11. The remaining 114 SNPs (64%) were scored with high fluorescent signals and were also polymorphic, although seven of them did not reproduced the polymorphism between Nipponbare and 93-11. For four markers, alleles could be scored only in *O. sativa* and *O. rufipogon* but not in the other species.

Several SNPs were found to be polymorphic between the parents of wild x cultivated CSSLs populations (Table 1) and the definition of sets of markers to screen for each population was initiated.

Training

Several PhD or Master students spent 3 months each year at Cornell University analyzing DNA samples: Laura Moreno from CIAT, Juan David Arbelaez from Fedearroz, Priscilla Rangel from EMBRAPA, Gustave Djedatin and Mamadou Cissoko from AfricaRice (ex-WARDA).

V. Deviations from workplan

There is no major deviation from the work plan. However, a few minor deviations need to be mentioned:

- The construction of the Universal Core Map was not planned in the original project. The construction of this tool was decided later on to facilitate the construction of CSSLs and the related activities were partially financed with external sources of funding. The results and outputs issued from this activity, we believe, represent a substantial added value to the project.
- Also, the decision was made to build a new *O. sativa* x *O. meridionalis* genetic map from a new population was taken due to the uncertainty in some genotypes identification stamps. We decided to redo the job in order to discard any even minor source of mistake in the subsequent selection of chromosomal segments.
- The same strategy was adopted for the two *O. sativa* x *O. glaberrima* populations in order to recover lost fragments and validate the genetic map.
- *O. barthii* CSSLs: 1) The project was seriously delayed during the year 2009 before of the departure of the staff in charge of the implementation of the project. The position was open and advertises and it took some time to find an adequate person. 2) We had low to no germination rate of CSSLs. It was decided to solve this issue before sowing the remaining material which is kept at -20C. A literature reviewed was made to review available rice seed germination protocols. Contact was also made with scientists familiar with the germination issue in rice. A protocol for rice seeds germination was selected. This protocol is being implemented on the parents of the CSSLs. It was also decided that the viability of seeds will be tested follow by the germination test using the new protocol.
- Overall, we estimate that we were about six months to one year late on the CSSL development activity. This was mainly due to difficulties in handling some logistical aspects, i.e., re-directing some initial activities from WARDA to CIAT, or delays related to visa issues or other logistical issues.
- Drought stress screenings: it has not been possible to carry out drought stress screenings at Fedearroz experimental station in good conditions. However, a collaboration (external to this project) with Thaura Ghneim's group (IVIC, Venezuela) that aims to screen 93 lines for various physiological parameters under drought stress from the Caiapo x MG12 cross should adequately complete the field screening, although the results cannot be considered as a GCP product.

VI. Data description and availability

Various data sets have been produced:

- Genotyping data (SSR markers) of 6 interspecific BC1F1 populations,
Format: Mapmaker/EXP compatible txt file
- Partial genotypic data of 5 BC2F2 populations
Format: CSSL Finder compatible Excel file
- Genotypic data for two advanced pre-CSSL populations from *O. sativa* x *O. glaberrima* crosses
Format: CSSL Finder compatible Excel file
- Six interspecific genetic maps aligned to the Nipponbare sequence
Format: MapDisto output txt file (Cmap-compatible format),
- Various sets of phenotypic data collected on two *O. sativa* x *O. glaberrima* ILs populations,
Format: Excel files
- Interspecific SNP locations and primers,
Format: Excel files

- SSR polymorphism for a collection of cultivated and wild rice accessions,
Format: Universal Core Map database (Excel-compatible)
- SSR genomic location of the Universal Core Genetic Map
Format: Universal Core Map database (Excel-compatible)
All of these data will be published on the GCP central database as soon as they have been valorized by publications.

VII. Conclusion

Although some deviations have occurred from the initial work plan, the main goals for this project have been achieved. As a summary:

· We developed six populations of CSSLs (Chromosome Segment Substitution Lines) that bear introgressions from the AA-genome rice species *O. glaberrima, O. barthii, O. meridionalis, O. rufipogon* and *O. glumaepatula*. These populations will constitute a valuable tool for genetic analyses and will allow us to identify key genomic regions that are associated to agronomically important traits.

· We developed a *Universal Core Genetic Map* for rice. This map has already been demonstrated as a very useful tool to help at designing introgression populations, particularly in the case of interspecific crosses. It is based on microsatellite markers that we discovered and choose with the help of several bioinformatic packages, including some that we develop at CIAT.

· A database, *Paddy Map*, was created and is available online (http://mapdisto.free.fr). This database aims to provide means to easily and quickly choose a series of genetic markers to be used to genotype a population derived from a specific cross.

· Five first-generation backcross segregating populations were genotyped and five interspecific genetic maps were developed from these data. These maps will be useful to assess the recombination rates for every wild species we use and will facilitate the localization of important genes or QTLs. All the maps we generated were based on the Universal Core Genetic Map.

· In order to fully exploit the information given by the genetic mapping analyses carried out using crosses that involve the *O. glaberrima* species, we collaborated with the Arizona Genomics Institute to develop a *library of Bacterial Artificial Chromosomes* (BAC) for this species. The library is available to the international community of plant genomicists, and it will constitute the basis of *positional cloning* approaches to identify and characterize important genes for *O. glaberrima* (this work was also supported by USAID funds).

· In parallel to this project, we designed a computer program that helps geneticists at creating CSSL populations. The program is called *CSSL Finder* at is available for download as freeware at http://mapdisto.free.fr/.

· A bioinformatic tool to facilitate the discovery of single-nucleotide polymorphisms (SNPs) was set up.

· Seven students and four research assistants were trained.

· Four students from Africa and Latin America do shuttle research between their respective centers and Cornell University.

· The international collaboration between several ARIs, CG centers and NARS was strengthened.

· Several publications are in preparation.

We expect the outputs of this project to provide very useful tools to the scientific community, for gene identification in wild species and pre-breeding purposes.

It is worth to mention that this project is now strongly connected to the OMAP project (Rod Wing, Arizona Genomics Institute). Indeed, BAC libraries for four of the parental

wild accessions involved in this project have been or will be constructed soon at AGI. Also, LGDP (Olivier Panaud's group) have developed a BAC library for the TOG 5681 (*O. glaberrima*) accession. Having this genomic resource available will add a great value to the populations of CSSLs developed here, as rice geneticists will have access to complete kits for gene identification and positional cloning in wild rice species.

VIII. References

Ahn SN, JP Suh, CS Oh, SJ Lee, and HS Suh (2002) Development of introgression lines of weedy rice in the background of Tongil-type rice. Rice Genetics Newsletter 19:14

Albar, L., M.-N. Ndjiondjop, Z. Essibak, A. Berger, A. Pinel, M. Jones, D. Fargette and A. Ghesquière, 2003. Fine genetic mapping of a gene required for rice yellow mottle virus cell to-cell movement. Theo. Appl. Genet 107: 371–378.

Aluko G., C. Martinez, J. Tohme, C. Castano, C. Bergman, J.H. Oard 2004. QTL Mapping of Grain Quality Traits From the Interspecific Cross Oryza sativa x O. glaberrima .Theor.Appl. Genet.109:630-639

Brondani, C, N. Rangel, V. Brondani, and E. Ferreira, 2002. QTL mapping and introgression of yield-related traits from Oryza glumaepatula to cultivated rice (Oryza sativa) using microsatellite markers. Theo. Appl. Genet 104: 1192–1203

Doi K, Sobrizal, K Ikeda PL Sanchez, T Kurakazu, Y Nagai and A Yoshimura (2003) Developing and evaluating rice chromosome segment substitution lines. Proc. IRRI Conference, Sept. 16-19, 2002. Beijing, China. p. 275-287

Enriquez, EC, DS Brar, TL Rosario, M Jones, GS Khush (2000) Production and characterization of doubled haploids from anther culture of the F1s of Oryza sativa L. x O. glaberrima Steud. Rice Genet News 17:26-27

Eshed, Y. and D. Zamir (1995). "An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL." Genetics 141: 1147-1162.

Frey KJ, EG Hammond and PK Lawrence (1975) Inheritance of oil percentage in interspecific crosses of hexaploid oats. Crop Sci 15:94-95

Ghesquière, A., J. Séquier, et al. (1997). "First steps towards a rational use of African rice, Oryza glaberrima, in rice breeding through a 'contig line' concept." Euphytica 96: 31-39.

Gur A and D Zamir (2004) Unused natural variation can lift yield barriers in plant breeding. Public Library of Science (in press)

Hash CT (1999) Contiguous segment substitution lines: New tool for elite pearl millet hybrid parental lines enhancement http://www.dfid-psp.org/projects/7382.htm

Jander G, SR Norris, SD Rounsley, DF Bush, IM Levin and RL Last (2002) Arabidopsis map-based cloning in the post-genomic era. Plant Phys 129:440-450

Koumproglou R, TM Wilkes, P Townson, XY Wang, J Beynon, HS Pooni, HJ Newbury (2002) STAIRS: A new genetic resource for functional genomic studies of Arabidopsis. The Pl. J. 31:355-364

M. Lorieux, G. Reversat, S. X. Garcia Diaz, C. Denance, N. Jouvenet, Y. Orieux, N. Bourger, A. Pando-Bahuon, A. Ghesquière. 2003. Linkage mapping of Hsa-1Og, a resistance gene of African rice to the cyst nematode, Heterodera sacchari. Theo. Appl. Genet 107:691-696

Moncada, M.P., C.P. Martínez, J. Tohme, E. Guimaraes, M. Chatel, J. Borrero, H. Gauch, Jr. and S.R. McCouch, 2001. Quantitative trait loci for yield and yield components in an Oryza sativa x Oryza rufipogon BC2F2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Monforte AJ, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the Lycopersicon hirsutum genome in a L. esculentum genetic background: A tool for gene mapping and gene discovery. Genome 43: 803–813

Pestsova EG, Borner A, Roder MS (2002) Development of wheat D-genome introgression lines assisted by microsatellite markers. In: Hernández P, Moreno MT, Cubero JI, Martin A, editors. Proceedings of the Fourth International Triticeae Symposium; 2001 September; Cordoba, Spain. Published in Cordoba by Junta de Andalucia. Consejera de Agricultura y Pesca, pp. 207–210.

Pestsova EG, Borner A, Roder MS (2003) Application of microsatellite markers to develop Triticum aestivum–Aegilops tauschii defined introgression lines. In: Borner A., Snape JW, Law CN, editors. Proceedings of the Twelfth International EWAC (European Wheat Aneuploid Co-operative) Workshop; 2002 July 1–6; Norwich, United Kingdom, John Innes Centre, publisher. pp 32-35

Ramsay LD, Jennings DE, Bohuon EJR, Arthur AE, Lydiate DJ, et al. (1996) The construction of a substitution library of recombinant backcross lines in Brassica oleracea for the precision mapping of quantitative trait loci (QTL). Genome 39: 558–567

Rick CM (1976) Natural variability in wild species of Lycopersicon and its bearing on tomato breeding.

Agraria 30: 249–510
Rick CM (1983) Conservation and use of exotic tomato germplasm. Conservation and Utilization of exotic germplasm to improve varieties. The 1983 Plant Breeding Research Forum: Conservation and utilization of exotic germplasm to improve varieties. Brown WL, editor. August 9-11, 1983, Heber Springs, Arkansas: Des Moines, Iowa, Pioneer Hybrid Int'l Inc., publisher. pp. 147–167.
Sears E (1956) The transfer of leaf-rust resistance from Aegilops umbellulata to wheat. Brookhaven Symp Biol 9: 1–22
Sobrizal K, Ikeda K, Sanchez PL, Doi K, Angeles ER, et al. (1996) Development of Oryza glumaepatula introgression lines in rice, O. sativa L. Rice Genet Newsl 16: 107
Sun et al 2001 Theo. Appl. Genet 102:157-162
Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277: 1063–1066
Thomson M, Tai T, McClung A, Xai X-H, Hinga M, Lobos K, Xu Y, Martinez P, McCouch S (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between Oryza rufipogon and the Oryza sativa cultivar Jefferson. Theor Appl Genet 107:479–493
Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, et al. (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, Oryza rufipogon. Genetics 150: 899–909.
Yano M (2001) Genetic and molecular dissection of naturally occurring variation. Curr Opin Plant Biol 4: 130–135
Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983-9
Zhu, Y.-L, Q.-J. Song, D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young, and P.B. Cregan. 2003. Single nucleotide polymorphisms (SNPs) in soybean. Genetics 163: 1123-1134

IX. Project evaluation (Table 1)

Table 1. **Project evaluation**

Evaluation criteria *(see notes directly below)*	**PI evaluation (%)**
1. Scientific achievements	**80**
2. Deviation from original workplan	**80**
3. Quality of raw data produced	**90**
4. Accessibility and organisation of raw data produced	**100**
5. Quality of non-data outputs produced	**75**
6. Availability of non-data outputs produced	**75**
7. Network building and communication	**80**
8. Report quality	**75**
9. Overall rating on project performance	**75**

Appendix A—Modified: Activities, quantifiable outputs and key products

Project title: Exploring Natural Genetic Variation: Developing Genomic Resources and Introgression Lines for Four AA Genome Rice Relatives					
Lead institute, Principal Investigator: IRD/CIAT, M.Lorieux; CIAT, J. Tohme					
Objective 1: Creation of six CSSL populations using AA genome rice species as donors					
Activities	**Quantifiable outputs**	E	FA	PA	NA
1. Develop CSSL populations using wild rice species as donors	1. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. meridionalis* 2. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. rufipogon* 3. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. glumaepatula* 4. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. barthii*		X X X	X	
2. Complete the development of two *O. sativa* x *O. glaberrima* CSSL populations	5. A CSSL populations of interspecific lines bearing introgressed genome segments from the AA genome cultivated species *O. glaberrima* in the *japonica* cultivated background 6. A CSSL populations of interspecific lines bearing introgressed genome segments from the AA genome cultivated species *O. glaberrima* in the *indica* cultivated background	X		X	
Objective 2: Development of a Universal Core Genetic Map for Rice					
Activities	**Quantifiable outputs**	E	FA	PA	NA
1. Define and validate a series of *anchors* along the rice genome that provide wide polymorphism across the wild rice species (AA genome)	7. A set of 125 anchors represented by three SSR loci each. At least one SSR per anchor is polymorphic for any of the worked crosses of this project 8. A database of polymorphic markers eases the definition of the best set of SSRs to be used for a specific cross	X X			
Objective 3: Development of a SNP kit for genotyping					
Activities	**Quantifiable outputs**	E	FA	PA	NA
1. Search for SNPs in stress-related genes, define primers and validate them for polymorphisms in the worked crosses	9. A set of 125 SNP markers based on the single base extension technology. These markers can be multiplexed in batches of 30 markers. The polymorphisms are validated for all the six			X	

	worked populations				
Objective 4: Analysis of *O. sativa* x *O. glaberrima* CSSLs for drought response					
Activities	**Quantifiable outputs**	E	FA	PA	NA
1. To screen the *O. sativa* x *O. glaberrima* CSSLs for drought response in order to identify alleles from *O. glaberrima*	10. Lines with superior behavior under stress (drought) conditions 11. Chromosomal locations of QTLs for drought tolerance from *O. glaberrima*		X	X	

Final products to be used by plant scientists outside the project:

1. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. meridionalis*
2. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. rufipogon*
3. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. glumaepatula*
5. A CSSL populations of interspecific lines bearing introgressed genome segments from the AA genome cultivated species *O. glaberrima* in the *japonica* cultivated background
6. A CSSL populations of interspecific lines bearing introgressed genome segments from the AA genome cultivated species *O. glaberrima* in the *indica* cultivated background
7. A set of 125 anchors represented by three SSR loci each. At least one SSR per anchor is polymorphic for any of the worked crosses of this project
8. A database of polymorphic markers eases the definition of the best set of SSRs to be used for a specific cross
10. Lines with superior behavior under stress (drought) conditions

Appendix B. Timeline

		Year 1												Year 2												Year 3											
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Objective 1	Activity 1																																				
	Activity 2																																				
Objective 2	Activity 1																																				
Objective 4	Activity 1																														6						

Year 4											
37	38	39	40	41	42	43	44	45	46	47	48
											1,2,3
					5,6						
	7	8									
					10						

Quantifiable Outputs:

1. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. meridionalis*
2. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. rufipogon*
3. A CSSL population of interspecific lines bearing introgressed genome segments from the AA genome wild species *O. glumaepatula*
5. A CSSL populations of interspecific lines bearing introgressed genome segments from the AA genome cultivated species *O. glaberrima* in the *japonica* cultivated background
6. A CSSL populations of interspecific lines bearing introgressed genome segments from the AA genome cultivated species *O. glaberrima* in the *indica* cultivated background
7. A set of 125 anchors represented by three SSR loci each. At least one SSR per anchor is polymorphic for any of the worked crosses of this project
8. A database of polymorphic markers eases the definition of the best set of SSRs to be used for a specific cross
10. Lines with superior behavior under stress (drought) conditions

Appendix E. Data format, availability and accessibility

Various data sets have been produced:
- Genotyping data (SSR markers) of 6 interspecific BC1F1 populations,
Format: Mapmaker/EXP compatible txt file
- Partial genotypic data of 5 BC2F2 populations
Format: CSSL Finder compatible Excel file
- Genotypic data for two advanced pre-CSSL populations from *O. sativa* x *O. glaberrima* crosses
Format: CSSL Finder compatible Excel file
- Six interspecific genetic maps aligned to the Nipponbare sequence
Format: MapDisto output txt file (Cmap-compatible format),
- Various sets of phenotypic data collected on two *O. sativa* x *O. glaberrima* ILs populations,
Format: Excel files
- Interspecific SNP locations and primers,
Format: Excel files
- SSR polymorphism for a collection of cultivated and wild rice accessions,
Format: Universal Core Map database (Excel-compatible)
- SSR genomic location of the Universal Core Genetic Map
Format: Universal Core Map database (Excel-compatible)
All of these data will be published on the GCP central database as soon as they have been valorized by publications.