ANNUAL REPORT
COMPREHENSIVE RESEARCH ON RICE
January 1, 2012 – December 31, 2012

PROJECT TITLE: Application of Molecular Marker-Assisted Selection to Rice Improvement

PROJECT LEADER: Thomas H. Tai, Research Geneticist, USDA-ARS Plant Sciences, UCD

PRINCIPAL UC INVESTIGATORS:
Thomas H. Tai, Research Geneticist, USDA-ARS, Plant Sciences, UCD

COOPERATORS:
Kent S. McKenzie, Director, Rice Experiment Station, Biggs
Luca Comai, Professor, UC Davis Genome Center, UCD
Virgilio C. Andaya, Plant Breeder, Rice Experiment Station, Biggs
Farman Jodari, Plant Breeder, Rice Experiment Station, Biggs
Jeffrey J. Oster, Plant Pathologist, Rice Experiment Station, Biggs

LEVEL OF 2012 FUNDING: $30,000

OBJECTIVES AND EXPERIMENTS CONDUCTED, BY LOCATION, TO ACCOMPLISH OBJECTIVES:
The overall objective is to integrate molecular genetic approaches and conventional breeding methods to develop improved germplasm for the California rice industry. Primary emphasis is on the development of DNA markers that can be used to predict the presence or absence of traits of interest (e.g. disease resistance, cold tolerance, grain quality) and the application of these markers via molecular marker-assisted selection to expedite the identification of useful germplasm and streamline the breeding of improved varieties.

In order to employ DNA markers, marker-trait associations must be established (i.e. the value of a marker in predicting a trait must be determined). Basic genetic studies have resulted in the identification of markers for many important traits. Several of these markers are based on differences in the DNA of specific genes. Differences (or polymorphisms) that are directly responsible for the characteristic in question are sometimes referred to as perfect markers as they are always (perfectly) associated with the trait. Genes underlying important traits in rice such as grain quality, yield, grain size, fertility, etc. have been identified. The major objectives of our research in 2012 were to 1) develop and apply a next-generation sequencing (NGS) strategy to identify markers and generate highly detailed genetic fingerprints of California rice varieties and germplasm; and 2) develop genetic mapping and mutant populations of rice to facilitate trait and gene discovery. Work on this project was conducted in the USDA-ARS rice genetics lab,
greenhouses, and other research facilities at UC Davis. In addition, some mapping population development work was conducted in the nursery at the Rice Experiment Station.

Specific 2012 objectives included:

1) **Next-generation genotyping (NGG) of rice varieties and breeding germplasm:** We will continue to develop and apply a next-generation sequencing (NGS) strategy to identify markers and generate highly detailed genetic fingerprints of California rice varieties and germplasm.

2) **Sequencing of expressed rice genes from California varieties:** We will employ a method called RNA-seq which will enable us to sequence thousands of genes from individual rice varieties. We will target specific tissues, such as the developing panicles, in order to focus on genes expressed during panicle and seed formation that are important for grain yield, morphology, and quality.

3) **Development and screening of rice populations for genetic analysis of agriculturally important traits:** We will continue to develop genetic mapping and mutant populations for genetic studies of important traits including but not limited to improved yield (total and milling) and stress tolerance. Screens for some traits of interest will be initiated.

4) **Evaluation of cold tolerance and characterization of gene expression during cold stress:** We will conduct further testing to identify screening conditions for distinguishing between cold blanking tolerant and sensitive rice varieties and examine rice gene expression under these conditions.

5) **Genetic fingerprinting of high yielding breeding lines and varieties:** High yield breeding lines from each RES breeding program will be subjected to NGG to determine if common DNA regions can be identified. Focus will be on lines containing the stem rot resistant line 87Y550 in their pedigrees.

**SUMMARY OF 2012 RESEARCH (MAJOR ACCOMPLISHMENTS) BY OBJECTIVES:**

1) **Next-generation genotyping (NGG) of rice varieties and breeding germplasm:** Single nucleotide polymorphisms (SNPs) are differences between individuals at the level of the four nucleotides (chemical blocks; represented by A, C, G, T) that make up DNA molecules (chromosomes). As noted in the 2011 RB-3 Annual Report, there two approaches to developing and applying SNP markers: 1) sequencing of two or more reference varieties to identify SNP markers which are then used to genotype or fingerprint additional varieties, germplasm, or mapping populations; and 2) *de novo* identification and genotyping by sequencing which involves sequencing each individual of interest and simultaneously identifying markers and fingerprinting the individuals (i.e. varieties, germplasm, etc.). While it is still too expensive and impractical to sequence the entire genome of each individual of interest, sequencing a representative portion of the genome is sufficient to provide a large number of markers while allowing many individuals to be analyzed together. One technical challenge is to ensure that the same portion of the genome in all the individuals is being examined so that comparisons can be made (Fig. 1). There are a number of methods for this and in 2011 we began work on using Restriction Enzyme Site Comparative Analysis (RESCAN) to sequence and
fingerprint California varieties. RESCAN was developed by Prof. Luca Comai and colleagues at the UC Davis Genome Center (Monson-Miller et al. 2012).

Fig. 1 SNP marker discovery and genotyping by next-generation DNA sequencing. DNAs are extracted and subjected to complexity reduction which reduces the amount and selects the same fraction of DNA from each individual. Samples are prepared and pooled together for sequencing and data are subjected to analyses to extract marker information.

In 2012, we completed a study of 45 California varieties in order to evaluate our methodology for complexity reduction and examine the utility of the RESCAN approach to identify markers in closely-related materials (Fig. 2-4).

Fig. 2. Distribution of SNP markers over the twelve rice chromosomes. The x-axis represents the length of the chromosomes and the y-axis indicates the number of SNP markers detected. Although complexity reduction is employed to reduce the amount of DNA to be sequenced, the coverage of the markers is sufficient to detect regions that distinguish individuals. This is important for identifying genes controlling traits of interest.
Fig. 3 SNP marker haplotype maps of California varieties generated by RESCAN. Varieties grouped by grain type: short (top), medium (middle) long (bottom). Blue regions match Nipponbare, the standard sequence reference. Red regions indicate differences between a variety and the reference. Similar patterns indicate relatedness and common regions may harbor traits that have been selected for by breeders. White regions are missing data and probably the result of lower amounts of DNA from those varieties. Improvements in technique and sequencing capacity will likely address the missing data issue.

Fig. 4. Genetic relatedness of the California varieties based on 4,629 SNP markers. This unrooted neighbor-joining tree shows the genetic relationships between the short (red) and medium (black) grain temperate japonicas and the long (blue) grain tropical japonica varieties. The numbers indicate varieties as shown above in Fig. 3. Three major clusters of branches from the central axis are observed, one consisting entirely of long grain, one with short and medium, and one with all three grain types.
We also employed RESCAN to identify regions in the rice genome that may confer the ability of rice plants to recover from cold stress at the early vegetative stage. Analysis of advanced backcross (AB) lines, derived by backcrossing a recombinant inbred line (MIb-71) with seedling cold tolerance from M-202 to a cold susceptible variety IR-50, revealed that in addition to the previously reported tolerance responses, some AB lines showed differences in their ability to recover from cold stress (Fig. 5). Although both MIb 4853-9 and MIb 6885-2 exhibited the same tolerance responses associated with the presence of the resistance QTLs qCTS4 (intermittent cold stress tolerance) and qCTS12 (constant cold stress tolerance), clear differences were observed in the ability of these lines to recover (i.e. resume growth) following constant cold stress with MIb 6885-2 exhibiting better recovery. Using RESCAN, we were able to generate high resolution SNP marker maps of MIb 4853-9 and MIb 6885-2 and identify candidate regions of interest regarding the cold stress recovery trait (Fig. 6). These results demonstrate the utility of RESCAN-mediated genotyping (fingerprinting).

Fig. 5. Differential recovery after cold stress in advanced backcross (AB) lines. AB lines MIb 4853-9 and MIb 6885-2 a) two week old seedlings (2-3 leaf stage) before cold stress; b) after exposure to 10°C for 10 days; c) after 15 day recovery period; d) time course of growth during recovery period. Days after cold stress on x-axis, plant height (cm) on y-axis. The mean and standard deviation from three replicates (24 plants per replicate) are shown.
Fig. 6. SNP marker genotyping of M-202, IR50, MIb 6885-2 and MIb 4853-9 reveals regions of introgression of M-202 into IR50. Seven major M-202 introgressions (black) were detected on 5 chromosomes as indicated with an asterisk (*). One introgression on chromosome 3 differed in size between the two MIb lines while two of the three introgressions on chromosome 6 were unique to one of the two lines. Regions in gray are monomorphic (i.e. M-202 and IR50 DNA cannot be distinguished) and regions in white represent missing data.

At the time the sequencing was performed for these studies, the 45 California individuals were sequenced using a platform which generated approximately 100 million sequences (also called sequence reads or reads) of about 50 bases for roughly $1000. Currently, next-gen sequencing has improved to the point where about 300 million reads of about 100 bases can be generated for about $1500. The utility of RESCAN and similar approaches along with the trends of increasing productivity of next-gen sequencing

2) **Sequencing of expressed rice genes from California varieties:** Originally the RNA-seq approach was to be employed to examine gene expression in selected California varieties. Pilot experiments with Korean cooperators indicated that RNA-seq, in particular the computational analyses required, would be too costly. In addition, approaches relying on gene expression are dependent on evaluating materials at the appropriate time and under
the appropriate conditions in order to ensure the genes of interest are identified. Another strategy which is not dependent on selection of the correct conditions for expression was pursued in order to identify differences in genes that may reflect functional differences and can be used as markers. This approach is called exome sequencing and involves sequencing the part of genes that are translated into proteins (Fig. 7). Much like RESCAN, exome sequencing is based on complexity reduction; however, the advantage is that the DNA being analyzed encodes genes and therefore differences identified by exome sequencing are more likely to be functionally relevant than the majority of SNP markers detected using RESCAN. One disadvantage is that exome sequencing requires additional effort and resources to prepare the DNA.

Fig. 7. Exome sequencing approach involves isolating or capturing the part of the genome that encodes expressed DNA. Genes consist of DNA, only a portion of which actually encodes the proteins that underlie traits of interest. In exome sequencing, DNA probes that match these expressed sequences (exons, shown in light blue and red boxes on top line) are used as “bait” to capture exons in a chemical process called DNA hybridization. These captured exons are sequenced and analyzed for SNPs. The genome consists of all DNA including parts of genes such as the promoter, UTR (untranslated region), and intron sequences which do not encode the protein. Conserved motifs are regions of a protein that are shared among species and differences in these regions are likely to have an impact on protein function. Image from L. Comai.

Exome sequencing was performed on three primary ancestral varieties of the short and medium grain California rices: Caloro, Colusa, and Lady Wright. In addition, one “modern” Southern long grain (Cypress) and M-204 were sequenced for comparison. Analysis of the exome sequences is currently underway. In addition to employing exome sequencing to examine different varieties, several M-204 mutants were sequenced to evaluate the number of mutations resulting from various treatments. Some of the initial findings are reported under the section on mutant population development.

3) **Development of rice populations for genetic analysis of agriculturally important traits:** We continued the development of California mapping populations through
generation advance (i.e. selfing to produce true breeding lines) and development of mutant populations in M-204.

Mapping Populations: For 2012, work continued on developing mapping populations derived from crossing California rice varieties using the seed single-descent method to advance generations and ultimately produce fixed (true breeding) lines. These structured (i.e. derived from known genetic crosses) populations are shown in Table 1. All the mapping populations appear to be segregating for various traits (e.g., grain size/shape and number, yield). Transgressive segregation/variation (i.e. when lines exhibit traits outside the range of the parents) for heading date and other traits was observed. Generation advance was carried out primarily in UC Davis greenhouses and the UC Davis rice/wetlands outdoor facility. Because the F6 generation of the M2036 population, which was produced in 2011, had insufficient seeds for field-based studies, planned genotyping of these lines using next-gen sequencing was delayed and will be performed in 2013 on the F7 lines which will be evaluated in the field. In addition to the advanced generation populations, F1 seeds from several crosses from the RES were provided by Dr. V. C. Andaya. One to two F1 plants from ~200 crosses were planted for F2 seeds, which may be used directly to map traits of interest or to develop additional mapping populations. No panicle rice mites were detected in 2012. Seed treatment policies established by UC Davis remain in place and continue to affect the speed and efficiency of research.

Table 1. California rice mapping populations under development in 2012

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MS2041</td>
<td>M-204/S-301</td>
<td>7 plants ranging from 700 to 1000 F2 seeds per plant</td>
<td>Generation advance to F5 (485 lines); F5 plants are currently in greenhouse</td>
</tr>
<tr>
<td></td>
<td>F1 plants (seed from Dr. V. Andaya)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM3014</td>
<td>S-301/M-204</td>
<td>8 plants ranging from 800 to 1000 F2 seeds per plant</td>
<td>Generation advance to F5 (487); F5 plants are currently in greenhouse</td>
</tr>
<tr>
<td></td>
<td>F1 plants (seed from Dr. V. Andaya)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM3016</td>
<td>S-301/M-206</td>
<td>290 F2 plants, harvested single panicles of F3 seeds</td>
<td>Generation advance to F6 (286 lines); grown at Davis and RES</td>
</tr>
<tr>
<td></td>
<td>F2 plants (seeds from Dr. V. Andaya)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2036</td>
<td>M-203/M-206</td>
<td>294 F2 plants, harvested single panicles of F3 seeds</td>
<td>Generation advance to F6 (~50 late/very late maturing lines) and F7 (234 lines; Davis and RES for bulking seeds for field evaluation in 2013)</td>
</tr>
<tr>
<td></td>
<td>F2 plants (seeds from Dr. V. Andaya)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Mutant Populations:** Generation advance and preliminary molecular characterization of the M-204 mutant populations was performed. Table 2 shows the number of M2 families that were planted this season. About 3-4 M2 seeds per family (i.e. per M1 plant) were sown in 50 cell plug flats and thinned to 1 or 2 plants per family (per cell) and grown to maturity. M3 seeds are being harvested. Due to growth conditions the number of M3 seeds per M2 plant was limited (~1 panicle per plant; estimate < 20 seeds). Seeds are sufficient for continued generation advance and some screening for traits of interest; however, additional M2 seeds from each family will be planted in 2013. Some observations of M2 seedlings were made although growth conditions and time constraints did not permit thorough visual evaluation.

Table 2. M-204 M2 mutants planted in 2012 for M3 seed production (generation advance).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>M0 seeds</th>
<th>M2 families</th>
<th>M2 planted</th>
</tr>
</thead>
<tbody>
<tr>
<td>350 Grays Cs137</td>
<td>~5000</td>
<td>1440</td>
<td>616</td>
</tr>
<tr>
<td>400 Grays Cs137</td>
<td>~5000</td>
<td>755</td>
<td>394</td>
</tr>
<tr>
<td>1 mM NaN3</td>
<td>~5000</td>
<td>4346</td>
<td>4262</td>
</tr>
<tr>
<td>3 mM NaN3</td>
<td>~5000</td>
<td>4540</td>
<td>3941</td>
</tr>
</tbody>
</table>

In addition to advancing mutant lines to the M3 generation, DNA samples from a small number of M2 plants from each of the four treatments (3 plants for the 350 Gy, 400 Gy, and 1 mM NaN3 and 4 plants for the 3 mM NaN3) were subjected to exome sequencing to evaluate the number of mutations present in the genome. Preliminary analysis of the exome sequence data indicates a relatively low mutation density overall with both populations derived from gamma-irradiation exhibiting on average around 0.4 mutations per Mb of diploid genome (rice genome ~380 Mb). The NaN3 treatments yielded about 0.5 mutations for 1 mM and 0.8 mutations for 3 mM. Thus, the NaN3 chemical treatment produced more mutations. The lower mutation densities in the gamma-irradiation derived M2 are in contrast to the fertility and growth exhibited by the M1 plants which were significantly reduced (indicative of strong effect of the treatment) in comparison to the untreated controls and the chemical treatments. Although the mutation densities were lower than the NaN3 plants, preliminary analysis of the gamma-derived plants did reveal the presence of deletions in some of these M2 plants (Fig. 8). Deletions are mutations in which pieces of a chromosome are missing. Physical agents such as gamma rays are thought to produce more deletions (i.e. larger changes to the DNA) than chemical agents which mainly generate point mutations, which are changes involving a single base pair of DNA (i.e., A, G, C, T) in which the base pair is changed or one is added/deleted.

Higher mutation densities are of interest for reverse genetics (i.e. screening for mutations by evaluating DNA); however, for forward genetics (i.e. screening directly for traits of interest) and plant breeding, lower mutation densities are suitable with the caveat that higher numbers of individuals (particularly the M1 and M2 generation) may need to be evaluated. Lower mutation densities also increase the chance that little or no backcrossing of a mutant of interest will be needed to remove background mutations which could negatively affect variety performance.
Fig. 8. Overlay plots of sequence read coverage from rice chromosome 7 of gamma-irradiation derived M-204 mutants. Relative read coverage is plotted on the y-axis and consecutive chromosome bins (groups of reads represented by black dots) are on the x-axis. Regions with well-below the average number of reads over several bins are indicative of deletions. Deletions were detected for chromosome 7 of two different gamma mutants (350 Gy treatment); one of about 590 kb (panel a, row 2200) and one of about 100 kb (bottom panel b, row 4800). Two middle panels are Lady Wright (panel c) and one 400 Gy mutant (panel d) which do not exhibit significant clusters of below average read coverage. Karyotype analysis and image by I. Henry (Comai Lab).

Both molecular and visual evaluation (Fig. 9) of the M₂ plants have confirmed the presence of mutations in all the populations and evaluation of the M₃ generation for traits of interest will be initiated in 2013. The estimated mutation densities suggest that additional M₂ plants should be grown to expand the number of mutations for trait evaluation. Alternative mutagenic treatments may be employed to generate populations with higher mutation densities.
4) **Evaluation of cold tolerance and characterization of gene expression during cold stress:** No major accomplishments to report on evaluation of cold tolerance of diverse rice germplasm (Silewah, Yoneshiro, Leng Kwang, Somewake. Mustukogane, Tatsumimochi, Hayayuki, Thangone, Pratao, and Mitak). Progress in this area has been slowed by lack of adequate cold testing facilities for blanking and by change in direction from RNA-seq method to exome sequencing. Crosses of some of these varieties have been made and crosses from the RES involving M-104 and cold sensitive varieties were grown in Davis for F2 seeds. These F2 may be evaluated and/or utilized to develop fixed lines for mapping.

5) **Genetic fingerprinting of high yielding breeding lines and varieties:** No major accomplishments to report for 2012. Still in discussions with RES breeders about the rice lines that will be analyzed using either RESCAN or exome sequencing. With the demonstrated success of both methods, genotyping of these RES lines as well as those with stem rot resistance will be the top priority for 2013.

Literature Cited:

PUBLICATIONS OR REPORTS:
None at this time. Two publications submitted in 2012 (one accepted pending revision).

CONCISE GENERAL SUMMARY OF CURRENT YEAR’S RESULTS:
A next-generation DNA sequencing method called Restriction Enzyme Site Comparative Analysis (RESCAN) was successfully used to identify molecular markers and generate high resolution DNA fingerprints of forty-five California varieties. This work demonstrated the utility
of the approach which will now be implemented to analyze much larger sets of rice germplasm including mapping populations for breeding and genetic studies. Emphasis will be on RES breeding lines and germplasm for yield (total and milling) and stem rot tolerance. Another sequencing strategy focusing on the portion of the genome which encodes proteins (i.e. expressed genes; exome) was also employed and characterization of the resulting DNA sequence data is underway. Preliminary results from the exome sequencing approach were used to estimate the mutation density of the four M-204 mutant populations currently being developed. M\textsubscript{3} generation seeds representing the majority of the M-204 mutant populations were produced for screening and for further generation advancement (i.e. production of fixed lines). Results of the exome sequencing indicate that more M\textsubscript{2} seed should be planted for M\textsubscript{3} seeds in order to improve the probability of identifying useful mutants. Additional mutagenesis may also need to be performed. Mapping population development continued and seeds from one population derived from a cross of M-203 and M-206 were increased to facilitate trait evaluations in 2013 in conjunction with DNA marker analysis using RESCAN.