ANNUAL REPORT
COMPREHENSIVE RESEARCH ON RICE
January 1, 2011 – December 31, 2011

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
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 2011 FUNDING: $60,000

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 molecular (DNA) markers that can be used to predict the presence or absence of a trait 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 DNA marker in predicting a trait must be determined). Basic genetic studies have resulted in the identification of DNA 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 2011 were to 1) initiate DNA marker development and gene characterization by next-generation sequencing of California rice varieties and breeding materials; and 2) develop genetic mapping and mutant populations of rice to facilitate trait and gene discovery. All work on this project was conducted in the USDA-ARS rice genetics lab, greenhouses, and rice research facility at UC Davis.

Specific 2011 objectives included:

1) **Next-generation sequencing of modern and ancestral California varieties to identify DNA markers:** Using the high throughput DNA sequencing capacity available at the UC Davis Genome Center, we will sequence DNA of selected modern and ancestral California varieties to obtain high resolution DNA fingerprints and identify DNA markers for use in mapping important rice genes. In addition, DNA markers will be converted to a form that may be used with low cost instrumentation and reagents.
2) **Sequencing of expressed rice genes from California varieties:** This work will expand on our efforts from 2009 and 2010 to sequence genes involved in agricultural important traits in rice. In 2011, 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 of rice populations for genetic analysis of agriculturally important traits:** We will continue to develop rice populations for genetic studies of important traits including but not limited to milling yield and cold tolerance. In addition to advancing lines for recombinant inbred line populations, we will also initiate new crosses for base-broadening and development mutant populations for screening of desirable traits such as stress tolerance and higher yield.
4) **Evaluation of cold tolerance and characterization of gene expression during cold stress:** Several rice germplasm accessions from Japan, Indonesia, and other countries have been identified which may possess superior cold tolerance at the seedling and reproductive (booting) stages of development. These accessions will undergo further testing in 2011. In cooperation with K. McKenzie, selected accessions will be tested in the Rice Experiment Station cold tolerance screening greenhouse. In addition, samples of developing panicles from tolerant and sensitive varieties undergoing cold stress will be taken for RNA-seq to examine what genes are expressed.
5) **Genetic fingerprinting of high yielding breeding lines and varieties:** High yield breeding lines (irrespective of quality and other traits) from each RES breeding program (F. Jodari, K. McKenzie, V. Andaya) will be obtained and these will be subjected to NGS to develop high resolution DNA fingerprints. This work will provide a foundation for efforts to identify genetic loci involved in high yields under temperate conditions.

SUMMARY OF 2011 RESEARCH (MAJOR ACCOMPLISHMENTS) BY OBJECTIVES:

1) **Next-generation sequencing of modern and ancestral California varieties to identify DNA markers:** 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). Until the past 5 years or so, DNA markers were mostly based on indirectly detecting differences such as SNPs. The development of DNA sequencing methods and instruments (collectively called platforms) has enabled direct detection of SNPs which allows for distinguishing very closely-related individuals such as the varieties and breeding materials that are the foundation of the RES breeding projects. Next-generation sequencing is a generic term for the most recent

platforms for DNA sequencing which are characterized by the production of huge amounts of sequence data, relatively low costs, and dependence on rigorous computational/statistical analyses (also known as bioinformatics). There are two approaches to developing and applying SNP markers:

Sequencing of two or more reference varieties: In this approach, SNPs are identified by sequencing reference varieties and then chemical tests (assays) are developed to screen individuals of interest. Using this approach, a Japanese group headed by Dr. Masahiro Yano (QTL Genomics Research Center, NIAS, Tsukuba, Japan) has developed a set of SNP markers by sequencing the varieties Koshihikari, Eiko, and Rikuu132 and comparing them to the previously sequenced reference variety Nipponbare. All are temperate japonicas, the same type of rice as grown in California. To determine the usefulness of these markers, we sent DNA from 16 varieties to Dr. Yano for analysis using a core set of markers (n = 768) which detect differences among Japanese varieties (Fig. 1). The results of the analysis of the California accessions are shown in Fig. 2.

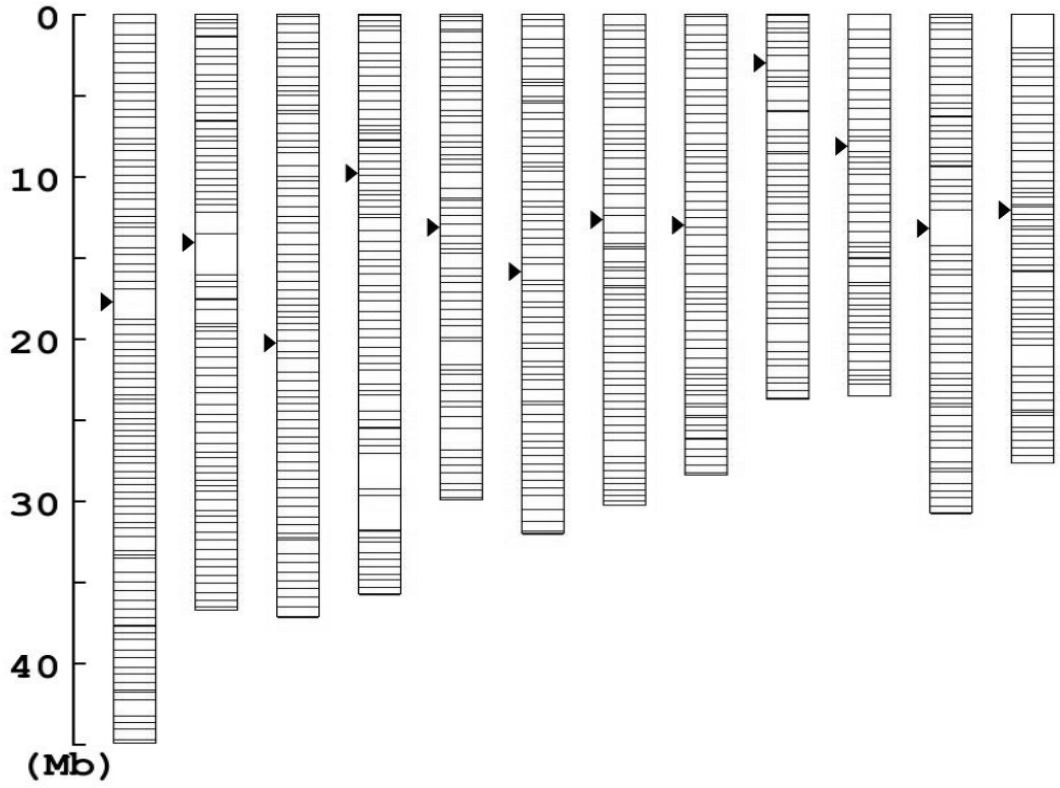

Fig. 1. Distribution of Japanese temperate japonica SNP markers over the twelve chromosomes of rice. Scale to the left is in megabases (Mb). One megabase is 1 million bases (nucleotides). The rice genome consists of about 390 Mb. Arrows mark the centromeric regions of the chromosomes.

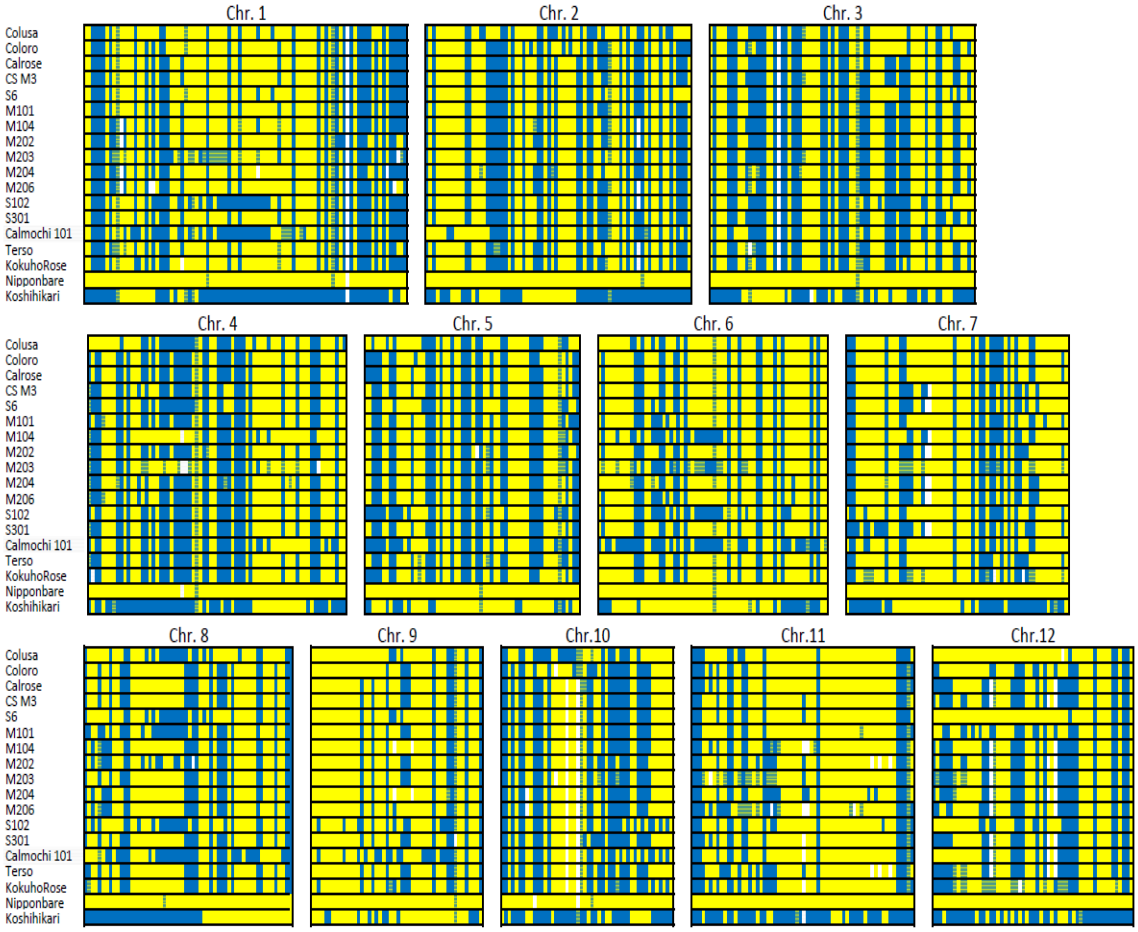

Fig. 2. Comparison of California varieties using SNP markers. Varieties are listed at the left (reference Japanese varieties Nipponbare and Koshihikari are included on the bottom). An overview of each chromosome is shown. More closely-related varieties show more similar patterns. For example, see chromosome 1 patterns of S-102 and Calmochi-101.

A more detailed look at differences between varieties is shown in Fig. 3 which depicts the comparison of M-206, M-204 and S-301. M-206 was derived from the cross S-301/M-204. SNP marker analysis revealed that there are specific regions on various chromosomes of M-206 that appear to be derived from one or the other parent (unique M-204 contributions marked in blue, unique contributions from S301 in red) and interestingly there were regions that are unique to M-206 (green). On the basis of these Japanese SNP markers, there appear to be four relatively large regions in M-206 that do not seem to be derived from either parent. This result might be explained by different seed sources (seeds from the original plants used to make the S-301/M-204 cross are not available) or the possibility that additional germplasm might have been introduced during the breeding process. The size of the various regions is dependent on the number of SNP markers. For example, the large blocks shown in yellow are regions where the markers tested could not distinguish the three varieties.

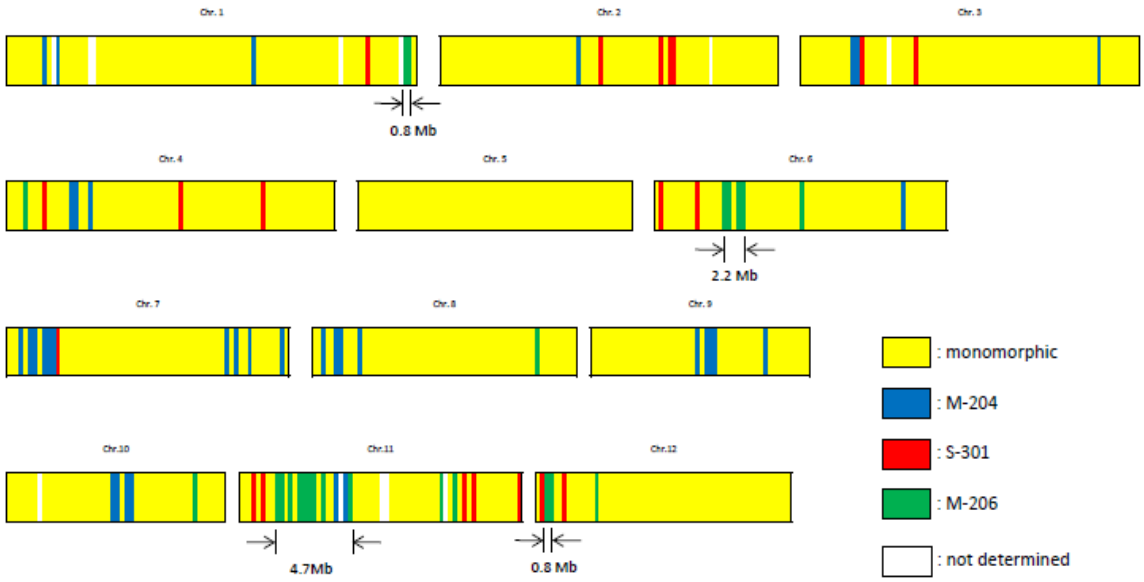

Fig. 3. Comparison of the chromosomes of the M-206, M-204 and S-301 varieties using Japanese SNP markers. Rectangular boxes represent each chromosome starting with 1 at the top left and reading across from left to right. The color legend indicates regions of the M206 chromosomes that appear to be derived from each parent and some regions that are unique to M-206 (four of which appear to be relatively large as indicated by arrows). Regions in which the markers used could not detect differences (yellow) or where there were insufficient marker data (white) are also shown.

De novo identification and genotyping by sequencing: This alternative approach involves sequencing each individual of interest. In this way, the identification and assaying of the SNP marker is carried out at the same time. It is still too expensive and impractical to sequence the entire genome (all the chromosomes) of each individual, but sequencing a small part of each chromosome is more than sufficient to provide extremely high numbers of markers. The technical challenge is to ensure that the same part of each chromosome in all the individuals is being examined. This is done through a process of selection based on restriction enzymes that break DNA molecules into pieces of specific size followed by isolation of a subset of those pieces based on their size (referred to as fractionation). In 2011, we employed a version of this method called Restriction Enzyme Site Comparative Analysis (RESCAN) to initiate characterization of California varieties developed by or associated with the RES. RESCAN was developed by Prof. Luca Comai and his colleagues at the UC Davis Genome Center.

RESCAN involves two major components: 1) DNA preparation from individuals of interest and sequencing; 2) computational analysis of the sequence data and generation of genotypes (fingerprints) for use in identifying regions of the genome associated with target traits. The RESCAN pipeline (workflow) is shown in Fig. 4 (courtesy L. Comai). Component 1 is shown in black and component 2 is shown in green. Component 1 is largely carried out in the lab with the exception of the sequencing which is performed at

the UC Davis Genome Center core facilities which are equipped with the necessary next-generation sequencing instruments. The methods/protocols are largely developed for component 1. Component 2 involves processing the sequence data that has been generated into user-friendly formats for interpretation and for mapping, in conjunction with trait data, genes controlling important traits. Component 2 requires the development and implementation of computer algorithms (programs), some of which are publicly available (shown in blue) and some which have been developed in-house by the Comai lab (shown in red). Component 2 is still under development by the Comai lab with regard to increasing the capacity to handle larger datasets and providing a user-friendly interface.

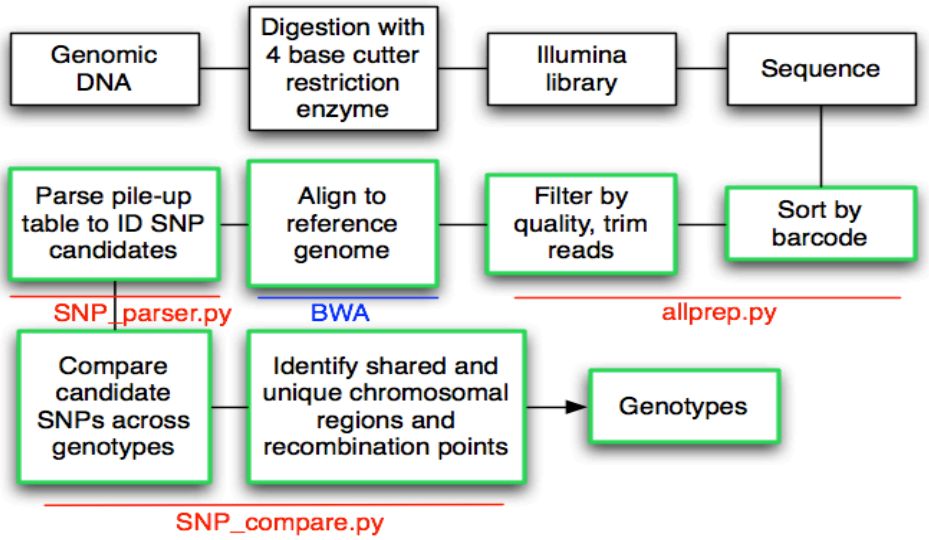

Fig. 4. Restriction Enzyme Site Comparative Analysis (RESCAN) pipeline for SNP marker genotyping by sequencing.

Forty-five California varieties (Colusa, Caloro, Calrose, CS-M3, CS-S4, M5, S6, Calrose76, M7, M9, Calmochi-201, L-201, M-101, M-301, S-201, Calmochi-202, M-302, M-401, M-201, L-202, Calmochi-101, M-202, A-301, M-102, M-203, S-101, M-103, S-301, L-203, M-204, A-201, L-204, S-102, Calhikari-201, Calmati-201, L-205, M-402, M-104, M-205, M-206, M-207, Calamylow-201, Calmati-202, L-206, and M-208) were subjected to the RESCAN method. Data for the entire set is still being analyzed. Preliminary analysis of M-206, S-301, and M-204 is shown in Fig. 5. Over 9,000 SNP markers (12X the number of Japanese markers) were identified. As with the Japanese markers, RESCAN-derived SNPs clearly show regions of the M-206 genome that were different from the parental varieties S-301 and M-204. These include parts of chromosome 1, 4, 6, 10, 11, and 12. A similar comparison of the California varieties M-203 and M-206 is shown in Fig. 6. In this case, over 11,000 SNP markers were detected between the two varieties with chromosomes 7, 8 and 11 showing the most widespread differences. Markers distinguishing M-203 and M-206 are of interest due to our work on developing the M2036 mapping population (seed objective 3 results) which we intend to use to investigate milling yield and stability among other traits.

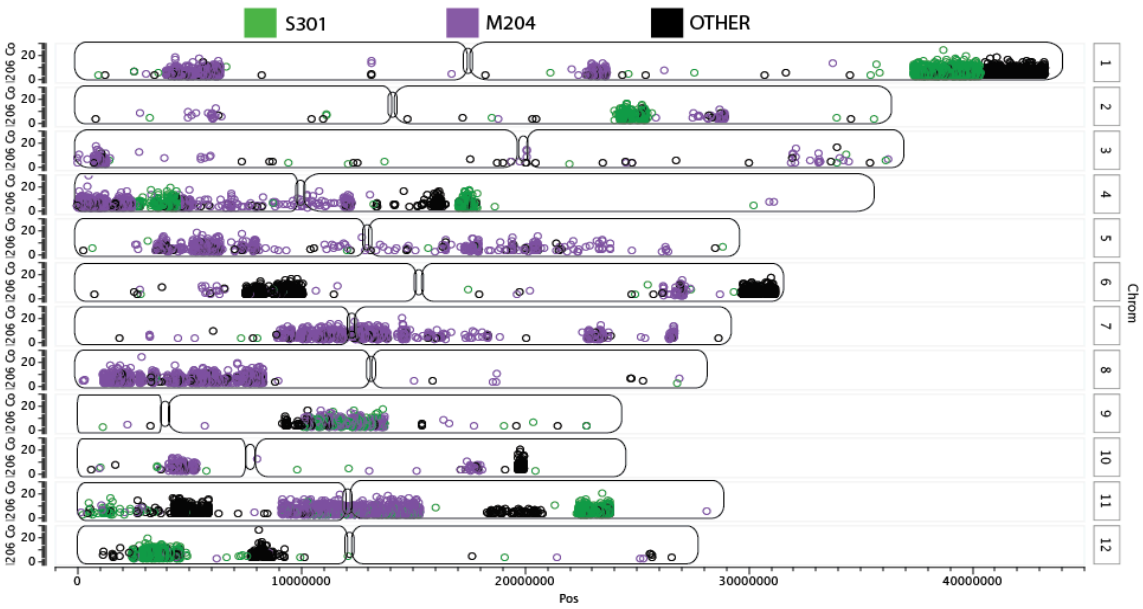

Fig. 5. RESCAN of California varieties M-206, S-301, and M-204 reveals shared and unique regions. Colored circles represent SNP markers along the length (x-axis) of each chromosome (indicated to the right). SNP markers in M-206 that are shared with (or derived from) S-301 (green) or M-204 (purple) or neither parent (black) are indicated. The number of times a given SNP marker was detected by sequencing (the more times the more confidence that the marker is real) is shown at the left. Regions of the chromosome where no circles are shown are indicative of a lack of distinguishing SNP markers.

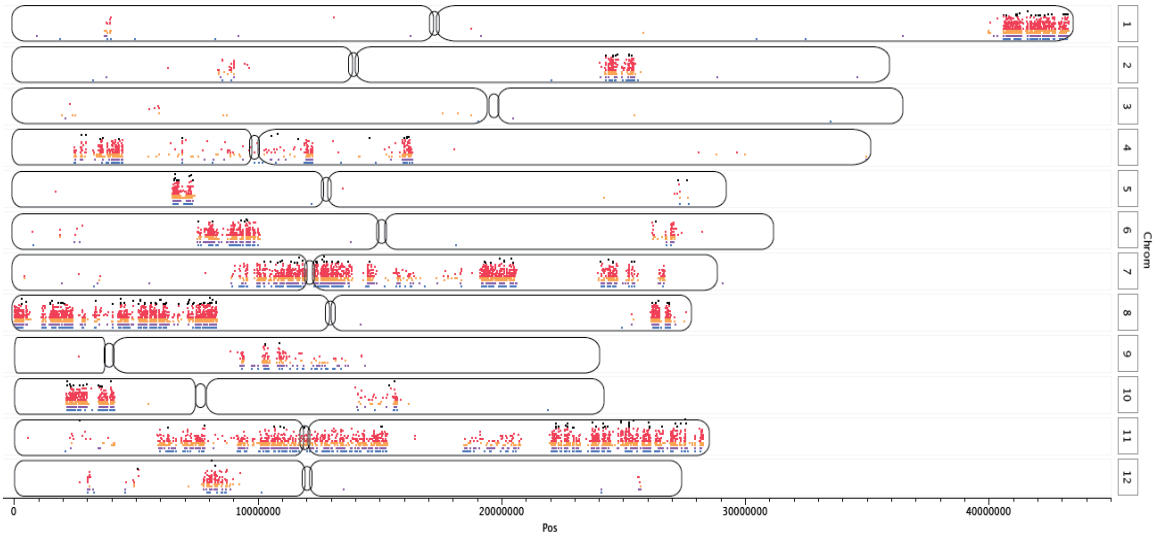

Fig. 6. RESCAN of M-203 and M-206. SNPs (~11,000) are indicated by colored dots on the chromosomes. The colors correspond to coverage (number of times the SNPs were sequenced/detected and therefore the confidence that those SNPs are real). Coverage from low to high (blue < purple < orange < red < black).

Our results for 2011 demonstrate that the RESCAN method can be employed to provide high resolution genotyping/fingerprinting of California varieties. The ability to distinguish the derivatives of the narrow germplasm base represented by these varieties and the relative ease of generating thousands of markers is indicative of the power of this approach for gene mapping and isolation. Additional work is necessary to increase the capacity and user-friendliness of the computational component of the RESCAN pipeline. It is expected that this will be accomplished during the first half of 2012 (including the analysis of the remaining California varieties sequenced in 2011) and will facilitate the expansion of this method to analyze hundreds of lines including those from the mapping populations described in objective 3 and the high-yielding lines that are the focus of objective 5.

2) **Sequencing of expressed rice genes from California varieties:** We are working on using the RNA-seq method to examine gene expression in selected California varieties. There are no major accomplishments to report at this time. It is expected that this work will be addressed in the first half of 2012.

3) **Development of rice populations for genetic analysis of agriculturally important traits:** We continued the development of California (temperate japonica rice) mapping populations through generation advance (i.e. selfing to produce true breeding lines) and initiated development of mutant populations in a California rice variety relevant to the RES breeding programs.

Mapping Populations: For 2011, 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. Emphasis was placed the M2036 population, which will be used to examine milling yield and milling yield stability in cooperation with the RES. The M2036 population was advanced to the F_6 generation (although the number of lines was reduced due to delayed heading and seed maturation). The M2036 F_6 lines are being genotyped using NGS (see objective 1) and we plan to grow these lines at the RES in 2012 to perform preliminary characterization of heading date, yield (total and milling), seed shape/size, and other agronomic traits.

All the mapping populations appear to be segregating for various traits (e.g., grain size/shape and number, yield). Transgressive segregration/variation (i.e. when lines exhibit traits outside the range of the parents) for heading date and other traits was observed. Evaluation of these mapping population lines in replicated trials under field and greenhouse conditions is needed to confirm these observations. Generation advanced was carried out in UC Davis greenhouses. No panicle rice mites were detected in 2011. The Animal and Plant Health Inspection Service (APHIS) has reclassified panicle rice mite (*Steneotarsonemus spinki*) as a non-quarantine pest. 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 2011

Population	Origin (2008)	Original lines	2011
MS2041	M-204/S-301 F_1 plants (seed from Dr. V. Andaya)	7 plants ranging from 700 to 1000 F_2 seeds per plant	Generation advance to F_4 (485 lines)
SM3014	S-301/M-204 F_1 plants (seed from Dr. V. Andaya)	8 plants ranging from 800 to 1000 F_2 seeds per plant	Generation advance to F_4 (487)
SM3016	S-301/M-206 F_2 plants (seeds from Dr. V. Andaya)	290 F_2 plants, harvested single panicles of F_3 seeds	Generation advance to F_5 (286 lines)
M2036	M-203/M-206 F_2 plants (seeds from Dr. V. Andaya)	294 F_2 plants, harvested single panicles of F_3 seeds	Generation advance to F_5 (285 lines) and F_6 (234 lines)

Mutant Populations: In previous years, small mutant populations have been developed under the RB-3 project using California varieties (e.g. S-102, Terso). In 2011, we initiated a new effort to produce a larger, more diverse mutant population. In consultation with Dr. Kent McKenzie, we selected the variety M-204 to produce the populations. In addition to being used in the RES breeding projects, M-204 is in the pedigree of the California mapping populations we are developing. Selection of M-204 should increase the chance of useful mutants being integrated into the RES breeding projects and may also complement trait and gene discovery using the rice mapping populations described in Table 1. To generate mutant populations, both physical and chemical agents were employed.

Physical agent: M-204 seeds were treated with gamma-irradiation from a Cesium-137 (Cs^{137}) source at the Center for Health and the Environment, UC Davis. A range of radiation doses were applied (250, 300, 350, 400, 500, and 600 Grays) to 35 g (~ 1000 seeds) samples in order to obtain information on the response of M-204 to this treatment. For all treatments, dry seeds were used except for the 350 Grays dose which was used to treat both dry seeds and a 35 g lot of seeds that had been soaked in water for 16 hours prior to irradiation. Two larger lots (165 g or ~5,000 seeds each) were treated with doses of 350 and 400 Grays based on casual observations of the effect of these treatments on the varieties M-202 and IR50, an indica from the Philippines (T. Tai).

Chemical agent: To generate additional diversity, a second agent, the chemical sodium azide (NaN_3) was also used to treat M-204 seeds. Two concentrations of sodium azide (1 mM and 3 mM) were used to treat ~5,000 seeds each. Doses were based on the literature and previous experience with the japonica rice variety Kitaake (T. Tai). In addition, small samples (100 seeds) of the previously gamma-irradiated M-204 seeds (35 g lots) were treated with 1 mM sodium azide to examine the effects of this double treatment.

To determine the effects of various doses of the gamma-irradiation and sodium azide, seeds from each treatment were planted in the greenhouse and data were collected on germination/survival, height, and fertility (seed set). These measurements provide an indication of the effectiveness of the treatment when compared with untreated controls. Data are still being collected and analyzed. Germination and height information are shown in Figs. 7 and 8.

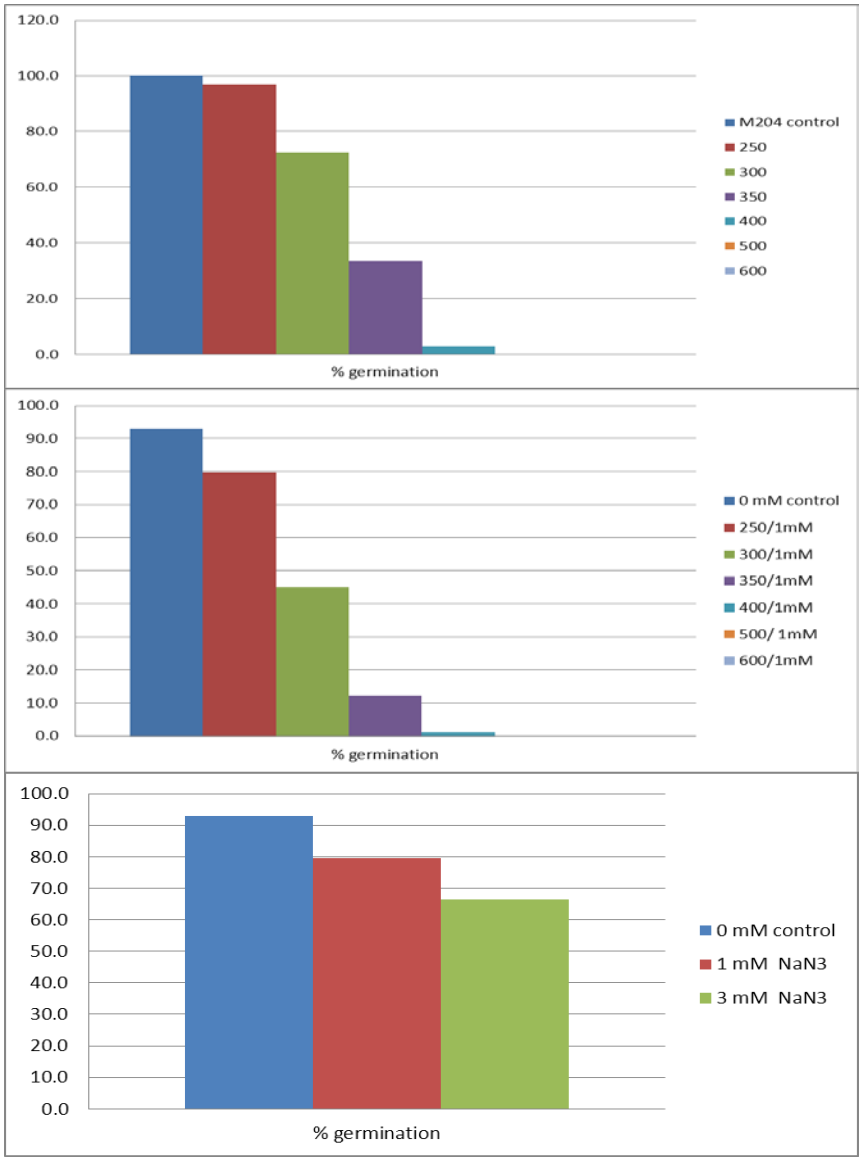

Fig. 7. Germination following various mutagenic treatments of dry seeds of M-204. Effects of gamma-irradiation only (top), gamma-irradiation followed by 1 mM sodium azide (middle), and sodium azide only (bottom) are shown (n = 98). The germination (%) is shown along the y-axis. Treatments are indicated by color (legends at right). There were no survivors of the 600 Grays seed treatment.

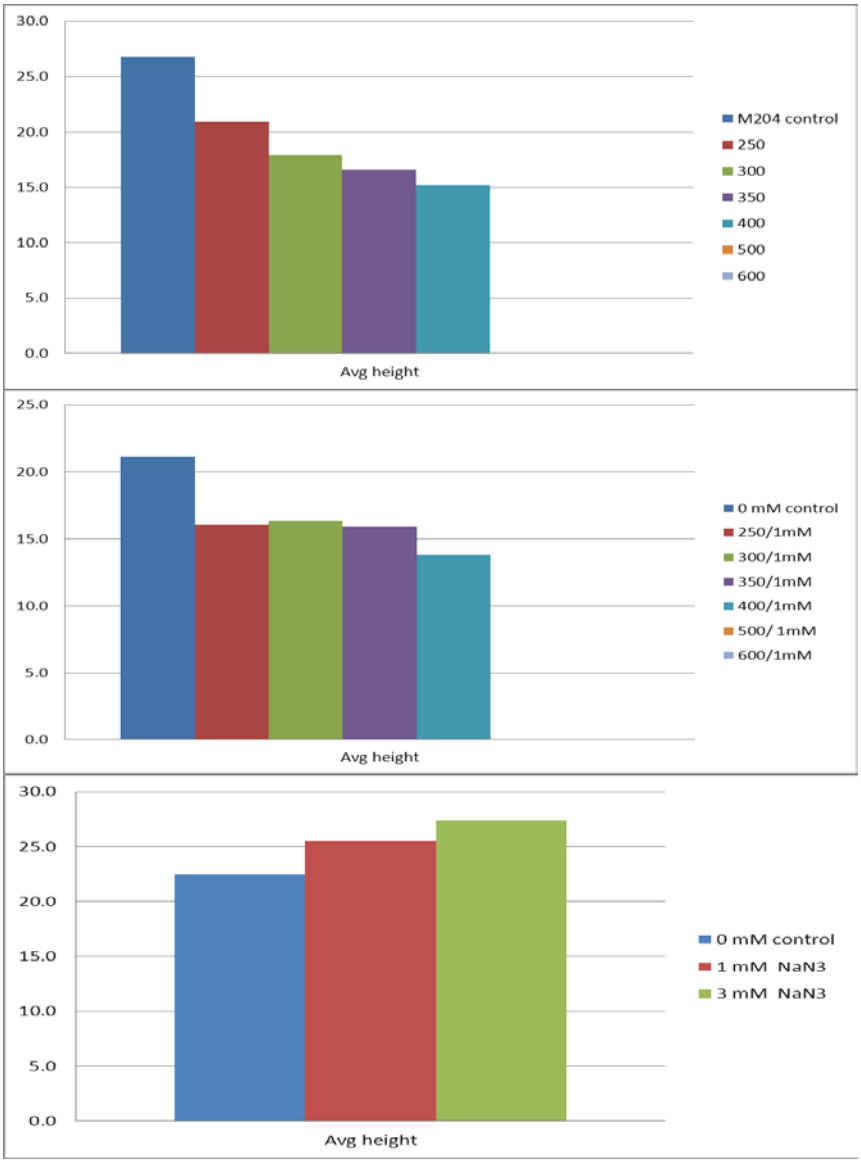

Fig. 8. Average height of M-204 M_1 plants from various mutagenic treatments. Effects of gamma-irradiation only (top), gamma-irradiation followed by 1 mM sodium azide (middle), and sodium azide only (bottom) are shown. The average heights (in cm) of the plants from Fig. 7 are shown along the y-axis. Treatments are indicated by color (legends at right). There were no survivors of the 600 Grays seed treatment.

As expected, the germination of seeds was inversely correlated with the dose of the mutagen applied. In the case of M_1 plant height, increasing doses of gamma-irradiation resulted in decreasing height. In marked contrast, M_1 plants derived from sodium azide-treated seeds were taller than the controls and height increased with increasing dose. In addition to analyzing seed set from these plants, M_2 seeds will be grown for phenotypic (visual observation) and genotypic (sequencing to assess mutation density) evaluation in the first half of 2012.

In order to generate larger populations for screening of mutant phenotypes of interest in 2012, four treatments were arbitrarily selected in 2011 for large-scale production. The treatments were: gamma-irradiation (350 and 400 Grays) and sodium azide (1 and 3 mM). Results of this work are shown in Table 2.

Table 2. Large M-204 mutant populations (M2 families) generated in 2011.

Treatment	M_0 seeds	M_1 plants	M_2 families
350 Grays Cs^{137}	~5000	2254*	1323
400 Grays Cs^{137}	~5000	2058*	729
3 mM NaN_3	~5000	4704	4130
1 mM NaN_3	~5000	~4500	TBD

*Represents a subset of M_1 seeds sown. These were grown in greenhouse. Additional plants were grown outdoors and counts are still in progress (TBD – to be determined).

4) **Evaluation of cold tolerance and characterization of gene expression during cold stress:** Several rice varieties from the U.S. collection were tested for reproductive cold tolerance Silewah, Yoneshiro, Leng Kwang, Somewake (2 accessions), Mustukogane, Tatsumimochi, Hayayuki, Thangone, Pratao, M-103, Calmochi-101, M-201, Calmati-201 in the RES cold blanking greenhouse in collaboration with Dr. Kent McKenzie. These accessions were selected as previous reports in the literature and observations by the RES indicated that these were tolerant to cold-induced blanking (with the exception of M-201 and Calmati-201 which were selected as susceptible checks). Evaluation in the RES greenhouse was not successful as severe blanking was observed and there were inconsistencies in the behavior of the various standard checks used by the RES. It was noted that most materials blanked with the exception of M-104 derivatives. Conditions for assessment of cold blanking tolerance need further refinement. As a result, this objective could not be addressed in 2011. This objective will be pursued in 2012.

5) **Genetic fingerprinting of high yielding breeding lines and varieties:** No major accomplishments to report for 2011. Currently in discussions with RES breeders about the rice lines that will be analyzed using the approach being employed in objective 1. It is expected that this work will be addressed in the first half of 2012.

PUBLICATIONS OR REPORTS:

None at this time.

CONCISE GENERAL SUMMARY OF CURRENT YEAR'S RESULTS:

A next-generation sequencing approach was taken to evaluate California rice varieties. This method, called Restriction Enzyme Site Comparative Analysis (RESCAN), compares very favorably to other DNA marker methods including SNP analysis using previously characterized

markers from Japan. Work continues on analyzing sequence data from the California varieties. Improvements to the work flow for this genotyping (DNA marker analysis) method with emphasis on computational analysis and development of a user-friendly interface are underway in cooperation with UC Davis colleagues. The RESCAN method will be applied to additional rice germplasm (i.e. varieties, breeding lines, mapping lines, and mutants) in 2012 including the high-yielding lines derived from the stem rot resistant RES breeding line 87Y550. In addition to DNA marker work, we continued to develop genetic mapping populations derived from RES varieties. The parents of these populations are from a narrow genetic base within the temperate japonica rice subspecies and the genotyping by sequencing (RESCAN) work will facilitate our use of these populations to map genes important for traits such as milling yield and stability. Work to develop mutant populations in a useful genetic background was initiated in 2011. Using both gamma-irradiation and the DNA modifying-chemical sodium azide, seeds of the variety M-204 were treated to produce mutant populations which are undergoing preliminary characterization. These populations will be screened to identify useful variants in 2012 and will also undergo further development to produce fixed mutant lines.