2007

Identification of molecular markers and association mapping of selected loci associated with agronomic traits in rice

Suresh Babu Kadaru
Louisiana State University and Agricultural and Mechanical College, skadar1@lsu.edu

IDENTIFICATION OF MOLECULAR MARKERS AND ASSOCIATION MAPPING OF SELECTED LOCI ASSOCIATED WITH AGRONOMIC TRAITS IN RICE

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in
School of Plant, Environmental and Soil Sciences

By
Suresh Babu Kadaru
B.Sc.(Ag), Acharya N G Ranga Agricultural University, 1997
M.Sc.(Biotech), Tamil Nadu Agricultural University, 2000

August 2007

I dedicate this work to my beloved wife Sudha.

ACKNOWLEDGEMENTS

I express my sincere gratitude to my major advisor, Dr. James H. Oard for giving me the opportunity to pursue PhD study in School of Plant, Environmental and Soil Sciences, LSU, Baton Rouge. I would like to express my appreciation for all the support and guidance that he had given to me since January 2003. I particularly appreciate his efforts to strengthen my English writing skills. I also thank all my committee members, Drs. Milton C. Rush, Charles J. Monlezun, Don R. LaBonte, and Gerald O. Myers for their constructive suggestions during my dissertation preparation. I would like to state that my PhD study would have not been possible without the financial support from Biotechnology Education for Students and Teachers program, LSU AgCenter. I am very grateful to Dr. Richard Tulley for working over by annual research budgets and work permits. I can never forget the help that he had offered during my outpatient surgery in 2003. I found Dr. Svetlana Oard as very supporting during my four and half year stay in the Wilson Laboratories, LSU. I thank Dr. Nengyi Zhang for his help and advices for the Chapter 2 of this manuscript. I also thank my good friend, Dr. WeiQing Zhang for his assistance in the herbicide resistance rice study. I found him as a very candid personality and appreciate his timely help at several occasions. I also thank Dr Manjit Kang for the inspiration and assistance in shaping my professional career.

I would also like to admit that my PhD endeavor might have not been possible without the constant encouragement and support from my wife and parents. All these years, I have always cherished my daughter's amusing actions. I express gratitude to all my LSU friends who made my stay in Baton Rouge as a wonderful experience. Finally, I thank almighty for keeping my family safe and healthy during this study endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Discriminate Analysis as a procedure was evaluated to select molecular markers associated with complex traits in US rice germplasm. Markers for percent head rice, percent total rice, and grain yield were identified with high levels of correct classification that mapped within or near traditional Quantitative Trait Loci (QTL).

Mixed model-regression procedure to identify molecular markers that predict phenotypic variance associated with four agronomic traits was created and validated in two distinct rice inbred populations. Main and epistatic effects were identified by standard hypothesis testing and Bayesian information criteria in a multivariate format. The new procedure increased power and enhanced prediction ability of markers in validation samples from both populations.

A new SNP discovery and genotyping protocol referred to as Alternative Ecotilling has identified four previously reported and 14 new SNPs in the *alk* and *waxy* genes among 57 accessions based on comparisons with sequencing results. The new procedure has been published in 2006 in the journal *Plant Molecular Biology Reporter.*

Application of haplotype-specific markers in exon 7 of the *BAD2* gene for marker-assisted identification and introgression of the aroma gene in U.S. rice was evaluated. Aromatic/non-aromatic phenotypes were consistent with corresponding marker haplotypes for all progeny tested which shows the potential of this procedure for marker assisted breeding of new aromatic varieties.

Similarly, an allele-specific PCR assays were developed to distinguish between homozygous and heterozygous imazethapyr-resistant $S_{653}D$ and $G_{654}E$ SNP alleles of the rice *ALS* gene. Field collections were successfully screened for the presence of $S_{653}D$ SNP, and F_2 progeny lines of natural CL 121 x red rice outcrosses were screened for the presence of the $G_{654}E$

SNP. These assays were proven successful and are currently used for detection of outcrossing and seed purity for the LSU AgCenter Rice Breeding Project.

CHAPTER 1 GENERAL INTRODUCTION

1.1 Association Mapping for Complex Traits

Modern genomic research through DNA sequencing efforts has laid the foundation to determine the role of selected genes that affect human health and economic productivity in plants and animals. Early efforts in genomics focused on traits governed by simple inheritance via one or two dominant nuclear genes (Botstein and Risch, 2003). However, many important life-history and fecundity characteristics in both humans and plants are controlled primarily by multiple genes that interact in varying degrees with the environment. A major effort for discovery of genes affecting human health has focused on development of a "haplotype map" (http://www.hapmap.org) that defines inherited blocks of molecular markers or haplotypes across the genome. Similar strategies have been reported recently using candidate genes in plants (Remington et al., 2001; Thornsberry et al., 2001; Olsen et al., 2004). Linkage disequilibrium (LD) mapping is used in both instances as the tool of choice to detect functional associations between haplotypes and selected traits of interest (Flint-Garcia et al., 2003). However, the LD strategy may not adequately account for variations in selection pressure, population structure, recombination rate and mating pattern that ultimately gave rise to high rates of false positives and lack of reproducibility across different populations (Weiss and Terwilliger, 2000; Terwilliger et al., 2002; Page et al., 2003; Pennisi, 2003). However, no consensus has been reached on the optimal approach for mapping complex traits (Risch, 2000; Botstein and Risch, 2003).

Mcharo et al. (2004) and Zhang et al. (2005) demonstrated the potential of Discriminant Analysis (DA), a multivariate statistical tool, as a possible alternative to LD association mapping technique in plants. Zhang et al. (2005) identified potential makers for 12 different agronomic traits using the molecular and phenotypic data of a collection of 218 diverse rice lines. The new

DA-selected markers were found to be mapped within or near traditional Quantitative Trait Loci (QTL) on the Cornell 2001 map (www.gramene.org). However, the results presented in this study were based on the phenotypic data of the plant material that was not evaluated in multiple environments. Hence, the potential of DA procedure to identify marker-trait associations in a narrow germplasm base rice collection and grown in multiple environments must be investigated.

Yu et al. (2006) proposed the mixed model procedure to account for spurious associations generated by population structure and familial relationships. Sucessful application of the mixed model procedure as an association genetic technique for candidate marker/gene identification was demonstrated in Maize (Yu et al., 2006), barley (Rostoks et al., 2006) and potato (Malosetti et al., 2007). However, Parisseaux and Bernardo (2004) argued that the mixed model analysis was primarily useful in identifying markers associated with triats with large effects and candidate gene discovery. Complex triat association mapping generally involves simultaneous use of more than one marker each explaining a portion of the variaion with in the trait. In this context, use of multiple regresson technique might be a good choice, as it facilitates several methods of selection of variables that can yield an optimal model (a combination of marker effects) with highest predictability (R^2 value). In addition, the inclusion of epistatic interactions occurring between alleles in QTL studies and association studies was emphasized by a number of rice researchers (Cao et al., 2001; Liao et al., 2001; Yu et al., 2002; Mei et al., 2003; Cui, 2005; Fan et al., 2005; Cui et al., 2006; Wan et al., 2006).

1.2 Successes and Challenges for Rice Breeding and Genomics Research

Rice is one of the major agricultural commodities in Louisiana with nearly 540,000 acres planted in 2003 (Childs, 2003) producing an estimated farm gate value of ~ $198,000,000

(http://www.lsuagcenter.com/agsummary/progressreport.aspx). Rice is also the primary staple food for nearly half of the world's population, thus there is a need for increasing production to fulfill the needs of an exploding global population. It is estimated that by the year 2025 nearly 4 billion, mostly poor, will consume rice as a basic food (http://www.knowledge bank.irri.org/factsheets/default.htm). In contrast to the world outlook, the U.S. rice industry experienced a reduction of 28.5 million cwt in milled rice exports in 2003 (Childs, 2004), and much of this reduction was due to milling quality of the rice grains. Thus, efforts to increase the rice grain yield and quality through genomics and breeding programs assume greater significance (Goff et al., 2004) that will ultimately benefit Louisiana and U.S. rice farmers.

Because milling yield and other complex characters pose a formidable challenge to rapid varietal development, there has been a growing emphasis on marker aided selection (MAS) to complement rice improvement programs. Apart from phenotypic data, MAS requires the identification of dependable molecular marker systems, generation or exploitation of variation within the crop of interest using DNA-based markers, and a means to link markers with economically important traits.

Public U.S. rice breeding programs from five rice-growing states (AR, LA, MS, MO and TX) participate annually in the cooperative Uniform Rice Nursery (URN). Each year a set of 200 elite rice inbred lines representing potential new varieties are evaluated in each state for plant height, maturity, grain yield, milling yield, and percent amylose content. Phenotypic data from these trials represent a valuable database from which to identify marker-trait associations for MAS and breeding programs in each of the participating states.

Aside from the practical aspects, rice is considered a model for genomic research in cereal grasses due to its relatively small genome size (~ 430 Mbp), diploid pairing, ability to take

up and express foreign genes, and the existence of large genetic stock centers in the U.S., Asia, Africa, and South America. These desirable attributes as a model plant and important food crop has led to the creation in 1997 of the International Rice Genomic Sequencing Project (IRGSP) (http://rgp.dna.affrc. go.jp /IRGSP/index.html) that produced a draft sequence of the *japonica* rice genome in 2002 which continues to post updates and progess reports on their website. Phase 3 sequences of chromosomes 1, 4, and 10 have been completed by IRGSP. The Beijing Genomics Institute in China and its collaborators have released a draft sequence of the corresponding *indica* genome including comparative single nucleotide polymorphism (SNP), complementary DNA (cDNA) and other data between the two subspecies (http://rise. genomics .org .cn/rice2 /index .jsp). In the U.S., the Rice Genome Project of The Institute of Genomic Research (TIGR; www.tigr.org) has contributed ~ 55 Mbp of DNA sequence to public databases and posted a tentative assembly of all 12 chromosomes (www.tigr.org). The Gramene website (www.gramene.org) serves as an information source for > 2000 mapped rice simple sequence repeat (SSR) markers, numerous annotated QTLs for different traits, comparative maps, and mutants.

International efforts to sequence the rice genome have laid the foundation to determine the location and function of genes for both basic and applied research interests. Certain genes with large effects in rice such as *Hd1* controlling for head rice (Yano et al., 2000) have been previously identified by positional cloning strategies, however, no one general approach to identifying genes that govern complex quantitative traits with moderate or small effects has been established. The data mining methods and results obtained during the proposed research will not only contribute to rice genomics in general, but also will complement other approaches such as microarray technology, proteomics, gene knockout studies, and RNAi-mediated gene silencing.

This will be accomplished by rapid identification of candidate markers and genes for evaluation and confirmation via modern technologies.

1.3 Single Nucleotide Polymorphisms (SNPs) in the Rice *Waxy* and *Alk* Loci

There are five basic types of DNA-based marker systems that include restricted fragment length plymorphisms (RFLPs), random amplification polymorphic differences (RAPDs), amplified fragment length plymorphisms (AFLPs), microsatellite or SSR markers and SNPs. Among these the SNP markers can maximally detect and exploit the variation between any two individuals of a given species. There are many SNPs reported in cereal crops such as rice, barley, maize etc. (Issiki et al., 1998; Bundock et al., 2004; Till et al., 2004) and their potential utility in plant functional genomics has been reported (Henikoff et al., 2003; Feltus et al., 2004). A genome-wide SNP identification effort has recently been published from two publicly available *indica* and *japonica* genome sequences *(*Feltus et al., 2004; Shen et al., 2004). Although additional research is needed, the SNP databases should prove invaluable for identifying polymorphisms in selected loci.

The amylose content of rice is mainly responsible for eating and cooking quality and has been reported to be governed primarily by the *waxy* (*Wx*) locus (Zhou et al., 2003; Bao et al., 2004; Yakanama et al., 2004). Two alleles at this locus (Wx^a and Wx^b) differ by a SNP at the first exon-intron donor splice site junction (Bao et al., 2004; Yamanaka et al., 2004). The Wx^a allele (AGGT) was shown to be predominant in non-waxy *indica* cultivars, whereas the Wx^b allele (AGTT) was common to the non-waxy *japonica* varieties (Issiki et al., 1998; Yamanaka et al., 2004). Amylose content was also found to be associated with the *alk* (alkali-spreading score, an indicator of the temperature at which the rice grain becomes gelatinous during cooking (McKenzie et al., 1983; Sano 1984) locus on chromosome 6. In addition, two SNP markers in

exon 8 of the *alk* gene have been characterized and found to be associated with percent amylose content (Fjellstrom et al., 2004).

1.4 Ecotilling via the CEL 1 Assay

Genetic variation can be generated by saturated mutagenesis (McCallum et al., 2000, Sallaud et al., 2003, Till et al., 2003, Henikoff et al., 2004) or exploited by natural existing variation in a population by the "Ecotilling" procedure (Comai et al., 2004). Ecotilling involves the identification of single or multiple nucleotide differences among a given set of plant material against a reference line. SNPs are particularly useful in studies dealing with narrow germplasm such as those in U.S. rice breeding programs. There are many tools available for the identification and validation of SNPs, but most are laborious, involving aligning of multiple sequences (cDNAs, ESTs or direct sequence of candidate genes), or using various homology search algorithms. Most SNP validation techniques are based on the polymerase chain reaction (PCR) that are combined with the use of costly equipment and chemicals (Pacey-Miller et al., 2003; Schmid et al., 2003). These validation techniques also require tedious primer designing and optimization steps. The CEL 1 endonuclease identifies potential SNPs via *in vitro* digestion of heteroduplex, double-stranded DNA molecules at the mismatch positions (Oleykowski et al., 1988; Till et al., 2004). SNP polymorphisms are detected by gel-based systems that reveal size differences between the reference and candidate fragments. The CEL 1 assay is a rapid, multi-allele detecting and semi high-throughput fine mapping technique (Comai et al., 2004). Utility of the CEL 1 nuclease assay in *Arabidopsis thaliana* was demonstrated by Colbert et al. (2001) and Comai et al. (2004), but the procedure required use of fluorescently tagged primers and an expensive SNP detection system. Therefore, a cost-effective alternative without involving expensive tagged primers and detection systems is desirable. Because the DNA sequence of the

rice genome of both *japonica* (Nipponbare) and *indica* (93-11) rice cultivars are publicly available (Sasaki et al., 2000; Yu et al., 2002), this information could be exploited for identification of targeted loci and design of primers for the CEL 1 assay. Two alternatives for the CEL 1 assays include development of a genome wide SNP map with markers distributed at a regular physical interval on all 12 rice chromosomes or characterization of SNPs in selected targeted gene/loci. The latter approach will be applied to grain yield and quality traits during the course of this research.

1.5 Aromatic Rice

The demand for premium priced aromatic rice in countries such as the United States, Canada, the Middle East, Europe and Australia, identifies the growing consumer preference for aromatic rice (Cordeiro et al., 2000; Jin et al., 2003). Currently, traditional aromatic rice growing countries such as Thailand, India and Pakistan are the only major international exporters for the premium "Jasmine" and "basmati" rice. The aroma in "Jasmine" and "basmati" rice varieties is mainly due to accumulation of 2-acetyl-1-pyrroline in leaf and seed tissues of the plant (Buttery et al., 1983). Bradbury et al. (2005a) reported that a stretch of mutations (a SNP haplotype) in the exon7 of the *fgr* gene, which encodes for the betaine aldehyde dehydrogenase (BAD2) enzyme, are responsible for the aroma in rice. During the transcription of the *BAD2* gene, this haplotype would encode for a premature stop signal resulting in the production of a non-functional truncated BAD2 enzyme (Bradbury et al., 2005a). This truncated BAD2 enzyme is deficient in three conserved protein motifs needed for its substrate binding activity and subsequently results in the accumulation of 2-acetyl-1-pyrroline (Bradbury et al., 2005a). Thus, detection of the haplotype alleles in the *BAD2* gene would enable discrimination between aromatic and non-aromatic rice and thus assist marker-assisted introgression of the aromatic trait into local rice

varieties.

The aromatic rice breeding program at the Rice Research Station, Crowley, Louisiana, primarily involves cooking rice grains for distinguishing the aromatic rice from non-aromatic rice (Sha et al., 2000). However, the use of this method is limited by the need for technical expertise and is low throughput. Recently, Bradbury et al. (2005b) demonstrated the utility of allele-specific PCR amplification assay for the *BAD2* gene in Australian temperate *japonica* aromatic and non-aromatic germplasm. However, there is a need for validating the use of *BAD2* gene haplotype based markers in a marker-assisted selection program for US aromatic rice varieties which are mainly derived from topical *japonica* and *indica* germplasm.

1.6 Red Rice Weed Control by ALS-inhibiting Herbicides

Red rice (*Oryza sativa* L.) is the most problematic weed in rice fields causing significant yield losses in the U.S. (Gealy et al., 2003). At the seedling stage this weed is virtually indistinguishable from commercial white rice, and control of the red rice weed in rice fields has been until recently, a near impossible task (Gealy et al., 2003). Nevertheless, herbicides such as imidazolinones (imazethapyr) that inhibit acetohydroxy synthase (AHAS) or acetolactate synthase (ALS) activity can provide effective control of this noxious weed in rice fields (Steele et al., 2002). AHAS/ALS is one of the key enzymes in the biosynthetic pathway of the branched chain amino acids and is encoded by the acetolactate synthase *ALS* gene (Tan et al., 2006). Commercial Clearfield rice varieties were developed by inducing two mutations in the *ALS* gene at the 1880 bp and 1883 bp positions, causing $S_{653}D$ and $G_{654}E$ substitutions in the normal or wild-type ALS gene product (Tan et al., 2005; Tan et al., 2006). The altered ALS enzyme of the Clearfield rice varieties fails to bind with imidazolinone herbicides, thus conferring resistance (Tan et al., 2005; Tan et al., 2006).

Nearly 30% of the total rice cultivated area in Louisiana was planted with the imazethapyr-resistant CL161 variety in 2005 (Gealy et al., 2006). However, increasing dependency on imazethapyr use and the possible gene-flow of herbicide resistance gene to red rice via outcrossing are two major concerns pertaining to Clearfield rice cultivation. The red rice genome has 12 chromosomes as that of cultivated rice and due to this genetic similarity, a high outcrossing frequency of up to 0.17% between red rice and rice has been observed by (Zhang et al., 2005). Furthermore, many examples of rice x red rice hybridization events (Chen et al., 2004; Messeguer et al., 2004; Song et al., 2004; Wang et al., 2006) and transfer of herbicide resistance from cultivated rice to red rice biotypes (Estorminos et al., 2002 ; Madsen et al., 2002; Gealy et al., 2003) have been reported. Considering the evidence of crop x wild hybridization and gene-flow between rice and red rice, there is an urgent need for the development of tools which can monitor the outcrossing events. DNA based herbicide resistance assay techniques have received more emphasis than conventional herbicide resistance assays such as pollen germination, leaf disc and AHAS enzyme activity assays (Corbett and Tardiff, 2006). However, application of DNA based techniques for assaying imazethapyr resistance in Clearfield rice has not been demonstrated. Genotyping of $S_{653}D$ and $G_{654}E$ SNP alleles in Clearfield x red rice hybrids through allele-specific PCR could provide direct evidence for the *ALS* gene transfer from Clearfield rice to red rice.

1.7 Research Objectives

(1) Evaluate the potential of Discriminate Analysis (DA) procedure to detect informative molecular markers associated with plant height, grain yield and quality traits (percent amylose content, percent head rice and percent total rice) in the 2000 URN inbred lines.

(2) Create and evaluate a mixed model-regression procedure that identifies main and epistatic effects by standard hypothesis testing and Bayesian information criteria in a multivariate format for agronomic traits evaluated in the 2000 URN trial

(3) Develop a simple, rapid, efficient, and cost-effective alternative to standard Ecotilling for SNP discovery and genotyping in rice that can be easily adapted to small or medium-sized laboratories.

(4) Develop and evaluate PCR-based assays for high-throughput SNP screening of aromatic rice.

(5) Develop and evaluate PCR-based assays for high-throughput SNP screening of imazethapyr herbicide resistant rice.

CHAPTER 2 IDENTIFICATION OF MOLECULAR MARKERS ASSOCIATED WITH GRAIN YIELD, QUALITY AND PLANT HEIGHT IN RICE USING DISCRIMINANT ANALYSIS

2.1 Introduction

2.1.1 Mapping of Quantitative Trait Loci (QTL)

The majority of traits related to fecundity and adaptation in plants and animals are governed by multiple genes that interact in varying degrees to changes in the environment that produce a continuous phenotypic response. For QTL mapping in plants, the initial task typically requires screening potential parents for polymorphic molecular markers and the subsequent production of segregating or recombinant inbred populations. Loci or intervals are then defined on pre-existing genetic maps that are linked with a trait of interest by single-factor ANOVA (Jermstad et al., 2003), regression (Wang et al., 2004), interval (Lincoln et al., 1992) or other standard mapping procedures. For complex quantitative traits in rice, ≥ 300 recombinant inbred lines are generally evaluated that require three to four years to develop. Moreover, relatively few meiotic events in F_2 or recombinant inbred lines limit the power of linkage analysis to effectively dissect traits governed by multiple loci, and examination of genetic diversity in diploids is restricted to only two alleles segregating per locus (Flint-Garcia et al., 2003). Near-isogenic lines have been used to identify and clone genes with large effect in rice (Yano et al., 2000) and *Arabidopsis thaliana* (Johanson et al., 2000) related to flowering, but this approach is time consuming and may not be efficient for complex loci with moderate or small effects. Production of large segregating or intermating populations can promote recombination, but substantial investments in time, labor, and financial resources over multiple generations are required.

2.1.2 Association/ LD Mapping in Humans

Due to the limited power and resolution of traditional QTL mapping research, association

or LD mapping was established based on the nonrandom association of different alleles across two or more loci in a population (Hill and Robertson, 1968). Association mapping reportedly enjoys increased precision and resolution over interval or QTL methods by capturing information contained in multiple recombination events over time in natural or selected populations (Ewens and Spielman, 2001). LD is traditionally measured between pairs of loci to calculate differences in observed and expected haplotype frequencies (Garcia et al., 2003). This approach has been used extensively in human studies for the discovery of markers and genotypes that underlie simple Mendelian traits with subsequent fine mapping and positional cloning of genes for certain disorders such as cystic fibrosis (Kerem et al., 1989), Alzheimer's disease (Martin et al., 2000), psoriasis (Trembeth et al., 1997), and colorectal cancer (Nishisho et al., 1991; Wooster et al., 1995). However, the majority of human diseases such as diabetes, stroke, heart disease, depression, and asthma are affected by multiple genes and environmental conditions. To identify those factors affecting complex human disorders and other traits, the International HAPMAP project was created in 2002 as an international collaboration among scientists in six countries including the U.S. (http://www.hapmap. org/abouthapmap.html.en). This project aims to identify single nucleotide polymorphism (SNP) loci inherited together in small chromosomal blocks or haplotypes across the genome to facilitate LD approaches to gene identification and characterization. However, a general lack of results and reproducibility has led some scientists to question this approach, the future impact of the HAPMAP project and the LD methodology in general (Couzin, 2002; Hirschhorn et al., 2002; Trikalinos et al., 2004).

2.1.3 LD Studies in Plants

It is important to emphasize that LD mapping research in plants has used the same general methods described above for human populations. Association genetic mapping in plants

has to date been conducted primarily in maize and *Arabidopsis thaliana*. LD values and patterns in maize were found to vary substantially among populations with different selection and developmental histories (Labate et al., 2000; Tenaillon et al., 2001; Clark et al., 2004). Rapid decline in LD was observed within a 1500 bp intragenic region of four candidate genes for height and flowering among a diverse collection of maize inbred lines (Remington et al., 2001; Thornsberry et al., 2001). SSR markers produced larger LD regions than SNP markers across the genome. Using elite maize inbred lines, Ching et al. (2002) found, in contrast to diverse germplasm, little or no reduction in LD values over a 300-500 bp range for 18 candidate genes. *A. thaliana* as a self pollinating species produced allelic associations over a much greater region (~ 250 kb) than in outcrossing species such as maize (Hagenblad and Nordborg, 2002; Nordborg et al., 2002). LD mapping was used recently to identify a serine substitution in the candidate CRY2 photoperiod receptor gene in *A. thaliana*, presumably responsible for the A(S) early flowering phenotype (Olsen et al., 2004).

Garris et al. (2003) characterized LD in the candidate region of *xa5,* a recessive gene conferring race-specific resistance to bacterial blight disease in rice. Thirteen segments from a 70-kb candidate region in 114 landrace accessions were sequenced along with five additional segments from an adjacent 45-kb region in resistant accessions. The results showed significant LD up to 100 kb between sites that suggested genome-wide scanning may be feasible for markers that are associated with simple and complex traits. The candidate gene approach was recently employed in LD mapping of QTLs for disease and maturity traits in tetraploid potato (Gebhardt et al., 2004; Simko et al., 2004).

2.1.4 Preliminary Studies using Discriminant Analysis for Marker-trait Associations in Rice

Discriminant Analysis (DA) is a multivariate statistical procedure first developed by

Fisher (1936) that involves the creation of two "training samples" derived from, in the case of the proposed research, selected inbred or recombinant lines with contrasting phenotypic values. From DNA profiles of all lines included in the experiment, markers are identified by DA that best differentiate between the training samples. An error rate, referred to as "percent correct classification", is calculated to measure ability of the markers to correctly assign individual lines to the training samples. With high levels of correct classification, an association between marker(s) and phenotype or plant trait is inferred.

DA has been used in plant research for diversity analysis of wild emmer wheat (Fahima et al., 2002), identification of drought-tolerant Kentucky bluegrass cultivars using morphological criteria (Ebdon et al., 1998) and to estimate position and effects of QTLs in simulated full and half-sib families (Gilbert and Le Roy, 2003). Microarray expression profiling studies have utilized DA to identify genes and gene clusters associated with human diseases (Mendez et al., 2002; DePrimo et al., 2003; Kari et al., 2003; Musumarra et al., 2003) and to detect protein coding regions in genomic sequences (Zhang, 1998; Zhang et al., 2002). Finally, the DA procedures was recently used to accurately assign unrelated sweet potato clones using AFLP markers to groups defined by high and low dry matter content (Mcharo et al., 2004).

The potential of DA has been investigated, along with complementary procedures described in this proposal, to identify SSR markers putatively associated with grain yield and quality characteristics in rice (Zhang et al., 2005; Kadaru et al., unpublished results). In the first of two cooperative studies with Drs. Xu and McCouch of Cornell University, a total of 218 U.S. and Asian inbred lines were grown in single-row plots in 1996 and 1997 near Alvin, TX. Three measurements per line were collected for 12 traits that included 1000 grain weight, tiller number, grain length-width ratio, and all known components of yield and grain quality.

DNA profiles were obtained for all 218 lines using 60 SSR and 114 RFLP markers selected randomly over the rice genome by Drs. Xu and McCouch. To evaluate DA for marker-trait associations, the following procedures were carried out (1) Transform phenotypic data if necessary to normal distribution by log, square root or other methods (2) Use 1, 2, or 3 standard deviations of trait distribution to create user-defined "training samples". For molecular data analysis: (1) Transform raw marker data to identify individual alleles (2) Fill in missing marker data using the Multiple Imputation procedure (SAS Institute, ver. 9.0) (3) Perform molecular analysis of variance (AMOVA, Excoffier et al. 1992) of marker profiles to test differences between training samples using Arlequin software (Schneider et al., 2002) (4) Identify potential population structure by genetic distance (www.powermarker.net) or model-based (www.stats.ox.ac.uk/ ~pritch/home.html) methods (5) Perform parametric Discriminant Analysis (PROC STEPDISC, SAS Institute, ver. 9.0) to identify marker(s) that best differentiate training samples within each subpopulation (6) Use "nonparametric method" within DISCRIM procedure (SAS Institute, ver. 9.0) to perform "K-nearest-neighbor" classification of inbred lines into pre-defined groups) and (7) Calculate percent correct classification with "crossvalidate" option within PROC DISCRIM procedure (SAS Institute, ver. 9.0). SSR and RFLP markers were located on the Rice-Cornell SSR 2001-1 and Rice-Cornell RFLP 2001-2 genetic maps (www.gramene.org).

However, to perform the DA analysis Zhang et al. (2005) have used a rice collection that was evaluated at a single location and had a wide germplasm base. In this scenario, DA successfully identified markers that can distinguish between high and low classes of 12 different agronomic traits. Therefore, the potential of DA procedure to identify marker-trait associations in

a narrow germplasm base rice collection, which were also evaluated in multiple environments, was investigated in this current study.

2.2 Materials and Methods

2.2.1 Plant Material

The first plant material consisted of 192 elite rice inbred lines that were evaluated in five rice-growing states (AR, LA, MS, MO and TX) during the year 2000 under the state cooperative Uniform Rice Nursery (URN) trials. Details for the second set of plant material consisting of 57 diverse rice lines are given in section 4.2.1.

2.2.2 Phenotypic Data

The phenotypic data consisted of observations collected in the year 2000 for the following traits in 192 elite URN lines across five US states (AR, LA, MS, MO and TX). Rice grain yield (measured in pounds / acre), percent amylose content (proportion of amylose to amylopectin content of a rice grain), percent total rice (ratio of weights of hulled whole and broken rice grains to that of total de-hulled rice grain sample, expressed as a percentage) and percent head rice (percentage ratio of milled whole rice grains to that of total de-hulled rice grains). The data for percent amylose content were obtained from the AR and TX locations.

2.2.3 Molecular Data

The molecular data consisted of SSR marker profiles for these same lines using 95 SSR primer sets, provided by Drs. Xu and McCouch, Plant Breeding Department, Cornell University. The marker genotypic data were converted into allele data containing 579 alleles with an average of six alleles/locus. Heterozygote individuals were treated as missing data. Missing data across these 579 alleles were imputed by using the multiple imputations procedure (PROC Mi) in SAS software v. 9.1.0. A total of five separate imputations were computed for the 579 allele data, and

the five datasets sets were used for the subsequent DA analysis.

2.3 Results and Discussion

2.3.1 Percent Amylose Content

The rice *waxy* locus is well characterized with known DNA sequence and SNP markers (Olsen and Purugganan, 2002; Fellstom et al., 2004). If the DA method described in this proposal exhibits sufficient power and precision, successful identification of markers associated with percent amylose content should be possible. A total of 192 US lines from the 2000 AR and TX URN databases were used to detect potential SSR marker-amylose content associations. Using the DA procedure, the well known RM190 marker within the *waxy* gene (Chen et al., 2004), along with three additional loci (RM231, RM25 and RM225) were selected by DA as genetic factors contributing to percent amylose content in the 2000 URN germplasm for AR and TX (Table 2.1). The RM190, RM231, RM25 and RM225 loci have been previously mapped to chromosome 6 (6.7 cM), chromosome 3 (15.7 cM), chromosome 8 (52.2 cM) and chromosome 6 (26.2 cM), respectively, on the Cornell Rice SSR 2001 map (www.gramene.org). The DA-selected locus RM225 was found within a reported Qualitative Trait Locus (QTL) for amylose (Septiningish et al., 2002; 0-33.6cM on Cornell Rice SSR 2001 map) and also within the *amy6* QTL (Aluko et al., 2004; 6.7-37.0cM on Cornell Rice SSR 2001 map). Thus, these newly reported SSR DA markers could be potential candidate loci for percent amylose content.

As a further validation, the DA selected markers were evaluated with an additional set of plant material consisting of 57 U.S. and Asian inbred lines with known percent amylose content. Markers RM25, RM225, RM231 and RM190 from our previous DA analysis were selected for fingerprinting of the above lines. SSR profiles were generated for all the 57 lines (Table 2.2) and regressed with their percent amylose content data. Examples of SSR profiles generated for 38 lines are shown in Figure 2.1.

Table 2.1 Discriminant Analysis -selected markers for percent amylose content

	State		
	Arkansas	Texas	Combined
DA markers / Individual R^2 value[1] 15% Training Sample	RM225_132_2[4]/0.5859	RM190_122_5/0.586	RM225_132_2/0.581
	RM190_122_5/0.5054	RM437_252_1/0.2365	RM190_122_5/0.5676
	RM510_119_2/0.374	RM234_135_2/0.2094	RM190_120_4/0.1154
	RM416_113_3/0.0752	RM214_154_8/0.1489	RM72_159_3/0.0357
	RM418_280_4/0.0554	RM3430_211_6/0.113	RM162_205_3/0.0357
	RM72_159_3/0.0363	RM422_387_8/0.1084	RM229_127_5/0.0357
	RM340_117_2/0.0363	RM475_185_1/0.0928	RM109_100_12/0.0175
	RM162_205_3/0.0339	RM190_120_4/0.0782	RM7_167_2/0.0175
	RM273_209_3/0.0339	RM481_171_11/0.0729	RM21_153_8/0.0175
	RM7_167_2/0.0178	RM1359_166_7/0.0377	
	RM3912_205_6/0.0178	RM437_274_5/0.0374	
	RM231_193_7/0.0178	RM623_348_4/0.0185	
	RM239_141_1/0.0178	RM7_167_2/0.0185	
	RM3912_207_7/0.0178	RM214_142_5/0.0185	
	RM16_217_9/0.0178	RM149_245_6/0.0176	
	RM481_208_16/0.0178	RM279_162_5/0.0176	
	RM109_100_120.0167	RM162_205_3/0.0172	
	RM169_168_30.0167	RM481_156_6/0.0065	
	RM317_161_6/0.0044	RM72_189_8/0.0031	
	RM482_186_3/0.0044		
Combined R^2 value[2]	0.6055	0.5829	0.6026
% Correct Classification[3]	100	100	95.55
DA markers /R^2 value[1] 5% Training Sample	RM231_181_2/0.8182	RM25_139_2/0.6481	RM190_122_5/0.6667
	RM21_155_9/0.0431	RM317_161_6/0.1307	RM317_161_6/0.1111
Combined R^2 value[2]	0.3607	0.1407	0.1575
% Correct Classification[3]	100	100	100

[1] Individual R^2 values calculated from Pearson correlation coefficient; [2] Combined R^2 value calculated from multiple regression (PROC REG, SAS Institute, ver. 9.0); [3] Percent correct classification were calculated by leave-one-out validation with in the training samples; [4] The first part of the DA marker denotes the SSR marker, the second part represents the observed allele size in bp and the third part stands for the allele number of the SSR locus.
All the sets of markers identified by DA procedure are reordered based on their individual R^2 values, not by relative contribution to the discriminant rule. Individual R^2 values are calculated for respective the Training samples only whereas the Combined R^2 values are calculated considering all individuals.

Table 2.2 Summary of percent amylose content and DA-selected SSR marker allele sizes for 57 rice lines

Accession	NPGS/GRIN Number	Origin	Amylose Class[a]	RM25[d]	RM190[d]	RM225[d]	RM231[d]
Bolivar	PI 628791	USA	High	142[e]	107[e]	119[e]	181[e]
Cocodrie	PI 6063631	USA	High	144	125	133	192
Cheniere	ND[b]	USA	High	144	123	133	192
Cypress	PI 9700184	USA	Intermediate	144	123	133	192
Dixiebelle	PI 595900	USA	High	140	107	133	181
Fortuna	PI 275448	USA	Low	140	119	128	192
Fortuna Moredo	PI 431075	USA	ND[b]	140	121	118	187
Francis	PI 632447	USA	Intermediate	144	123	134	181

Table 2.2 (continued)

Accession	NPGS/GRIN Number	Origin	Amylose Class[a]	RM25 [d]	RM190[d]	RM225[d]	RM231[d]
Golden Steve	PI 612579	USA	Low	137	119	128	185
Jackson	PI 572412	USA	Intermediate	144	123	130	181
J-85	PI 595927	USA	Low	142	118	138	185
Kokubelle	PI 612581	USA	(Intermediate)[c]	144	123	135	181
Lacrosse	PI 389966	USA	Low	140	123	138	181
LaGrue	PI 568891	USA	Intermediate	144	123	135	181
L-205	PI 608664	USA	High	144	107	137	181
Maxwell	PI 612582	USA	(Low)[c]	140	119	129	192
Millie	PI 538354	USA	Intermediate	144	125	140	181
Neches	PI 633972	USA	Glutinous	144	125	137	181
Sierra	PI 633623	USA	High	144	108	137	181
Tor Tora	PI 431150	USA	ND	144	115	140	181
Tebonnet	PI 487195	USA	Intermediate	144	104	140	187
Tsuri Mai	PI 612580	USA	ND	140	123	130	189
TX 2172	TX 2172	USA	Glutinous	148	121	132	183
TX 3043	TX 3043	USA	High	140	109	138	181
TX 4175	TX 4175	USA	High	140	109	138	181
Waxy M101	PI 506223	USA	Glutinous	140	125	132	189
Wells	PI 612439	USA	Intermediate	144	115	142	181
EPAGRI 106	ND	Brazil	ND	140	111	137	181
Gui Chow	ND	China	High	142	109	123	185
Hsuan Jha	PI 160829	China	High	142	109	140	185
E Che Goo	PI 389570	China	High	144	109	123	185
TeQing	PI 536047	China	High	144	109	124	192
ZHE733	PI 629016	China	High	142	109	123	185
Nipponbare	PI 514663	Japan	Low	140	123	135	192
Yuukara	PI 341937	Japan	Low	140	123	137	192
Chong Kuc Tae	CI 12284	S Korea	Glutinous	140	121	132	181
Dawebyan	PI 222405	Myanmar	High	142	113	143	187
HB 1	ND	Philippines	Glutinous	133	121	145	183
IR 29	PI 393986	Philippines	Glutinous	144	121	143	185
IR 532-1-33	PI 388332	Philippines	High	144	107	123	168
IR 1561-243-5-6	PI 385340	Philippines	High	144	107	142	168
Chin Feng Hsuch	PI 389048	Taiwan	High	142	109	140	183
Hung Chu Shien	PI 389073	Taiwan	High	142	109	142	192
Taipei Woo Co	PI 294397	Taiwan	High	144	109	140	181
Dhariyal	PI 297569	Bangladesh	High	130	107	123	185
IR 36	PI 408586	Philippines	High	144	109	140	185
ARC 10764	PI 373576	India	High	130	104	124	185
Basmati	PI 173923	India	Intermediate	140	118	143	181
Ratna	PI 413980	India	High	142	107	124	171
Achhame	PI 400028	Nepal	Intermediate	144	119	124	189
Kakani2	PI 400020	Nepal	Intermediate	144	118	124	189
Fine Mushkan	PI 385765	Pakistan	High	127	121	124	185
Hansraj	PI 385815	Pakistan	High	127	111	124	185
Palman	PI 385814	Pakistan	High	127	104	124	185
Sufaida	PI 385819	Pakistan	High	127	109	124	185
Mad/S	PI 385323	Rwanda	High	142	109	140	192

[a] Amylose class where apparent amylose content falls into the following categories: Glutinous = 0 to 5%, Low = 5 to 19%, Intermediate = 19 to 23%, and High > 23%. [b] ND = no data available. [c] Data as provided on US Plant Variety Protection description of accession. [d] Denotes the name of the SSR marker. [e] Represents the obsereved band size in bp for the SSR locus.

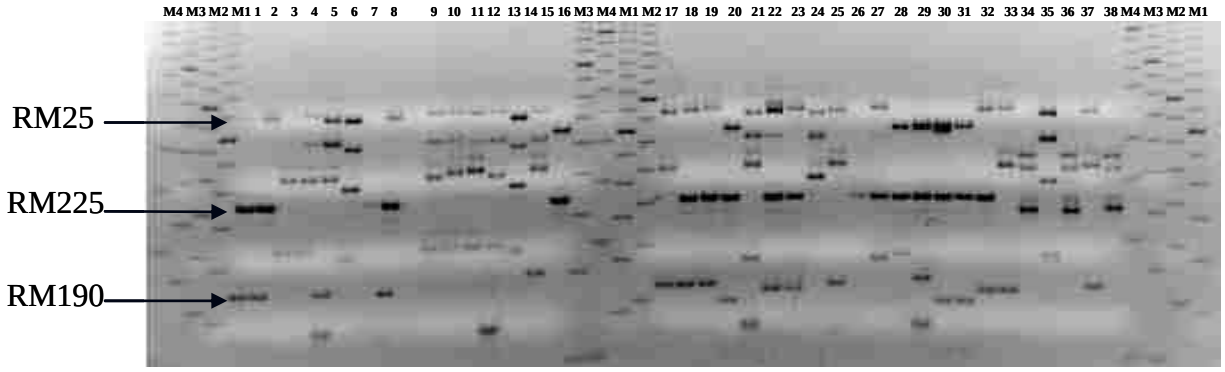


Figure 1 Resolving of RM190, RM225 and RM25 SSR marker PCR amplified products on 6% non-denature PAGE by multiplex-loading. M1 = first loading of Sigma PCR 20 bp low ladder, M2, M3 and M4 = subsequent 3 loadings of Sigma PCR 20 bp low ladder at 60 minute interval.

Out of 45 observed alleles from RM190, RM25, RM225 and RM231 markers, 38 alleles explained 89.9% (Pearson Correlation coefficient $R^2 = 0.899$; P value of <0.001) of the observed variation in percent amylose content among the 57 lines (Table 2.2). The best multiple regression models with dependent variables ranging from 1-10 along with individual adjusted R^2 values were determined (Table 2.3). The best single independent allele that could explain maximum variation was RM225_132 (Adj-$R^2 = 0.2477$, P value = <0.0001). The RM231 locus was also identified in the AR and TX germplasm as contributing to percent amylose content (Table 2.1, Arkansas lines, 5% training sample).

2.3.2 DA Analysis for Percent Head Rice, Percent Total Rice, Plant Height, Heading Date and Grain Yield

The potential of Discriminate Analysis (DA) as an association genetics tool to select markers in adapted US rice germplasm was explored. All the steps required for DA procedure were performed from data of 192 elite US inbred lines planted in year 2000 and candidate SSR

Table 2.3 Multiple regression Adj-R^2 values obtained for the DA-selected alleles associated with percent amylose content among 57 rice lines using stepwise selection from 1 to 10 variable models

No. of alleles in the model	Adj-R^2 value	Calculated p-value	Order of the DA-selected marker alleles in the model
1	0.2477	<.0001	Intercept RM225_132
2	0.3309	<.0001	Intercept RM225_132 RM225_145
3	0.4466	<.0001	Intercept RM190_109 RM225_132 RM225_145
4	0.5552	<.0001	Intercept RM190_107 RM190_109 RM225_132 RM225_145
5	0.5999	<.0001	Intercept RM190_107 RM190_109 RM225_132 RM225_145 RM25_127
6	0.6291	<.0001	Intercept RM190_107 RM190_109 RM225_132 RM225_145 RM25_127 RM25_140
7	0.6581	<.0001	Intercept RM190_107 RM190_109 RM225_132 RM225_133 RM225_145 RM25_127 RM25_140
8	0.6773	<.0001	Intercept RM190_107 RM190_108 RM190_109 RM225_132 RM225_133 RM225_145 RM25_127 RM25_140
9	0.7051	<.0001	Intercept RM190_107 RM190_108 RM190_109 RM225_132 RM225_133 RM225_137 RM225_145 RM25_127 RM25_140
10	0.7262	<.0001	Intercept RM190_107 RM190_108 RM190_109 RM190_113 RM225_132 RM225_133 RM225_137 RM225_145 RM25_127 RM25_140

Table 2.4 Individual R^2 values for 10 best variables/alleles calculated using simple linear regression

DA-selected SSR marker allele	Adj-R^2	p-value
RM225_132	0.2477	<.0001
RM25_140	0.1117	0.0083
RM190_109	0.0994	0.0123
RM25_142	0.0769	0.025
RM190_121	0.0751	0.0265
RM190_107	0.0735	0.0278
RM225_145	0.0661	0.0353
RM25_148	0.0661	0.0353
RM25_133	0.0661	0.0353
RM231_183	0.0592	0.0438

markers for percent head rice, percent total rice and gross yield across five states (AR, LA, MS, MO and TX) were identified. DA markers were identified for each trait for all the 5% training samples (TS), 15% TS and 30% TS. The state summaries of the DA-selected markers for 15% and 5% TS are given in Tables 2.5-2.10.

DA markers produced high levels of percent correct classification and R^2 values indicating potential value of this approach in marker-assisted selection. The DA-selected markers were primarily location-specific, most likely emphasizing different environment conditions prevailing in these four states.

2.3.3 Comparison of DA and QTL results

Many QTLs were reported in the past for the traits percent head rice, percent total rice (Tan et al., 2001; Mei et al., 2002; Septiningsih et al., 2003; Aluko et al., 2004) and grain yield (Xiao et al., 1996a; Xiao et al., 1996b; Tan et al., 1997; Li et al., 2000; Ishimaru et al., 2001; Zhuang et al., 2001; Cai et al., 2002; Hittalmani et al., 2002; Hua et al., 2002; Lafitte et al., 2002; Venuprasad et al., 2002; Xing et al., 2002; Cui et al., 2003; Hittalmani et al., 2003; Hua et al., 2003; Ishimaru et al., 2003; Li et al., 2004). New DA markers identified for percent head rice, percent total rice, plant height, heading date and grain yield were compared with these traditional (QTL) loci with the Rice Cornell SSR 2001 genetic map (Figure 2.2, 2.3). Mapping all these QTLs has clearly shown the overlapping regions among the grain quality traits and yield, which was also evident in the DA analysis. DA markers mapped within or nearby previously reported QTLs emphasizing the robustness of the DA procedure (Table 2.6, 2.7). Certain DA markers were found distant to the reported QTLs, suggesting potential new markers for the corresponding traits.

Table 2.5 Discriminant Analysis-selected markers for percent head rice, percent total rice, plant height, heading date and grain yield from 192 lines of 2000 URN field trials across all five states

		Trait				
		Percent Head Rice	Percent Total Rice	Plant Height	Heading Date	Grain Yield
Combined	DA markers / Individual R^2 value[1] 15% Training Sample	[4]RM106_287_1/0.2307 RM341_142_4/0.2279 RM481_156_6/0.1835 RM437_274_5/0.105 RM475_185_1/0.1049 RM120_184_4/0.0684 RM279_162_5/0.0676 RM13_149_5/0.0316 RM250_173_8/0.0287 RM481_219_18/0.0161	RM437_274_5/0.2465 RM279_158_3/0.0959 RM55_234_4/0.0769 RM144_256_11/0.0644 RM169_168_3/0.0514 RM333_165_2/0.0204	RM431_254_4/0.8556 RM109_96_8/0.0481 RM5_126_7/0.0308 RM232_153_5/0.0236 RM214_142_5/0.0151	RM510_119_2/0.2694 RM248_81_3/0.2593 RM25_139_2/0.1488 RM25_141_3/0.1335 RM623_348_4/0.0794 RM149_242_3/0.0781 RM3912_195_5/0.0473 RM333_189_8/0.0462 RM482_189_4/0.0303 RM214_142_5/0.0303 RM341_139_3/0.0149 RM475_235_5/0.0149 RM178_115_1/0.0149 RM333_204_11/0.0149 RM231_187_5/0.0149 RM184_217_3/0.0149 RM5752_138_3/0.0024 RM162_240_11/0.0011 RM1167_175_3/0.0004 RM72_189_8/0	RM25_139_2/0.1029 RM498_217_4/0.0756 RM341_174_7/0.0634 RM481_171_11/0.0517 RM481_159_7/0.0438 RM315_137_2/0.0376 RM184_204_1/0.0197 RM25_145_6/0.0195 RM316_196_2/0.0141 RM408_117_1/0.013 RM1167_171_1/0.0119 RM21_129_2/0.0094 RM144_256_11/0.006 RM109_94_6/0.0052 RM169_194_5/0.0045 RM109_99_11/0.004 RM142_237_2/0.0032 RM5_114_5/0.0011 RM210_151_7/0.0005 RM16_217_9/0.0002
	Combined R^2 value[2]	0.2977	0.0965	0.3463	0.3549	0.2653
	% Correct Classification[3]	100	100	100	99.99	99.96
	DA markers / Individual R^2 value[1] 5% Training Sample	RM437_252_1/0.4571 RM1189_176_2/0.102 RM214_146_6/0.0026	RM437_274_5/0.2465 RM279_158_3/0.0959 RM55_234_4/0.0769 RM144_256_11/0.0644 RM169_168_3/0.0514 RM333_165_2/0.0204	RM431_254_4/0.8264 RM109_96_8/0.0455	RM3912_193_4/0.0654 RM144_253_10/0.0342 RM273_209_3/0.0258 RM162_236_9/0.0245 RM341_142_4/0.0157 RM149_241_2/0.0135 RM181_239_1/0.0103 RM422_385_7/0.0064 RM333_161_1/0.0047	RM341_142_4/0.2747 RM437_274_5/0.2525 RM431_250_2/0.1905 RM109_100_12/0.1111
	Combined R^2 value[2]	0.0743	0.0965	0.1168	0.2251	0.1766
	% Correct Classification[3]	100	100	100	100	100

[1] Individual R^2 values calculated from Pearson correlation coefficient; [2] Combined R^2 value calculated from multiple regression (PROC REG, SAS Institure, ver. 9.0); [3] Percent correct classification were calculated by leave-one-out validation with in the training samples; [4] The first part of the DA marker denotes the SSR marker, the second part represents the observed allele size in bp and the third part stands for the allele number of the SSR locus. All the sets of markers identified by DA procedure are reordered based on their individual R^2 values, not by relative contribution to the discriminant rule. Individual R^2 values are calculated for respective the Training samples only whereas the Combined R^2 values are calculated considering all individuals.

Table 2.6 Discriminant Analysis-selected markers for percent head rice, percent total rice, plant height, heading date and grain yield from 192 lines of 2000 URN field trials evaluated in Arkansas

		Trait				
		Percent Head Rice	Percent Total Rice	Plant Height	Heading Date	Grain Yield
Arkansas	DA markers / Individual R^2 value[1] 15% Training Sample	[4]RM250_177_10/0.202 RM418_283_5/0.1296 RM21_139_4/0.1011 RM3431_150_2/0.0799 RM149_241_2/0.0667 RM333_165_2/0.0664 RM234_141_3/0.0565 RM214_146_6/0.038 RM1189_176_2/0.0333 RM228_117_6/0.0188 RM144_261_13/0.0164 RM169_172_4/0.0164 RM109_95_7/0.0026	RM279_160_4/0.2961 RM16_167_1/0.1148 RM149_240_1/0.092 RM279_162_5/0.0842 RM169_166_2/0.0546 RM478_202_3/0.0229 RM478_236_8/0.0175 RM149_243_4/0.011 RM403_239_1/0.0017	RM431_254_4/0.4613 RM5_126_7/0.0647 RM109_96_8/0.0356 RM229_127_5/0.0356 RM144_261_13/0.0175 RM284_146_3/0.0175 RM232_153_5/0.0175 RM109_87_2/0.0154 RM169_164_1/0.0154 RM214_142_5/0.0154 RM474_275_8/0.0154 RM1359_166_7/0.0068 RM481_219_18/0	RM3430_211_6/0.1904 RM478_200_2/0.1579 RM3912_195_5/0.0991 RM5_128_8/0.0942 RM162_236_9/0.0699 RM232_141_2/0.068 RM120_186_5/0.0579 RM431_254_4/0.0566 RM482_189_4/0.0447 RM21_151_7/0.0293 RM481_219_18/0.0293 RM149_241_2/0.0224 RM1189_188_7/0.0187 RM431_242_1/0.0144 RM7_167_2/0.0144 RM333_204_11/0.0144 RM3431_148_1/0.0028 RM149_244_5/0.0016 RM17_183_2/0.0001 RM181_241_2/0.0001	RM279_164_6/0.2233 RM136_100_3/0.1557 RM317_165_7/0.09 RM3431_148_1/0.0842 RM119_170_4/0.0357 RM25_145_6/0.0347 RM623_348_4/0.034 RM144_253_10/0.0309 RM234_135_2/0.0252 RM228_117_6/0.0175 RM190_105_1/0.006
	Combined R^2 value[2]	0.3249	0.2806	0.351	0.3746	0.2503
	% Correct Classification[3]	100	100	100	100	100
	DA markers / Individual R^2 value[1] 5% Training Sample	RM3912_191_3/0.264 RM418_283_5/0.1835 RM120_184_4/0.1173 RM190_113_3/0.1107 RM312_94_1/0.1107 RM169_164_1/0.053	RM279_160_4/0.56 RM279_162_5/0.1316 RM481_219_18/0.0833 RM210_151_7/0.0014	RM431_254_4/0.5483 RM109_96_8/0.0496 RM284_146_3/0.0496 RM169_164_1/0.0417	RM279_164_6/0.3968 RM481_219_18/0.1429 RM250_175_9/0.0667	RM250_173_8/0.64 RM316_196_2/0.0526 RM214_146_6/0.0526
	Combined R^2 value[2]	0.0915	0.1196	0.17	0.1333	0.0726
	% Correct Classification[3]	100	100	100	100	100

[1] Individual R^2 values calculated from Pearson correlation coefficient; [2] Combined R^2 value calculated from multiple regression (PROC REG, SAS Institute, ver. 9.0); [3] Percent correct classification were calculated by leave-one-out validation with in the training samples. [4] The first part of the DA marker denotes the SSR marker, the second part represents the observed allele size in bp and the third part stands for the allele number of the SSR locus. All the sets of markers identified by DA procedure are reordered based on their individual R^2 values, not by relative contribution to the discriminant rule. Individual R^2 values are calculated for respective the Training samples only whereas the Combined R^2 values are calculated considering all individuals.

Table 2.7 Discriminant Analysis-selected markers for percent head rice, percent total rice, plant height, heading date and grain yield from 192 lines of 2000 URN field trials evaluated in Louisiana

		Trait				
		Percent Head Rice	Percent Total Rice	Plant Height	Heading Date	Grain Yield
Louisiana	DA markers / Individual R^2 value[1] 15% Training Sample	[4]RM341_142_4/0.3783 RM279_158_3/0.3422 RM202_176_4/0.1168 RM55_234_4/0.0341 RM149_250_8/0.0341 RM120_186_5/0.0203 RM144_253_10/0.0201	RM296_125_2/0.2804 RM279_158_3/0.2469 RM433_223_2/0.2074 RM481_171_11/0.1984 RM341_142_4/0.1902 RM109_97_9/0.1115 RM317_165_7/0.0883 RM409_91_6/0.0558 RM161_165_1/0.032	RM431_254_4/0.36 RM120_180_2/0.0517 RM3431_160_5/0.0421 RM161_185_7/0.0421 RM214_115_2/0.0338 RM161_180_4/0.0338 RM109_96_8/0.0207 RM144_247_8/0.0207 RM16_184_6/0.0166 RM228_113_4/0.0166 RM1167_177_4/0.0156 RM481_165_9/0.0001 RM5_106_1/0.0001	RM437_252_1/0.2639 RM3430_211_6/0.1809 RM317_161_6/0.159 RM149_241_2/0.1289 RM279_164_6/0.0977 RM109_98_10/0.0786 RM149_240_1/0.045 RM116_279_3/0.0365 RM119_167_3/0.0365 RM162_201_1/0.0236 RM623_334_2/0.0194 RM178_117_2/0.0184 RM104_234_3/0.0184 RM214_148_7/0.0159 RM210_159_10/0.013 RM149_245_6/0.0102 RM1189_174_1/0.009 RM149_243_4/0.007 RM162_226_7/0.0011 RM162_205_3/0.0002	RM481_168_10/0.2083 RM437_254_2/0.1812 RM120_184_4/0.1664 RM228_111_3/0.1544 RM316_196_2/0.0753 RM3431_160_5/0.0545 RM416_112_2/0.0545 RM5_106_1/0.0545 RM181_239_1/0.0448 RM481_219_18/0.0357 RM161_181_5/0.0258 RM1167_177_4/0.0252 RM317_165_7/0.0252 RM420_203_4/0.0175 RM234_151_4/0.0175 RM109_87_2/0.0175 RM5_112_3/0.0128 RM109_98_10/0.0061
	Combined R^2 value[2]	0.1925	0.2194	0.3096	0.3437	0.3633
	% Correct Classification[3]	100	100	100	100	100
	DA markers / Individual R^2 value[1] 5% Training Sample	RM341_142_4/0.5844 RM149_250_8/0.1467 RM144_253_10/0.0005	RM341_142_4/0.7162 RM162_205_3/0.0294 RM109_87_2/0.0294	RM437_252_1/0.6481 RM118_162_3/0.2083 RM7_173_4/0.0946 RM1359_158_4/0.0749 RM144_253_10/0.0749 RM119_167_3/0.0118 RM190_120_4/0.0003	RM437_252_1/0.3529 RM279_160_4/0.2667 RM190_122_5/0.2667 RM3431_160_5/0.1895 RM161_180_4/0.0571	RM149_240_1/0.5385 RM1167_177_4/0.12 RM482_192_5/0.0526
	Combined R^2 value[2]	0.1085	0.0823	0.1117	0.2166	0.1083
	% Correct Classification[3]	100	100	100	100	100

[1] Individual R^2 values calculated from Pearson correlation coefficient; [2] Combined R^2 value calculated from multiple regression (PROC REG, SAS Institure, ver. 9.0); [3] Percent correct classification were calculated by leave-one-out validation with in the training samples; [4] The first part of the DA marker denotes the SSR marker, the second part represents the observed allele size in bp and the third part stands for the allele number of the SSR locus. All the sets of markers identified by DA procedure are reordered based on their individual R^2 values, not by relative contribution to the discriminant rule. Individual R^2 values are calculated for respective the Training samples only whereas the Combined R^2 values are calculated considering all individuals.

Table 2.8 Discriminant Analysis-selected markers for percent head rice, percent total rice, plant height, heading date and grain yield from 192 lines of 2000 URN field trials evaluated in Mississippi

		Trait				
		Percent Head Rice	Percent Total Rice	Plant Height	Heading Date	Grain Yield
Mississippi	DA markers / Individual R^2 value[1] 15% Training sample	[4]RM231_191_6/0.2389 RM181_241_2/0.1381 RM21_139_4/0.1282 RM250_177_10/0.101 RM475_185_1/0.0948 RM118_162_3/0.0573 RM149_254_11/0.0547 RM229_117_2/0.0385 RM474_259_5/0.0379 RM225_142_6/0.0331 RM340_114_1/0.0253 RM184_215_2/0.014 RM55_227_2/0.0132 RM17_157_1/0.0131 RM109_87_2/0.0131 RM162_212_6/0.0131 RM341_139_3/0.0131 RM437_270_4/0.0125 RM72_159_3/0.0125 RM317_140_1/0.0083 RM190_113_3/0.0061 RM104_238_4/0.0047 RM136_100_3/0.0041	RM279_162_5/0.1631 RM437_274_5/0.1299 RM214_146_6/0.0916 RM284_142_1/0.0718 RM7_173_4/0.0667 RM273_199_1/0.0605 RM149_245_6/0.0408 RM109_95_7/0.0317 RM431_250_2/0.03 RM171_328_2/0.0172 RM433_221_1/0.0165 RM478_236_8/0.011 RM210_151_7/0.0086 RM234_141_3/0.0054 RM169_168_3/0.0008 RM1189_188_7/0.0002 RM162_236_9/0	RM431_254_4/0.6776 RM1167_177_4/0.09 RM13_149_5/0.0434 RM317_153_4/0.0213 RM144_235_4/0.0213 RM481_202_15/0.0213 RM422_381_5/0.0213 RM112_123_1/0.017 OSR13_98_2/0.017 RM210_141_3/0.0148 RM25_143_4/0.0084 RM341_156_5/0.0084 RM210_153_8/0.0042 RM5_126_7/0.0034 RM25_139_2/0.0016	RM225_132_2/0.3486 RM623_348_4/0.0761 RM333_189_8/0.0761 RM136_103_5/0.0307 RM517_264_3/0.0243 RM25_148_8/0.0164 RM109_90_3/0.0164	RM25_139_2/0.2609 RM250_177_10/0.1765 RM161_179_3/0.0985 RM184_215_2/0.0943 RM481_216_17/0.075 RM3431_160_5/0.0545 RM475_194_2/0.0545 RM431_254_4/0.0216 RM279_160_4/0.0193 RM482_186_3/0.0185 RM214_111_1/0.0175 RM5752_159_5/0.0175 RM149_256_13/0.0175 RM474_275_8/0.0175 RM120_176_1/0.0143 RM3912_191_3/0.0048 RM474_261_7/0.0023 RM338_182_2/0.0023 RM408_127_5/0.0011 RM228_105_2/0
	Combined R^2 value[2]	0.4252	0.3656	0.2654	0.235	0.4111
	% Correct Classification[3]	100	100	100	100	100
	DA markers / Individual R^2 value[1] 5% Training sample	RM106_287_1/0.3025 RM317_165_7/0.1613 RM482_192_5/0.1613 RM104_222_1/0.1232 RM1359_166_7/0.0909 RM190_113_3/0.0313	RM7_175_5/0.3656 RM317_161_6/0.1361 RM109_97_9/0.0649 RM408_117_1/0.029 RM5752_159_5/0.029	RM431_250_2/0.8462 RM475_185_1/0.0368	RM181_239_1/0.4152 RM475_199_4/0.3667 RM225_142_6/0.1618 RM206_147_5/0.1618 RM232_141_2/0.0764 RM420_203_4/0.0764 RM144_244_7/0.0404	RM231_191_6/0.375 RM109_95_7/0.1111 RM1189_174_1/0.1111 RM340_117_2/0.0526
	Combined R^2 value[2]	0.1613	0.162	0.1175	0.1835	0.0782
	% Correct Classification[3]	100	100	100	100	100

[1] Individual R^2 values calculated from Pearson correlation coefficient; [2] Combined R^2 value calculated from multiple regression (PROC REG, SAS Institute, ver. 9.0); [3] Percent correct classification were calculated by leave-one-out validation with in the training samples; [4] The first part of the DA marker denotes the SSR marker, the second part represents the observed allele size in bp and the third part stands for the allele number of the SSR locus. All the sets of markers identified by DA procedure are reordered based on their individual R^2 values, not by relative contribution to the discriminant rule. Individual R^2 values are calculated for respective the Training samples only whereas the Combined R^2 values are calculated considering all individuals.

Table 2.9 Discriminant Analysis-selected markers for percent head rice, percent total rice, plant height, heading date and grain yield from 192 lines of 2000 URN field trials evaluated in Missouri

		Trait				
		Percent Head Rice	Percent Total Rice	Plant Height	Heading Date	Grain Yield
Missouri	DA markers / Individual R^2 value[1] 15% Training Sample	Phenotypic data is not available	Phenotypic data is not available	[4]RM431_254_4/0.4033 RM431_250_2/0.2954 RM486_99_3/0.2906 RM118_158_1/0.1512 RM1189_190_8/0.1505 RM55_232_3/0.0982 RM1167_177_4/0.0728 RM409_91_6/0.0701 RM214_148_7/0.0478 RM109_96_8/0.0422 RM162_240_11/0.0199 RM162_201_1/0.0174 RM422_387_8/0.0152 RM232_153_5/0.0136 RM475_196_3/0.0136 RM279_160_4/0.0109 RM232_155_6/0.0076 RM7_175_5/0.0025 RM161_181_5/0.0007 RM481_168_10/0.0001	RM190_122_5/0.3324 RM474_257_4/0.1099 RM21_153_8/0.0862 RM1167_177_4/0.0502 RM623_348_4/0.0415 RM5_106_1/0.0414 RM136_103_5/0.0414 RM3431_160_5/0.0414 RM418_277_3/0.0414 RM3430_211_6/0.0351 RM5_128_8/0.0203 RM120_188_6/0.0203 RM161_183_6/0.0203 RM162_201_1/0.0175 RM228_113_4/0.0175 RM232_153_5/0.0175 RM120_182_3/0.0135 RM72_189_8/0.0048 RM149_245_6/0.0027 RM1189_190_8/0.0001	RM623_350_5/0.1984 RM229_129_6/0.1673 RM279_158_3/0.164 RM7_173_4/0.1081 RM149_241_2/0.1067 RM623_348_4/0.0741 RM250_173_8/0.0433 RM3431_158_4/0.0252 RM416_113_3/0.0175 RM162_201_1/0.0175 RM422_399_11/0.017 RM317_165_7/0.0023
	Combined R^2 value[2]			0.4181	0.3546	0.2705
	% Correct Classification[3]			99.99	100	100
	DA markers / Individual R^2 value[1] 5% Training Sample	Phenotypic data is not available	Phenotypic data is not available	RM431_254_4/0.5819 RM109_96_8/0.0803 RM149_245_6/0.0451	RM478_200_2/0.2899 RM3912_193_4/0.2222 RM475_199_4/0.16 RM232_141_2/0.1026 RM162_201_1/0.0494	RM228_111_3/0.2525 RM7_173_4/0.1905 RM210_143_4/0.1765 RM171_344_4/0.1111 RM293_202_3/0.0526
	Combined R^2 value[2]			0.1528	0.1185	0.1087
	% Correct Classification[3]			100	100	100

[1] Individual R^2 values calculated from Pearson correlation coefficient; [2] Combined R^2 value calculated from multiple regression (PROC REG, SAS Institure, ver. 9.0); [3] Percent correct classification were calculated by leave-one-out validation with in the training samples; [4] The first part of the DA marker denotes the SSR marker, the second part represents the observed allele size in bp and the third part stands for the allele number of the SSR locus. All the sets of markers identified by DA procedure are reordered based on their individual R^2 values, not by relative contribution to the discriminant rule. Individual R^2 values are calculated for respective the Training samples only whereas the Combined R^2 values are calculated considering all individuals.

Table 2.10 Discriminant Analysis-selected markers for percent head rice, percent total rice, plant height, heading date and grain yield from 192 lines of 2000 URN field trials evaluated in Texas

		Trait				
		Percent Head Rice	Percent Total Rice	Plant Height	Heading Date	Grain Yield
Texas	DA markers / Individual R^2 value[1] 15% Training Sample	[4]RM623_334_2/0.3872 RM3912_193_4/0.1512 RM109_95_7/0.1174 RM408_127_5/0.1113 RM341_142_4/0.1102 RM437_274_5/0.1096 RM312_94_1/0.0987 RM422_399_11/0.0619 RM206_195_14/0.0144 RM202_176_4/0.0052 RM1189_188_7/0.0001	RM437_274_5/0.169 RM408_127_5/0.1451 RM316_196_2/0.118 RM181_249_5/0.069 RM1167_171_1/0.0659 RM1359_166_7/0.0586 RM418_283_5/0.0243 RM214_154_8/0.0048 RM72_189_8/0.0007	RM431_254_4/0.7389 RM623_350_5/0.36 RM486_99_3/0.2516 RM1167_177_4/0.0536 RM21_129_2/0.0223 RM478_226_7/0.0185 RM232_139_1/0.0185 RM171_324_1/0.0185 RM106_293_3/0.0182 RM109_87_2/0.0172 RM3912_207_7/0.0172 RM72_186_7/0.0148 RM475_199_4/0.0134 RM225_138_4/0.0079 RM333_189_8/0.0068 RM293_200_2/0.0056 RM279_158_3/0.0049 RM5_126_7/0.0032 RM55_227_2/0.0031 RM234_141_3/0.0022	RM478_200_2/0.1837 RM3912_193_4/0.1787 RM317_161_6/0.173 RM623_348_4/0.0659 RM341_156_5/0.0632 RM120_186_5/0.0632 RM1189_174_1/0.0382 RM149_241_2/0.0343 RM420_182_1/0.0322 RM409_85_2/0.0274 RM25_148_8/0.0204 RM162_242_12/0.0204 RM190_113_3/0.0204 RM210_137_1/0.0199 RM1359_166_7/0.0066 RM312_96_2/0.0028 RM510_111_1/0.0023 RM474_275_8/0.0022 RM21_147_5/0.0022 RM340_114_1/0.0016	RM3912_195_5/0.2083 RM486_105_4/0.1623 RM250_169_6/0.1154 RM623_348_4/0.1154 RM315_132_1/0.1147 RM316_196_2/0.095 RM409_91_6/0.0873 RM16_167_1/0.0584 RM210_137_1/0.0556 RM144_250_9/0.0175 OSR13_94_1/0.0012
	Combined R^2 value[2]	0.3109	0.2205	0.335	0.3765	0.3525
	% Correct Classification[3]	100	100	100	100	100
	DA markers / Individual R^2 value[1] 5% Training Sample	RM481_156_6/0.3422 RM279_158_3/0.2853 RM409_85_2/0.2424 RM3431_148_1/0.2381 RM284_142_1/0.1736 RM293_198_1/0.1686 RM474_259_5/0.1071 RM341_174_7/0.0862 RM232_155_6/0.0305 RM317_140_1/0.0081	RM25_145_6/0.2582 RM3431_148_1/0.1452 RM109_97_9/0.0955 RM21_151_7/0.0547 RM422_387_8/0.0198 RM120_182_3/0.0014	RM431_250_2/0.825 RM109_87_2/0.055	RM478_212_6/0.5295 RM1189_174_1/0.1136 RM119_148_1/0.064 RM333_177_4/0.0521 RM210_141_3/0.025 RM21_157_10/0.0139 RM478_206_5/0.0124 RM119_167_3/0.0097 RM231_191_6/0.008 RM239_144_3/0.0024	RM250_169_6/0.4286 RM498_213_2/0.1905 RM144_253_10/0.0526
	Combined R^2 value[2]	0.1801	0.0946	0.1226	0.1317	0.154
	% Correct Classification[3]	100	100	100	100	100

[1] Individual R^2 values calculated from Pearson correlation coefficient; [2] Combined R^2 value calculated from multiple regression (PROC REG, SAS Institute, ver. 9.0); [3] Percent correct classification were calculated by leave-one-out validation with in the training samples; [4] The first part of the DA marker denotes the SSR marker, the second part represents the observed allele size in bp and the third part stands for the allele number of the SSR locus. All the sets of markers identified by DA procedure are reordered based on their individual R^2 values, not by relative contribution to the discriminant rule. Individual R^2 values are calculated for respective the Training samples only whereas the Combined R^2 values are calculated considering all individuals.

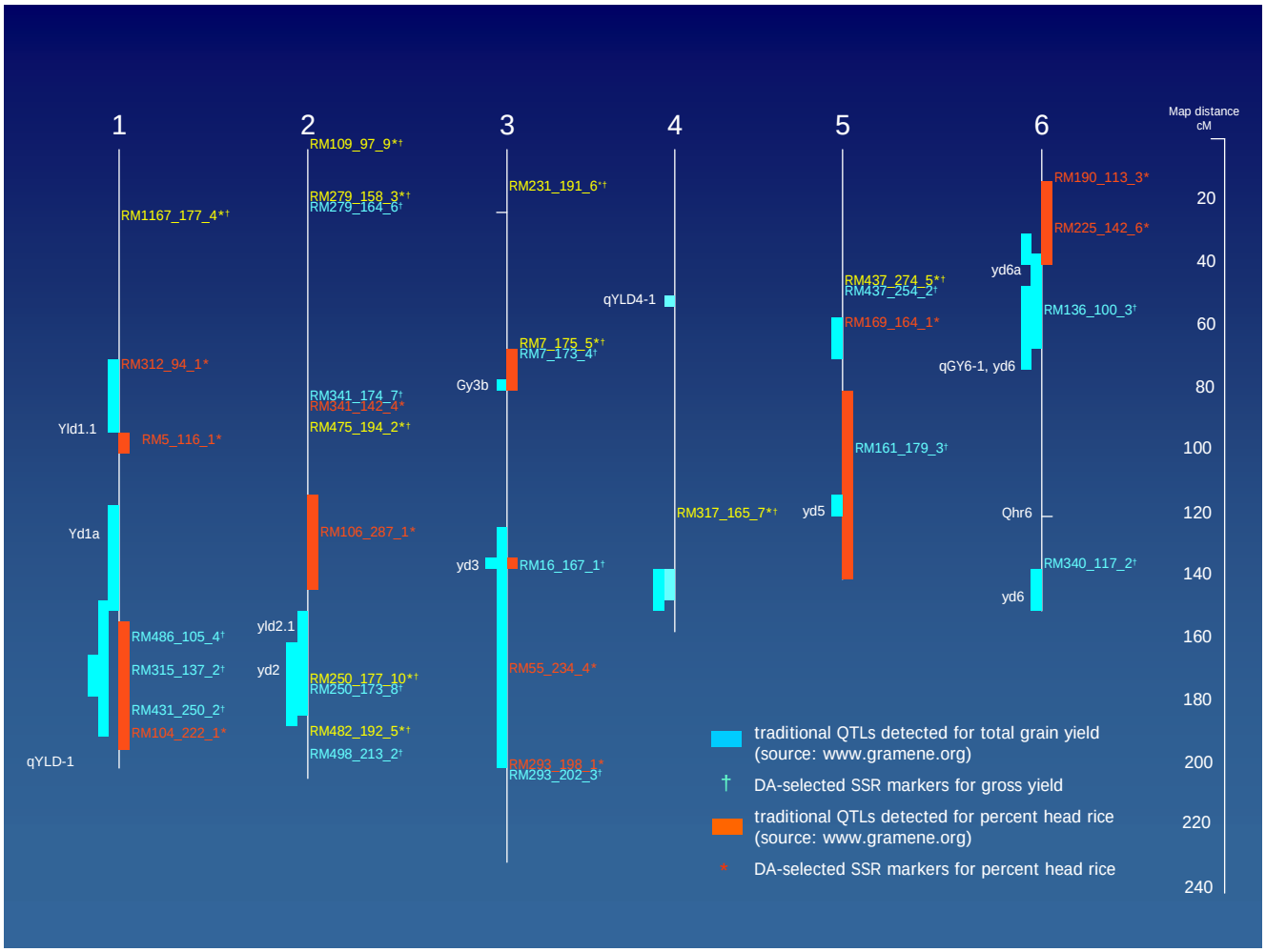


Figure 2.2 Chromosomal positions of DA selected markers and traditional QTLs for grain yield, percent head rice and percent total rice.

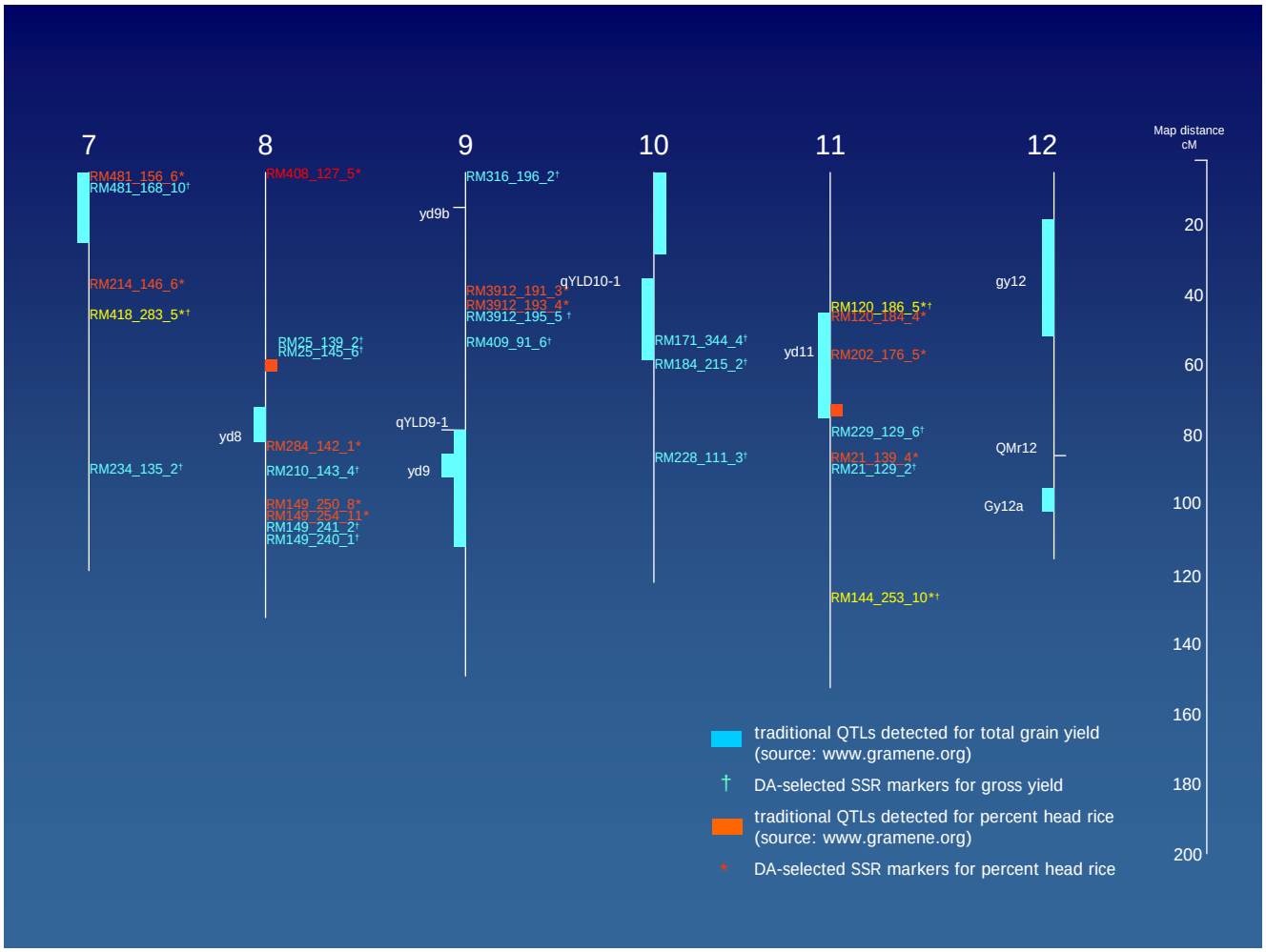


Figure 2.3 Chromosomal positions of DA selected markers and traditional QTLs for grain yield, percent head rice and percent total rice.

Table 2.11 Summary of chromosomal positions of traditional QTLs and new DA-selected markers on Cornell SSR 2001 map for Grain yield

Linkage group (QTL name)	Map	Position (cM)	Flanking molecular markers	Position of markers (cM)	Position (cM) on Cornell SSR_2001	Cited Reference	Remarks	DA markers shown on map	Position of DA markers
1	Rice-Cornell 9024/LH422 RI QTL 1996	90.8-120.7	RG173 RZ276	120.7 103.6	66.4 85.4	Xiao et al., 1996	Complete QTL not defined	RM312_94_1*	71.6
1 (yld1.1)	Rice-Cornell V20A/Oruf QTL 1998	53.3-102.7	RZ276 RM5	50.0 79.0	85.4 94.9	Xiao et al., 1996	Complete QTL not defined	RM5_116_1* RM5_112_3†	79.0 79.0
1	Rice-Cornell IR64/Azu DH QTL	154.8-194.1	RM212 RM414	148.7 191.1	145.6 191.1	Lafitte et al., 2002	Complete QTL not defined	RM315_139_3Φ RM315_137_2†	165.3 165.3
1 (qYLD1)	Rice-Cornell IR64/Azu DH QTL	175.5-177.9	RM315 RM431	165.3 178.3	165.3 178.3	Hittalmani et al., 2002	Complete QTL not defined	RM315_139_3Φ RM315_137_2†	165.3 165.3
2 (yd2)	Rice-CNHZAU Zh97/Ming63 RI QTL 2002	0.0-31.8	RM240 RM213	0.0 31.8	158.9 186.4	Xing et al., 2002	Complete QTL defined	RM482_192_5†	187.5
3	Rice-JRGP Nip/Kas F2 QTL 2000	73.5-97.1	V142 V8	80.0 86.8	119.7 198	Ishimaru et al., 2001	Complete QTL not defined	RM55_234_4* RM55_217_1Φ RM416_112_2†	168.2 168.2 191.6
5	Rice-JRGP Nip/Kas F2 QTL 2000	49.4-67.2	Cen5	53.2-54.6	52.5-70.0	Ishimaru et al., 2001	Complete QTL not defined	RM169_172_4*	57.9
6 (yd6)	Rice-Cornell 9024/LH422 RI QTL 1996	132.9-142.3	RG653 RG433	128.9 142.3	133.5 149.7	Li et al., 2000	Complete QTL not defined	RM340_117_2Φ†	133.5
7	Rice-JNIG W1944/Peik QTL 2002	0.0-25.0	RG128 RZ387	1.5 25.0	0.0 25.0	Cai et al., 2002	Complete QTL not defined	RM481_171_11Φ RM481_168_10† RM481_219_18† RM481_202_15† RM481_216_17Φ	3.2 3.2 3.2 3.2 3.2
10 (qYLD10-1)	Rice-IRRI IR64/Azu DH QTL 2003	25.8-42.4	RG257 RG241	25.8 42.4	33.2 58.3	Hittalmani et al., 2003	Complete QTL defined	RM5752_129_2* RM184_215_2†	11.0 58.3
11 (yd11)	Rice-CNHZAU Zh97/Ming63 RI QTL 2002	100.1-127.8	RG118 RM209	127.8 87.7	38.9 73.9	Hua et al., 2002	lower tail is not clearly defined	RM120_186_5* RM120_184_4† RM202_176_4* RM287_101_2Φ	41.7 41.7 - 54.0

Traditional QTLs for total grain yield, percent head rice are from web source: www.gramene.org. * - DA-selected SSR markers for gross yield, † - DA-selected SSR markers for percent head rice, and Φ - DA-selected SSR markers for percent total rice

Table 2.12 Summary of chromosomal positions of traditional QTLs and new DA markers on Cornell SSR 2001 map for percent head rice

Linkage group (QTL name)	Map	Position (cM)	Flanking markers	Position (cM) of markers	Position (cM) on Cornell SSR_2001	Reference	Remarks	DA markers Shown on map	Position of DA markers
1	Rice-Cornell IR64/IRG105 QTL 2003	191.2-227.1	RM265 RG331	191.2 227.1	155.9 194.1	Septiningsih et al., 2003	Complete QTL not defined	RM315_139_3Φ RM315_137_2†	165.3
5	Rice-Cornell IR64/IRG105 QTL 2003	116.8-156.2	RM430 RM334	116.8 156.2	78.7 141.8	Septiningsih et al., 2003	Complete QTL defined	RM161_165_1Φ	96.9

Traditional QTLs for total grain yield, percent head rice are from web source: www.gramene.org. † - DA-selected SSR markers for percent head rice, and Φ - DA-selected SSR markers for percent total rice.

2.4 References

Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard JH (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. Theor Appl Genet 109: 630-639.

Ayres NM, McClung, AM, Larkin PD, Bligh HF, Jones CA, Park WD (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94: 773-781.

Bao J, Kong X, xie T, Xu L (2004) Analysis of genotype and enveronmental effects on rice starch. 1. Apparent amylose content pasting, viscosity, and gel texture. J Agri Food Chem 52: 6010-6016.

Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nat Genet Suppl 33: 228-238.

Bradbury LMT, Fitzgerald TL, Henry RJ, Jin Q, Waters DLE (2005a) The gene for fragrance in rice. Plant Biotech. J. 3: 363-370.

Bradbury LMT, Henry RJ, Jin Q, Reinke RF, Waters DLE (2005b) A perfect marker for fragrance genotyping in rice. Mol. Breeding 16: 279-283.

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

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the Isa gene of barley. Theor Appl Genet 109: 543-51.

Buttery RG, Ling LC (1983) Cooked rice aroma and 2-acetyl-1-pyrroline. J. Agric. Food Chem. 31: 823-826.

Cai W, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. Theor Appl Genet 104: 1217-1228.

Cao G, Zhu J, He C, Gao Y, Yan J, Wu P (2001) Impact of epistasis and QTL x environment interaction on the developmental behavior of plant height in rice (*Oryza sativa* L.). Theor Appl Genet 103: 153-160.

Chen LJ, Lee DS, Song ZP, Suh H, Lu BR (2004) Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. Annals of Botany 93: 67-73.

Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc Natl Acad Sci USA 10: 700-707.

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) High-throughput screening for induced point mutations. Plant Physiol 126: 480-4.

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J 37: 778-86.

Couzin S (2002) New mapping project splits the community. Science 296: 1391-1393.

Cui H, Peng B, Xing Z, Yu B, Xu G, Zhang Q (2003) Molecular dissection of the genetic relationships of source, sink and transport tissue with yield traits in rice. Theor Appl Genet 106: 649-658.

Corbett CL, Tardif FJ (2006) Detection of resistance to acetolactate synthase inhibitors in weeds with emphasis on DNA-based techniques: a review. Pest Manag Sci 62: 584-597.

Cordeiro GM, Christopher MJ, Henry RJ, Reinke RF (2002) Identification of microsatellite markers for fragrance in rice by analysis of rice genome sequence. Mol. Breed. 9: 245-250.

Childs, Nathan (2004) Rice Situation and Outlook Yearbook. Market and Trade Economics Division, Economic Research Service, U.S. Dept. Agriculture.

Ching A, Caldwell KS, Jung M, Dolan, M, Smith, OS, Tingey S, Morgante M, Rafalski A (2002) SNP frequency, haplotye structure and linkage disequilibrium in elite maize inbred lines. Biomed Central Genetics 3: 19-32.

Cui Y, Wu J (2005) Statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTLs. Genet Res 86: 65-75.

Cui Y, Wu J, Shi C , Littell RC, Wu R (2006) Modeling epistatic effects of embryo and endosperm QTL on seed quality traits. Genet Res 87: 61-71.

DePrimo SE, Wong LM, Khatry DB, Nicholas SL, Manning WC, Smolich BD, O'Farrell AM, Cherrington JM (2003) Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. BMC Cancer 3:3.

Ebdon JS, Petrovic AM, Schwager SJ (1998) Evaluation of Discriminant Analysis in ident-ification of low and high-water use Kentucky bluegrass cultivars. Crop Sci 38: 152-157.

Estorminos LE Jr, Gealy DR, Dillon TL, Baldwin FL, Burgos RR, Tai TH (2002) Determination of hybridization between rice and red rice using four microsatellite markers. Proc South Weed Sci Soc 55: 197-198.

Fahima T, Roder MS, Wendehake K, Kirzhner VM, Nevo E (2002) Microsatellite polymorphism

in natural populations of wild emmer wheat, Triticum dicoccoides, in Israel. Theor Appl Genet 104: 17-29.

Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang Q (2005) The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. Theor Appl Genet 110: 1445-52.

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14: 1812-9.

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7: 179-188.

Fljnt-Garcia, SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Ann Rev Plant Biol 54: 357-374.

Gealy DR, Mitten DH, Rutger JN (2003) Gene flow between red rice (*Oryza sativa*) and herbicide resistant rice (*O. sativa*): Implications for weed management. Weed Tech. 17: 627-645.

Gealy DR, Yan WG, Rutger JN (2006) Red rice (*Oryza sativa*) plant types affect growth, coloration, and flowering characteristics of first- and second-generation crosses with rice. Weed Tech 20: 839-852.

Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza satva* L.). Genetics 165: 759-769.

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schuler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol Breed 13: 93-102.

Gilbert H, Le Roy P (2003) Comparison of three multitrait methods for QTL detection. Genet Sel Evol 35: 281-304.

Goff SA, Salmeron JM. (2004) Back to the future of cereals. Genomic studies of the world's major grain crops, together with a technology called marker-assisted breeding could yield a new green revolution. Sci Am 291: 42-9.

Henikoff S, Comai L (2003) Single-nucleotide mutations for Plant functional genomics. Annual Rev Plant Biol 54: 375-401.

Henikoff S, Till BJ, Comai L. TILLING (2004) Traditional mutagenesis meets functional genomics. Plant Physiol 135: 630-636.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38: 226-231.

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic associatin studies. Genet Med 4: 46-61.

Hittalmani S, Huang N, Courtois B, Venuprasad R, Shashidhar HE, Zhuang JY, Zheng KL, Liu G F, Wang GC, Sidhu JS, Srivantaneeyakul S, Singh VP, Bagali PG, Prasanna H C, McLaren G, Khush GS (2003) Identification of QTL for growth and grain yield related traits in rice across nine locations of Asia. Theor Appl Genet 107: 679-690.

Hittalmani S, Shashidhar HE, Bagali PG, Huang N, Sidhu JS, Singh VP, Khush GS (2002) Molecular mapping of quantitative trait loci for plant growth, yield and yield related traits across three diverse locations in a doubled haploid rice population. Euphytica 125: 207-214.

Hua J, Xing Y, Wu W, Xu C, Sun X, Yu S, Zhang Q (2003) Single locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Nat Aca Sci USA 100: 2574-2579.

Hua JP, Xing YZ, Xu CG, Sun XL, Yu SB, Zhang Q (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162: 1885-1895.

Ishimaru K, Yano M, Aoki N, Ono K, Hirose T, Lin SY, Monna L, Sasaki T, Ohsugi R (2001) Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor Appl Genet 102: 793-800.

Ishimaru K (2003) Identification of a locus increasing rice yield and physiological analysis of its function. Plant Physiol 133: 1083-1090.

Issiki M, Morino K, Okagaki RJ, Wressler SR, Izawa T, Shimamoto K (1998) A naturally occuring functional allele of the rice waxy locus has a GT to TT mutation at the 5'splice site of the first intron. Plant J 15: 133-138.

Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC, Neale DB (2003) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. Genetics 165: 1489-1506.

Jin Q, Waters DLE, Cordeiro GM, Henry RJ, Reinke RF (2003) A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (*Oryza sativa* L.). Plant Sci. 165: 359-364.

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in Arabidopsis flowering time. Science 290: 344-347.

Kari L, Loboda A, Nebozhyn M, Rook AH, Vonderhied EC, Nichols C, Virok D, Chang C, Horng WH, Johnston J, Wysocka M, Showe MK, Showe LC (2003) Classification and prediction of survival in patients with the leukemic phase of cutaneoius T cell lymphoma. J Exp Med 197: 1477-1488.

Labate JA, Lamkey KR, Lee M, Woodman W (2000) Hardy-Weinberg and linkage equilibrium estimates for random mated populations in the BSSS and BSCB1 reciprocal recurrent selection program. Maydica 45: 243-255.

Lafitte HR, Courtois B, Arraudeau M (2002) Genetic improvement of rice in aerobic systems: Progress from yield to genes. Field Crops Res 75: 171-190.

Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q, Yuan L, McCouch SR (2004) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. Genome 47: 697-704.

Li XH, Xu CG, Gao YJ, Yu SB, Zhang Q, Li JX, Tan YF (2000) Analyzing quantitative trait loci for yield using a vegetatively replicated F2 population from a cross between the parents of an elite rice hybrid. Theor Appl Genet 101: 248-254.

Li Z, Thomas TL (1998) PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. Plant Cell 10: 383-398.

Liao CY, P Wu, B Hu, Yi KK (2001) Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. Theor Appl Genet 103: 104-111.

Listgarten J, Damaraju S, Poulin B, Cook L, Dufour J, Driga A, Mackey J, Wishart D, Greiner R, Zanke B (2004) Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. Clin Cancer Res 10: 2725-2737.

Madsen KH, Valverde BE, Jensen JE (2002) Risk assessment of herbicide-resistant crops: A Latin American perspective using rice (*Oryza sativa*) as a model. Weed Tech 16: 215-223.

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA. (2007) A Mixed-Model Approach to Association Mapping Using Pedigree Information With an Illustration of Resistance to Phytophthora infestans in Potato.Genetics 175: 879-889.

Martin ER, Gilbert JR, Lai EH, Riley J, Rogala AR, Slotterbeck BD, Sipe CA, Grubber JM, Warren LL, Conneally PM, Saunders AM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM (2000) Analysis of association at single nucleotide polymorphisms in the APOE region. Genomics 63: 7-12.

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol 123: 439-442.

McKenzie KS, Rutger JN (1983) Genetic analysis of amylose content, alkali spreading score,

and grain dimensions in rice. Crop Sci 23: 306-311.

Mcharo M, Labonte D, Oard JH, Kays SJ, McLaurin WJ (2004) Linking quantitative traits with AFLP markers in sweetpotato using Discriminant Analysis. Acta Hort 637: 285-293.

Mei H, Luo L, Guo L, Wang Y, Yu X, Ying C, Li Z (2002) Molecular mapping of QTLs for rice milling yield traits. Acta Genetica Sinica 29: 791-797.

Mei HW, Luo LJ, Ying CS, Wang YP, Yu XQ, Guo LB, Paterson AH, Li ZK (2003) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. Theor Appl Genet 107: 89-101.

Mendez, MA, Hodar C, Vulpe C, Gonzalez M, Cambiazo V (2002) Discriminant analysis to evaluate clustering of gene expression data. FEBS Lett 522: 24-28.

Messeguer J, Marfa V, Catala MM, Guiderdoni E, Mele E (2004) A field study of pollen-mediated gene flow from Mediterranean GM rice to conventional rice and the red rice weed. Mol Breeding 13: 103-112.

Msumarra G, Barresi V, Condorelli DF, Scire S (2003) A bioinformatics approach to the identification of candidate genes for the development of new cancer diagnostics. Biol Chem 384: 321-327.

Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, Hedge P (1991) Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 253: 665-669.

Nordberg, M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30: 190-193.

Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Mutation detection using a novel plant endonuclease. Nucleic Acids Res 26: 4597-4602.

Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weinig C, Schmitt J, Purugganan MD (2004) Linkage disequilibrium mapping of Arabidopsis CRY2 flowering time alleles. Genetics 167: 1361-1369.

Pacey-Miller T, Henry R (2003) Single-nucleotide polymorphism detection in plants using a single-stranded pyrosequencing protocol with a universal biotinylated primer. Anal Biochem 317: 166-170.

Page, GP, George V, Go RC, Page P, Allison D (2003) Are we there yet?: Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. Am. J. Hum. Genet 73: 711-719.

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet (2004) 109: 508-514.

Pennisi E (2003) A closer look at SNPs suggests difficulties. Science 281: 1787-1792.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci 98: 1147-1184.

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405: 847-856.

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Genetics 103: 18656-18661.

Sallaud C, Meynard D, van Boxtel J, Gay C, Bes M, Brizard JP, Larmande P, Ortega D, Raynal M, Portefaix M, Ouwerkerk PB, Rueb S, Delseny M, Guiderdoni E (2003) Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. Theor Appl Genet 106: 1396-13408.

Sano Y (1984) Differential regulation of waxy gene expression in rice endosperm. Theor Appl Genet 68: 467-473.

Sasaki T, Burr B (2000) International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Nature 3: 138-141.

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Res 13: 1250-1257.

Septiningsih EM, Trijatmiko KR, Moeljopawiro S, McCouch SR (2003) Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon.* Theor Appl Genet 107: 1433-1441.

Sha XY, Linscombe SD, Bearb KF, Howard AM, Theunissen BW, Hoffpauir HL, Cramer SW (2000) Evaluation of specialty rice progenies for aroma. 92th Annual Research Report: Rice Research Station. Crowley: Louisiana Agricultural Experiment Station 55-58.

Shen YJ, Jiang H, Jin JP, Zhang ZB, Xi B, He YY, Wang G, Wang C, Qian L, Li X, Yu QB, Liu HJ, Chen DH, Gao JH, Huang H, Shi TL, Yang ZN (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. Plant Physiol 135: 1198-11205.
Simko I, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004) Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor Appl Genet 108: 217-224.

Song ZP, Lu BR, Zhu YW, Chen JK (2003) Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157: 657-665.

Steele GL, Chandler JM, McCauley GN (2002) Control of red rice (*Oryza sativa*) in imidazolinone-tolerant rice (*O sativa*). Weed Tech 16: 627-630.

Tan S, Evans R, Singh B (2006) Herbicidal inhibitors of amino acid biosynthesis and herbicide tolerant crops. Amino Acids 30:195-204.

Tan S, Evans RR, Dahmer ML, Singh BK, Shaner DL. (2005) Imidazolinone-tolerant crops: history, current status and future. Pest Manag Sci 61:246-257.

Tan YF, Sun M, Xing YZ, Hua JP, Sun XL, Zhang QF, Corke H (2001) Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. Theor Appl Genet 103: 1037-1045.

Tan ZB, Shen LS, Yuan ZL, Lu CF, Chen Y, Zhou KD, Zhu LH (1997) Identification of QTLs for ratooning ability and grain yield traits of rice and analysis of their genetic effects. Acta Agronomica Sinica 23: 289-295.

Tenaillon et al., (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98: 9161-9166

Terwilliger JD, Haghighi F, Hiekkalinna TS, Goring HH (2002) A bias-ed assessment of the use of SNPs in human complex traits. Curr Opin Genet Dev 12: 726-734.

Thornsberry JM, Goodman MM, Doebkey J, Kresovich S, Nielsen D (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28: 286-289.

Till BJ, Colbert T, Tompa R, Enns LC, Codomo CA, Johnson JE, Reynolds SH, Henikoff JG, Greene EA, Steine MN, Comai L, Henikoff S (2003) High-throughput TILLING for functional genomics. Methods Mol Biol 236: 205-220.

Till BJ, Burtner C, Comai L, Henikoff S (2004) Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res 32: 2632-2641.

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4: 12.

Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, Ioannidis JPA (2004) Establishment of genetic associations for complex diseases is independent of early study findings. Eur J Hum Genet 12: 762-769.

Venuprasad R, Shashidhar HE, Hittalmani S, Hemamalini GS (2002) Tagging quantitative trait loci associated with grain yield and root morphological traits in rice (*Oryza sativa* L.) under

contrasting moisture regimes. Euphytica 128: 293-300.

Wang F, Yuan QH, Shi L, Qian Q, Liu WG, Kuang BG, Zeng DL, Liao YL, Cao B, Jia SR. (2006) A large-scale field study of transgene flow from cultivated rice (*Oryza sativa*) to common wild rice (*O. rufipogon*) and barnyard grass (*Echinochloa crusgalli*). Plant Biotechnol J. 4:667-676.

Wang ZY, Zheng FQ, Shen GZ, Gao JP, Snustad DP, Li MG, Zhang JL, Hong MM (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. Plant J 7: 613-622.

Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with SNPs? Nat Genet 26: 151-157.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G (1995) Identification of the breast cancer susceptibility gene BRCA2. Nature 378: 789-792.

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L (1996) Genes from wild rice improve yield. Nature 384: 223-224.

Xiao JH, Li JM, Yuan LP, Tanksley SR (1996) Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a sub specific rice cross. Theor Appl Genet 92: 230-244.

Xing Z, Tan F, Hua P, Sun L, Xu G, Zhang Q (2002) Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet 105: 248-257.

Yamanaka S, Nakamura I, Watanabe KN, Sato Y (2004) Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theor Appl Genet 108: 1200-1204.

Yano M, Katayose Y, Shikari M, Yamanouchi U, Monna L (2000) Hd1 a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene (*CONSTANS*). Plant Cell 12: 2473-2483.

Yu J et al., (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296: 79-9.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203-208.

Yu SB, Li JX, Xu CG, Tan YF, Li XH, Zhang Q (2002) Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. Theor Appl Genet 104: 619-

625.

Zhang CT, Wang J, Zhang R (2002) Using a Euclid distance discriminant method to find protein coding genes in the yeast genome. Comput Chem 26: 195-206.

Zhang MQ (1998) Identification of protein-coding regions in *Arabidopsis thaliana* genome based on quadric discriminant analysis. Plant Mol Biol 37: 803-806.

Zhang NY, Linscombe S, Oard J (2003) Out-crossing frequency and genetic analysis of hybrids between transgenic glufosinate herbicide-resistant rice and the weed, red rice. Euphytica 130: 35-45.

Zhang, N, Xu Y, Akash M, McCouch S, Oard JH (2004) Identification of candidate markers associated with agronomic traits in rice using Discriminant Analysis. Theor Appl Genet 110:721-729.

Zhou PH, Tan YF, He YQ, Xu CG, Zhang Q (2003) Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. Theor Appl Genet 106: 326-331.

Zhuang JY, Fan YY, Wu JL, Xia YW, Zheng KL (2001) Comparison of the detection of QTL for yield traits in different generations of a rice cross using two mapping approaches. Acta Genetica Sinica 28: 458-464.

CHAPTER 3 VALIDATION OF MIXED MODEL-REGRESSION PROCEDURE FOR ASSOCIATION GENETICS IN RICE[†]

3.1 Introduction

Completion of the rice genome sequencing project (Takashi et al., 2005) will serve as a powerful springboard for functional characterization of rice genes by a variety of methods that include identity and validation of DNA markers associated with complex traits. Standard QTL mapping approaches such as Composite Interval Mapping (Zeng, 1994) can be used, but power and precision may be compromised by limited recombination in segregating/recombinant inbred lines and by relatively small sample size of most mapping populations (Kearsey and Farquhar, 1998; Beavis, 1998). Moreover, low predictive performance has been reported in different mapping studies when markers were first selected in one population and then evaluated in separate test samples (Beavis, 1994; Beavis, 1998; Melchinger et al., 1998; Mei et al., 2003; Sillanpaa and Auranen, 2004). Resampling and cross validation methods have been proposed to obtain unbiased estimates of QTL position and effect for marker-assisted selection (Beavis, 1994; Utz et al., 2000; Schon et al., 2004).

3.1.1 Kinship Relationships

Kinship describes the probability that two homologous genes are identical by decent in a given sample. However, kinship relationships have not been considered in most plant mapping or marker-assisted selection strategies. Mixed models using variance component approaches that account for kinship estimates have been exploited in animal research for over two decades (Henderson, 1984; George et al., 2000). Nagamine and Haley (2001) extended the mixed model of Henderson to detect QTL by interval mapping in animal systems. Parisseaux and Bernardo

[†] The MR procedure analysis for the population I was carried out by Samuel A. Ordonez Jr, School of Plant, Environmental and Soil Sciences, LSU.

(2004) developed a mixed model for hybrid crops incorporating effects for general combining ability of markers associated with agronomic traits. Arbelbide et al. (2006) developed a mixed model for self pollinating plants that accounted for multiple location effects and kinship based on pedigree records. Arbelbide and Bernardo (2006) applied single and multiple marker analyses in the mixed model format for candidate loci and genes associated with bread quality traits in wheat (*Triticum aestivum* L.).

3.1.2 The TASSEL Software Program

The TASSEL software program (http://www.maizegenetics.net) incorporates population structure and kinship estimates into a mixed model for association genetics of unrelated individuals (Yu et al., 2006). However, the mixed model has not been extensively explored in selfing species such as rice. The TASSEL mixed model was used recently in association studies of a complex agronomic trait in barley (Rostoks et al., 2006). Epistasis was postulated to impact the ability to detect marker-trait associations for the selected population of inbred varieties. Zhao et al. (2007) found that the TASSEL mixed model correctly identified some, but not all major candidate genes related to flowering in Arabidopsis, and the method was not sufficiently sensitive to identify additional loci with minor effects.

3.1.3 Hypothesis Testing in Complex Trait Mapping

The "model selection" approach, based on information criteria such as Bayesian Information Criterion (BIC; Schwarz, 1978) and Akaike Information Criterion (AIC; Akaike, 1974), has been investigated to address selection bias present in standard QTL mapping techniques (Ball, 2001; Piepho and Gauch, 2001; Bogdan et al., 2004; Bogdan and Doerge, 2005). The model selection strategy proposes to identify the fewest number of variables that minimize BIC or other information criteria as opposed to standard hypothesis testing to build the

optimal predictive model. Model selection was reported to be superior to Composite Interval Mapping in simulated studies (Broman and Speed, 2002).

3.1.4 Association Genetics

Association genetics is an alternative strategy to standard QTL methods that is routinely used in human studies (Newton-Cheh and Hirschhorn, 2005), and one that is gaining support in the plant research community (Hayes and Szucs, 2006). The principal advantage of this approach, generally referred to as "linkage disequilibrium" mapping, is based on the ability to rapidly query informative regions of the genome among unrelated individuals that have generated numerous meiotic events over multiple generations. Linkage disequilibrium studies have been conducted for various marker-trait associations in maize (Yu and Buckler, 2006), rice (Garris et al., 2003), potato (Simko et al., 2006), barley (Kraakman et al., 2004; Malysheva-Otto et al., 2006; Rostoks et al., 2006) and wheat (Breseghello and Sorrells, 2006), but few studies have validated results in separate test populations.

3.1.5 Population Structure

Spurious associations between genotype and phenotype caused by population stratification must be detected among unrelated individuals in association studies to reduce Type I errors. Clustering techniques are one approach to identify stratified populations. For example, the model-based clustering "Structure" software program identifies putative population structure and assigns individuals to subgroups based on genotype frequencies (Pritchard et al., 2000).

3.1.6 Significance of Epistatic Interactions

Epistatic interactions between alleles at different loci in rice have been reported to exert considerable influence on different characters such as hybrid vigor (Yu et al., 1997; Goodnight, 1999; Li et al., 2001), cooking quality (Fan et al., 2005), plant height and heading date (Yu et al.,

2002), panicle number (Liao et al., 2001) and other complex traits (Cao et al., 2001; Mei et al., 2003). QTL models have therefore been developed to account for epistasis in rice and other species (Bogdan et al., 2004; Cui and Wu, 2005; Cui et al., 2006; Wan et al., 2006), so it is advantageous to include an epistatic component in robust models developed for association genetics.

The multivariate discriminant analysis procedure was previously evaluated to identify markers associated with agronomic traits among a diverse collection of U.S. and Asian inbred rice lines (Zhang et al., 2005). Consideration of population structure and estimation of missing data resulted in selection of markers that mapped within previously identified QTL regions for 12 complex traits.

The first objective of our current research was to evaluate the mixed model for ability to predict phenotypic variance of four complex agronomic traits in two distinct inbred populations of rice. The second objective focused on the creation and validation of a mixed model-regression (MR) procedure for prediction ability of selected markers in separate test samples.

3.2 Materials and Methods

3.2.1 Plant Material and Phenotypic Data Collection

Two distinct collections of inbred lines representing diverse and narrow germplasm were evaluated separately in this study. The first collection, referred to as Population I and described in Zhang et al. (2005), was comprised of a random group of 218 diverse lines with 56% of U.S. origin and the remaining from Asian sources. The phenotypic data (plant height, heading date, tiller number) and genotypic profiles for the current study were obtained from Zhang et al. (2005). U.S. public rice breeders in Crowley, LA, Beaumont, TX, Stuttgart, AR, Stoneville, MS, and Portageville, MO conduct a replicated field plot trial each year of common elite breeding

lines and varieties representing a narrow germplasm base. All 192 inbred lines of the trial, referred to as Population II in this study, were planted from March to April, 2000 at each of the five locations above in two to four replicated six-row plots, 2.0 m x 1.4 m, in a randomized complete block design. Standard agronomic practices at each location were implemented to minimize weed and insect damage for maximum grain yield. The center four rows of each plot were used to collect data for plant height and heading date in the same manner as Population I. In addition, data for grain yield at 12% moisture for all states of Population II and amylose content for TX and AR were collected. Phenotypic data were transformed if necessary to a normal distribution by log transformation and averaged across replications within each state to compute mean and variances along with analysis of variance (ANOVA) using PROC MIXED, SAS Institute, v. 9.0.

3.2.2 Molecular Marker Analyses

The initial molecular marker data used for Population I was described by Zhang et al. (2005). Heterozygous and rare (< 0.07%) alleles were excluded for the current study, reducing the number of marker alleles in Population I from 1153 to 309. For Population II, 97 single sequence repeat (SSR) markers, evenly spaced over the 12 chromosomes, generated a total of 579 alleles with an average of six alleles/locus. Rare alleles were removed as above, but heterozygous loci were retained to provide an adequate number of marker alleles (235) for the final analysis. PROC ALLELE, SAS Genetics, SAS Institute v. 9.1.4, was used to estimate polymorphism information content (PIC), level of heterozygosity and allelic diversity.

3.2.3 Creation of Training and Validation Samples

The “Complete Sample” consisted of the entire collection of inbred lines for Population I (n=218) and Population II (n=192). Complete Samples were randomly partitioned into 80% and

20% subsamples by the "partition fraction (validate=0.2)" statement of PROC GLMSelect, SAS v 9.1.3, to generate the Training and Validation Samples, respectively. The Training and Validation Samples of Population I consisted of 177 and 41 individuals, respectively, while Population II included 161 individuals in the Training Sample and 31 individuals in the Validation Sample. Detection of potential population structure in the Complete Sample was carried out by the "Structure" software program (http://pritch.bsd. uchicago.edu/structure.html).

3.2.4 TASSEL/Mixed Model Analyses

Phenotypic and marker datasets from Training Samples of each population were used with the "Simple" (S), "Structure" (Q), "Kinship" (K), and "Full" (Q+K) models in version 1.9.6 of the TASSEL software program (http://www.maizegenetics.net/ index.php?page= bioinformatics/tassel/downloads.html) to identify potential marker-trait associations. Kinship estimates were obtained from TASSEL for both populations. Negative values obtained were set to zero, implying no relationship. All markers selected by a model at the $p < 0.05$ level for each trait in the Training Sample were evaluated by TASSEL in the Validation Sample of both populations. Pooled R^2 values for the Q+K mixed model were obtained by summing partial R^2 values of individual marker alleles identified in the Validation Samples.

3.2.5 Mixed Model-Regression (MR) Procedure

In the first step of the MR procedure, the four statistical models of the TASSEL program described above were tested for their ability to identify candidate marker alleles associated with each of the four agronomic traits in the Training Sample. In step two, significant markers identified in the Training Sample by each of the four TASSEL models were subsequently evaluated in the Validation Sample by Stepwise and Forward methods of PROC GLMSelect. Both the Stepwise and Forward methods were assessed using the CHOOSE, SELECT and STOP

options in all combinations with Bayesian Information Criterion (BIC), Coefficient of Variation (CV), Adjusted R^2 (ADJRSQ), or SL selection criterion = 0.15, with and without consideration of epistasis. For each selected marker-trait combination, a total of 76 different GLMSelect models were evaluated, generating a total of 1,140 and 2,432 models that were assessed in the Validation Samples of Population I and II, respectively. The model that produced the highest adjusted R^2 value for a given trait in the Validation Sample was considered the "optimal" MR procedure.

3.2.6 Statistical Models of TASSEL and the MR Procedure

Four models of the TASSEL software program (Simple (S), Structure (Q), Kinship (K) and Mixed (Q+K) were evaluated in the current study and are described in detail by Yu et al. (2006). For the MR procedure, markers selected by TASSEL in the Training Samples were evaluated in the Validation Samples with four models, designated 'MR', 'MR-E', 'MR-Q', and 'MR-QE' using PROC GLMSelect. The 'MR' model, similar to the 'S' model in TASSEL considered main fixed effects (SSR/RFLP markers) as follows:

$$y = X\beta + S\alpha + \varepsilon \qquad (1)$$

where y = vector of experimental trait (phenotypic) values; β = vector of all fixed effects excluding molecular marker effects and population structure effects; α = vector that included information of more than one molecular marker effect, excluding their interaction effects; ε = vector of residual effects. The total number of effects and their order in α were determined by selection criteria in PROC GLMSelect. The X and S coefficients represented incidence matrices for β and α vectors consisting of 0s and 1s. The 'MR-E' model was identical to the 'MR' model above except the α vector included information for more than one molecular marker effect along with their two-way interaction (epistatic) effects.

The ‘MR-Q’ model added a population structure term to equation 1 above as follows:

$$y = X\beta + S\alpha + Q\nu + \varepsilon$$

where α = vector for more than one molecular marker effect, excluding their interaction effects, ν = vector of population structure effects and Q = design matrix for the ν vector. The ‘MR-EQ’ model was identical the ‘MR-Q’ model except that the α vector not only included the information of more than one molecular marker effect, but also their two-way interaction (epistatic) effects.

3.3 Results

Descriptive statistics of the four traits evaluated in the Complete, Training, and Validation Samples of Population I and II are shown in Table 3.1. The Training and Validation Samples were generally representative of average values and the extent of phenotypic variability of the Complete Samples. Exceptions to this trend were the higher variance observed for tiller number in the Validation Sample vs. other samples in Population I and the smaller variance observed for heading date and amylose content in the Validation Sample of Population II. Means and variances were greater as expected for plant height and heading date among the diverse lines represented by Population II vs. the narrow germplasm of elite lines in Population II. Grain yield and amylose content of Population II were within the expected range of current U.S. commercial cultivars. Molecular variation as measured by PIC and allelic diversity were consistent across the Complete, Training, and Validation Samples for both populations (data not shown). As expected, mean variation of the diverse Population I for PIC (0.47, range 0.46-0.49) and allelic diversity (0.52, range 0.51-0.53) was greater for corresponding values in the narrow germplasm of Population II (PIC = 0.34, range 0.31-0.36; allelic diversity = 0.38, range 0.35-0.39). Surprisingly, the average level of heterozygosity in the narrow germplasm of Population II (0.05)

was five-fold greater than the corresponding heterozygosity in the diverse lines of Population I. This unexpected finding may be due to greater segregation at loci in the early generation breeding lines of Population II vs. the later generation lines represented primarily by varieties and fixed lines of Population I.

The pairwise kinship estimates as determined by the TASSEL program for individuals in the Training Samples of Population I and II are shown in Figure 3.1. The vast majority of individuals showed little or no apparent relationship in the diverse germplasm of Population I which was similar to estimates of kinship detected in a large collection of maize inbreds (Yu et al., 2006). In contrast, higher levels of relatedness were observed in the narrow germplasm of Population II.

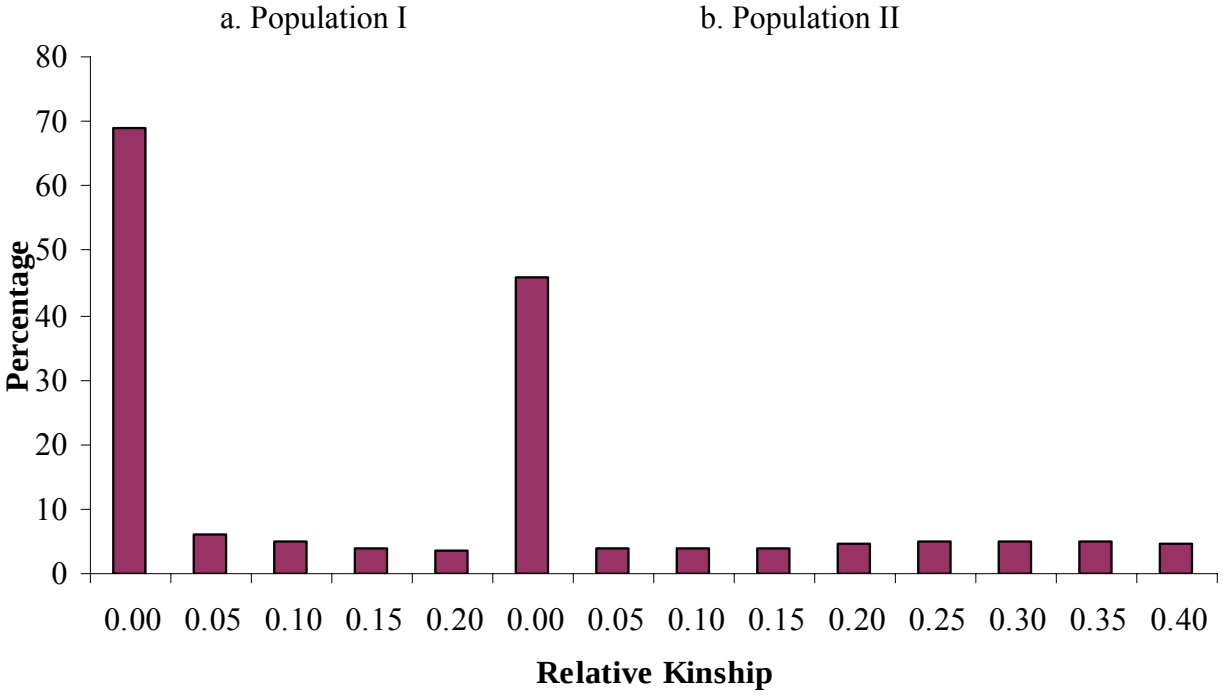


Figure 3.1 Pairwise kinship estimates from Training Samples of Populations I and II

Table 3.1 Means, variances and ranges for plant height, heading date, tiller number, grain yield and amylose content of Complete, Training, and Validation Samples in Population I and II

	Complete Sample[a]			Training Sample[b]			Validation Sample[c]		
Trait	Mean	Variance	Range Min-Max	Mean	Variance	Range Min-Max	Mean	Variance	Range Min-Max
Population I									
Plant height (cm)	106	426	63-185	107	423	63-185	105	450	70-165
Heading date (d)	96	84	75-129	96	83	75-129	95	90	81-123
Tiller number	12	69	4-50	11	63	4-50	12	98	5-47
Population II									
Plant height (cm)[d]	98	177	83-256	98	197	83-256	100	71	86-120
Heading date (d)[d]	85	14	71-96	85	15	71-97	85	9	79-90
Grain yield (kg/ha)[d]	7881	693189	4477-9759	7880	702159	4477-9759	7890	668360	6250-9642
Amylose content (%)[e]	19	17	0-26	19	19	0-26	20	9	12-25

[a] Complete Sample comprised of 218 inbred lines in Population I and 192 individuals in Population II
[b] Training Sample comprised of 177 inbred lines in Population I and 161 individuals in Population II
[c] Validation Sample comprised of 41 inbred lines in Population I and 31 individuals in Population II
[d] Mean values obtained across five states of LA, TX, AR, MS and MO.
[e] Mean values obtained across two states of TX and AR

3.3.1 TASSEL/Mixed Model Analysis

The full mixed model in TASSEL identified numerous markers at the $p < 0.05$ level presumably associated with the four agronomic traits in the Training Samples of Population I and II (data not shown). Markers previously reported to be associated with traits evaluated in this study were also detected by the mixed model. For example, the RM190 microsatellite, commonly used to classify rice into different amylose content classes (Bao et al., 2006), was the top marker, as judged by p values, from the mixed model in both populations. For plant height in Population I, three of the top five markers mapped inside or ≤ 5 cM from published QTLs based on the Gramene (gramene.org) website (data not shown). Markers in the three corresponding loci were also found associated with plant height in the same population using Discriminant Analysis by Zhang et al. (2005). The mixed model detected different markers that mapped within published QTL regions for plant height in Population II (data not shown). Similar results were found for the remaining traits evaluated in both populations of this study.

As expected, the contribution of individual markers as revealed by partial R^2 values from the mixed model in the Training and Validation Samples varied depending on the trait (data not shown). Nevertheless, the role of individual markers was found to be small in all cases. This result was not unexpected for complex traits, but a small effect was found even for a marker with a known large impact such as RM190.

Because the predictive ability of individual markers selected by the mixed model was found to be low, I applied all significant markers from the Training Sample to the Validation Sample for mixed model analysis. Partial R^2 values were pooled across individual significant markers detected in the Validation Sample at the $p < 0.05$ level. Moderate prediction rates were obtained after combining contributions from individual markers for plant height in both

populations (Table 3.2). Individual R^2 values were low and varied from 0.13 to 0.21 in Population I and 0.12 to 0.24 in Population II. The mixed model performed poorly for heading date in both populations where no significant markers were detected. In contrast, a combined high prediction rate for tiller number was produced in Population I while individual markers explained phenotypic variation that ranged from 0.11 to 0.25. Only moderate levels of prediction were detected for grain yield (range of individual markers = 0.08 to 0.20) and amylose content (range of individual markers = 0.01 to 0.15) with Population II.

Figure 3.2 shows the P-value and power plots of the four different models evaluated in TASSEL for all traits in both populations, except for amylose content. Type I error rates as revealed by P-values for both populations were consistently the highest in every case with the Simple Model (Figure 3.2a-c, g-i). For all traits in Population I, the full mixed Q+K model was no better in reducing Type I errors when compared to the model accounting for only Q population structure effects. A different trend was observed with the narrow Population II in that the Q+K model was the most effective in reducing type I errors for all traits. Amylose content produced the same outcome (data not shown). Importantly, P-value plots revealed substantial Type I error rates for all traits in each population, regardless of the model used. Reducing error in the Training Sample by applying more stringent p values < 0.001 and a False Discovery Rate (FDR) threshold of 0.05 were also attempted, but no markers were selected under these conditions in the Validation Sample.

The power of each model generated by TASSEL as a function of marker effect is shown in Figures 3.2 and 3.3. The plots revealed that the Q structure and Q+K full models in Population I exhibited similar power that was greater than the Simple (S) and Kinship (K) models (Figure 3.2d-f). The greatest contrast among models for power was observed for the complex trait of

Table 3.2 Optimal values produced by mixed model and Mixed model-regression (MR) procedure for Adjusted R^2, Root Mean Square Error (MSE), Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC), Average error sum of squares (ASE) and Predicted Residual Sum of Squares (PRESS) values for plant height, heading date, tiller number and grain yield in Validation Samples of Population I and II

	Plant height			Heading date			Tiller number		
A		MR Procedure			MR Procedure			MR Procedure	
Population I	Mixed Model	No Epistasis	Epistasis	Mixed Model	No Epistasis	Epistasis	Mixed Model	No Epistasis	Epistasis
Adjusted R^2	0.57[a]	0.70[a]	0.91	0.00	0.72	0.90	1.52	0.84	0.92
Root MSE	nd[b]	0.12	0.07	nd	3.96	2.31	nd	3.54	2.52
BIC	nd	-146.21	-183.10	nd	145.48	109.79	nd	119.48	91.58
AIC	nd	-165.06	-210.52	nd	123.20	80.66	nd	109.20	81.30
ASE	nd	0.01[c]	0.00	nd	10.70	3.12	nd	10.71	5.42
PRESS	nd	0.76	0.21	nd	748.52	86.30	nd	755.51	321.72
Markers of V.S.[d]	6	10	15	0	12	16	12	5	5
Markers of T,S.[e]	44	44	44	35	35	35	55	55	55

	Plant height			Heading date			Grain Yield		
B		MR Procedure			MR Procedure			MR Procedure	
Population II	Mixed Model	No Epistasis	Epistasis	Mixed Model	No Epistasis	Epistasis	Mixed Model	No Epistasis	Epistasis
Adjusted R^2	0.62[a]	0.68[a]	0.95	0.00	0.27	0.63	0.71	0.57	0.96
Root MSE	nd[b]	0.01	0.00	nd	0.01	0.01	nd	0.02	0.00
BIC	nd	-165.81	-200.96	nd	-157.34	-166.87	nd	-166.38	-210.17
AIC	nd	-173.45	-212.96	nd	-160.18	-171.59	nd	-172.47	-227.23
ASE	nd	0.00[c]	0.00	nd	0.00	0.00	nd	0.00	0.00
PRESS	nd	0.08	0.04	nd	0.00	0.00	nd	0.02	0.00
Markers of V.S.[d]	3	6	10	0	2	4	8	4	13
Markers of T.S.[e]	19	30	30	17	16	16	22	28	28

[a] Adjusted R^2 values for MR procedure calculated by SAS PROC GLM Select. R^2 values for full Q+K mixed model calculated by summing partial R^2 values of individual marker alleles identified in Validation Sample by TASSEL software program.
[b] No data collected; [c] ASE values rounded off to the nearest two decimal points.
[d] Number of selected markers/variables from Validation Sample.
[e] Number of selected marker alleles from Training Sample.

Population I

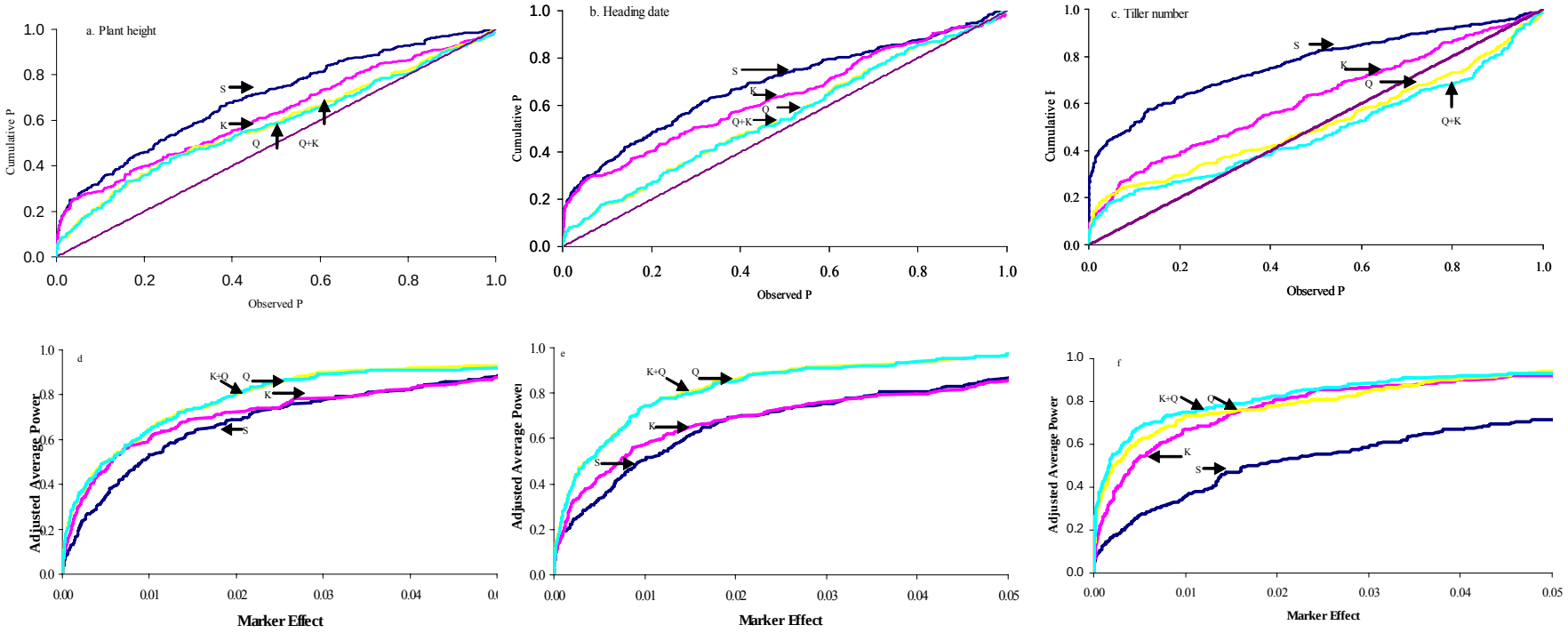


Figure 3.2 Type I error rates generated by simple (S), kinship (K), structure (Q), and full mixed (K+Q) models for plant height (a) heading date (b) and tiller number (c) in Population I. Adjusted average power of different models shown for plant height (d) heading date (e) and tiller number (f) in Population I.

Population II

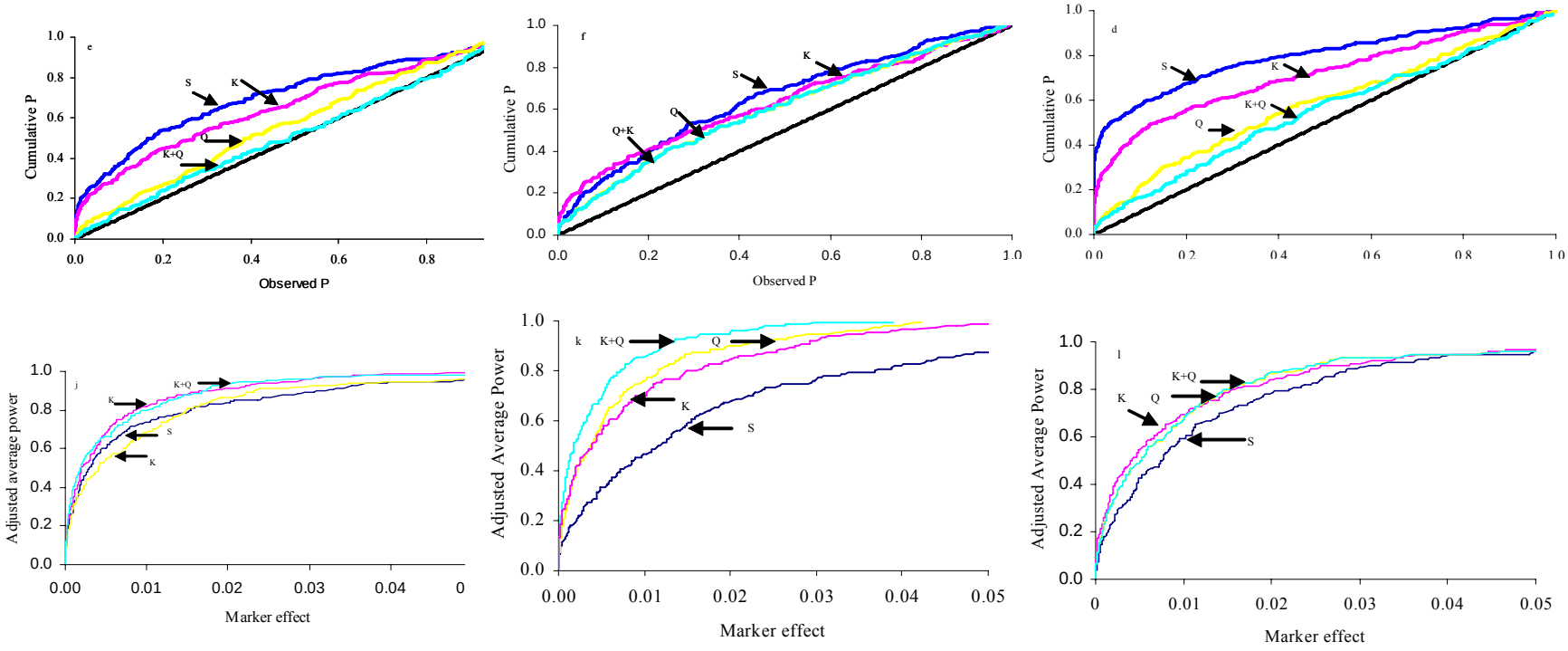


Figure 3.3 Type I error rates shown for plant height (g) heading date (h), and grain yield (i) in Population II. Adjusted average power of different models shown for plant height (j) heading date (k), and grain yield (l) in Population II. Straight diagonal line indicates perfect model fit with no Type I errors.

tiller number. The greatest power for heading date in Population II was produced with the full Q+K model (Figure 3.3k). Similar results were found for plant height (Figure 3.3j), grain yield (Figure 3.3l) and amylose content (data not shown).

3.3.2 Validation Results of the MR Procedure

Table 3.2 summarizes the "optimized" adjusted R^2 and information criteria values obtained by the MR procedure with and without epistasis for agronomic traits evaluated in the Validation Samples of Population I and II. When the MR procedure was implemented without consideration of epistasis for plant height and heading date in Validation Samples of Population I, prediction values were detected only at modest levels. Surprisingly, a moderately high prediction value was produced for tiller number, even when epistasis was ignored. The highest prediction values for all traits in the Validation Sample of Population I were observed when epistasis was incorporated into the MR procedure. Consideration of epistasis in MR markedly improved prediction by ~ 18 to 20% for plant height and heading date, although ability of epistasis to enhance prediction was modest for tiller number. Smaller values for Bayesian criteria (BIC, AIC), standard measures of variation (Root MSE, ASE), and leave-one-out cross validation (PRESS) provided additional evidence as to the value of including epistasis in the MR analysis. I found that the use of BIC, AIC or adjusted R^2 as variable selection criteria in Population I generally resulted in higher predictive ability with fewer variables vs. the standard F-value hypothesis testing approach (data not shown). However, no one model was optimal for all traits in Population I, and in several instances, identical values were obtained from similar, but different models. As was the case with the mixed model described above, the MR procedure identified variables in Population I that mapped to published QTL regions for all traits (data not shown).

MR analyses of the four traits evaluated across locations in Validation Samples of Population II are presented in Table 3.2. Consistent with the diverse germplasm of Population I, a moderate prediction level by MR without epistasis was observed for plant height in Population II, but a substantial improvement of ~ 25% was found when epistasis was considered. In the case of heading date, prediction ability was poor without epistasis, and only a moderate rate could be produced with epistasis by MR which was in contrast to corresponding high prediction rates obtained in Population I.

The economically important grain yield character in the Validation Sample of Population II followed the trend of plant height in that a modest prediction rate was produced by MR without epistasis, whereas a substantial gain in prediction was obtained when epistasis was included in the analysis (Table 3.2). The relatively high adjusted R^2 value for yield was accompanied by a moderate number of selected variables which was not surprising given the complex nature of this trait. In contrast, a high prediction value with few selected variables was detected for amylose content with or without epistasis by MR in the Validation Sample of Population II (data not shown). As was the case for the mixed model, RM190 was the top marker selected by the MR procedure. In general, the best MR procedures with high predictive ability for traits in the Validation Samples of Population II were generated with the Q+K mixed model in the first step coupled with the MR-E model using stepwise regression and epistasis in the second step. The use of BIC, CV, and ADJRSQ selection criteria generally outperformed the standard hypothesis testing SL option for ability to generate high prediction rates. All information criteria values of BIC, AIC, ASE, and PRESS detected in the Validation Samples of Population II were consistent, as in Population I, with improved MR models when epistasis was included in the analysis.

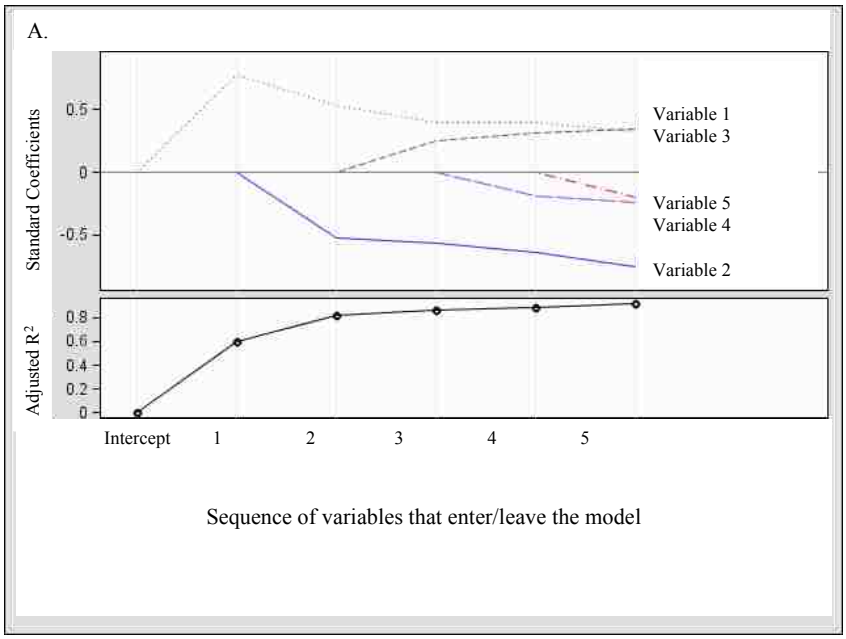
Figure 3.4a displays the sequence and contribution of selected variables by the GLM Select portion of the MR procedure for tiller number in Population I. The selected variables showed both positive and negative contributions that varied over different steps in development of the model. Variables that were selected in the first two steps appeared to play major roles in the final model which was reflected in the adjusted R^2 values that rapidly increased up to the second sequence, but saw little improvement from steps three to five. Improving prediction values by using only variables with positive coefficients was also attempted, but this approach failed perhaps because epistatic variables with both positive and negative coefficients often contained the same marker alleles (data not shown).

Figure 3.4b depicts the "coefficient evolution" panel for grain yield in the Validation Sample of Population II. While more variables were selected for grain yield vs. tiller number, model development for both traits followed similar trends. For example, early variable selection played a major role for both traits as seen by coefficient and adjusted R^2 values. Both positive and negative variables for each trait contributed to the final "optimized" model.

3.4 Discussion

Creation of a mixed model-regression procedure for association genetics in rice that was validated in two separate inbred populations is reported in this study. The mixed model component of our two-step approach proved useful for detection of individual known and candidate markers associated with published QTL regions. Similar results have been observed with mixed models in maize (Yu et al., 2006). The mixed model by itself may therefore prove useful for fine mapping of individual loci with large effects and gene discovery efforts in association genetics as suggested by Parisseaux and Bernardo (2004).

Our mixed model analyses of marker-trait associations showed an inherent advantage of



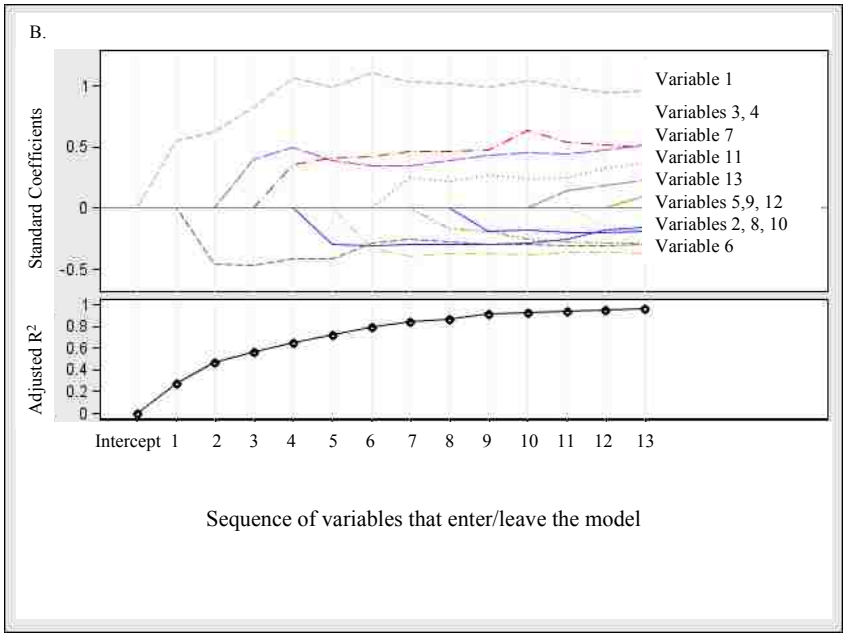


Figure 3.4 Coefficients of selected variables and adjusted R^2 values for tiller number and grain yield by mixed model-regression procedure (MR) as a function of when variables enter and leave the model

kinship estimates to reduce Type I errors in the narrow germplasm consistent with a recent study in maize (Yu et al., 2006). However, no such advantage was observed when mixed models considered kinship with the wide germplasm base of Population I. The Structure Q model was apparently sufficient to account for the majority of variation in relatedness among the inbred lines. The mixed model analyses found that individual contributions of each selected marker were small and could not be used in a practical setting for marker assisted selection. The selected markers were also evaluated by the TASSEL/mixed model as a group in the Validation Sample, but combined predictive ability was generally not satisfactory (i.e. $R^2 < 0.90$) for applied breeding purposes.

The idea that there was still inherent value captured from markers identified in the mixed model that could be exploited in a multivariate format was also tested. The conventional multivariate regression model without variable selection proved fruitless, so the mixed model-regression MR approach described in this study was created and evaluated. Key features of the MR procedure during this study were the following: (1) Detection of informative markers that mapped to published QTL regions by the mixed model in step one (2) Creation of Training Samples and further evaluation of selected markers in Validation Samples. High rates of Type I error detected in the Training Sample by the mixed model may have been mitigated in part by subsequent testing in the Validation Sample that enriched for unbiased estimators as suggested by previous studies (Beavis, 1994; Utz et al., 2000; Schon et al., 2004). (3) Inclusion of epistasis in the GLM Select portion of step two. Greater than 95% of the selected variables by MR were epistatic, demonstrating the importance of intergenic interactions for complex traits as proposed by Carlborg and Haley (2004) and (4) Use of Bayesian criteria, adjusted R^2, and other information criteria along with standard F tests to identify candidate variables in the GLM Select

portion of the MR procedure. Only moderate predictive ability was found for heading date in the narrow germplasm of Population II. This may have been due to a relatively low amount of variation in the Validation Sample vs. greater variation accompanied by a higher prediction rate (0.79) detected in the Training Sample

It is noteworthy that a high frequency of rare alleles was detected both in the diverse and narrow germplasm collections of Population I and II. This allelic architecture arose, not as a result of artificial selection pressure as one might suspect even in the narrow germplasm of Population II, but due to high allelic diversity among lines in both populations where rare alleles were common. The TASSEL program eliminated the rare alleles for evaluation as conventional genetic wisdom advocates, but this action diluted the number of alleles to 27%-40% in the original two populations that most likely reduced the power of the MR procedure. A positive perspective is that even greater predictive ability of MR may be likely with minimal occurrence of rare alleles, but this possibility must be confirmed in other samples. Results from this study provide strong evidence that the MR procedure should be further explored as a robust strategy to identify molecular variables associated with complex traits in rice and other plants.

3.5 References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans. Automat. Contrl AC-19: 716-723.

Arbelbide M, Yu J Bernardo R (2006) Power of mixed-model QTL mapping from phenotypic, pedigree and marker data in self-pollinated crops. Theor Appl Genet 112: 876-884.

Ball RD (2001) Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian Information Criterion. Genetics 159: 1351-1364.

Bao JS, Corke H, Sun M (2006) Microsatellites, single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.) Theor Appl Genet 113: 1185-1196.

Beavis WD (1994) The power and deceit of QTL experiments. Proc Annu Corn Sorghum Res Conf 49: 250-266.

Beavis WD (1998) QTL analyses: power, precision, and accuracy, pp. 145-162 in Molecular Dissection of Complex Traits, edited by A H Paterson, CRC Press, New York.

Bogdan M, Doerge RW (2005) Biased estimators of quantitative trait locus heritability and location in interval mapping. Heredity 95: 476-484.

Bogdan M, Ghosh JK, Doerge RW (2004) Modifying the Schwarz Bayesian Information Criterion to locate multiple interacting quantitative trait loci. Genetics 167: 989-999.

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172: 1165-77.

Broman KW, Speed TP (2002) A model selection approach for the identification of quantitative trait loci in experimental crosses. J R Statist Soc 64: 641-656.

Cao G, Zhu J, He C, Gao Y, Yan J, Wu P (2001) Impact of epistasis and QTL x environment interaction on the developmental behavior of plant height in rice (*Oryza sativa* L.). Theor Appl Genet 103: 153-160.

Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nature Reviews Genetics 5: 618-625.

Cui Y, Wu J, Shi C , Littell RC, Wu R (2006) Modeling epistatic effects of embryo and endosperm QTL on seed quality traits. Genet Res 87: 61-71.

Cui Y, Wu J (2005) Statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTLs. Genet Res 86: 65-75.

Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang Q (2005) The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. Theor Appl Genet 110: 1445-52.

Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). Genetics 165: 759-69.

George AW, Visschler PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156: 2081-2092.

Goodnight C J (1999) Epistasis and heterosis, pp. 59-67 in The Genetics and Exploitation of Heterosis in Crops, edited by J. G. Coors and S. Pandey. American Society of Agronomy, Crop Science Society of America and Soil Science Society of America, Madison, WI.

Hayes P, Szucs P (2006) Disequilibrium and association in barley: Thinking outside the glass. Proc Nat Acad Science 103: 18385-18386.

Henderson CR (1984) Application of linear models in animal breeding, Univ. of Guelph, Ontario Kearsey MJ, Farquhar AG (1998) QTL analysis in plants: where are we now? Heredity 80:137-142.

Kraakman ATW, Niks RE, Van den Berg P, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168: 435-446.

Li ZK, Luo LJ, Mei HW, Wang DL, Shu Q Y, Tabien R, Zhong DB, Ying C S, Stansel JW, Khush G S, Paterson AH (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics 158**:** 1737-1753.

Liao CY, P Wu, B Hu, KK Yi (2001) Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. Theor Appl Genet 103: 104-111.

Malysheva-Otto LV, Ganal MW, Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genet 7:6.

Mei HW, Luo L J, Ying CS, Wang YP, Yu XQ, Guo LB, Paterson AH, Li Z K (2003) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. Theor Appl Genet 107: 89-101.

Melchinger AE, Utz F, Schon CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149: 383-403.

Nagamine Y, Haley CS (2001) Using the mixed model for interval mapping of quantitative trait loci in outbred line crosses. Genet Res 77: 199-207.

Newton-Cheh C, Hirschhorn JN (2005) Genetic association studies of complex traits: design and analysis issues. Mutation Research 573: 54-69.

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet (2004) 109: 508-514.

Piepho HP, Gauch HG (2001) Marker pair selection for mapping quantitative trait loci. Genetics 157: 433-444.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. Genet. Res 67: 175-186.

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Genetics 103: 18656-18661.

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2000) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167: 485-498.

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6: 461-464.

Sillanpaa MJ, Auranen K (2004) Replication in studies of complex traits. Ann Hum Genet 68: 646-657.

Simko I, Haynes KG, Jones RW (2006) Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. Genetics 73: 2237-2245.

Southey BR, Fernando RL (1998) Controlling the proportion of false positives among significant results in QTL detection. Proc. 6th World Congr. Genet. Appl. Livest. Prod. 26: 221-224.
Takashi M et al., (2005) The map-based sequence of the rice genome. Nature 436: 793-800.

Utz HF, Melchinger, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 54: 1839-1849.

Wan XY, Wan JM, Wan L, Jiang JK, Wang HQ, Zhai JF, Weng HL, Wang CL, Lei JL, Wang X, Zhang X, Cheng ZJ, Guo XP (2006) QTL analysis for rice grain length and fine mapping of an identified QTL with stable and major effects. Theor Appl Genet 112: 1258-1270.

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17: 155-60.

Yu, J, Pressoir G, Briggs WH, Vroh BI, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203-208.

Yu SB, Li JX, Xu CG, Tan YF, Gao YG (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 94: 9226-9231.

Yu SB, Li JX, Xu CG, Tan YF, Li XH, Zhang Q (2002) Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. Theor Appl Genet 104: 619-625.

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.

Zhang N, Xu Y, Akash M, McCouch S, Oard JH (2005) Identification of candidate markers associated with agronomic traits in rice using Discriminant Analysis. Theor Appl Genet 110: 721-729.

Zhao K, Aranzana MJ, Kim S, Lister C, Schindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. Plos Genet 3: e4.

CHAPTER 4 ALTERNATIVE ECOTILLING PROTOCOL FOR RAPID, COST-EFFECTIVE SNP DISCOVERY AND GENOTYPING IN RICE (*ORYZA SATIVA* L.)[*]

4.1 Introduction

Variation at the single nucleotide level has been successfully associated with many important human diseases (Xu et al., 2004; Greb et al., 2005; Litonjua et al., 2005; Shi et al., 2005) with subsequent use in clinical diagnostics (Kuppuswamy et al., 1991) and development of novel drugs (Waschke et al., 2005). SNPs are widely used in molecular evolutionary studies that focus on the origin and diversity of various plant and animal species (Olsen and Purugganan 2002). A genome-wide SNP identification effort in rice has recently been published from two publicly available *indica* (93-11) (http://rise.genomics.org.cn/rice/index2.jsp) and *japonica* (Nipponbare) (http://rgp.dna.affrc.go.jp/) genome sequences (Feltus et al., 2004; Shen et al., 2004). SNP mutations in the rice *alk* gene that encodes for soluble starch synthase (SSSIIa) have been shown to alter the amylose content in grains (Umemoto et al., 2004). SNP mutations in the rice *alk* gene that encodes for soluble starch synthase (SSSIIa) have been shown to be associated with altered SSSIIa enzyme activity, cooking quality, and amylopectin properties in rice (Umemoto et al., 2002; Fjellstrom et al., 2004; Umemoto et al., 2004). Fjellstrom et al. (2004) detected two additional SNPs in exon 8 of the *alk* gene associated with cooking quality. The *waxy* gene encodes for the granule-bound starch synthase enzyme, and the first exon-intron donor splice site was found to interfere with normal mRNA splicing, leading to low amylose production in the grain (Isshiki et al., 1998; Yamanaka et al., 2004).

The potential utility of SNPs in plants to elucidate gene function and regulation has been proposed (Feltus et al., 2004; Henikoff et al., 2004). Conventional SNP discovery is typically

[*] Reprinted with permission from the editor of the "Plant Molecular Biology Reporter" journal.

carried out by direct alignment of sequences obtained from whole genomes (Feltus et al., 2004; Shen et al., 2004), genes (Olsen and Purugganan, 2002), or cDNAs/ESTs (Grivet et al., 2003; Morales et al., 2003; Chaves et al., 2005). Standard methods for SNP typing include allele specific polymerase chain reaction (Moutou et al., 2001; Hayashi et al., 2004), single nucleotide primer extension (Xiong et al., 1998; Russom et al., 2003), cleaved amplified polymorphic sequence (CAPS) (Thiel et al., 2004), single-strand conformational polymorphism (Sato et al., 2003), pyrosequencing (Ronagi, 2001), and heteroduplex analysis by denaturing high performance liquid chromatography (dHPLC) (Giordano et al., 1999). However, these methods require expensive equipment set up, complex PCR primer design, and often experience a high rate of false positives (Comai et al., 2004).

The TILLING and Ecotilling methods were developed as high-throughput systems to simultaneously identify and genotype mutations that exist in mutagenized or natural populations by heteroduplex analysis using single-stand specific (sss) nucleases such as CEL I nuclease (Oleykowski et al., 1998) and mung bean nuclease (Colbert et. al., 2001; Comai et al., 2004; Till et al., 2004a). The strength of these methods lies in their ability to reduce sequencing costs associated with verification of SNP haplotype analysis in large populations (Comai et al., 2004; Gilchrist et al., 2005). TILLING and Ecotilling have been used to study DNA variation in *Arabidopsis thaliana* (Till et al., 2003a; Comai et al., 2004; Henikoff et al., 2004), maize (Till et al., 2004b), wheat (Slade et al., 2005), and *Drosophilia melanogaster* (Winkler et al., 2005). In addition, the Surveyor^{TM} Mutation Discovery platform from Transgenomic, Inc., NE uses a heteroduplex mismatch detection system with CEL I nuclease and reverse phase HPLC for mutation detection. The results to date show that these techniques can be successfully carried out in a well established laboratory dedicated to high-throughput analysis that requires a large initial

investment in equipment and subsequent purchase of relatively expensive chemical reagents. However, this approach may not be suitable for research programs with limited funds that cannot purchase costly equipment or regents as required by TILLING/ Ecotilling. This is particularly true where SNP analysis is a component, but not the major emphasis of a research program. Moreover, the single strand-specific endonucleases used in Ecotilling/TILLING were reported to exhibit simultaneous and competing exo-nuclease activity during mismatch cleaving reactions (Kroeker et al., 1976; Till et al., 2004a). This unwanted exonuclease activity of the single strand-specific nucleases is believed to reduce sensitivity of this method (Till et al., 2004a).

To address these issues, a simple, rapid, efficient, and cost-effective alternative to standard Ecotilling for SNP discovery and genotyping in rice that can be easily adapted to small or medium-sized laboratories was developed. Utility of the modified Ecotilling approach for SNP discovery and genotyping was demonstrated by evaluation of a 922 bp region of the *alk* gene and a 472 bp fragment of *waxy* gene among a large diverse group of rice accessions from 13 countries.

4.2 Materials and Methods

4.2.1 Plant Material and DNA Isolation

A total of 57 diverse inbred rice accessions representing low, intermediate, and high amylose classes were selected for the present study (Table 4.1). The selected materials represent a wide geographical sampling from 13 countries and three continents. (28 lines from USA, 1 line from Brazil, 8 lines from NE Asia, 8 lines from SE Asia, 11 lines from S Asia, and 1 line from Africa). This collection also contains the cultivar Nipponbare to serve both as the reference and internal negative control. All accessions, except EPAGRI 106 from Brazil and HB1 from Dr. Rush, Louisiana State University, were obtained from the USDA National Plant Germplasm

collection. Amylose class and starch gelatinization temperature data were obtained from the USDA-ARS-Germplasm Resources Information Network and USDA-ARS Rice Research Unit, USDA Grain Quality Laboratory, Beaumont, TX. The varieties Cocodrie, Cypress, and Ratna were included twice as internal controls. Four to five plants from each accession were grown to the three leaf stage in the greenhouse, and 25 mg samples from single leaf tissue from each plant were bulked (100 mg) to isolate genomic DNA using the GenElute Plant Genomic DNA kit (Sigma-Aldrich, MO). DNA stocks at 2.5 ng/µl were prepared for each accession. Pools of genomic DNA, each consisting of 8 accessions with equal amounts of DNA, were assembled from these working stocks and used to evaluate sensitivity of the alternative Ecotilling protocol. The last genomic pool consisted of only 4 individuals.

4.2.2 Primer Design for *Alk* and *Waxy* Gene Regions

4.2.2.1 SNP Discovery

Primers were designed for the *alk* and *waxy* genes using Primer3 v.4 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) with sequence information obtained from GenBank accessions AP003509 and AP002542, respectively. For validation of SNP discovery by this protocol, a set of primers namely, alkF1 (5'-GTG GGG TTC TCG GTG AAG AT-3') and alkRn (5'-AAG CAA GAG GCA AAC AGC TC-3') were designed to amplify a 922 bp DNA fragment (from bp 4041 to 4963) of the *alk* gene (Figure 4.1a). A second set of primers namely, waxyF3 (5'-TGC ATC TTT CAT TGC TCG TT-3') and waxyHR (5'- TGC TTC ACT TCT CTG CTT GTG-3') were also designed to amplify a 472 bp DNA region (from bp 1655 to 2127 positions) within the first intron of the *waxy* gene (Figure 4.1b).

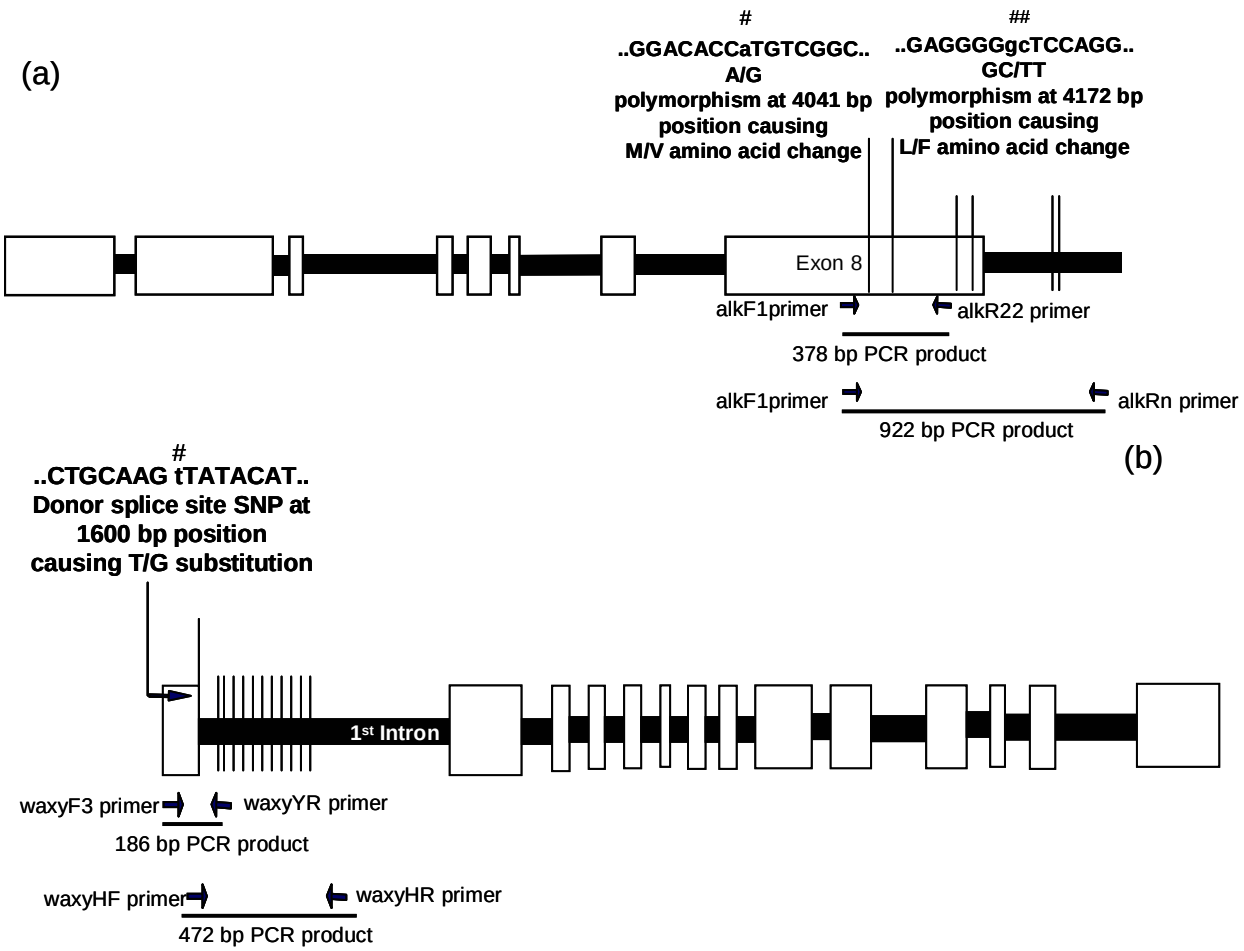


Figure 4.1 (a) Diagram of exons (white boxes) and introns (black boxes) of rice *alk* gene showing location of six SNPs in exon 8. The long vertical bars depict locations of two known SNPs, whereas the short vertical bars show the locations of four newly discovered SNPs by the alternative Ecotilling method. The # symbols are placed over the SNPs in lower case. Positions of two PCR primer sets are designated with arrows and the horizontal line represents the amplified product containing the two SNPs. (b) Location of the SNP at donor splice site of intron 1 in *waxy* gene and the primers designed to amplify the 186 and 472 bp products.

4.2.2.2 SNP Genotyping

One primer set designed to amplify a 378 bp targeted fragment in exon 8 of the *alk* gene consisted of the reverse primer, alkR22 (5'-CCA TTG GTA CTT GGC CTT GA-3') and the alkF1 primer (described in the previous section). This DNA region contains two mutations, reported by Fjellstrom et al. (2004), at positions 4041 and 4172

bp (Figure 4.1a). The second primer set consisted of the forward waxyF3 primer (5'-TGC TTC ACT TCT CTG CTT GTG-3') and the reverse waxyYR primer (5'-TTT CCA GCC CAA CAC CTT AC-3') designed to amplify a 186 bp DNA fragment bracketing one reported SNP (1600 bp position) at the donor splice site of the *waxy* gene (Isshiki et al., 1998) (Figure 4.1b).

4.2.3 Polymerase Chain Reaction (PCR)

A total of 5 ng of genomic DNA from each accession or pooled sample was used to amplify the target region in a 10 µl PCR reaction containing 1 X polymerase buffer, 2.0 mM each dNTP mix (GeneAmp, Applied Biosystems, CA), 0.08 µM of each primer, 5 % DMSO, and 0.2 U of Optimase (Transgenomic, Inc., NE) or AmpliTaq (Perkin Elmer, NJ) polymerase enzyme. PCR reactions were performed using the iCycler (Bio-Rad, CA). The thermocycle profile used to amplify the 922 bp *alk* fragment was 95^0C - 4 min, 33 cycles of (95^0C - 30 s, 62^0C - 15 s, 72^0C - 60 s) and 72^0C - 5 min. PCR amplifications for the 378 bp fragment of the *alk* gene, and the 186 bp and 472 bp *waxy* fragments were carried out using 95^0C - 4 min, 30 cycles of (95^0C - 30 s, 57.2^0C - 30 s, 72^0C - 30 s) and 72^0C - 5 min.

4.2.4 Mega-Gel Preparation

4.2.4.1 SNP Discovery

The Mega-Gel Dual High-Throughput Vertical Electrophoresis Unit (C.B.S. Scientific Company, CA) (Wang et al., 2003) was used in the present study. Non-denaturing gels (5.0 % (w/v) acrylamide/bis-acrylamide (19:1), 0.5X TBE buffer (110 mM Tris, 90 mM Boric acid, 2.5 mM EDTA, pH 8.0), 0.07% (w/v) ammonium persulfate, 0.08% (w/v) TEMED) were prepared in a 50 (L) x 22 (W) x 1.5 (T) cm format

for evaluating products of PCR/sss nuclease assays. The gels were prepared without Gel Wrap® casting system gaskets.

4.2.4.2 SNP Genotyping

To carry out SNP genotyping using the alternative Ecotilling protocol, 6.5% (w/v) acrylamide/bis-acrylamide PAGE gels were prepared as described in the above SNP discovery section.

4.2.5 Alternative Ecotilling Using CEL I Nuclease

Commercial standard electrophoresis grade CEL I or Surveyor nuclease (Transgenomic, Inc., NE) was used to genotype the SNPs in the *alk* and *waxy* genes. A 100 ng aliquat of PCR product (5-6.5 µl) from the queried rice line (individual or pooled DNA) was mixed with 100 ng (5-6.5 µl) of the PCR product from the Nipponbare cultivar. Mixtures were denatured and annealed to form heteroduplex DNA molecules using the iCycler as per the Surveyor^{TM} Mutation Discovery kit protocol (Transgenomic, Inc., NE). This 1:1 PCR product mix was then made to a final 15 µl reaction volume by adding 1.5 µl of 10X reaction buffer, 0.5 µl of enhancer solution (Transgenomic, Inc., NE), and 0.5 µl of CEL I nuclease. The reactions were incubated at 45^0C for 15 min and then stopped by adding 3 µl of 6x stopping dye (0.0625% Xylene cyanol, 30% glycerol solution, 135 mM EDTA). Entire samples were then immediately loaded onto a freshly prepared gel along with the DNA markers. Fifty ng of a 20 bp DNA marker (Sigma-Algrich, MO) premixed with the 1.5 µl of 10X CEL I nuclease buffer (1 X concentration) were loaded per well. Gels were run at 300 v for 90 min and then stained for 10 min using 200 mL of 0.5 X TBE, pH 8.0, containing a 1:10000 dilution of SYBR Gold nucleic acid gel stain (Molecular Probes, OR) followed by two washings using 500 ml of

distilled water. Images of gels were obtained with a KODAK GL100 system using the 535 nm WB 50 optical band-pass filter. Bands were analyzed for SNP genotyping using KODAK 1D 3.0 software. SNP genotypes observed at both *alk* and *waxy* loci were scored as band present (1) or absent (0) for each individual or pooled sample.

4.2.6 Alternative Ecotilling Using Mung Bean Nuclease

To examine the flexibility of this alternative protocol, mung bean nuclease (New England Biolabs, MA) was also used for SNP genotyping. The 10 X mung bean nuclease reaction buffer (100 mM $MgSO_4$, 2 mM $ZnSO_4$, 200 mM Bis-Tris pH 6.5, 0.02% Triton X-100, and 0.002 mg/ml BSA) needed for the single-strand specific nuclease activity was prepared as per Till et al. (2004a). The experimental protocol followed was similar to that of CEL I nuclease described previously except that the reactions were incubated at 65^0C for 30 min.

4.2.7 Standard Ecotilling Assay

The original Ecotilling assay was also carried out for both *alk* and *waxy* gene regions. The eight-fold genomic DNA pools of 57 rice accessions were queried against Nipponbare DNA for both these gene regions as per the International Rice Research Institute website protocol on the 4300 LI-COR gel analysis system (http://www.knowledgebank.irri.org/microarray2004/docs/ Lab_session_EcoTILLING_protocol.doc). The 5' end IRDye 700 modification was carried out for the alkF1 and waxyF3 primers, whereas alkR22 and waxyHR primers were labeled with IRDye 800 at their 5'ends. The denaturation and annealing steps were carried out as per the Surveyor^{TM} Mutation Discovery kit protocol. The LI-COR images were then manually scored for SNP analysis.

4.2.8 DNA Sequencing and Alignment

For verification of the alternative Ecotilling SNP discovery and genotyping results, both the *alk* and *waxy* gene PCR products of all 57 accessions and those of the reference Nipponbare cultivar were sequenced. All sequencing reactions were performed at the LSU Genomics Core Facility, Pennington Biomedical Research Center. The sequencing reactions were carried out using BigDye® Terminator v3.1 Cycle Sequencing Kit with 5 to 20 ng of the gel-purified PCR product (Zymoclean™ Gel DNA Recovery Kit, Zymo Research, CA) and were analyzed using the ABI 3100 DNA Analyzer. Sequences were then aligned using ClustalX 1.8 software (http://bips.u-strasbg.fr/fr/Documentation/ ClustalX/) to locate SNPs. All mismatches in the alignments were manually traced back to their sequencing quality (phred) scores before evaluating the modified Ecotilling results.

4.3 Results and Discussions

4.3.1 SNP Discovery

The PCR amplified products of eight genomic DNA pool templates of 57 rice lines were analyzed against that of Nipponbare cultivar as per the alternative Ecotilling protocol for both *alk* gene and *waxy* gene regions. For the 922 bp PCR heteroduplex analysis of the *alk* gene, new SNPs were identified in all genomic DNA pools (Figure 4.2). Therefore, all individual members from all the 8-fold pools were sequenced and SNPs were characterized and revealed four new SNPs at bp positions 4525 (A/G), 4541 (A/G), 4693 (G/A) and 4695 (C/T), along with two previously reported SNPs at positions 4041 (A>G) and 4172 (GC>TT) by Fjellstrom et al. (2004) (Figure 4.2).

A small cluster of 10 SNPs and one indel in intron 1 comprising a 472 bp region

of the *waxy* gene was also evaluated by alternative Ecotilling that reproducibly detected polymorphism in this region including the indel, but it did not clearly identify the individual SNPs (data not shown). Instead, multiple digestion products of similar size were observed as a smear on the gel, so in this case sequencing would be necessary to resolve each variation. Subsequent sequencing of 30 randomly selected lines for this region confirmed the alternative Ecotilling results. Presence of clusters for SNP discovery is not considered a problem, since the average distance between individuals SNPs in the rice genome is estimated to be ~ 500 bp (Feltus et al., 2004). Standard Ecotilling using IR Dye labeled primers was also carried out for this same *waxy* gene region, and as expected all the bands were clearly separated (data not shown). Even in such a case, sequencing was also necessary to properly score the *waxy* SNP cluster in pools, but not the individual, DNA templates of the standard Ecotilling protocol.

4.3.2 SNP Genotyping

The alternative Ecotilling protocol using CEL I nuclease was performed individually for 57 rice accessions against Nipponbare (1:1) for the two SNPs in the exon 8 region of the *alk* gene (Figure 4.3). The genotyping was carried out by analyzing the modified Ecotilling images supplemented with the aligned sequence information of all observed representative haplotype samples and the reference accession. The (A/G) polymorphism was not observed in the reference Nipponbare accession, but was detected in the remaining 56 lines (Table 4.1). A relatively small proportion of the accessions (13/57; 23%) contained the GC/TT mutation which may not be surprising given the presence of two substitutions at this locus. Nevertheless, a moderately high percentage of U.S. accessions (8/20; 40%) contained the GC/TT polymorphism which was four-fold

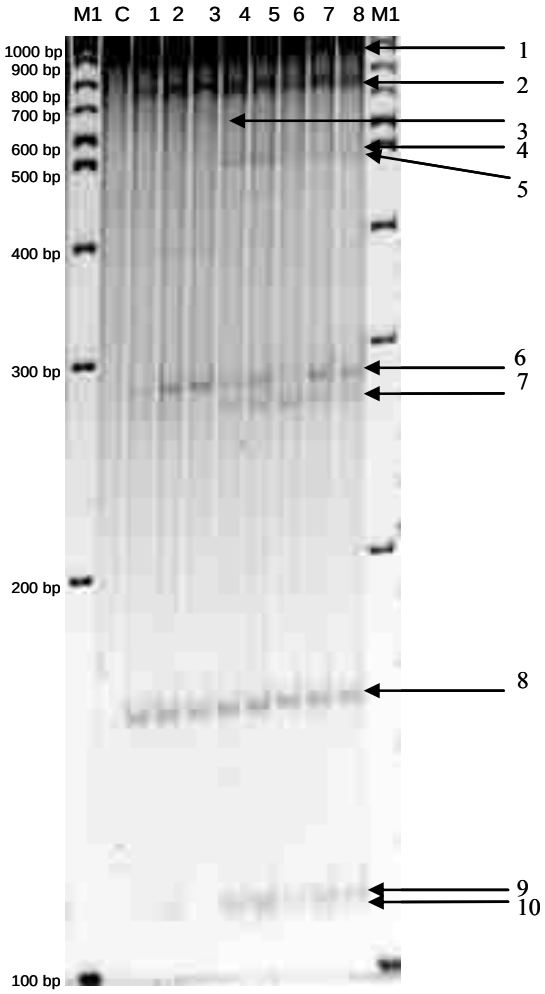


Figure 4.2 SNP genotyping using pools (8 samples combined) of genomic DNA for 922 bp exon 8 region of the *alk* gene. The order of individual rice accessions of the eight-sample DNA pools is the same as that denoted in Figure 4.4. The last genomic pool consists of only 4 individuals. Lane C is the Nipponbare/Nipponbare homoduplex control; Lanes 1 to 8 are heteroduplex digestion products of individual pools using CEL I; M1 = 100 bp marker. Arrow 1 points to CEL I nuclease undigested 922 bp PCR product mix; arrow 2 - 766 bp DNA fragment for the first A/G SNP; arrow 3 - 635 bp DNA fragment for GC/TT SNP; arrow 4 - 521 bp DNA band derived from digestion of fragments 3 and 9; arrow 5 - 490 bp DNA band derived from digestion of fragments 2 and 7; arrow 6 - 287 bp DNA fragment for the GC/TT SNP and 283 bp DNA fragment for the second A/G SNP; arrow 7 - 266 bp DNA fragment for the third A/G SNP; arrow 8 - 156 bp DNA fragment for the first A/G SNP; arrow 9 - 114 bp DNA fragment for the G/A SNP; arrow 10 - 112 bp DNA fragment for the C/T SNP.

more frequent than germplasm originating from southeast Asia (Myanmar, Philippines, Taiwan, Bangladesh, India, Nepal, Pakistan).The modified Ecotilling genotyping results were compared with sequence information from all 57 accessions and found to be in complete agreement. All accessions containing the GC/TT mutation displayed low gelatinization temperature, although nine accessions with low gelatinization temperature did not display this mutation (Table 4.1).

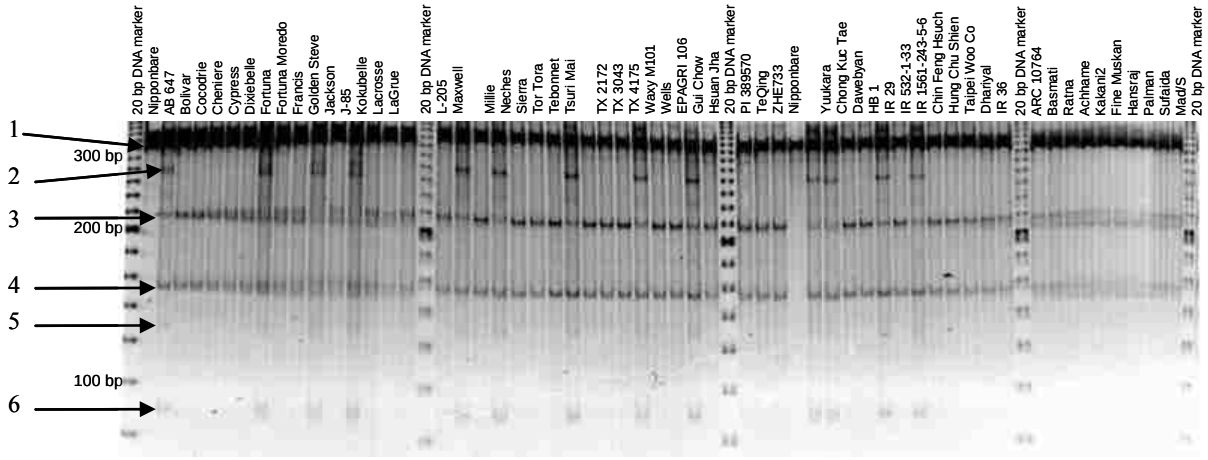


Figure 4.3 Modified Ecotilling of two SNPs in exon 8 region of the *alk* gene for 57 accessions using CEL I nuclease. Lanes 1 - 57 are CEL I nuclease digests of PCR heteroduplex molecules of 57 accessions with Nipponbare; Lane 2 contains CEL I nuclease digest of Nipponbare/Nipponbare PCR homoduplex; Arrow 1 points to CEL I nuclease undigested 378 bp PCR product mix; arrow 2 - 287 bp DNA fragment corresponding to GC/TT SNP; arrow 3 - 221 bp DNA fragment for the A/G SNP; arrow 4 - 156 bp DNA fragment for the A/G SNP; arrow 5 - 131 bp DNA band derived from digestion of fragments 2 and 3; arrow 6 - 91 bp DNA fragment for GC/TT SNP.

Alternative Ecotilling using mung bean nuclease in the *waxy* gene for the 57 accessions revealed that a majority (46/57; 81%) carried the T/G SNP at the donor splice (Figure 4.4). All five accessions from China carried this SNP while both the accessions from Japan, including Nipponbare did not (Table 4.1). Inspection of the G/A SNP,

Table 4.1 SNP genotypes in *alk* and *waxy* genes of 57 rice accessions using the Alternative Ecotilling protocol

Accession	NPGS/GRIN Number	Origin	Gel. Temperature[a]	Amylose Class[b]	Exon 8 of *alk* gene		Intron 1 of *waxy* gene	
					SNP A/G	SNP GC/TT	SNP T/G	SNP G/A
AB 647	ND[c]	USA	Low	High	1	1	1	1
Bolivar	PI 628791	USA	Intermediate	High	1	0	1	1
Cocodrie	PI 6063631	USA	Intermediate	High	1	0	1	0
Cheniere	ND	USA	Intermediate	High	1	0	1	0
Cypress	PI 9700184	USA	Intermediate	Intermediate	1	0	1	0
Dixiebelle	PI 595900	USA	Intermediate	High	1	0	1	1
Fortuna	PI 275448	USA	Low	Low	1	1	0	0
Fortuna Moredo	PI 431075	USA	ND	ND	1	0	1	1
Francis	PI 632447	USA	Intermediate	Intermediate	1	0	1	0
Golden Steve	PI 612579	USA	Low	Low	1	1	0	0
Jackson	PI 572412	USA	Intermediate	Intermediate	1	0	1	0
J-85	PI 595927	USA	Low	Low	1	1	0	0
Kokubelle	PI 612581	USA	(Intermediate)[d]	(Intermediate)[d]	1	0	1	0
Lacrosse	PI 389966	USA	Intermediate	Low	1	0	1	0
LaGrue	PI 568891	USA	Intermediate	Intermediate	1	0	1	0
L-205	PI 608664	USA	Intermediate	High	1	0	1	1
Maxwell	PI 612582	USA	(Low)[d]	(Low)[d]	1	1	0	0
Millie	PI 538354	USA	Intermediate	Intermediate	1	0	1	0
Neches	PI 633972	USA	Low	Glutinous	1	1	1	0
Sierra	PI 633623	USA	Intermediate	High	1	0	1	1
Tor Tora	PI 431150	USA	ND	ND	1	0	1	0
Tebonnet	PI 487195	USA	Intermediate	Intermediate	1	0	1	1
Tsuri Mai	PI 612580	USA	ND	ND	1	1	0	0
TX 2172	TX 2172	USA	Intermediate	Glutinous	1	0	0	0
TX 3043	TX 3043	USA	Intermediate	High	1	0	1	1
TX 4175	TX 4175	USA	Intermediate	High	1	0	1	1
Waxy M101	PI 506223	USA	Low	Glutinous	1	1	0	0
Wells	PI 612439	USA	Intermediate	Intermediate	1	0	1	0
EPAGRI 106	ND	Brazil	ND	ND	1	0	1	0
Gui Chow	ND	China	Low	High	1	1	1	1
Hsuan Jha	PI 160829	China	Intermediate	High	1	0	1	1

Table 4.1 (continued)

Accession	NPGS/GRIN Number	Origin	Gel. Temperature[a]	Amylose Class[b]	Exon 8 of *alk* gene		Intron 1 of *waxy* gene	
					SNP A/G	SNP GC/TT	SNP T/G	SNP G/A
E Che Goo	PI 389570	China	Intermediate	High	1	0	1	1
TeQing	PI 536047	China	Intermediate	High	1	0	1	1
ZHE733	PI 629016	China	Intermediate	High	1	0	1	1
Nipponbare	PI 514663	Japan	Low	Low	0	0	0	0
Yuukara	PI 341937	Japan	Low	Low	1	1	0	0
Chong Kuc Tae	CI 12284	S Korea	Low	Glutinous	1	1	1	1
Dawebyan	PI 222405	Myanmar	Low	High	1	0	1	1
HB 1	ND	Philippines	Intermediate	Glutinous	1	0	0	0
IR 29	PI 393986	Philippines	Low	Glutinous	1	1	0	0
IR 532-1-33	PI 388332	Philippines	Intermediate	High	1	0	1	1
IR 1561-243-5-6	PI 385340	Philippines	Low	High	1	1	1	1
Chin Feng Hsuch	PI 389048	Taiwan	Low	High	1	0	1	1
Hung Chu Shien	PI 389073	Taiwan	Low	High	1	0	1	1
Taipei Woo Co	PI 294397	Taiwan	Intermediate	High	1	0	1	1
Dhariyal	PI 297569	Bangladesh	Intermediate	High	1	0	1	1
IR 36	PI 408586	Philippines	Intermediate	High	1	0	1	1
ARC 10764	PI 373576	India	Low	High	1	0	1	1
Basmati	PI 173923	India	Intermediate	Intermediate	1	0	1	0
Ratna	PI 413980	India	Low	High	1	0	1	1
Achhame	PI 400028	Nepal	Low	Intermediate	1	0	1	1
Kakani2	PI 400020	Nepal	Low	Intermediate	1	0	1	0
Fine Mushkan	PI 385765	Pakistan	Low	High	1	0	1	0
Hansraj	PI 385815	Pakistan	Intermediate	High	1	0	1	1
Palman	PI 385814	Pakistan	Intermediate	High	1	0	1	1
Sufaida	PI 385819	Pakistan	Intermediate	High	1	0	1	1
Mad/S	PI 385323	Rwanda	Intermediate	High	1	0	1	1

[a] Gel. temperature = starch gelatinization temperature class as estimated by alkali spreading value documented in USDA-ARS-GRIN or USDA-ARS Rice Research Unit records. [b] Amylose class where apparent amylose content falls into the following categories: Glutinous = 0 to 5%, Low = 5 to 19%, Intermediate = 19 to 23%, and High > 23%. [c] ND = no data available. [d] Data as provided on US Plant Variety Protection description of accession.
'0' indicates presence of the wild type or reference haplotype for a given SNP, whereas '1' indicates presence of the mutant or alternate haplotype for the same SNP. Note: the order of rice accessions is rearranged as per their geographical origin, but not as per the order in which the alternative Ecotilling assay was carried out.

located 85 bp downstream of the donor slice site, showed that half of the accessions (30/57) contained this polymorphism. All of the 30 accessions with this SNP also carried the T/G SNP at the donor splice site. The sequence information from the 57 accessions was once again in complete agreement with the alternative Ecotilling results. The G/A polymorphism occurred at the 5' end of a group of 10 SNPs and in an insertion/deletion (Olsen and Purugganan, 2002) that spanned a region of 472 bp. The T/G polymorphism was associated with an intermediate or high amylose class in those 40 accessions with measured amylose data except for the Chong Ku Tae accession. In addition, a G/A polymorphism was associated with the high amylose class in 26 of the 29 accessions carrying this mutation (Table 4.1).

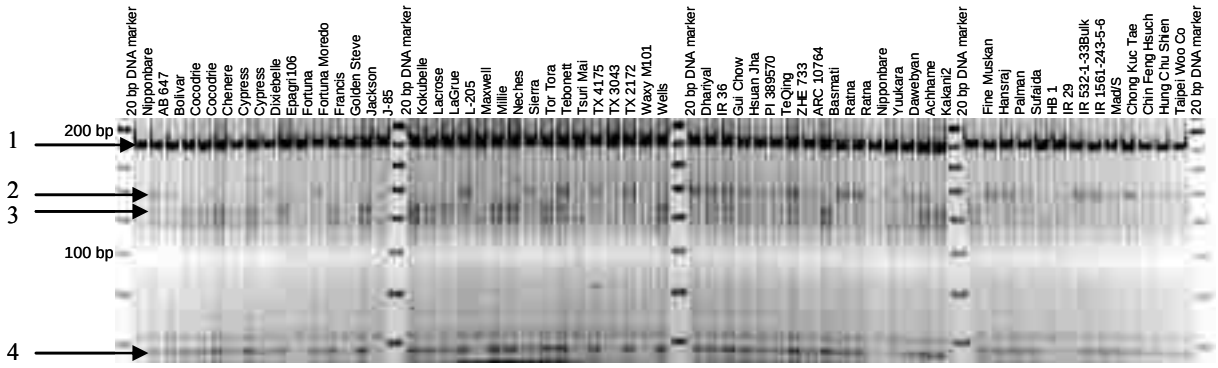


Figure 4.4. Modified Ecotilling of *waxy* locus for 57 accessions using mung bean nuclease. Lanes 1 - 57 are CEL I nuclease digests of PCR heteroduplex molecules of 57 accessions with Nipponbare. The M lane - 100 bp marker; Lanes 2 and 45 - CEL I nuclease digest of Nipponbare / Nipponbare homoduplex; Arrow 1 points to undigested 186 bp PCR product mix; arrow 2 - 140 bp DNA fragment for G/A mutation; arrow 3 - 131 bp DNA fragment for T/G donor splice site SNP; arrow 4 - 56 bp DNA fragment for T/G donor splice site SNP. Three accessions (Cocodrie, Cypress, and Ratna) were duplicated as controls in lanes 5, 7 and 45, and accessions are not in the same order as Table 4.1. All 30 US accessions were alphabetically ordered, followed by alphabetical ordering of the remaining accessions according to their country of origin.

4.3.3 Sensitivity of the Alternative Ecotilling Method

The eight-fold pools of genomic DNA prepared from the 57 accessions were also genotyped for the *alk* gene region, and the results obtained were consistent with the individual

SNP results for all members of the pools (Figure 4.5c). All SNPs were faithfully detected and signal or band intensity was comparable to that of individual SNP typing shown in Figure 4.5a, b. All genomic DNA pooled samples were assayed for the *alk* region via the standard Ecotilling method as per the IRRI protocol. Results from the standard method were consistent with information obtained previously by alternative Ecotilling and sequence information (data not shown).

The CEL I nuclease and mung bean (sss) endonucleases were used separately to evaluate flexibility of the alternative Ecotilling protocol (Figure 4.5a, b). Four different amounts of commercial CEL I nuclease (1 µl, 1/2 µl, 1/3 µl, 1/4 µl) and mung bean nuclease concentrations (10 U or 1 µl, 5 U or 1/2 µl, 3.33 U or 1/3 µl and 2.5U or 1/4 µl) were tested. Use of 1 µl of commercial CEL I nuclease or 1 µl of commercial mung bean nuclease was found to produce the highest band intensities. However, use of 0.33 µl CEL I nuclease or 0.5 µl mung bean nuclease gave reproducible and unequivocal genotyping of SNPs. In fact, all results presented in this paper for SNP genotyping of the *alk* and *waxy* loci were carried out using the reduced concentrations.

4.3.4 Time and Cost Analyses

Time requirements for both modified and standard Ecotilling procedures were determined for the following steps: PCR setup and running, template denaturation, CEL I nuclease and mung bean nuclease reaction setup and digestion, gel preparation, setup, mounting, and pre-run, comb/membrane loading, gel running (200-1000 bp fragments), staining, imaging and data collection. All the steps of the modified Ecotilling procedure could be completed in 7.5 hr, including data collection and analysis. Each gel illustrated in this paper was generated and analyzed in a single day. In contrast, the standard Ecotilling protocol (in our hands), required

11.5 hr to complete over a two day period. DNA precipitation and volume reduction, software data input, and denaturation of samples and size standards were additional steps required for standard Ecotilling. Finally, only a single digital image was needed to genotype SNPs with alternative Ecotilling, whereas two separate image files were required to collect and analyze data with standard Ecotilling.

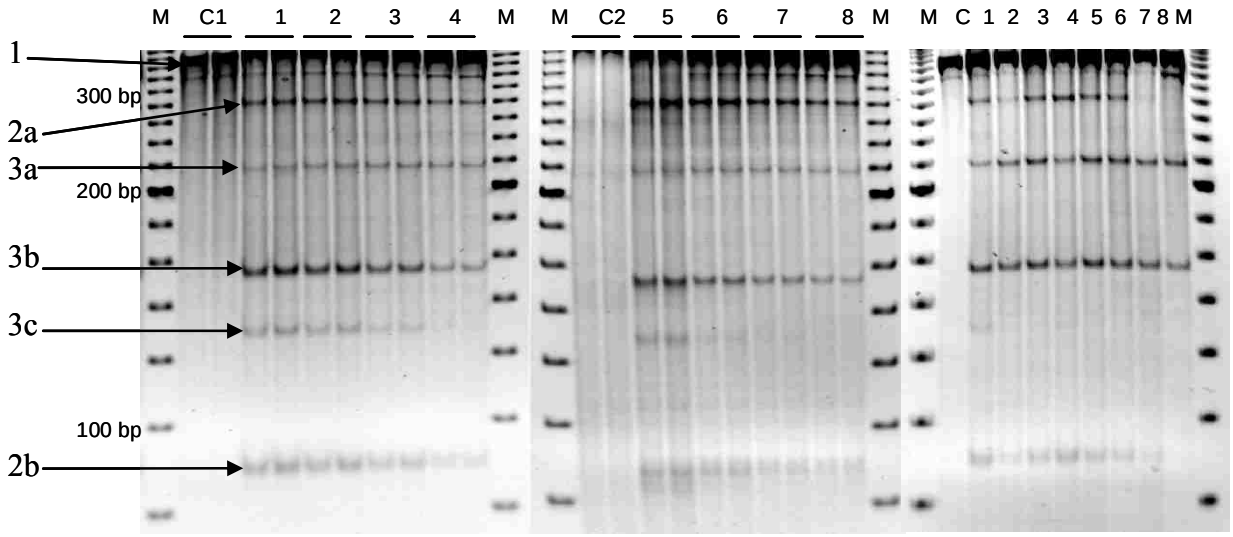


Figure 4.5 Modified Ecotilling for SNP detection in exon 8 of *alk* gene carried out with varying concentrations of CEL I nuclease and mung bean nucleases. All treatments were replicated twice as indicated with a horizontal line above corresponding lanes. Panel a: Lanes C1,2 - PCR homoduplex digests of Nipponbare cultivar using 1 µl of CEL I nuclease (negative control); Lanes 1 through 8 - PCR heteroduplex digests of Waxy M101and Nipponbare using 1 µl, 1/2 µl, 1/3 µl and 1/4 µl of CEL I nuclease, respectively. Arrow 1 points to CEL I nuclease undigested 378 bp PCR product mix; arrow 2 - 287 bp DNA fragment for A >G SNP; arrow 3 - 221 bp DNA fragment for the GC/TT SNP; arrow 4 - 156 bp DNA fragment for the GC/TT SNP; arrow 5 - 131 bp DNA band derived from digested fragments 2 and 3; arrow 6 - 91 bp DNA fragment for A/G SNP; Panel b: Lanes C1, 2 - PCR homoduplex digest of Nipponbare using 1 µl of mung bean nuclease (negative control); Lanes 1 through 8 - PCR heteroduplex digests of Waxy M101 and Nipponbare using 1 µl, 1/2 µl, 1/3 µl and 1/4 µl of mung bean nuclease, respectively. Panel c: SNP genotyping using pools (8 samples combined) of genomic DNA for exon 8 region of *alk* gene. Lane C is homoduplex control; Lanes 1 to 8 are heteroduplex digestion products of individual pools using CEL I; M = 20 bp marker. The order of individual rice accessions for the eight-sample DNA pools is the same as that denoted in Figure 4.4.

Cost estimates for SNP analysis that included the following items were determined for both alternative and standard Ecotilling (Table 4.2): Genomic DNA isolation, pipette tips, sample tubes, PCR and PAGE gel reagents, labeled and unlabeled primers, gel buffer, loading dye, size standards, and CEL I nuclease and mung bean nuclease. Membrane combs were specific to standard Ecotilling as was SYBR Gold dye to modified Ecotilling. The cost/sample for standard Ecotilling was estimated to be $1.96 when using CEL I nuclease while alternative Ecotilling was $1.26, a saving of ~ 46%, primarily attributed to reduced costs of unlabeled primers and native PAGE gels. A reduction of $0.48/sample was possible in both procedures when commercial mung bean nuclease vs. CEL I nuclease was used. Even though some researchers may prefer to carry out direct sequencing at ~ $25 per 1000 bp, both strands, for SNP genotyping, the alternative Ecotilling method is justified considering the reduced costs and increased speed.

Table 4.2 Time requirements for different stages of alternative vs. standard Ecotilling

Modified Ecotilling		Standard Ecotilling	
Day 1		Day 1	
PCR reaction setup	30 min	PCR reaction setup	30 min
PCR	120 min	PCR	120 min
Denaturation	25 min	Denaturation	25 min
CEL I nuclease reaction setup	15 min	CEL I nuclease reaction setup	15 min
CEL I nuclease reaction	20 min	CEL I nuclease reaction	20 min
		Isopropanol precipitation	45 min
Gel preparation	30 min	Volume reduction by incubating at 85 ^{0}C	45 min
Gel setting	60 min	Subtotal	300 min (5 hrs)
Gel mounting	5 min	Day 2	
Sample loading into 100 well comb	5 min	Gel Preparation	45min
Gel running (500 bp)	90-120 min	Gel setting	60min
Gel staining and imaging	20min	Gel cleaning and mounting	30 min
		SAGA template preparation	15 min
		Pre-run	25 min
		Denaturation of size standards	5 min
		Loading samples onto 96 well TILLING membrane	30 min per plate
		Gel running (500bp)	180 min
		Subtotal	390 min (6.5 hrs)
Total	450 min (7.5 hr)	Total	690 min (11.5 hrs)

4.4 References

Chaves LD, Rowe JA, Reed KM (2005) Survey of a cDNA library from the turkey (*Meleagris gallopavo*). Genome 48: 12-17.

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) High-throughput screening for induced point mutations. Plant Physiol 126: 480-484.

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J 37: 778-786.

Fjellstrom RG, Chen M, Bergman CJ,, McClung AM (2004) Single nucleotide polymorphism markers at the rice *alk* locus controlling alkali spreading value. In: Rice Technical Working Group Meeting Proceedings, February 29-March 4, 2004, New Orleans, LA.

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14: 1812-1819.

Gilchrist EJ, Haughn GW (2005) TILLING without a plough: a new method with applications for reverse genetics. Curr Opin Plant Biol 8: 211-215.

Giordano M, Oefner PJ, Underhill PA, Cavalli Sforza LL, Tosi R, Richiardi PM (1999) Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. Genomics 56: 247-53.

Greb RR, Grieshaber K, Gromoll J, Sonntag B, Nieschlag E, Kiesel L, Simoni M (2005) A common single nucleotide polymorphism in exon 10 of the human follicle stimulating hormone receptor is a major determinant of length and hormonal dynamics of the menstrual cycle. J Clin Endocrinol Metab 10: 2004-2268.

Grivet L, Glaszmann JC, Vincentz M, Silva F da, Arruda P (2003) ESTs as a source for sequence polymorphism discovery in sugarcane: example of the *Adh* genes. Theor Appl Genet 106:190-197.

Hayashi K, Hashimoto N, Daigen M, Ashikawa I 2004. Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. Theor Appl Genet 108:1212-20.

Henikoff S, Till BJ, Comai L (2004) TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol 135: 630-636.

Isshiki M, Morino K, Nakajima M, Okagaki RJ, Wessler SR, Izawa T, Shimamoto K (1998) A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5' splice site of the first intron. Plant J 15:133-138.

Kroeker WD, Kowalski D, Laskowski M Sr (1976) Mung bean nuclease I. Terminally directed hydrolysis of native DNA. Biochemistry 15: 4463-7.

Kuppuswamy MN, Hoffmann JW, Kasper CK, Spitzer SG, Groce SL, Bajaj SP (1991) Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX), cystic fibrosis genes. Proc Natl Acad Sci USA 88: 1143-1147.

Litonjua AA, Belanger K, Celedon JC, Milton DK, Bracken MB, Kraft P, Triche EW, Sredl DL, Weiss ST, Leaderer BP, Gold DR (2005) Polymorphisms in the 5' region of the CD14 gene are associated with eczema in young children. J Allergy Clin Immunol 115: 1056-1062.

Morales M, Roig E, Monforte AJ, Arus P, Garcia-Mas J (2004) Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.). Genome 47: 352-360.

Moutou C, Gardes N, Rongieres C, Ohl J, Bettahar-Lebugle K, Wittemer C, Gerlinger P, Viville S (2001) Allele-specific amplification for preimplantation genetic diagnosis (PGD) of spinal muscular atrophy. Prenat Diagn 21: 498-503.

Oleykowski CA, Bronson Mullins CR, Godwin AK, Yeung AT (1998) Mutation detection using a plant endonuclease. Nucleic Acids Res 26: 4597-4602.

Olsen KM, Purugganan MD. (2002) Molecular evidence on the origin and evolution of glutinous rice. Genetics 162:941-950.

Qiu P, Shandilya H, D'Alessio JM, O'Connor K, Durocher J, Gerard GF (2004) Mutation detection using Surveyor nuclease. Biotechniques 36: 702-7.

Rahman MH, Jaquish B, Khasa PD (2000) Optimization of PCR protocol in microsatellite analysis with silver and SYBR stains. Plant Mol Biol Rep 18: 339-348.

Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. Genome Res 11: 3-11.

Russom A, Tooke N, Andersson H, Stemme G (2003) Single nucleotide polymorphism analysis by allele-specific primer extension with real-time bioluminescence detection in a microfluidic device. J Chromatogr A 1014: 37-45.

Sato Y, Nishio T (2003) Mutation detection in rice waxy mutants by PCR-RF-SSCP. Theor Appl Genet 107: 560-567.

Shen YJ, Jiang H, Jin JP, Zhang ZB, Xi B, He YY, Wang G, Wang C, Qian L, Li X, Yu QB, Liu HJ, Chen DH, Gao JH, Huang H, Shi TL, Yang ZN (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. Plant Physiol 135: 1198-1205.

Shi J, Zhang S, Tang M, Ma C, Zhao J, Li T, Liu X, Sun Y, Guo Y, Han H, Ma Y, Zhao Z (2005) Mutation screening and association study of the neprilysin gene in sporadic Alzheimer's

disease in Chinese persons. J Gerontol A Biol Sci Med Sci 60: 301-306.

Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, non-transgenic approach to wheat crop improvement by TILLING. Nature Biotechnology 23: 75-81.

Thiel T, Kota R, Grosse I, Stein N, Graner A (2004) SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. Nucleic Acids Res 32: 1-5.

Till BJ, Burtner C, Comai L, Henikoff S (2004a) Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res 32: 2632-2641.

Till BJ, Colbert T, Tompa R, Enns LC, Codomo CA, Johnson JE, Reynolds SH, Henikoff JG, Greene EA, Steine MN, Comai L, Henikoff S (2003a) High-throughput TILLING for functional genomics. Methods Mol Biol 236: 205-220.

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S (2003b) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res 13: 524-530.

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004b) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4: 12.

Umemoto T, Yano M, Satoh H, Shomura A, Nakamura Y (2002) Mapping of a gene responsible for the difference in amylopectin structure between *japonica*-type and *indica*-type rice varieties. Theor Appl Genet 104: 1-8.

Umemoto T, Aoki N, Lin H, Nakamura Y, Inouchi N, Sato Y, Yano M, Hirabayashi H, Maruyama S (2004) Natural variation in rice starch synthase IIa affects enzyme and starch properties. Functional Plant Biol 31: 671-684.

Wang D, Shi J, Carlson SR, Cregan PB, Ward RW, Diers BW (2003) A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci 43: 1828-1832.

Waschke KA, Villani AC, Vermeire S, Dufresne L, Chen TC, Bitton A, Cohen A, Thomson AB, Wild GE (2005) Tumor necrosis factor receptor gene polymorphisms in Crohn's disease: association with clinical phenotypes. Am J Gastroenterol 100: 1126-1133.

Winkler S, Schwabedissen A, Backasch D, Bokel C, Seidel C, Bonisch S, Furthauer M, Kuhrs A, Cobreros L, Brand M, Gonzalez-Gaitan M (2005) Target-selected mutant screen by TILLING in Drosophila. Genome Res 15: 718-723.

Xiong Z, Tsark W, Singer-Sam J, Riggs AD (1998) Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension. Nucleic Acids Res 26: 684-686.

Xu B, Arlehag L, Rantapaa-Dahlquist SB, Lefvert AK (2004) Beta2-adrenergic receptor gene single-nucleotide polymorphisms are associated with rheumatoid arthritis in northern Sweden. Scand J Rheumatol 33: 395-398.

Yamanaka S, Nakamura I, Watanabe KN, Sato Y (2004) Identification of SNPs in the *waxy* gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theor Appl Genet 108: 1200-1204.

CHAPTER 5 DEVELOPMENT AND APPLICATION OF HAPLOTYPE-SPECIFIC ASSAYS FOR GENOTYPING OF THE AROMA GENE IN RICE

5.1 Introduction

5.1.1 Market Potential of Aromatic Rice

A growing demand for aromatic rice has created new and expanding market opportunities in the United States, Canada, the Middle East and Europe (Cordeiro et al., 2000; Jin et al., 2003). International market value of "Jasmine" aromatic rice for Thailand in 2003 was $840 million while "basmati" aromatic rice produced $960 million for India and Pakistan in the same year (http://basmati.com/aromatic/index.shtml). In the U.S. nearly 12% of the total rice consumed is aromatic, primarily imported and consumed by the Asian-American community (Sha, 2005). In addition, the aromatic rice farmer often secures higher returns for his produce than the conventional rice growing farmer (Jin et al., 2003).

5.1.2 Aroma Detection Assays

The aroma characteristic in rice is primarily attributed to accumulation of 2-acetyl-1-pyrroline in leaf and seed tissues (Buttery et al., 1983). Determination of rice seed aroma by conventional abrasive, tasting, cooking or chemical methods is generally imprecise, time consuming, requires large amounts of sample materials, and often suffers from lack of agreement among laboratories (Lorieux et al., 1996; Widjaja et al., 1996; Cordeiro et al., 2002). Use of molecular markers as an alternative to conventional methods for distinguishing aromatic and non-aromatic genotypes has been evaluated in previous research (Ahn et al., 1992; Garland et al., 2000; Cordeiro et al., 2002; Jin et al., 2003). Ahn et al. (1992) mapped the aroma trait 4.5 cM from the restricted fragment length polymorphism (RFLP) marker RG28 on chromosome 8

(Garland et al., 2000). The simple sequence repeat (SSR) marker SCU015RM was detected 4 cM from the aroma trait by Cordeiro et al. (2002). Jin et al. (2003) subsequently reported the association of aroma with a C/T SNP marker located ~ 2 cM from the *fragrance* (*fgr)* gene. However, these molecular markers were not tightly linked with the aroma trait, so this approach to distinguish aromatic and non-aromatic rice genotypes was not successful (Jin et al., 2003).

5.1.3 The Aroma Gene

Bradbury et al. (2005a) reported that the recessive nature of aroma was controlled by the *fgr* gene on chromosome 8 which encodes for the enzyme, betaine aldehyde dehydrogenase (BAD2). Sequence analysis of the *fgr* or *BAD2* gene from 14 aromatic and 62 non-aromatic rice varieties revealed presence of two aromatic and non-aromatic haplotypes in exon 7, consisting of two A/T mutations (bp positions 6147, 6149, Genbank accession AP004463), an 8 bp indel (bp positions 6151 to 6158), followed by a C/T mutation at bp position 6159. The C/T SNP of this haplotype was postulated to generate a transcriptional stop signal during synthesis of the BAD2 enzyme that leads to production of a truncated protein (Bradbury et al., 2005a). The truncated non-functional BAD2 enzyme presumably lacks three conserved protein motifs needed for its substrate binding activity, resulting in accumulation of 2-acetyl-1-pyrroline in plant tissues and the aroma characteristic in rice Thus, presence or absence of the aroma SNP haplotype would be directly associated with presence or absence of the aromatic trait for marker-assisted introgression into rice varieties.

5.1.4 SNP Genotyping Assay for the Aromatic Rice

Although current SNP genotyping technologies promise accurate and high-throughput results, their utility is seriously limited by use of expensive reagents and detection equipment

(Hayashi et al., 2004). In this context, a number of inexpensive allele-specific SNP genotyping assays have been developed by several researchers (Ye et al., 2001; Soleimani et al., 2003; Zhang et al., 2003; Chiapparino et al., 2004; Bundock et al., 2006). However, these methods require use of complex PCR primer designing, cycling conditions and scoring strategies. In contrast, the two allele-specific SNP primer system with a common reverse primer (Zhang et al., 2003) has been employed in previous genotyping studies of single copy genes in diploid organisms such as rice (Hayashi et al., 2004). However, for combined genotyping of more than one SNP, this approach requires a restricted size of amplified PCR products within a specific range to avoid unambiguous discrimination. This study emphasizes a simple modification of the methods of Hayashi et al. (2004) that results in accurate scoring of PCR haplotype products on agarose gels that differ by only a single base pair.

Bradbury et al. (2006) reported a single tube allele-specific amplification method for genotyping aroma-associated haplotypes in rice. However, due to presence of a wide range of DNA bands (585 bp, 577 bp, 355bp and 257 bp) in the PCR amplified products generated by this method, multiple loadings of the PCR products, either on the agarose or the native PAGE gel is not possible. This difficulty will seriously limit the high-through put scaling ability for the SNP genotyping method described by Bradbury et al. (2006). In addition, the SNP genotyping by this method requires use of higher PCR reaction volumes than the conventional allele specific assays. Moreover, results from these studies were obtained primarily from Australian temperate *japonica* germplasm that may not be applicable for *indica* or tropical *japonica* lines.

The objective of this research was to develop and apply a haplotype-specific assay for genotyping of the aroma gene in US and Asian rice germplasm. The technique developed during this study involved use of unlabeled primers, optimization of PCR cycling conditions and a

simple assay using standard gel electrophoresis.

5.2 Materials and Methods

5.2.1 Plant Material and Genomic DNA Extraction

Three sets of lines from U.S. (Louisiana, California) and Asian (Thailand, India, Japan) sources were used to carry out the *BAD2* SNP haplotype study. The first set consisted of seven non-aromatic Japanese and U.S. inbred rice varieties, namely, Nipponbare, Cocodrie, Cheniere, Cypress, Francis, Trenasse, CL131, and 13 aromatic U.S. and Asian inbred varieties/lines, namely A201, A301, Basmati370, Calmati, Della, Dellamati, Dellrose, KDM-105, LA2131, LA2137, LA2140, LA2177 and LA0502183 (Figure 5.1a). For the haplotype assays described below, leaf samples from a single plant of each line or variety were collected in 2005 field plots at the Rice Research station, Crowley, LA. A total of 100 mg of leaf tissue/line was used to isolate genomic DNA using the GenElute Plant Genomic DNA kit (Sigma-Aldrich, MO), and ~ 2 to 4 ng from this genomic DNA was used to perform the haplotype assays.

The second set of material consisted of 50 breeding lines in the F_2 to F_7 generations developed at the Rice Research Station, Crowley, LA that tested positive for aroma by a standard cooking method (Sha et al., 2000) (Table 5.1). Collection of leaf samples and DNA isolation were the same as that described for the 20 varieties. The third set of materials consisted of four to five progeny derived in the subsequent generation from each of the 50 breeding lines. Each individual was grown in the greenhouse and leaf samples from 3-4 week seedlings of each progeny were collected for haplotype assays. Seeds from each progeny line were harvested, dried to ~ 12% moisture and scored for aroma as described by Sha et al. (2000). Seeds of the aromatic Dellrose and non-aromatic Cocodrie varieties served as positive and negative controls, respectively, for the experiment.

5.2.2 Haplotype-specific Primer Design

All primers were designed using the Primer3 software (Rozen and Helen 2000). Sequence information for the *BAD2* gene was obtained from Genbank accession no. AP004463. In case of the aromatic and non-aromatic haplotypes, the group of linked SNP and indel alleles occurred over five nucleotide bases (from bp positions 6147 to 6151) and over 13 nucleotide bases (from bp positions 6147 to 6159), respectively (Figure 5.1a). Using this information, four different haplotype-specific forward primer sets were designed and evaluated (results not shown). Unlike the Zhang et al. (2003) original protocol, current haplotype-specific forward primers did not contain a single target SNP allele at the last nucleotide base. Instead, the linked SNP and indel alleles were placed at three different locations within the 22-nt (aromatic) or 23-nt (non-aromatic) primer. Among all primers tested, the following set produced accurate discrimination between aromatic and non-aromatic haplotypes: aromatic specific primer, 5′-CTG GTA TAT ATT TCA GCT GAT C-3′, designed to contain 'T' alleles at bp positions 6147 and 6149 (see Figure 1a) for the first two SNPs, presence of an 8 bp deletion at bp positions 6151 through 6158, and a 'T' allele for the last SNP at bp 6159. Positions for the SNP and indel alleles associated with primer sequences are underlined. The non-aromatic forward primer 5′-AAA GAT TAT GGC TTC AGC TGA TC-3′ was designed with an 'A' allele for the second SNP at bp 6149 bp, no deletion at bp 6151 through 6158 and a 'C' allele for the third SNP at bp 6159 bp. The first "A" SNP at position 6147 was not included due to large Tm differences between the primer pairs. To reduce background amplification with other haplotypes, a nucleotide mismatch of C→A was introduced in the forward primers at the third base position upstream of the 3′ termini as described by Zhang et al. (2003) and Hayashi et al. (2004). No SNPs or mismatches were introduced into the common reverse primer, 5′-CCA GTG AAA CAG GCT GTC AA-3′.

The aromatic and non-aromatic specific primers were used to amplify expected 236 bp and 237 bp PCR products, respectively, from exon 7 of the *BAD2* gene (Figure 5.1b, c).

5.2.3 Haplotype-Specific Polymerase Chain Reaction (HS-PCR)

To evaluate each variety or line, two separate PCR reactions specific to each of the *BAD2* haplotypes were carried out using the respective forward primer and the common reverse primer. The 6.5 µl PCR mix consisted of the following: 2 µl of 2.5 ng μl^{-1} template genomic DNA mixed with 0.65 µl of 10x Taq Buffer (15 mM $MgCl_2$, Gene Amp, Applied Biosystems, CA), 0.52 µl of 2.5 mM dNTPs mix (Gene Amp, Applied Biosytems), 0.13 µl each of 20 µM forward and reverse primers and 0.05 µl of 5 U μl^{-1} AmpliTag DNA Polymerase (Applied Biosytems, CA). The following optimized stringent PCR cycling conditions were carried out using the BioRad iCycler: 95^{o}C - 2 min, 28 cycles of (95^{o}C - 12 s, 60^{o}C - 12 s, 72^{o}C - 12 s) and 72^{o}C - 5 min. For haplotype scoring, 5 µl of each PCR product were loaded onto a 2% agarose gel, run for 20 min at 7.5 V cm^{-1} in 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) and photographed after staining with ethidium bromide. To demonstrate high levels of specificity and absence of background, the remaining 1.5 µl of the PCR product was loaded onto a non-denaturing 6.0% (w/v) acrylamide/bis-acrylamide (19:1) poly-acrylamide gel, electrophoresed for 45 min in 0.5x TBE buffer (110 mM Tris, 90 mM Boric acid, 2.5 mM EDTA, pH 8.0) at 15 V cm^{-1} and stained with SYBRGreen (Molecular Probes, OR). Presence or absence of bands generated from each of the aromatic and non-aromatic specific PCRs were scored as 1 or 0, respectively, to identify corresponding homozygous or heterozygous haplotypes (Table 1). The 1, 0 and 1, 1 scores represent the dominant, non-aromatic haplotypes (homozygous and heterozygous states, respectively), whereas 0, 1 denotes presence of the recessive homozygous aroma haplotype.

5.2.4 DNA Sequencing and Alignment

For validation of genotyping results from the first set of 20 varieties, PCR products were gel-purified with the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, CA), and 25 to 50 ng were used as template for cycling reactions using the BigDye® Terminator v3.1 Cycle Sequencing Kit. PCR products of both aromatic and non-aromatic *BAD2* haplotypes were sequenced on both strands. Sequence data were aligned using the ClustalX 1.8 software (http://bips.u-strasbg.fr/fr/Documentation/ClustalX/) to locate SNPs and compare haplotype results.

5.3 Results and Discussions

5.3.1 Aroma Phenotypes and Haplotypes of 20 Varieties

The ability of the haplotype-specific assay to detect aromatic or non-aromatic haplotypes at the *BAD2* gene was first carried out in a blind study with 20 known inbred aromatic and non-aromatic varieties and lines (Figure 5.1a). In two separate PCR experiments, genomic DNA of each variety or line was used to generate haplotypes using each of the specific forward primers and the common reverse primer. Haplotypes were successfully scored with either agarose or PAGE gels, although the PAGE format showed distinct bands of greater intensity than the agarose method. Out of 20 varieties/lines genotyped, 13 aromatic types were homozygous for the expected 236 bp band. Both haplotypes for five of the aromatic varieties/lines (A301, Basmati 370, Calmati, Dellrose, and KDM-105) are shown in lanes 1 to 5 (non-aromatic) and the corresponding 1′ to 5′ (aromatic) (Figure 5.1b, c). The remaining 7 non-aromatic varieties were homozygous for the expected 237 bp band. Results for five varieties (Cheniere, Cocodrie, Cypress, Francis and Nipponbare) are shown in lanes 6 to 10 (non-aromatic) and 6′ to 10′ (aromatic) (Figure 5.1b, c). To validate genotyping results, PCR amplified products from exon 7

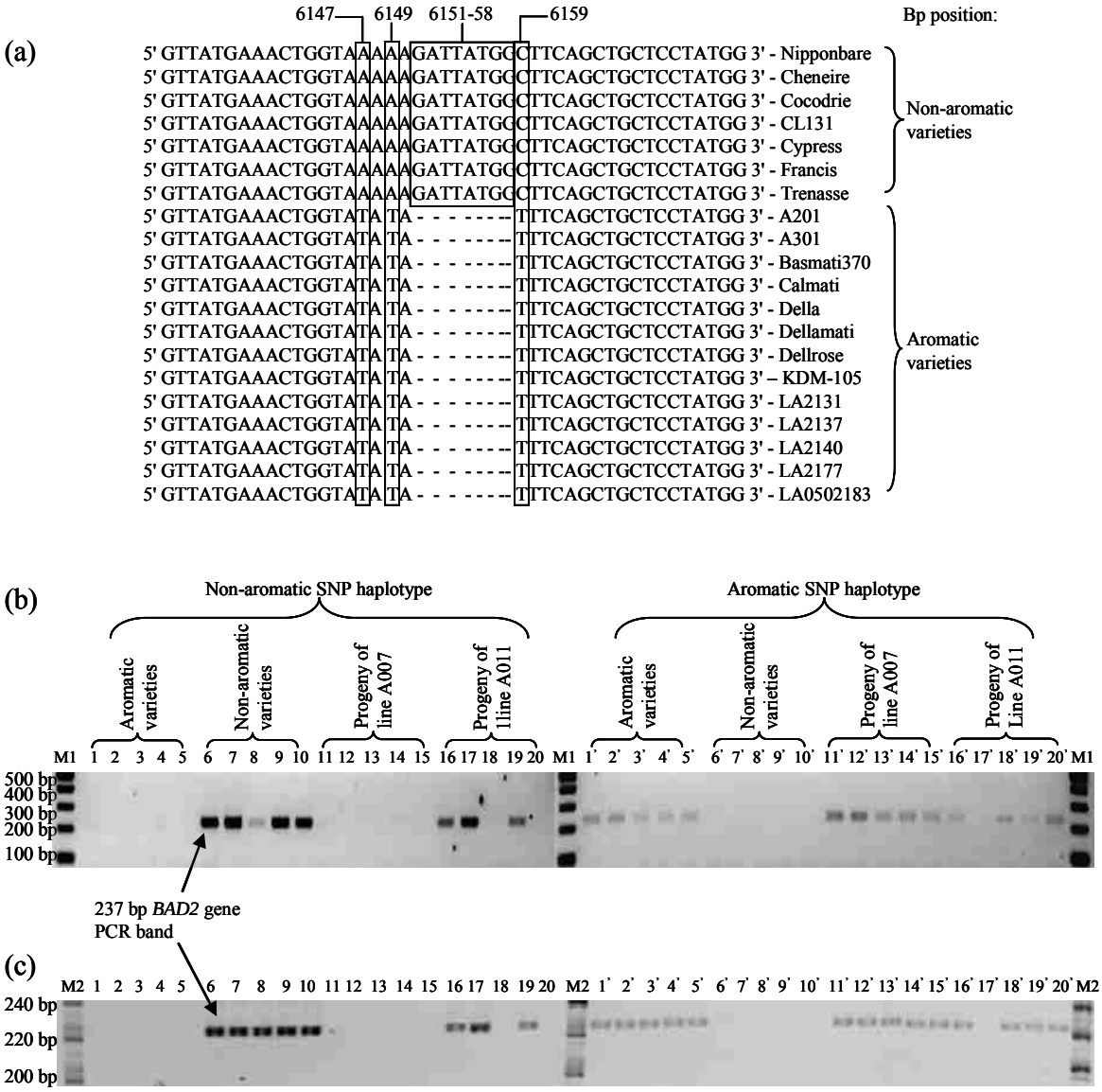


Figure 5.1 Haplotype-specific assays of aromatic and non-aromatic lines. Panel (a): Sequence alignment results for the exon 7 haplotypes of *BAD2* gene (bp positions 6131 - 6176) carried out for the 20 aromatic and non-aromatic varieties. Vertical boxes are used to highlight the observed SNP and indel alleles. Panel (b): haplotypes of varieties/lines observed on 2 % agarose gel; M1 = 100 bp DNA ladder (N.E. Biolabs); Lanes 1 - 5 and the corresponding 1′ - 5′ are A301, Basmati 370, Calmati, Dellrose, and KDM-105, respectively; Lanes 6 - 10 and the corresponding 6′ - 10′ are Cheniere, Cocodrie, Cypress, Francis and Nipponbare, respectively; Lanes 11 - 15 and the corresponding 11′ - 15′ are 5 progeny of breeding line A007 (see Table 1); Lanes 16 - 20 and the corresponding 16′ - 20′ are 5 progeny of breeding line A011 (not in the same order as in Table 1). Panel (c): haplotypes of lines as in Panel (b) observed on 6 % (w/v) native Mega-Gel PAGE gel; M2 = 20 bp PCR DNA marker (Sigma-Aldrich).

of the *BAD2* gene were sequenced for both stands and aligned for all 20 varieties/lines (Figure 5.1a). Sequencing results confirmed that all 13 types scored as aromatic showed the expected 8 bp deletion at positions 6151 through 6158 and three 'T' alleles at bp positions 6147, 6149, and 6169 that is consistent with results of Bradbury et al. (2005a). For all 7 non-aromatic varieties, sequencing revealed the presence of two 'A' SNP alleles at bp positions 6147 and 6149, no deletion at bp positions 6151 through 6158, and a 'C' SNP allele at bp position 6159 in agreement with Bradbury et al. (2005a). Recombinant haplotypes were not observed for any variety/line, corresponding to results of Bradbury et al. (2005a), suggesting that only two SNP haplotypes occur in exon 7 of the *BAD2* gene. Thus, there was complete correlation between sequencing alignment results and the *BAD2* haplotypes.

5.3.2 Aroma Phenotypes and Haplotypes of 50 Breeding Lines and Their Progeny

To further assess versatility of the aroma haplotype assay, I evaluated 50 U.S. rice breeding lines (second set) for association of the aroma trait and the SNP haplotypes in exon 7 of the *BAD2* gene (Table 5.1). All 50 lines were scored as aromatic as determined by the protocol of Sha et al. (2000). Because this breeding material may be actually segregating for aroma and the haplotypes observed from the first set of 20 varieties, all 50 breeding lines were grown in the greenhouse and advanced to the next generation that represents the third set of lines. Presence/absence of seed aroma in 4 to 5 progeny from each of the 50 lines was determined as before and recorded for this third set of material (Table 5.1). Out of the 226 progeny tested, a majority (203/226, ~90%) strongly expressed the aroma trait. Due to insufficient seed production, determination of aroma phenotype for 22 progeny with the aromatic haplotype and two progeny with non-aromatic haplotypes was not determined. Therefore, this material will not be discussed. Nevertheless, genotyping results presented in Table 5.1 show that 203 progeny

with the aromatic phenotype also exhibited the aroma haplotype (0, 1) while 23 non-aromatic progeny produced the dominant non-aromatic haplotypes (either 1, 0 or 1, 1). Thus, results demonstrate the expected association between haplotypes and observed aroma phenotypes for the 226 progeny tested. Sequencing of target haplotype regions in segregating individuals was not performed because ambiguous results ('N' reads) occur frequently at the SNP sites (Kadaru, unpublished results), an expected result when two SNP alleles occur in the same template sample from heterozygotes.

The introduction of a mismatch towards the 3′ terminus for the forward haplotype-specific primer substantially reduced background PCR amplification of other genotypes and produced reliable discrimination between both haplotypes. Substitution of the pyrimidine cytosine by the purine adenine at the third base position from the 3′ end eliminated background bands that otherwise would cause erroneous genotyping results. Figures 5.1b and 5.1c show that haplotype discrimination was unambiguous even when performed on moderately sensitive ethidium bromide stained agarose gels and as well as on highly sensitive SYBRGreen stained non-denaturing PAGE gels. The strategy to conduct separate PCR reactions and score bands on different gels resulted in accurate and easy determination of haplotypes/PCR products that differed by only one base pair. Moreover, this approach would allow correct scoring of haplotypes, a feature not possible by the method of Hayashi et al. (2004). Using the unlabeled primer set, a 96 well thermocycler, and a native PAGE gel format, 400 individual samples can be processed and analyzed in one day by a single operator. The cost for each haplotype assay was estimated to be ~ $0.15 using the equipment, reagents, and methods described in this study. Thus, this haplotype assay is particularly suitable for marker-assisted research projects desiring moderately high throughput on a limited budget.

Table 5.1 Phenotypes and SNP haplotypes of rice breeding lines and their progeny

Line	Pedigree	Progeny 1		Progeny 2		Progeny 3		Progeny 4		Progeny 5	
		Hap[a]	Phen[b]	Hap	Phen	Hap	Phen	Hap	Phen	Hap	Phen
A007[c]	Dellrose/3/Katy/NWBT//Jodon	0,1	-	0,1	a	0,1	a	0,1	a	0,1	a
A010	96INT/AR1188	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A011	CCDR/A301	1,0	na	1,1	-	1,1	na	0,1	a	0,1	a
A012	L202/Leah//Toro/3/IR67016	0,1	a	0,1	a	0,1	-	0,1	a	0,1	a
A013	J-85/Della/3/ L202/Leah//Toro	0,1	a	0,1	a	0,1	a	1,1	na	0,1	a
A015	Calmati	0,1	a	0,1	a	0,1	a	0,1	a	0,1	-
A016	CPRS//L201/RU7402003/3/BASMATI SUF AID PAK	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A018	902207X2/DG 1275	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A021	A201/SADARI TYPE[d]	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A022	A201/SADARI TYPE	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A023	A201/SADARI TYPE	0,1	a	0,1	a	0,1	a	0,1	-	0,1	a
A025	Dellrose/3/Katy/NWBT//Jodon /4/JSMN/DLLA	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A028	A201//ADAR/JODN/3/CPRS	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A030	CPRS/LGRU//97 KDM X2-5	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A031	Dellrose/3/Katy/NWBT//Jodon	1,1	na	0,1	a	0,1	a	0,1	-	0,1	a
A032	Dellrose/3/Katy/NWBT//Jodon	0,1	a	0,1	a	0,1	-	0,1	a	1,0	na
A033	97 KDMX2-1/Wells	0,1	a	0,1	a	0,1	a	0,1	-	0,1	a
A034	CCDR/LGRU//97 KDMX2-5	0,1	a	0,1	a	0,1	a	0,1	a	0,1	-
A035	CCDR/LGRU//97 KDMX2-5	0,1	a	0,1	a	0,1	a	0,1	a	0,1	-
A037	DLMT/B8462T3-710//DMSI/3/RSMT/4/Wells	0,1	a	0,1	-	0,1	a	0,1	a	0,1	a
A038	DLMT/B8462T3-710//DMSI/3/RSMT/4/Wells	0,1	a	0,1	a	0,1	a	0,1	a	0,1	-
A040	L202/Leah//Toro/3/IR67016	0,1	a	0,1	-	0,1	a	0,1	a	0,1	a
A041	DLMT/B8462T3-710//DMSI/3/RSMT/4/Wells	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A042	Dellmont/3/NWBT/KATY//L202	0,1	a	0,1	a	1,1	na	0,1	a	0,1	a
A043	Dellrose/3/Katy/NWBT//Jodon	1,1	na	1,1	na	0,1	a	1,1	na	1,0	na
A044	CPRS/Dellrose	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A045	Dellrose/3/Katy/NWBT//Jodon	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A046	DLMT/B8462T3-710//DMSI/3/RSMT	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A047	Dellrose/3/Katy/NWBT//Jodon	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A048	Dellrose/3/Katy/NWBT//Jodon	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A049	A201/SADARI TYPE	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A050	A201/SADARI TYPE	0,1	a	0,1	a	0,1	a	0,1	a	1,0	-
A051	A201/SADARI TYPE	0,1	a	0,1	a	0,1	-	0,1	a	0,1	a
A052	A201/SADARI TYPE	0,1	a	0,1	a	0,1	a	0,1	a	0,1	A
A053	A201/SADARI TYPE	0,1	a	0,1	a	0,1	a	0,1	a	0,1	-

Table 5.1 (continued)

Line	Pedigree	Progeny 1		Progeny 2		Progeny 3		Progeny 4		Progeny 5	
		Hap[a]	Phen[b]	Hap	Phen	Hap	Hap	Phen	Hap	Phen	Hap
A054	A201/SADARI TYPE	1,1	na	0,1	a	0,1	-	0,1	a	0,1	a
A055	A201/SADARI TYPE	1,0	na	0,1	a	1,1	na	0,1	a	0,1	a
A057	J-85/Della/3/RU9302065//LSBR-5/LMNT	0,1	a	0,1	-	0,1	a	0,1	a	0,1	a
A059	J-85/Della /3/RU9302065// LSBR-5/LMNT	0,1	a	0,1	a	0,1	a	0,1	-	0,1	a
A060	J-85/Della /3/RU9302065// LSBR-5/LMNT	0,1	a	0,1	-	0,1	a	0,1	a	0,1	a
A061	J-85/Della /3/RU9302065// LSBR-5/LMNT	0,1	-	1,1	na	0,1	a	0,1	a	0,1	a
A062	DLMT/3/NWBT/KATY/2/L202/4/ DLMT/B8462T3-710//DMSI	0,1	a	1,1	na	1,0	na	0,1	a	0,1	a
A063	Dellrose/3/Katy/NWBT//Jodon /4/A201	0,1	a	0,1	a	0,1	a	0,1	-	0,1	a
A064	Dellrose/3/Katy/NWBT//Jodon /4/J-85/Della	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A065	Dellrose/3/Katy/NWBT//Jodon /4/J-85/Della	0,1	-	0,1	a	1,1	na	1,1	na	0,1	a
A066	Dellrose/3/Katy/NWBT//Jodon /4/ J-85/Della	0,1	a	0,1	a	0,1	a	0,1	-	0,1	a
A067	Dellrose/3/Katy/NWBT//Jodon /4/ J-85/Della	1,1	na	1,1	na	0,1	a	0,1	a	0,1	a
A068	Dellrose/3/Katy/NWBT//Jodon /4/ J-85/Della	0,1	a	0,1	a	0,1	a	1,1	na	0,1	a
A069	Dellrose/3/Katy/NWBT//Jodon /4/ J-85/Della	0,1	a	0,1	a	0,1	a	0,1	a	0,1	a
A070	DLMT/B8462T3-710//DMSI/3/RSMT/4/Wells	0,1	a	1,1	na	0,1	a	1,0	na	0,1	A

Hap[a] = SNP haplotype phase results for the *BAD2* gene. Non-aromatic haplotype (237 bp band) results are designated first, followed by the aromatic haplotypes (236 bp band); 1 = presence of haplotype and 0 = absence of haplotype.
Phen[b] = aroma trait evaluated as per Sha et al. (2000) protocol; 'a' = aroma detected, 'na' = aroma not detected and '-' = data not obtained due to insufficient seed production.
[c] All 50 lines were scored as aromatic as determined by Sha et al. (2000).
[d] Different lines with the same pedigree are full sibs derived from the same cross.

5.4 References

Ahn SN, Bollich CN, Tanksley SD (1992) RFLP tagging of a gene for aroma in rice. Theor Appl Genet 84: 825-828.

Bradbury LMT, Fitzgerald TL, Henry RJ, Jin Q, Waters DLE (2005a) The gene for fragrance in rice. Plant Biotech J 3: 363-370.

Bradbury LMT, Henry RJ, Jin Q, Reinke RF, Waters DLE (2005b) A perfect marker for fragrance genotyping in rice. Mol Breeding 16: 279-283.

Bradbury LMT, Henry RJ, Jin Q, Reinke RF, Waters DLE (2006) Genotyping of the fragrance haplotype phase in rice. Plant & Animal Genomes XIV Conference, San Diego, CA, USA.

Bundock PC, Cross MJ, Shapter FM, Henry RJ (2006) Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. Theor Appl Genet 112: 358-65.

Buttery RG, Ling LC (1983) Cooked rice aroma and 2-acetyl-1-pyrroline. J Agric Food Chem 31: 823-826.

Chiapparino E, Lee D, Donini P (2004) Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR. Genome 47: 414-20.

Cordeiro GM, Christopher MJ, Henry RJ, Reinke RF (2002) Identification of microsatellite markers for fragrance in rice by analysis of rice genome sequence. Mol Breed 9: 245-250.

Garland S, Lewin L, Blakeney A, Reinke RF (2000) PCR based molecular markers for the fragrance gene in rice (*Oryza sativa* L.). Theor Appl Genet 101: 364-371.

Hayashi K, Hashimoto N, Daigen M, Ashikawa I (2004) Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. Theor Appl Genet 108: 1212-20.

http://basmati.com/aromatic/index.shtml - Internet based e-market for the global rice industry.

Jin Q, Waters DLE, Cordeiro GM, Henry RJ, Reinke RF (2003) A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (*Oryza sativa* L.). Plant Sci 165: 359-364.

Lorieux M, Petrov M, Huang N, Guiderdoni E, Ghesquiere A (1996) Aroma in rice: genetic analysis of quantitative trait. Theor Appl Genet 93: 1145-1151.

Rozen S, Helen JS (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology, 365-386. Humana Press, Totowa, NJ, USA.

Sha XY (2005) Researchers make progress on new aromatic rice varieties. Rice Research Station News. Crowley: Louisiana Agricultural Experiment Station 2: 4

Sha XY, Linscombe SD, Bearb KF, Howard AM, Theunissen BW, Hoffpauir HL, Cramer SW (2000) Evaluation of specialty rice progenies for aroma. 92th Annual Research Report: Rice Research Station. Crowley: Louisiana Agricultural Experiment Station 55-58.

Soleimani VD, Baum BR, Jhonson DA (2003) Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley. Plant Mol Bio Rep 21: 281-288.

Widjaja R, Craske JD, Wooton M (1996) Comparative studies on volatile components of non-fragrant and fragrant rices. J Sci Food Agric 70: 151-161.

Ye S, Dhillon S, Ke X, Collins AR, Day INM (2001) An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res 29: e88-8.

Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of allele-specific PCR markers for γ-gliadin alleles in *Triticum aestivum*. Theor Appl Genet 107: 130-138.

CHAPTER 6: DEVELOPMENT AND APPLICATION OF ALLELE-SPECIFIC PCR ASSAYS FOR IMAZETHAPYR HERBICIDE RESISTANCE IN RICE

6.1 Introduction

6.1.1 The Noxious Red Rice Weed

Red rice (*Oryza sativa* L.) is a common and notorious weed in rice fields throughout the world (Tan et al., 2005) causing losses up to $50 million annually in the southern United States alone (Gealy et al., 2003). The yield losses in cultivated rice fields are mainly due to competition for water, nutrients, sunlight, and poor milling quality of rice grains resulting from weed seed mixtures (Dilday et al., 1990). As the red rice biotypes are morphologically similar to cultivated rice and belong to the same genus and species as that of cultivated rice, managing this weed in rice fields has been a very difficult task (Gealy et al., 2003; Tan et al., 2005; Zhang, 2005).

6.1.2 Amino Acid Biosynthesis Inhibiting Herbicides

There are three main types of amino acid biosynthesis inhibiting herbicides namely, 1) EPSP synthase enzyme inhibiting herbicides such as glyphosate, 2) GS enzyme inhibiting herbicides such as glufosinate and 3) AHAS or ALS enzyme inhibiting herbicides such as imidazolinones (imazethapyr, imazamox etc), sulfonylureas, triazolopyramidines, pyrimidinyl-oxybenzoates, and sulfonylamino-carbonyl-triazolinones (Corbett and Tardif, 2006). ALS-inhibiting herbicides are increasingly used for control of weeds that mimic cultivated crops and for both broad spectrum weed control (Tan et al., 2005). Examples of mimic weeds controlled by imidazolinone herbicides are jointed goat grass (*Aegilops cylindrica* Host) in winter wheat fields (Ball et al., 1999), shattercane grass (*Soghum bicolor* (L) Moench) and johnson grass (*Sorghum halepens* (L) Pers) in maize (Krausz et al.,1998; Askew et al., 1999), wild mustard

(*Brassica kaber* (DC) LC Wheeler) in rape oilseed fields (Tan et al., 2005), and red rice (*Oryza sativa* L) in rice fields (Steele et al., 2002). Broad spectrum weed control by these ALS-inhibiting herbicides was reported in rice, sunflower, and rapeseed crops (Tan et al., 2005). All of the above examples of weed control were made possible by growing imidazolinone-resistant crop varieties and by application of imidazolinone herbicides. This system is commercially known as the Clearfield™ production system was developed by screening natural *AHAS* or *ALS* gene variants or by induced mutagenesis (Tan et al., 2006). Due to their non-GM/non-transgenic nature, imidazolinone-resistant crops have gained rapid acceptance by farmers and the commodity markets (Tan et al., 2006). Five imidazolinone-resistant crops namely, Clearfield maize, Clearfield wheat, Clearfield canola, Clearfield sunflower, and Clearfield rice constitute a major proportion of total commercial Clearfield cultivation in US, Canada and Europe (Tan et al., 2006). On the other hand, commercial Clearfield cultivation has disadvantages of increased herbicide dependency and possibility of natural transfer of herbicide resistance genes to weed species via cross pollination (Tan et al., 2005). Furthermore, in the last two decades, nearly 95 weed biotypes have been found to be resistant to these herbicides (Heap, 2007) suggesting concurrent evolution of resistant weed biotypes with increased use of the imidazolinone herbicides.

6.1.3 Gene-flow from Crop Species and Their Wild Relatives

Successful crop-wild relative hybridization events have been documented in the literature for cases such as canola (Jorgensen et al., 1994; Legere 2005), cotton (Brubacker et al., 1993; Dawson et al., 1996), pearl millet (Renno et al., 1997), radish (Campbell et al., 2006), sorghum (Arriola et al., 1996), sugar beet (Arnaud et al., 2003),

sunflower (Alexander et al., 2001; Mercer et al., 2006), and wheat (Guadagnuolo et al., 2001). Moreover, Ellstrand (2003) reported a list of 48 cultivars for which spontaneous hybridization with their wild relatives was documented and also gave estimates for spontaneous hybridization rates for 10 important crop species. With the introduction and adoption of genetically modified crops, numerous cases of transgene-flow into wild relatives and its ecological impacts were also discussed (Quist and Chapala, 2001; Poppy, 2004; Vacher et al., 2004; Andow, 2005; Zhang et al., 2005; Guadagnuolo et al., 2006). Particularly, the spread of herbicide resistance genes was demonstrated in at least two important crop species *viz*., canola (Reiger et al., 2001) and winter wheat (Perez-Jones et al., 2006). Reiger et al. (2001) have documented the pollen-mediated gene transfer between herbicide-resistant canola growing fields and surrounding conventional canola growing fields, while Perez-Jones et al. (2006) have reported introgression of an imidazolinone-resistance gene from winter wheat to jointed goatgrass.

6.1.4 Outcrossing among Cultivated Rice and Red Rice

Even in the case of rice, crop-wild relative hybridization events were reported by many authors (Chen et al., 2004; Messeguer et al., 2004; Song et al., 2004; Wang et al., 2006). Gene flow between cultivated rice and red rice has been reported in detail by Dillon et al. (2001), and Rong et al. (2004). Other reports include Estorminos et al. (2002), Gealy et al. (2003), Madsen et al. (2002) and Zhang et al. (2003). Studies by Estorminos et al. (2002) and Dillon et al. (2001) have revealed 0-0.05% outcrossing rate between the CL 2551 (CL 121) rice and AR red rice biotypes and identified putative imidazolinone-resistant CL 2551 x red rice hybrids. Zhang et al. (2003) established the transfer of glufosinate herbicide resistance from transgenic glufosinate-resistant rice to

red rice. Thus, there is a growing concern pf imazethapyr herbicide resistance acquired by red rice from commercial Clearfield rice cultivars to via cross pollination under field conditions (Tan et al., 2005; Zhang, 2005).

The earliest attempts of identification of crop-wild hybrids by random amplified molecular polymorphic DNA (RAPD) markers were carried out in sunflower (Whitton et al., 1997; Linder et al., 1998) and pearl millet (Arriola et al., 1996). Gealy et al. (2002) used Simple Sequence Repeat (SSR) or microsatellite markers for differentiating among US red rice, rice and hybrid populations. These authors reported that out of the 18 SSR makers they tested, four SSR markers namely RM215, RM234, RM251, and RM253 could reliably distinguish between red rice and rice. Using these SSR markers, Zhang et al. 2005 successfully confirmed outcrossing between Clearfield rice and red rice in Louisiana. Although this molecular marker analysis indicated the presence of outcrossing between Clearfield rice and red rice in chromosomes 3, 6 and 7, it did not show that the mutant ALS gene from Clearfield rice that conferred herbicide resistance (see below) was actually transferred to red rice.

6.1.5 The *ALS* Gene

Acetohydroxy synthase, encoded by the acetohydroxy synthase *AHAS/ALS* gene, is one of the key enzymes in the biosynthetic pathway of the branched chain amino acids namely, isoleucine, leucine and valine in plants and is the target for action of imidazolinone herbicides such as imazethapyr (Tan et al., 2006). Six point mutations in the *ALS* gene (either naturally existing or artificially induced) that confer resistance to this herbicide exist in crop plants and weed biotypes (Tranel and Wright, 2002; Tan et al., 2005; Corbett and Tardif, 2006; Tan et al., 2006). In the *ALS* gene of *Arabidopsis*

thaliana, positions of these six mutations correspond to amino acids Ala_{122}, Pro_{197}, Ala_{205}, Asp_{376}, Trp_{574} and Ser_{653} (Corbett and Tardif, 2006), and these point mutations are designated $A_{122}T$, $P_{197}H$, $A_{205}V$, $D_{376}E$, $W_{574}L$, and $S_{653}T$, respectively. In the case of the rice *ALS* gene, two 'G/A' transition mutations at 1880 bp and 1883 bp positions were reported and these correspond to Ser_{653} and Gly_{654} locations in *A. thaliana* (Tan et al., 2005; Tan et al., 2006). In the rice AHAS gene product, these point mutations are known to cause amino acid substitutions from serine to asparagine (S to D) and glycine to glutamic acid (G to E), respectively. Both these substitutions are reported to prevent binding of imidazolinone herbicides with the rice AHAS enzyme, thus conferring resistance to the imazethapyr (New path) herbicide (Tan et al., 2005; Tan et al., 2006). The commercial Clearfield^{TM} rice technology involves the use of these two *ALS* gene mutations (CL 121 and CL 141 varieties carry the $G_{654}E$ mutation, whereas the CL 161 variety has the $S_{653}D$ mutation) (Tan et al., 2005; Tan et al., 2006). All the Clearfield varieties have the A_{205} mutation which confers resistance to the imazethapyr herbicide.

6.1.6 ALS-inhibiting Herbicide Resistance Assays

The diagnostic tests that are devised for confirmation of ALS-inhibitor herbicide resistance can be broadly classified into three main categories namely, the conventional bioassays, the enzyme based tests and the DNA based methods (Corbett and Tardif, 2006). The conventional herbicide resistance bioassays such as the seedling bioassay, modified seedling assay (Cirujeda et al., 2001; Walsh et al., 2001), pollen germination assay (Ritcher and Powles, 1993), and leaf disc assays (Patzoldt and Tranel, 2002) are generally very effective, but are also labor intensive, time consuming and ineffective in elucidation of cross-resistance patterns (Corbett and Tardif, 2006). The ALS enzyme

activity based assays such as the acetoin accumulation assay (Hinz and Owen, 1995; Kwon and Penner, 1995; Hall et al., 1998), and Ketoacid Reductoisomerase (KARI) assays (Simpson et al., 1995; Lovell et al., 1996) are limited by the complex nature of enzyme extraction procedures and quick deterioration of extracted enzyme samples (Corbett and Tardif, 2006). Direct sequencing of the *ALS* gene, PCR-restriction fragment length polymorphism (PCR-RFLP), PCR amplification of specific alleles (PASA) and denaturing high-performance liquid chromatography (DHPLC) methods are the main DNA based assays developed for the ALS-inhibitor herbicides (Corbett and Tardif, 2006). Direct sequencing (McNaughton et al., 2005) and DHPLC (Siminszky et al., 2005) of the *ALS* gene in the target crop species or weed biotypes are the most informative methods of all DNA based herbicide resistance detection methods. The DHPCL method involves PCR amplification of the target *ALS* gene, preparation of heteroduplex DNA (by mixing wild plant PCR product with mutant plant PCR product and by heating and cooling of the mixture) and their separation on an HPLC column. However, use of these methods is restricted by expensive instruments, time and cost per sample, and high-through put factors (Corbett and Tardif, 2006). The PCR-RFLP method is similar to cleaved amplified polymorphic sequence (CAPS) technique, which involves PCR amplification of target *ALS* gene and subsequent digestion of these PCR products using specific restriction endonucleases. The PCR-RFLP assay was successfully employed for all the reported *ALS* gene mutations in the 6.1.5 subsection *viz*., $P_{197}H$ (Guttieri et al., 1992; Tan and Medd, 2002), $W_{574}L$ (Foes et al., 1999; Tan and Medd, 2002), $A_{122}T$ (Corbett and Tardif, 2006), $A_{205}V$ (Corbett and Tardif, 2006), $D_{376}E$ (Corbett, 2004) and $S_{653}T$ (Corbett and Tardiff, 2006). However, exact substitution

information of the nucleotide base at the mutation site cannot be determined by this method (Corbett and Tardif, 2006). The PASA method is a three primer (namely forward, reverse and middle) PCR set up system, where in the forward and reverse primers non-specific to the target SNP and the middle primer (also a reverse primer) is allele specific. Upon PCR amplification, the herbicide resistant plant DNA yields two bands, while the susceptible plant DNA will produce only one band (Corbett and Tardif, 2006). Application of the PASA, nested PASA, or multiplex PASA method for detection of ALS-inhibiting herbicide resistance was demonstrated by Corbett (2004), Patzoldt and Tranel (2002), Patzoldt and Tranel (2003), and Wagner et al. (2002).

6.1.7 SNP Based Assays in Clearfield Rice x Red Rice Outcrossing Assessment

Development of a robust SNP genotyping technique that can differentiate between the 'A' or 'G' allele at the 1880 bp (S_{653}D SNP) and 1883 bp (G_{654}E SNP) positions of the rice *ALS* gene would enable direct assessment of transfer of this gene from Clearfield rice to red rice. Even though the PASA technique described in the previous subsection is a single nucleotide polymorphism (SNP) based assay, its application in the rice crop was not reported. In addition, this method is seriously limited by its inability to effectively differentiate between the heterozygous and homozygous state of resistant alleles in crop or weed biotypes. Corbett and Tardif (2006) have argued that the intensity of the allele-specific (PASA) band in heterozygous resistant plant would be half that of homozygous resistant plant. However, differences in band intensities can also arise due to differences in initial template DNA concentration, purity of template DNA and even a bad reaction set up by the researcher. It is well known that the action of the *ALS* gene is semi-dominant in nature (Tan et al., 2006) and herbicide crop resistance is a function of

number of *ALS* gene copies. The present study is aimed at constructing a SNP marker assay that would not rely on allele-specific band intensity for distinguishing homozygous resistant (Clearfield rice) and heterozygous resistant (Clearfield rice x red rice) plants. The objective of this research was to develop a simple, reproducible, cost-effective, high-throughput method for screening of imazethapyr herbicide resistant Clearfield rice and Clearfield rice x red rice hybrids. Using the new *ALS* gene SNP assays, a total of 483 field-collections were screened for the presence of $S_{653}D$ SNP and another 145 F_2 progeny lines of natural CL 121 x red rice crosses were screened for the presence of the $G_{654}E$ SNP.

6.2 Materials and Methods

6.2.1 Plant Materials and Their Genomic DNA Isolation

Three different sets of plant collections comprising cultivated rice lines, red rice weed biotypes, and their hybrids were used for carrying out the present study. The first plant collection consisted of the rice varieties Cocodrie, CL 121, CL 141, CL 161, Cypress, LaGrue, Nipponbare, and progeny from CL 121 x red rice, and CL 161 x natural outcrosses. The leaf samples of these plants were collected from different rice growing locations across Louisiana during 2004 and 2005 by Dr. Weiqiang Zhang. Some of the Clearfield rice and red rice samples in this set were collected at more than one location. Genomic DNA from the leaf tissues of these plants were extracted by the procedure described in the next paragraph and were included in the subsequent *ALS* gene SNP assays either as positive or negative control samples.

The second set of plant material was used to conduct the *ALS* $G_{654}E$ SNP assay. Five previously identified naturally outcrossed Clearfield-red rice F_1 hybrids and their F_2

progeny, kindly provided by Dr. Weiqiang Zhang, were screened for the presence of the 'A' or 'G' allele in the *ALS* gene at the $G_{654}E$ locus. Seeds of these five natural outcrosses were planted in the field at Ben Hur Farm in 2004 and Newpath herbicide was applied at 140 gm/ha rate at the two to three leaf stage. A second application of herbicide was carried out at the same rate 20 days after the first application. Leaf tissues of ~30 herbicide resistant lines from each of the five progeny, totaling 145 F_2 lines were collected for genomic DNA extraction. A total of 100 mg of leaf tissue/line was used to isolate genomic DNA using the GenElute Plant Genomic DNA kit (Sigma-Aldrich, MO) and ~ 2-4 ng from this genomic DNA was used to perform the AS-PCR assay.

The third set of plant material was a field collection of 483 putative CL 161 x red rice hybrid plants collected by Dr. Weiqiang Zhang across 11 different locations in Louisiana in 2004. Only leaf samples from the plants which have the distinct red rice morphology were collected. This plant material was used for carrying out the *ALS* $S_{653}D$ SNP assay. Using 25 mg of leaf tissue of these lines, genomic DNA was isolated using the UltraClean-htp™ 96 Well Plant DNA Kit (MO BIO Laboratories, Inc., CA) as per the manufacturer's instructions.

6.2.2 Allele Specific Primer Design and Polymerase Chain Reaction (AS-PCR)

The *ALS* gene sequence information obtained from Genbank no. AP005841 was used to design allele-specific primers for the $G_{654}E$ and $S_{653}D$ SNPs. For each of these SNPs, two separate allele-specific forward primers were designed, of which, the last nucleotide base corresponded to one of the two possible alleles occurring in the *ALS* gene. To increase binding specificity of the allele-specific forward primers with their corresponding target SNP allele, an additional nucleotide mismatch was introduced at the

third or fourth base position upstream of the 3′ termini (Zhang et al., 2003; Hayashi et al., 2004). Both allele-specific primers employed the same reverse primer to amplify the SNP containing region of the *ALS* gene. For both SNPs, 16 different susceptible and resistant specific primers were designed and evaluated (data not shown) using the control/first set of plant material. The list of the allele specific primers which produced best discrimination between resistant and susceptible alleles for the $G_{654}E$ and $S_{653}D$ SNPs are shown in Table 6.1. For all the primer sets of Table 6.1, polymerase chain reaction (PCR) amplifications were performed using the following cycling conditions: 95°C - 2 min, 28 cycles of (95°C - 12 s, 60°C - 12 s, 72°C - 12 s) and 72°C - 5 min. Details for the PCR reaction set up, gel running, and the subsequent scoring procedure can be found in subsection 5.2.3.

Table 6.1 AS-PCR primer design and single nucleotide polymorphism (SNP) details for the *ALS* gene

Locus	Mutation details	Forward Primer		Reverse primer	PCR amplicon size
		Name and Sequence	Modification	Name and Sequence	
$G_{654}E$ SNP in *ALS* gene	'G' allele for the reported G/A mutation	ALS654SusF 5'-CTG CCT ATG ATC CCA AGG GG-3'	Artificial mismatch at third base from 3′ end (T was replaced by G)	ALSR3 5'-TGG GTC ATT CAG GTC AAA CA-3'	131 bp
	'A' allele for the reported G/A mutation	ALS654ResF 5'-CTG CCT ATG ATC CCA AGG GA-3'	Same as above	ALSR3 5'-TGG GTC ATT CAG GTC AAA CA-3'	131 bp
$S_{653}D$ SNP in *ALS* gene	'G' allele for the reported G/A mutation	ALS653SusF 5'-GTG CTG CCT ATG ATC CTA AG-3'	Artificial mismatch at fourth base from 3′ end (C was replaced by T)	ALSR3 5'-TGG GTC ATT CAG GTC AAA CA-3'	134 bp
	'A' allele for the reported G/A mutation	ALS653ResF 5'-GTG CTG CCT ATG ATC CTA AA-3'	Same as above	ALSR3 5'-TGG GTC ATT CAG GTC AAA CA-3'	134 bp

6.2.3 DNA Sequencing and Alignment

For validation of *ALS* gene SNP assay results, PCR products of the *ALS* gene fragments amplified from all control plant lines described in the first paragraph of 6.2.1 subsection were sequenced. For further details about template preparation, sequencing and alignment procedures, refer to subsection 5.2.4.

6.3 Results and Discussions

6.3.1 SNP Genotyping Results for the Control Set of Plants

As a pilot study, the first plant collection (described in the first paragraph of the subsection 6.2.1) were genotyped for the $G_{654}E$ and $S_{653}D$ SNPs. Each individual plant DNA was genotyped four times to identify the four alleles present at the $G_{654}E$ and $S_{653}D$ SNPs. For all lines the SNP genotyping results were consistent with their expected imazethapyr resistance pattern (based on their origin). The Cocodrie, CL 161, Cypress, LaGrue, Nipponbare, red rice tested negative for the $G_{654}E$ SNP, whereas CL 121, CL 141 and CL 121 x red rice hybrids tested positive for the $G_{654}E$ SNP (Figure 6.2). Similarly, Cocodrie, CL 121, CL 141, Cypress, LaGrue, Nipponbare, red rice tested negative for the $S_{653}D$ SNP, while CL 161 and CL 161 x red rice hybrids tested positive for the $S_{653}D$ SNP (Figure 6.3). Heterozygous resistant alleles in the case of CL 121 x red rice hybrids (for the $G_{654}E$ SNP) and CL 161 x red rice hybrids (for the $S_{653}D$ SNP) were faithfully genotyped (Figures 6.2 and 6.3). As a further validation of these allele-specific SNP genotyping results, the *ALS* gene from all these lines were sequenced and aligned to identify the SNP alleles at the $G_{654}E$ and $S_{653}D$ loci. A perfect correlation between the observed allele-specific SNP genotyping results and the sequencing alignment results was obtained. For both SNP loci, the alignment information revealed the presence of the homozygous and heterozygous alleles. An example of sequence alignment for 10 unique lines from the first plant collection is depicted in Figure 6.1.

6.3.2 The *ALS* $G_{654}E$ SNP Assay Results

For the *ALS* $G_{654}E$ SNP assay, five independent F_2 populations (~30 lines each) grown from previously identified natural CL 121 x red rice or CL 141 x red rice hybrids

Figure 6.1 Sequencing alignment results for the $G_{654}E$ and $S_{653}D$ SNP mutations in the rice *ALS* gene (from 1854 - 1910 bp positions) for 10 representative plant samples. The vertical boxes high lighten the alleles observed for the $G_{654}E$ and $S_{653}D$ SNPs.

were screened using the $G_{654}E$ SNP primer set. Figure 6.2 shows the DNA band pattern for the $G_{654}E$ mutation in the *ALS* gene (131 bp PCR amplified fragment) for 13 representative rice lines. Robust identification of heterozygous resistant alleles in CL 121 x red rice progeny lines (lanes 11, 25 and 14, 28) was observed. As expected, all five hybrids exhibited segregation of alleles at $G_{654}E$ SNP locus in their progenies. Of the 145 F_2 progenies screened, 59 rice lines were found to be heterozygous resistant and the remaining 86 were homozygous resistant at the *ALS* $G_{654}E$ SNP locus.

6.3.3 The *ALS* $S_{653}D$ SNP Assay Results

For the *ALS* $S_{653}D$ SNP assay, a field collection of 483 putative CL 161 x red rice hybrids were screened for the presence of the 'A' or 'G' allele indicated by the 134 bp DNA band. The allele specific genotyping results of the $S_{653}D$ SNP in *ALS* gene for 16 representative rice lines are shown in Figure 6.3. Of the 483 rice lines tested, 87 lines were identified as homozygous resistant and 339 lines identified as heterozygous resistant for the $S_{653}D$ SNP locus in the *ALS* gene.

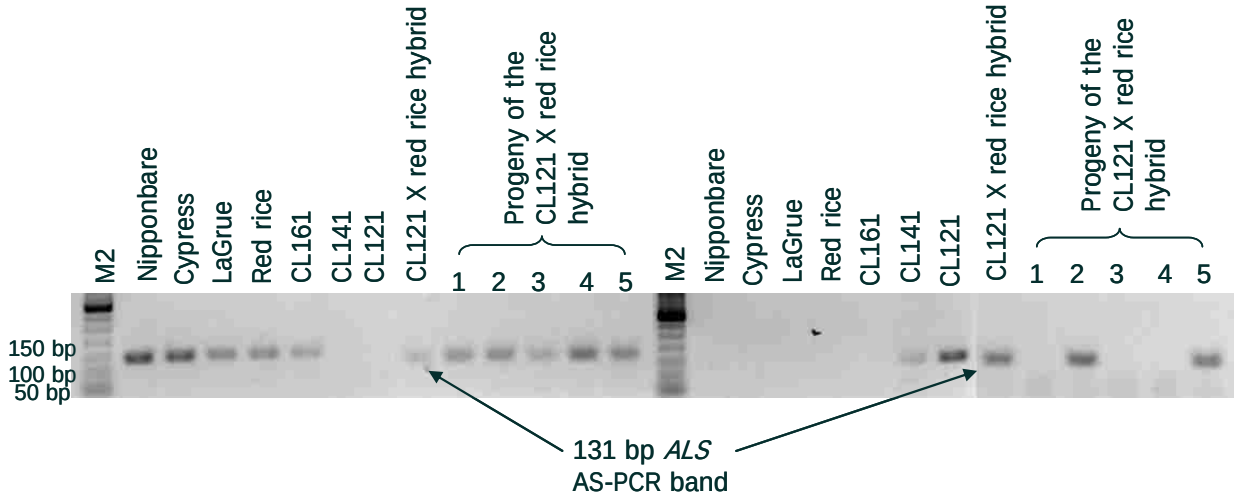


Figure 6.2 The *ALS* gene $G_{654}E$ SNP assay results for 13 representative rice lines on the 2% agarose gel. M2 = NEB 50 bp DNA marker.

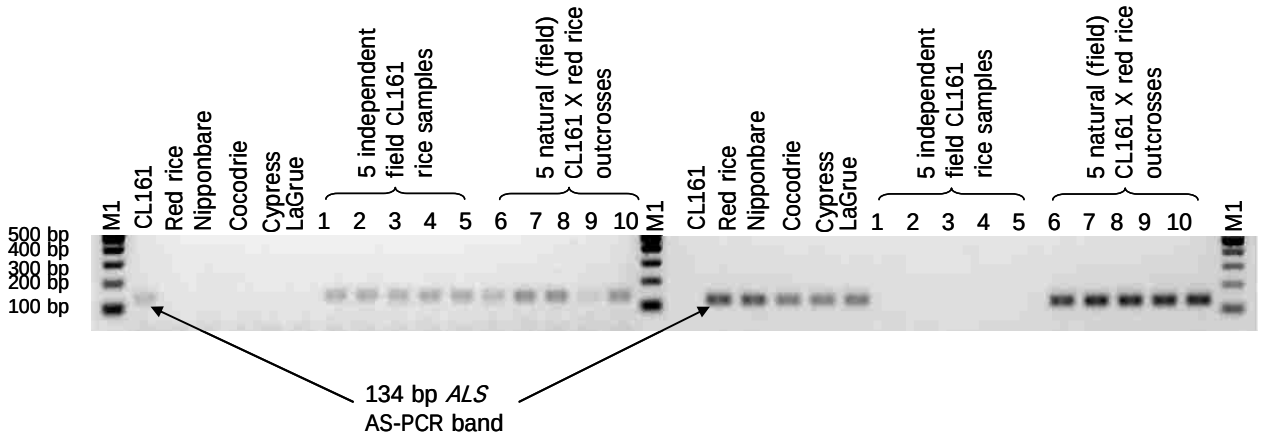


Figure 6.3 *ALS* $S_{653}D$ SNP assay results for 16 representative rice lines on the 2% agarose gel. M1 = NEB 100 bp DNA marker.

6.3.4 Validation of SNP Genotyping Results Using Micro Satellite Markers

To corroborate the *ALS* gene SNP genotyping assays (for both $G_{654}E$ and $S_{653}D$ SNPs), all the heterozygous imazethapyr-resistant samples (putative natural outcrosses) identified in the earlier analysis were tested using three simple sequence repeat (SSR)

markers, namely RM 180, RM 251 and RM 253. These SSR markers were reported to distinguish between the weedy red rice and cultivated rice cultivars (Gealy et al., 2002; Zhang, 2005). As expected, all these heterozygous individuals possessed both the red rice and the cultivated rice specific SSR bands (data not shown). The *ALS* gene in rice is physically located on chromosomes 2, whereas the SSR markers RM251, RM253 and RM180 are located on chromosomes 3, 6 and 7, respectively. Thus, SSR amplification results confirmed presence of recombinant DNA (at loci other than *ALS* gene) in the CL 161 x red rice hybrid material. These results indicate that the new *ALS* gene SNP assays developed in this study are a reliable method for assessment of out-crossing and subsequent transfer of *ALS* gene between Clearfield rice varieties and red rice.

Using a 96 well thermocycler, 400 individual samples can be feasibly genotyped and analyzed by the new *ALS* gene SNP assays in one day by a single operator. Pooling of genomic DNA samples was found to be not effective while performing the *ALS* or imazethapyr-resistant SNP assays. The cost/sample for each allele specific PCR was estimated to be ~ $0.15 (details not shown), and thus these assays are well suitable for laboratories with limited funding such as regional research stations.

6.4 References

Alexander HM, Cummings CL, Kahn L, Snow AA. (2001) Seed size variation and predation of seeds produced by wild and crop-wild sunflowers. Am J Bot 88:623-627.

Arnaud JF, Viard F, Delescluse M, Cuguen J (2003) Evidence for gene flow via seed dispersal from crop to wild relatives in *Beta vulgaris* (Chenopodiaceae): consequences for the release of genetically modified crop species with weedy lineages. Proc Biol Sci 270: 1565-71.

Arriola PE, Ellstrand NC (1996) Crop-to-weed gene flow in the genus Sorghum (Poaceae): spontaneous interspecific hybridization between Johnsongrass, *Sorghum halepense*, and crop sorghum, *S. bicolor*. Am J Bot 83:1153-1160.

Askew SD, Wilcut JW, Walls FR Jr (1999) Weed management in imidazolinone-tolerant and -resistant corn. Proc South Weed Sci Soc 52: 23.

Ball DA, Young FL, Ogg AG Jr (1999) Selective control of jointed goatgrass (*Aegilops cylindrica*) with imazamox in herbicide-resistant wheat. Weed Technol 13:77-82.

Brubaker CL, Koontz JA, Wendel JF (1993) Bidirectional cytoplasmic and nuclear introgression in the New World cottons, *Gossypium barbadense* and *G. hirsutum* (Malvaceae.) Am J Bot 80:1203-1208.

Chen LJ, Lee DS, Song ZP, Suh H, Lu BR (2004) Gene flow from cultivated rice (Oryza sativa) to its weedy and wild relatives. Annals of Botany 93: 67-73.

Cirujeda A, Recasens J, Taberner A (2001) A qualitative quick-test for detection of herbicide resistance to tribenuron-methyl in *Papaver rhoeas*. Weed Res 41: 523-534.

Corbett CL (2004) DNA-based diagnostic tests for the detection of acetolactate synthase-inhibiting herbicide resistance in *Amaranthus* sp. Msc Thesis. University of Guelph.

Corbett CL, Tardif FJ (2006) Detection of resistance to acetolactate synthase inhibitors in weeds with emphasis on DNA-based techniques: a review. Pest Manag Sci 62: 584-597.

Corbett CL, Tardif FL (2006) Detection of resistance to acetolactate synthase inhibitors in *Amaranthus* sp. Using DNA polymorphisms. Pestic Biochem Physiol (in press).

Dilday RH, Nastasi P, Smith RJ, Khodayari K (1990). Herbicide-tolerant germplasm in rice. In J. Janick, J.E. Simon (ital: eds.), Advances in new crops. Timber Press, Portland. 146-150.

Dillon TL, Talbert RE, Baldwin FL (2001) A three year overview of weed control in Clearfield rice. Proc South Weed Sci Soc 54: 43.

Ellstrand NC (2003) Current knowledge of gene flow in plants: implications for transgene flow. Philos Trans R Soc Lond B Biol Sci 358:1163-70.

Estorminos LE Jr, Gealy DR, Dillon TL, Baldwin FL, Burgos RR, Tai TH (2002) Determination of hybridization between rice and red rice using four microsatellite markers. Proc South Weed Sci Soc 55: 197-198.

Foes MJ, Liu LX, Vigue G, Stoller EW, Wax LM, Tranel PJ (1999) A kochia (*Kochia scoparia*) biotype resistant to triazine and ALS-inhibiting herbicides. Weed Sci 47: 20-27.

Gealy DR, Mitten DH, Rutger JN (2003) Gene flow between red rice (*Oryza sativa*) and herbicide resistant rice (*O. sativa*): Implications for weed management. Weed Tech 17: 627-645.

Gealy DR, Tai TH, Sneller CH (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Sci 50:333-339.

Guadagnuolo R, Clegg J, Ellstrand NC (2006) Relative fitness of transgenic vs. non-transgenic maize x teosinte hybrids: a field evaluation. Ecol Appl 16:1967-74.

Guadagnuolo R, Savova-Bianchi D, Felber F (2001). Gene flow from wheat (*Triticum aestivum* L.) to jointed goatgrass (*Aegilops cylindrical* Host), as revealed by RAPD and microsatellite markers. Theor Appl Genet 103: 1-8.

Guadagnuolo, R, Savova-Bianchi D, Keller-Senften J, Felber F (2001) Search for evidence of introgression of wheat (*Triticum aestivum* L.) traits into sea barley (*Hordeum murinum* s.str.Huds.) and bearded wheatgrass (*Elymus caninus* L.) in central and northern Europe, using isozymes, RAPD and microsatellite markers. Theor Appl Genet 103: 191-196.

Guttieri MJ, Eberlein CV, Mallorysmith CA, Thill DC, Hoffman DL (1992) DNA-sequence variation in domain a of the acetolactate synthase genes of herbicide-resistant and herbicide-susceptible weed biotypes. Weed Sci 40: 670-676.

Hall LM, Stromme KM, Horsman GP, Devine MD (1998) Resistance to acetolactate synthase inhibitors and quinclorac in a biotype of false cleavers (*Galium spurium*). Weed Sci 46: 390-396.

Hayashi K, Hashimoto N, Daigen M, Ashikawa I. (2004) Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. Theor Appl Genet 108: 1212-20.

Heap I (2007) The International Survey of Herbicide Resistant Weeds. Available: www.weedscience.com.

Hinz JRR, Owen MDK (1997) Acetolactate synthase resistance in a common waterhemp (*Amaranthus rudis*) population. Weed Tech 11: 13-18.

Jorgensen RB, Andersen B (1994) Spontaneous hybridization between oilseed rape (*Brassica napus*) and weedy *B. campestris* (Brassicaceae): a risk of growing genetically modified oilseed rape. American Journal of Botany 81: 1620-1626.

Krausz RF, Kapusta G (1998) Total postemergence weed control in imidazolinone-resistant corn (*Zea mays*). Weed Technol 12: 151-156.

Kwon CS, Penner D (1995) Response of a chlorsulfuron-resistant biotype of kochia-scoparia to ALS inhibiting herbicides and piperonyl butoxide. Weed Sci 43: 561-565.

Lovell ST, Wax LM, Simpson DM, McGlamery M (1996) Using the in vivo acetolactate synthase (ALS) assay for identifying herbicide-resistant weeds. Weed Tech 10: 936-942.

Madsen KH, Valverde BE, Jensen JE (2002) Risk assessment of herbicide-resistant crops: A Latin American perspective using rice (*Oryza sativa*) as a model. Weed Tech 16: 215-223.

McNaughton KE, Letarte J, Lee EA, Tardif FJ (2005) Mutations in ALS confer herbicide resistance in redroot pigweed (*Amaranthus retroflexus*) and Powell amaranth (*Amaranthus powellii*). Weed Sci 53: 17-22.

Mercer KL, Wyse DL, Shaw RG (2006) Effects of competition on the fitness of wild and crop-wild hybrid sunflower from a diversity of wild populations and crop lines. Evolution Int J Org Evolution 60:2044-55.

Messeguer J, Marfa V, Catala MM, Guiderdoni E, Mele E (2004) A field study of pollen-mediated gene flow from Mediterranean GM rice to conventional rice and the red rice weed. Mol Breeding 13: 103-112.

Noldin JA, Chandler JM, McCauley GN (1999). Red rice (*Oryza sativa*) Biology. I. Characterization of red rice ecotypes. Weed Technology 13: 12-18.

Oard J, Cohn MA, Linscombe S, Gealy DR, Gravios K (2000). Field evaluation of seed production, shattering, and dormancy in hybrid population of transgenic rice (*Oryza sativa*) and the weed, red rice (*Oryza sativa*). Plant science 157: 13-22.

Patzoldt WL, Tranel PJ (2002) Molecular analysis of cloransulam resistance in a population of giant ragweed. Weed Sci 50: 299-305.

Patzoldt WL, Tranel PJ (2003) Imidazolinone-specific resistance in twp Illinois tall waterhump (*Amaranthus tuberculatus*) biotypes. Proc Weed Sci Soc Am 43: 90.

Perez-Jones A, Mallory-Smith CA, Hansen JL, Zemetra RS (2006) Introgression of imidazolinone-resistance gene from winter wheat (*Triticum aestivum* L) into jointed goatgrass (*Aegilops cylindrica* Host). Theor Appl Genet 114: 177-186.

Poppy GM (2004) Geneflow from GM plant--towards a more quantitative risk assessment. Trends Biotechnol 22: 436-438.

Quist D, Chapela IH (2001) Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. Nature 414: 541-543.

Renno, J.F., T. Winkel, F. Bonnefous, G. Bezancon, 1997. Experimental study of gene flow between wild and cultivated Pennisetum glaucum. Can J Bot 75: 925-931.

Richter J, Powles SB (1993) Pollen expression of herbicide target site resistance genes in annual ryegrass (*Lolium rigidum*). Plant Physiol 102: 1037-1041.

Rieger MA, Lamond M, Preston C, Powels SB, Roush RT (2002) Pollen-mediated

movement of herbicide resistance between commercial canola crops. Science 296:2386-2388.

Rong J, Xia H, Zhu YY, Wang YY, Lu BR (2004) Asymmetric gene flow between traditional and hybrid rice varieties (*Oryza sativa*) indicated by nuclear simple sequence repeats and implications for germplasm conservation. New Phytologist 163: 439-445.

Siminszky B, Coleman NP, Naveed M (2005) Denaturing high-performance liquid chromatography efficiently detects mutations of the acetolactate synthase gene. Weed Sci 53: 146-152.

Simpson DM, Stoller EW, Wax LM (1995) An *in-vivo* acetolactate synthase assay. Weed Tech 9: 17-22.

Song ZP, Lu BR, Zhu YW, Chen JK (2003). Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157: 657-665.

Steele GL, Chandler JM, McCauley GN (2002) Control of red rice (*Oryza sativa*) in imidazolinone-tolerant rice (*O sativa*). Weed Tech 16: 627-630.

Tan MK, Medd RW (2002) Characterisation of the acetolactate synthase (ALS) gene of *Raphanus raphanistrum* L. and the molecular assay of mutations associated with herbicide resistance. Plant Sci 163: 195-205.

Tan S, Evans R, Singh B (2006) Herbicidal inhibitors of amino acid biosynthesis and herbicide tolerant crops. Amino Acids 30: 195-204.

Tan S, Evans RR, Dahmer ML, Singh BK, Shaner DL. (2005) Imidazolinone-tolerant crops: history, current status and future. Pest Manag Sci 61: 246-57.

Tranel PJ, Wright TR. (2002) Resistance of weeds to ALS-inhibiting herbicides: what have we learned. Weed Sci 50: 700-712.

Wagner J, Haas HU, Hurle K (2002) Identification of ALS inhibitor-resistant Amaranthus biotypes using polymerase chain reaction amplification of specific alleles. Weed Res 42: 280-286.

Walsh MJ, Duane RD, Powles SB (2001) High frequency of chlorsulfuron-resistant wild radish (*Raphanus raphanistrum*) populations across the Western Australian wheatbelt. Weed Tech 15: 199-203.

Wang F, Yuan QH, Shi L, Qian Q, Liu WG, Kuang BG, Zeng DL, Liao YL, Cao B, Jia SR. (2006) A large-scale field study of transgene flow from cultivated rice (*Oryza sativa*) to common wild rice (*O. rufipogon*) and barnyard grass (*Echinochloa crusgalli*). Plant Biotechnol J 4: 667-76.

Zhang BH, Pan XP, Guo TL, Wang QL, Anderson TA (2005) Measuring gene flow in the cultivation of transgenic cotton (*Gossypium hirsutum* L). Mol Biotechnol 31: 11-20.

Zhang NY, Linscombe S, Oard J (2003) Out-crossing frequency and genetic analysis of hybrids between transgenic glufosinate herbicide-resistant rice and the weed, red rice. Euphytica 130: 35-45.

Zhang W (2005) Risk assessment of the transfer of the imazethapyr herbicide resistance from Clearfield rice to red rice. Ph.D. Dissertation. Louisiana State University and A&M College, Baton Rouge, LA.

Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of allele-specific PCR markers for γ-gliadin alleles in *Triticum aestivum*. Theor Appl Genet 107: 130-138.

CHAPTER 7 SUMMARY AND CONCLUSIONS

7.1 Discriminant Analysis

This study was carried out to evaluate the potential of Discriminate Analysis (DA) procedure to detect informative molecular markers associated with percent amylose content and five agronomic traits among 192 inbred rice lines. The DA procedure identified marker sets for the complex traits even with narrow germplasm breeding material evaluated across five U.S. states. While some markers were common among two or more states, the majority of DA-selected alleles were location-specific, indicating strong GxE effects.

The DA procedure successfully identified new markers RM25, RM225, RM231, along with the known RM190 (*Waxy*) locus for percent amylose content. These DA-selected markers were successfully validated in the second set of 57 US and Asian rice lines, suggesting feasible application of simple inherited trait DA results across different plant populations. Thus, DA identified three new loci associated with amylose content that may be useful for marker-assisted selection in a diverse representation of rice accessions.

The DA-selected markers overall produced high levels of leave-one-out percent correct classification in the training samples and individual R^2 values, indicating potential value of this approach in identification of informative alleles for marker-assisted selection. In addition, the DA-selected markers were identified for percent amylose content, percent head rice, percent total rice, and grain yield that mapped on the rice genetic map within or near traditional Quantitative Trait Loci (QTL). Results from this study suggest that DA can successfully complement traditional methods to identify

markers associated with complex and economically important traits in rice. Dr. Don Labonte, LSU AgCenter, currently uses the DA method in his breeding program for marker assisted research of visus resistance and other traits.

7.2 Mixed Model-Regression Approach

I have created and evaluated a mixed model-regression procedure that identifies main and epistatic effects by standard hypothesis testing and Bayesian information criteria in a multivariate format for agronomic traits evaluated in a collection of elite Louisiana breeding lines. Validation of the procedure in a separate test samples indcates that additional research using the mixed model-regression approach is warranted.

7.3 Alternative Ecotilling

I have successfully developed a simple, rapid, efficient, and cost-effective alternative to standard Ecotilling for SNP discovery and genotyping in rice that can be easily adapted to small or medium-sized laboratories. Results from analysis of the *alk* and *waxy* loci demonstrate that modified Ecotilling is a reproducible, simple, rapid, and cost-effective approach for SNP discovery and genotyping in rice when compared to the standard Ecotilling method. Specifically, the Alternative Ecotilling SNP detection and analysis is consistent with standard Ecotilling and sequencing results over separate experiments representing a diverse geographical collection of accessions from Asia, Africa, and the Americas. The alternative Ecotilling protocol successfully identified expected associations between SNP variation and trait measurements. This feature was verified by the strong association between the *waxy* T/G SNP and amylose class and where all accessions having the *alk* GC/TT mutation displayed low gelatinization temperature. Preliminary data also indicates that the G/A SNP in the *waxy* gene intron

may be associated with increased rice starch pasting viscosity measurements (R.G. Fjellstrom, data not shown). In addition, a unique A/G SNP was detected in the *alk* gene of Nipponbare among the accessions that was verified by sequencing, and a G/A SNP at the *waxy* locus, unknown to us at the onset of this study. Alternative Ecotilling was equally effective in SNP detection from either an individual or a pooled sample, indicating that rapid genotyping of large populations is possible.

This study shows that modified Ecotilling does not require investment in expensive laboratory equipment or costly reagents such as florescent compounds for primer labeling. In contrast to standard Ecotilling, data collection, storage, and analysis by the alternative Ecotilling procedure do not involve extensive training or use of complex software programs written in different languages. Up to 400 individual or pooled samples can be realistically processed and analyzed in one day by a single operator. This means that 3200 individuals in an 8-fold pool format could be screened in a single day for SNP variation. Therefore, SNP analysis using alternative Ecotilling should be very suitable for laboratories with limited funding for various targeted research objectives in rice genomics, breeding, and evolutionary studies.

7.4 Haplotype Genotyping of the Aromatic Rice

A simple, rapid, and precise haplotype-specific assay for a targeted region of the *BAD2* gene that unambiguously distinguishes homozygous and heterozygous genotypes associated with seed aroma in rice was developed. Sufficient details were provided regarding the design and practical application of haplotype-specific primers for marker-assisted identification and introgression of the aroma gene from *tropical japonica* and *indica* sources. The present marker-based approach should also prove useful for

combining aroma as a recessive trait with other characteristics such as disease resistance in three-way forward crosses.

While conducting this research, Bradbury et al. (2005b) published a study that described a four-primer system to identify aromatic and non-aromatic rice. This system was also tested and found to correctly genotype the U.S. germplasm (data not shown). Due to size and range of multiple PCR products, the system developed by Bradbury et al. (2005b) would not be amenable for high throughput genotyping of rice for aroma. The method reported in this study, however, will remain useful for high throughput using a PAGE format.

7.5 Marker Development for Outcrossing among Clearfield Rice and Red Rice

I have successfully developed simple, rapid, relatively high-throughput and precise allele-specific SNP genotyping techniques for the *ALS* gene in rice which require use of only standard PCR and electrophoresis instruments. For the first time, a DNA based Imazethapyr resistance assay that could differentiate between the resistant Clearfield rice and susceptible red rice was demonstrated. Inaddition, this assay could effectively discriminate between the homozygous resistant and heterozygous resistant $S_{653}D$ and $G_{654}E$ SNP alleles. Using the new *ALS* gene SNP assays, a total of 483 field-collections were successfully screened for the presence of $S_{653}D$ SNP and another 145 F_2 progeny lines of natural CL 121 x red rice crosses were screened for the presence of the $G_{654}E$ SNP. Natural outcrosses between CL 121 x red rice and CL 161 x red rice were successfully identified in these palnt collections. In addition, amplification results of SSR markers that can differentiate red rice and cultivated rice, have confirmed presence of recombinant DNA (at loci other than *ALS* gene) in the CL 161 x red rice hybrid material.

Thus, this AS-PCR SNP genotyping results can be immediately and effectively applied for accurate assessment of out-crossing using Clearfield^{TM} rice technology and thus development of better weed management practices.

APPENDIX: PERMISSION LETTER

From: "Dr. Don P. Bourque" <dbourque@u.arizona.edu>
To: Suresh Kadaru <skadar1@lsu.edu>
CC:
Subject: Re: Permission to use my published work in dissertation
Date: Tuesday, March 13, 2007 3:14:59 PM

Dear Dr. Kadaru:

As representative of the ISPMB, and for the PMBR, I hereby grant you permission to use your published paper (cited below) as part of dissertation work

SB Kadaru, AS Yadav, RG Fjellstrom and JH Oard. 2006. Alternative Ecotilling protocol for rapid, cost-effective SNP discovery and genotyping in rice (*Oryza sativa* L.). ***Plant Molecular Biology Reporter*** 24:1-20.

Congratulations on completing your graduate work and best of luck for the future.

Sincerely,
Don P. Bourque

Suresh Kadaru wrote:

Dear Dr. Bourque,

I request you to kindly grant the permission to use the below cited publication as part of dissertation work.

SB Kadaru, AS Yadav, RG Fjellstrom and JH Oard. 2006. Alternative Ecotilling protocol for rapid, cost-effective SNP discovery and genotyping in rice (*Oryza sativa* L.). ***Plant Molecular Biology Reporter*** 24:1-20.

Regards,

Suresh B Kadaru
Graduate Student
Rice Genetics Lab
School of Plant, Environmental and Soil Sciences
104 Madison B. Sturgis Hall

Louisiana State University
Baton Rouge, LA 70803
Phone: Office:(225) 578 7864
Cell:(225) 202 0409

--
Don P. Bourque, Professor
Departments of Biochemistry and Molecular Biophysics
& Molecular and Cellular Biology
Editor, Plant Molecular Biology Reporter
536 Biosciences West Bldg.
1041 E. Lowell Street
Tucson AZ 85721-0088
(520) 621-7529
Fax (520) 621-1697
e-mail:dbourque@u.arizona.edu

VITA

Suresh Babu Kadaru was born in 1974 in Nidubrolu town of Andhra Pradesh, India. He had pursued his high school education in Hyderabad, Andhra Pradesh, India. He finished his bachelor's degree in Acharya N G Ranga Agricultural University, Rajendranagar Campus. After a short period of work experience, he joined Tamil Nadu Agricultural University, Coimbatore, India, and completed his master's degree in biotechnology. Before joining LSU, he worked for two years in National Research Center for Plant Biotechnology, Indian Agricultural Research Institute, India.