2012

Application of genetic and statistical tools for improvement of Louisiana rice

James Silva Garcia
Louisiana State University and Agricultural and Mechanical College

APPLICATION OF GENETIC AND STATISTICAL TOOLS FOR IMPROVEMENT OF LOUISIANA 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

The School of Plant, Environmental, and Soil Sciences

by
James Silva-Garcia
B.S. Statistics, Universidad del Valle, Colombia, 1994
Master of Applied Statistics, Louisiana State University, 2011
August 2012

To my loving wife ***Millis***, my beautiful daughters

Viviana, Isabella, and Ashley

To my parents, family, and friends

This dissertation is humbly dedicated….

ACKNOWLEDGEMENTS

First of all, to **God Almighty** for showing me his love and for proving that everything has its own appropriate time. Thank **God** for supporting me and my family for ever.

To Dr. James H. Oard for recruiting me, for his mentoring and guidance, and for teaching me how important dedication, discipline, and hard work are. On behalf of my whole family I express my gratitude to him.

To the members of my Graduate Committee, Dr. Stephen A. Harrison, Dr. Gerald O. Myers, Dr. Bin Li, and Dean's Representative, Dr. John C. Larkin. I thank them for their commitment, guidance, mentoring, and patience.

For allowing me to use the SSR marker data set I want to thank USDA, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas. I also want to express my gratitude to Dr. Donald E. Groth (LSU AgCenter) for sharing his sheath blight data. Similarly, I thank the RiceCAP project (USDA/CSREES grant 2004–35317–14867) for generation of sequence data, generation of the SB5 mapping population, and associated QTL data. I recognize the Louisiana Rice Research Board and The Mississippi Rice Promotion Board for partial funding of this research. I acknowledge Dr. Georgia Eizenga (USDA-ARS Dale Bumpers National Rice Research Center) for her generous gift of DNA isolated from two *O. nivara* accessions.

I thank Professor Luis Escobar for his tutoring and trustful friendship; to his wife, his daughters and son, and our friend Margarita for their unconditional willingness to help my loved ones. I thank Myriam C. Duque and Carlos A. Moreno for their friendship, coaching, training, and support.

I thank the School of Plant, Environmental, and Soil Sciences, to the Department of Experimental Statistics, and to all fellow students for their support and for sharing their knowledge and skills. To my lab mates Yamid, Dominique, Christian, and my former lab mate Samuel for their friendship and collaboration.

To my friend Stephanie Gil for becoming one of the closest friends my family and I ever had; for her being always ready to help whenever we were facing difficult times, and for the joyful and delightful moments we enjoyed with Steph and the family Gil.

Thanks to my uncle Fabian Garcia and his wife, to my aunt Luz M. Garcia, her husband Raul Ordonez, and my cosines Nayibe, Raul, and Karen for their help and the wonderful Christmas and vacations we shared together in Miami.

To my brother John A. Ramos for his love, support, and encouragement to keep working hard to quickly attain our goals.

I thank families Silva, Garcia, Muriel, Meneses, Lopez, and Arismendi for their believing and encouragement. I thank our friends Mauricio Martinez and Nelson Murillo for their constant support and love.

I will be forever grateful to my parents, brothers, and sisters for their love and for always expecting and wishing the best for me and my family. Overcoming all difficulties of this Ph.D. program and reaching the goal line would not had been possible without the love, continuous support, and patience of my loved wife Millis D. Arismendi. Finally, I want to thank my beautiful daughters Viviana, Isabella, and Ashley Sophia for their unconditional love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Breeding for grain quality traits and resistance to sheath blight (SB), a disease caused by *Rhizoctonia solani* Kuhn, are important objectives for the rice (*Oryza sativa* L.) industry. Grain quality traits and SB resistance play an important role in the economic prosperity of commercial rice markets. The objectives of our research were to: (1) Explore performance and stability for SB resistance among doubled-haploid (DH) lines of the SB2 mapping population using GGE biplots (2) Exploit whole genome sequences of 13 inbred lines to identify non-synonymous SNPs (nsSNPs) and candidate genes for SB resistance.

Genotype-by-environment interaction for SB analysis was performed using heritability-adjusted GGE (HA-GGE) biplot. DH lines were evaluated for two years in Louisiana and Arkansas; a single "mega-environment" was identified consisting of the four year-location combinations. HA-GGE biplot analyses identified 11 high and stable DH lines; five susceptible DH lines were also identified with greater stability than the susceptible parent used to develop the SB2 population. Material identified in this study represents a potential source of SB resistance for cultivar development.

Two filtering strategies were developed to identify nsSNPs between two groups of known resistant and susceptible lines. More than 200 genes with selected nsSNPs were assigned to 42 categories based on family/gene ontology. Individual alleles of 24 nsSNPs were evaluated by PCR whose presence/absence corresponded to known resistant/susceptible phenotypes of nine additional lines. "Resistant" alleles were detected in two accessions of *O. nivara* that suggests sources for resistance occur in additional *Oryza* sp. Results from this study provide a foundation for future marker-assisted breeding of rice for SB resistance.

CHAPTER 1 GENERAL INTRODUCTION

To ensure global food security, agricultural development is facing major challenges including the need to produce high yielding crops adapted to climatic changes and the identification of feedstock crops for biofuel production. These are challenges that encourage new approaches to plant breeding and functional genomics (Furbank and Tester, 2011). Currently, food sources for human population is relying primarily on 15 to 20 species (Chrispeels et al., 2003; Balick et al., 1997), including rice (*Oryza sativa* L.) as an economically important crop accounting for about 20% of the world population's caloric intake (Huang et al., 2012). For crops like rice, maize (*Zea mays*), and wheat (*Triticum aestivum*), annual increases in yield by breeding programs around the world are unable to meet projected demands (Furbank et al., 2009; Reynolds et al., 2009; Tester and Langridge, 2010), implying a required increase of at least 70% in cereal grain yields before 2050 (Furbank and Tester, 2011).

Breeding for high yielding varieties is not the only objective for rice breeders and producers, because rice grain quality traits (appearance, eating, cooking, and milling) command worldwide attention and play a crucial economic role as reported by Ordonez et al. in 2010 (refer to Appendix A for details). Factors as translucency of the endosperm and grain shape significantly impact the quality of appearance of rice grains (Juliano and Villareal, 1993; Unnevehr et al., 1992). Eating and cooking quality is determined mainly by apparent amylose content, a trait governed primarily by the *Waxy* (Wx) locus on chromosome 6 (Hao et al., 2009; Kepiro et al., 2008; Aluko et al., 2004; Septiningsih et al., 2003; Tan et al. 1999) and additional QTL of minor effect at various chromosomal locations (Aluko et al., 2004; Wan et al., 2004; Tan et al., 2001). Separate QTL studies by Fan et al. (2005) and Wan et al. (2004) showed that environment was a major source of variation for amylose content while epistasis played a minor role.

Whole grain or head rice, defined as the proportion of whole kernels that also includes broken kernels 75% to 80% of the whole rice grain, is a well known important component for establishing market value and the most important characteristic of overall milling quality. Due to the time-consuming effort of evaluating lines with multi-step procedures in replicated field plots and laboratory analyses, most of the reported QTL mapping studies were carried out at a single location in a single year and were able to identify numerous QTL with small effects across different chromosomal regions, but some investigations also detected QTL with major effects when evaluated at individual locations (Aluko et al., 2004; Kepiro et al., 2008; Lou et al., 2009). Four studies identified two-way QTL interactions on six chromosomes (Tan et al 2001; Septiningsih et al. 2003; Aluko et al. 2004; Lou et al., 2009). In a multi-environment trial by Lou et al. (2009), genotype-by-environment (GE) interaction was not significant for head rice and two-way interactions (epistasis) produced only a minor effect.

The time of flowering, also known as heading date, is considered a crucial factor in production of high quality rice grain (Fan et al., 2005; Tabien et al., 2009), and more than 100 QTLs associated with heading date have been identified (www.gramene.org). Certain QTLs (e.g., Hd1 located at the top of chromosome 6 reported by Yano et al., 2000) were recently shown to be directly involved in complex interactions for heading date and/or photoperiodic responses. A recent study suggested that expression of Hd3a, a major QTL on chromosome 6 regulated by Hd1, was also impacted by variation in temperature and day-length (Luan et al., 2009).

Advanced technologies has led to an increase in the number of markers at lower costs per data point (Eathington et al., 2007), which translates into a higher complexity of the statistical methods to analyze data for marker-assisted breeding programs. As an extension of quantitative

genetics models, the statistical basis for association genetic studies of complex traits in plants has been the general linear model (GLM) that assumes continuous response variables linearly associated with one or more fixed categorical variables such as DNA marker alleles.

The GLMSELECT procedure was released by the SAS Institute Inc. in 2008 as a tool to perform selection of effects in general linear models with capabilities to customize selection and stopping criteria from traditional and computationally efficient significant-level-based criteria to more computationally intensive validation-based criteria. This procedure was the main analytical tool used in Appendix A to identify candidate marker effects associated with two grain quality and one flowering trait by association mapping in a collection of elite tropical *japonica* lines evaluated at five different locations.

Similar to genotyping, phenotyping populations is a labor-intensive and costly component of the challenge of assembling the necessary genetic resources for the success of a breeding program because it needs to be done precisely through replicated trials across multiple environments and over a number of seasons (Furbank and Tester, 2011). One major challenge in the development of Sheath Blight (SB, a fungal disease caused by the pathogen *Rhizoctonia solani* Kuhn) resistant commercial rice has been the low repeatability of SB scores in field-plot and greenhouse evaluations due to variation in environmental conditions across years and locations, which translates into a highly significant contribution of the genotype-by-environment (GE) interaction effect (Oard and Groth, unpublished data, 2010).

The SB2 mapping population of 322 doubled-haploid (DH) lines (Chu et al., 2006) developed during the RiceCAP project (www.uark.edu/ua/ricecap) was derived from a cross between the resistant parent MCR10277 (Rush et al., 2006) and the susceptible parent Cocodrie

(Linscombe et al., 2000). To select stable genotypes with high levels of resistance, exploring the potential of different approaches to identify candidate rice lines with high and stable levels of SB resistance is required. Since its invention in 1971 by Gabriel, genotype plus GE interaction (GGE) biplots have been used to interpret GE effects that impact performance and stability of agronomic traits. Although GGE biplots have been used primarily for yield data, this methodology has been also useful for analyzing disease resistance data. Examples include identification of wheat (*Triticum aestivum* L.) lines with stable resistance to the powdery mildew pathogen *Blumeria graminis* f. sp. *tritici* (Lillemo et al., 2010). Similarly, biplots were used to select elite wheat lines resistant to Fusariam head blight caused by *Fusarium graminearum* (Kadariya et al., 2008) and to identify barley (*Hordeum vulgare L.*) lines showing resistance to net blotch caused by *Drechslera teres f. sp. maculata* (Yan and Falk, 2002).

Analysis and interpretation of GE interactions can also be accomplished by other analytical methods like the "median polish" method (Tukey, 1977), Hühn's nonparametric methods (Nassar and Hühn, 1987; Truberg and Hühn, 2000), and the additive main effects and multiplicative interaction (AMMI) method (Gauch, 1992). Median polish was recently implemented to identify stable resistance for two important diseases in wheat (Arraiano and Brown, 2006; Lillemo et al., 2010) and to identify sunflower (*Helianthus annuus* L.) lines resistant to *Phoma macdonaldii* Boerema (Darvishzadeh et al., 2007). Hühn's nonparametric methods have been exploited primarily for stability analyses of yield and associated traits in various crops (Hassanpanah and Chakherchaman, 2010; Sabaghnia et al., 2006; Scapim et al., 2000; Lillemo et al., 2010). Similar to GGE biplot, the AMMI method uses two-dimensional graphical displays to evaluate GE; AMMI has been used to evaluate host-pathogen interactions of rice-*Xanthomonas oryzae pv. oryzae* (Nayak et al., 2008), of tulip (*Tulipa* sp.)-*R. solani*

interactions (Schneider and Van den Boogert, 1999), and stability of faba beans (*Vicia faba* L.) for resistance to *Orobanche* sp. (Flores et al., 1996). A modified GGE biplot approach proposed by Yan and Holland (2010) was used in Chapter 2 to explore performance and stability for SB resistance among double-haploid lines of the SB2 rice population developed as part of the RiceCAP project.

An important contribution of the RiceCAP project was completion of the whole genome sequencing of 13 rice lines (including *japonica* and *indica* germplasm) that were selected to represent elite breeding material that is used in modern varietal development in the U.S. and Asia. Genomic DNA was isolated from each of the 13 lines and sent to the National Center for Genome Resources (NCGR) where the Illumina GA IIx platform was used to perform WGS, and SNP calling. In spite of several research efforts that have been reported (Channamal-likarjuna et al., 2010; Kim et al., 2003; Liu et al., 2004; Liu et al., 2009; Maruthasalam et al., 2007; Pinson et al., 2005; Prasad and Eizenga, 2008; Shah et al., 2009; Tan et al., 2005; Venu et al., 2007; Wang et al., 2010; Zhao et al., 2008; Zuo et al., 2008), the routine use of marker-assisted selection to enhance SB resistance in commercial rice cultivars has not been reported.

The advent of next-generation sequencing has been proposed as a rapid, cost effective alternative to Sanger sequencing for identification of candidate genes and variants underlying simple and even complex traits (Hobert, 2010; Teer and Mullikin, 2010). Whole genome sequencing (WGS) of one or a few individuals has recently identified single or multiple variants associated with different Mendelian disorders in humans (Rios et al., 2010; Roach et al., 2010; Sobreira et al., 2010; Tong et al., 2010; Lupski et al., 2010). Similar progress has been made with whole-exome sequencing to uncover rare or recessive variants in humans causing different diseases or adaptations to different environments (Bilguvar et al., 2010; Krawitz et al., 2010; Ng

et al., 2010a, b; Walsh et al., 2010; Yi et al., 2010). Xie et al. (2010) recently used WGS of recombinant inbred lines of rice at low coverage to construct a linkage map of about 209 K SNPs that successfully identified a known QTL associated with grain width. A similar WGS strategy for chromosome segment substitution lines allowed identification of a QTL containing the *sd1* locus for plant height (Xu et al., 2010). A genomic DNA library enriched for genic sequences in rice was recently constructed followed by deep sequencing that revealed approximately 2,600 SNPs between an *indica* and a *tropical japonica* line (Deschamps et al., 2010).

As previously stated, in addition to QTL mapping for SB resistance by Liu et al. (2009), RiceCAP completed WGS of 13 rice lines in cooperation with NCGR (Scheffler et al., unpublished data); sample variant reports provided by NCGR were used in Chapter 3 to develop two strategies, consisting of a consecutive series of filtering steps, to identify candidate genes for SB resistance. Because non-synonymous SNPs (nsSNPs) have been reported to play a role in the function and evolution of plant resistance (Fu et al., 2010; Ling et al., 2009; McNally et al., 2009; Song et al., 1995) that may complement microarray or other gene expression studies, identification of this type of genomic variant was the main goal.

1.1 Research Objectives

(1) To explore performance of and stability for SB resistance among DH lines of the SB2 population using GGE biplot and other methods.

(2) To use sequence data of 13 lines to identify nsSNPs and corresponding candidate genes for SB resistance.

1.2 References

Aluko G., Martinez C., Tohme J., Castano C., Bergman C., Oard J.H. (2004) "QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* x *O. glaberrima*." *Theoretical and Applied Genetics* **109**, 630–639.

Arraiano L.S., Brown J.K.L. (2006) "Identification of isolate-specific and partial resistance to *Septoria tritici* blotch in 238 European wheat cultivars and breeding lines." *Plant Pathology* **55**, 726–738**.**

Balick M.G. (1997) *Plants, People, and Culture: the Science of Ethnobotany*, vol. 9. Scientific American Library. New York, NY, USA. 2nd Edition.

Bilgüvar K., Oztürk A.K., Louvi A., Kwan K.Y., Choi M., Tatli B., Yalnizoğlu D., Tüysüz B., Cağlayan A.O., Gökben S., Kaymakçalan H., Barak T., Bakircioğlu M., Yasuno K., Ho W., Sanders S., Zhu Y., Yilmaz S., Dinçer A., Johnson M.H., Bronen R.A., Koçer N., Per H., Mane S., Pamir M.N., Yalçinkaya C., Kumandaş S., Topçu M., Ozmen M., Sestan N., Lifton R.P., State M.W., Günel M. (2010) "Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations." *Nature* **467**, 207–210.

Channamallikarjuna V., Sonah H., Prasad M., Rao G.J.N., Chand S., Upreti H.C., Singh N.K., Sharma T.R. (2010) "Identification of major quantitative trait loci qSBR11–1 for sheath blight resistance in rice." *Molecular Breeding* **25**, 155–166.

Chrispeels M.J., Sadava D.E. (2003) *Plants, Genes, and Crop Biotechnology*, vol. 24. Jones and Bartlett. Boston, MA, USA. 2nd Edition.

Chu Q.R., Linscombe S.D., Rush M.C., Groth D.E., Oard J., Sha X., Utomo H.S. (2006) "Registration of a C/M Doubled Haploid Mapping Population of Rice." *Crop Science* **46**, 1416.

Darvishzadeh R., Dechamp-Guillaume G., Hewezi T., Sarrafi A. (2007) "Genotype-isolate interaction for resistance to black stem in sunflower (*Helianthus annuus*)." *Plant Pathology* **56**, 654–660.

Deschamps S., La Rota M., Ratashak J.P., Biddle P., Thureen D.A., Luck S., Beatty M., Nagasawam N., Michael L., Llaca V., Sakai H., May G., Lightner J., Campbell M.A. (2010) "Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina Genome Analyzer." *The Plant Genome* **3**, 53–68.

Eathington S.R., Crosbie T.M., Edwards M.D., Reiter R.D., Bull J.K. (2007) "Molecular markers in a commercial breeding program." *Crop Science* **47**, S154–S163**.**

Fan C.C., Yu X.Q., Xing Y.Z., Xu C.G., Luo L.J., 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 population. *Theoretical and Applied Genetics* **110**, 1445–1452.

Flores F., Moreno M.T., Martinez A., Cubero J.I. (1996) "Genotype-environment interaction in faba bean: comparison of AMMI and principal coordinate model." *Field Crops Research* **47**, 117–127.

Fu Y.B., Peterson G.W., McCallum B.D., Huang L. (2010) "Population-based resequencing analysis of improved wheat germplasm at wheat leaf rust resistance locus Lr21." *Theoretical and Applied Genetics* **121**, 271–281.

Furbank R.T., Tester M. (2011) "Phenomics – technologies to relieve the phenotyping bottleneck." *Trends in Plant Science* **16**, 635–644.

Furbank R.T., von Caemmerer S., Sheehy J., Edwards G. (2009) "C_4 rice: a challenge for plant phenomics." *Functional Plant Biology* **36**, 845–856.

Gabriel K.R. (1971) "The biplot graphic display of matrices with application to principal component analysis." *Biometrika* **58**, 453–467.

Gauch H.G. (1992) "Statistical analysis of regional yield trials: AMMI analysis of factorial designs." Eksevuerm Amsterdam, the Netherlands.

Hao W., Zhu M-Z., Gao J-P., Dun S-Y., Lin H-X. (2009) "Identification of quantitative trait loci for rice quality in a population of chromosome segment substitution lines." *Journal of Integrative Plant Biology* **51**, 500–512.

Hassanpanah D., Chakherchaman S.H.A. (2010) "Estimating rank and stability of cultivars by non-parametric analysis." *Research Journal of Environmental Sciences* **2**,173–179.

Hobert O. (2010) "The impact of whole genome sequencing on model system genetics: get ready for the ride." *Genetics* **184**, 317–319.

Huang X., Zhao Y., Wei X., Li C., Wang A., Zhao Q., Li W., Guo Y., Deng L., Zhu C., Fan D., Lu Y.,Weng Q., Liu K., Zhou T., Jing Y., Si L., Dong G., Huang T., Lu T., Feng Q., Qian Q., Li J., Han B. (2012) "Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm." *Nature Genetics* **44(1)**, 32–41.

Juliano B.O., Villareal C.P. (1993) "Grain quality evaluation of world rices." International Rice Research Institute, Manila.

Kadariya M., Glover K.D., Mergoum M., Osborne L.E. (2008) "Biplot analysis of agronomic and *Fusarium* head blight resistance traits in spring wheat." *Journal of Crop Improvement* **22**, 147–170.

Kepiro J.L., McClung A.M., Chen M.H., Yeater K.M., Fjellstrom R.G. (2008) "Mapping QTLs for milling yield and grain characteristics in a tropical *japonica* long grain cross." *Journal of Cereal Science* **48**, 477–485.

Kim J.K., Jang I.C., Wu R., Zuo W.N., Boston R.S., Lee Y.H., Ahn I.P., Nahm B.H. (2003) "Co-expression of a modified maize ribosome-inactivating protein and a rice basic chitinase gene in transgenic rice plants confers enhanced resistance to sheath blight." *Transgenic Research* **12**, 475–484.

Krawitz P.M., Schweiger M.R., Rödelsperger C., Marcelis C., Kölsch U., Meisel C., Stephani F., Kinoshita T., Murakami Y., Bauer S., Isau M., Fischer A., Dahl A., Kerick M., Hecht J., Köhler S., Jäger M., Grünhagen J., de Condor B.J., Doelken S., Brunner H.G., Meinecke P., Passarge E., Thompson M.D., Cole D.E., Horn D., Roscioli T., Mundlos S., Robinson P.N. (2010) "Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome." *Nature Genetics* **2**, 827–829.

Lillemo M., Singh R.P., van Ginkel M. (2010) "Identification of stable resistance to powdery mildew in wheat based on parametric and non-parametric methods." *Crop Science* **50**, 478–485.

Ling K.S., Harris K.R., Meyer J.D., Levi A., Guner N., Wehner T.C., Bendahmane A., Havey M.J. (2009) "Non-synonymous single nucleotide polymorphisms in the watermelon eIF4E gene are closely associated with resistance to zucchini yellow mosaic virus." *Theoretical and Applied Genetics* **120**, 191–200.

Linscombe S.D., Jodari F., Bollich P.K., Groth D.E., White L.M., Chu Q.R., Dunand R.T., Sanders D.E. (2000) "Registration of 'Cocodrie' rice." *Crop Science* **40**, 294.

Liu M., Sun Z.X., Zhu J., Xu T., Harman G.E., Lorito M.. (2004) "Enhancing rice resistance to fungal pathogens by transformation with cell wall degrading enzyme genes from *Trichoderma atroviride.*" *Journal of Zhejian University SCIENCE* **5**, 133–136.

Liu G., Jia Y., Correa-Victoria F.J., Prado G.A., Yeater K.M., McClung A., Correll J.C. (2009) "Mapping quantitative trait loci responsible for resistance to sheath blight in rice." *Phytopathology* **99**, 1078–1084.

Lou J., Chen L., Yue G.Y., Lou Q., Mei H., Xiong L. (2009) "QTL mapping of grain quality traits in rice. *Journal of Cereal Science* **50**, 145–151.

Luan W., Chen H., Fu Y., Si H., Peng W., Song S., Liu W., Hu G., Sun Z., Xie D., Sun C. (2009) "The effect of the crosstalk between photoperiod and temperature on the heading-date in rice." *PLos ONE* **4(6)**: e5891. Doi;10.1371/journal.pone0005891.

Lupski J.R., Reid J.G., Gonzaga-Jauregui C., Deiros D.R., Chen D.C.Y., Nazareth L., Bainbridge M., Dinh H., Jing C., Wheeler D.A., McGuire A.L., Zhang F., Stankiewicz P., Halperin J.J., Yang C., Gehman C., Guo D., Irikat R.K., Tom W., Fantin N.J., Muzny D.M., Gibbs R.A. (2010) "Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy." *The New England Journal of Medicine* **362**, 1181–1191.

Maruthasalam S., Kalpana K., Kumar K.K., Loganathan M., Poovannan K., Raja J.A., Kokiladevi E., Samiyappan R., Sudhakar D., Balasubramanian P. (2007) "Pyramiding transgenic resistance in elite *indica* rice cultivars against the sheath blight and bacterial blight." *Plant Cell Reports* **26**, 791–804.

McNally K.L., Childs K.L., Bohnert R., Davidson R.M., Zhao K., Ulata V.J., Zeller G., Clark R.M., Hoen D.R., Bureau T.E., Stokowski R., Ballinger D.G., Frazer K.A., Cox D.R., Padhukasahasram B., Bustamante C.D., Weigel D., Mackill D.J., Bruskiewich R.M., Ratsch G., Buell C.R., Leung H., Leach J.E. (2009) "Genomewide SNP variation reveals relationships among landraces and modern varieties of rice." *Proceedings of the National Academy of Sciences of the United States of America* **28**, 12273–12278.

Nassar R., Hühn M. (1987) "Studies on estimation of phenotypic stability: Tests of significance for non parametric measures of phenotypic stability." *Biometric* **43**, 45–53.

Nayak D., Bose L.K., Reddy P.R., Nayak P. (2008) "Host-pathogen interaction in rice-bacterial blight pathosystem." *Journal of Plant Protection Research* **48**, 371–384.

Ng S.B., Buckingham K.J., Lee C., Bigham A.W., Tabor H.K., Dent K.M., Huff C.D., Shannon P.T., Jabs E.W., Nickerson D.A., Shendure J., Bamshad M.J. (2010a) "Exome sequencing identifies the cause of a Mendelian disorder." *Nature Genetics* **42**, 30–35.

Ng S.B., Bigham A.W., Buckingham K.J., Hannibal M.C., McMillin M.J., Gildersleeve H.I., Beck A.E., Tabor H.K., Cooper G.M., Mefford H.C., Lee C., Turner E.H., Smith J.D., Rieder M.J., Yoshiura K., Matsumoto N., Ohta T., Niikawa N., Nickerson D.A., Bamshad M.J., Shendure J. (2010b) "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome." *Nature Genetics* **42**,790–793.

Ordonez Jr. S.A., Silva J., Oard J.H. (2010) "Association mapping of grain quality and flowering time in elite *japonica* rice germplasm." *Journal of Cereal Science* **51**, 337–343.

Pinson S.R.M., Capdevielle F.M., Oard J.H. (2005) "Confirming QTLs and finding additional loci conditioning sheath blight resistance in rice using recombinant inbred lines." *Crop Science* **45**, 503–510.

Prasad B., Eizenga G.C. (2008) "Rice sheath blight disease identified in *Oryza* spp. accessions." *Plant Disease* **92**, 1503–1509.

Reynolds M., Foulkes M.J., Slafer G.A., Berry P., Parry M.A.J., Snape J.W., Angus W.J. (2009) "Raising yield potential in wheat." *Journal of Experimental Botany* **60**, 1899–1918.

Rios J., Stein E., Shendure J., Hobbs H.H., Cohen J.C. (2010) "Identification by whole genome resequencing of gene defect responsible for severe hypercholesterolemia." *Human Molecular Genetics* **19**, 4313–4318.

Roach J.C., Glusman G., Smit A.F.A., Huff C.D., Hubley R., Shannon R.T., Rowen L., Pant K.P., Goodman N., Bamshad M., Shendure J., Drmanac R., Jorde L.B., Hood L., Galas D.J. (2010) **"**Analysis of genetic inheritance in a family quartet by whole-genome sequencing." *Science* **328**, 636–639.

Rush M.C., Lindberg G.D. (1996) "Rice disease research." *Rice Journal* **77**, 49–52.

Sabaghnia N., Dehghani H., Sabaghpour S.H. (2006) "Nonparametric methods for interpreting genotype x environment interaction of lentil genotypes." *Crop Science* **46**, 1100–1106.

SAS Institute Inc. 2008. SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc.

Scapim C.A., Oliveira V.R., Braceini A.L., Cruz C.D., Andrade C.A., Vidigal M.C.G. (2000) "Yield stability in Maize (*Zea mays* L.) and correlation among the parameters of the Eberhart and Russell, Lin and Binns and Huehn models." *Genetics and Molecular Biology* **23**, 387–393.

Schneider J.H.M., Van den Boogert P.H.J.F. (1999) **"**Exploring Differential Interactions between *Rhizoctonia solani* AG 2-t Isolates and Tulip Cultivars." *Plant Disease* **83**, 474–481.

Septiningsih E.M., Trijatmiko K.R., Moeljopawiro S., McCouch S.R. (2003) "Identification of quantitative trait loci for grain quality in and advanced backcross population derived from the Oryza sativa variety IR64 and the wild relative *O. rufipogon*." *Theoretical and Applied Genetics* **107**, 1433–1441.

Shah J.M., Raghupathy V., Veluthambi K. (2009) "Enhanced sheath blight resistance in transgenic rice expressing an endochitinase gene from Trichoderma virens." *Biotechnology Letters* **31**, 239–244.

Sobreira N.L.M., Cirulli E.T., Avramopoulos D., Wohler E., Oswald G.L., Stevens E.L., Ge D., Shianna K.V., Smith J.P., Maia J.M., Gumbs C.E., Pevsner J., Thomas G., Valle D., Hoover-Fong J.E., Goldstein D.B. (2010) "Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene." *PLoS Genetics* **6**(6), e1000991. doi:10.1371/journal.pgen.1000991.

Song W.Y., Wang G.L., Chen L.L., Kim H.S., Pi L.Y., Holsten T., Gardner J., Wang B., Zhai W.X., Zhu L.H., Fauquet C., Ronald P. (1995) "A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21." *Science* **270**, 1804–1806.

Tabien R.E., Samonte S.O.P.B., Tiongco E.R. (2009) "Relationship of milled grain percentages and flowering-related traits in rice." *Journal of Cereal Science* **49**, 122–127.

Tan Y.F., Li J.X., Yu S.B., Xing Y.Z., Xu C.G., Zhang Q. (1999) "Three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63." *Theoretical and Applied Genetics* **99**, 642–648.

Tan Y.F., Sun M., Xing Y.Z., Hua J.P., Sun X.L., Zhang Q.F. 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." *Theoretical and Applied Genetics* **103**, 1037–1045.

Tan C.X., Ji X.M., Yang Y., Pan X.Y., Zuo S.M., Zhang Y.F., Zou J.H., Chen Z.X., Zhu L.H., Pan X.B. (2005) "Identification and marker-assisted selection of two major quantitative genes controlling rice sheath blight resistance in backcross generations." *Chinese Journal of Genetics and Genomics* **32**, 399–405.

Teer J.K., Mullikin J.C. (2010) "Exome sequencing: the sweet spot before whole genomes." *Human Molecular Genetics* **19**, 145–151.

Tester M., Langridge P. (2010) "Breeding technologies to increase crop production in a changing world." *Science* **327**, 818–822.

Tong P., Prendergast J.G.D., Lohan A.J., Farrington S.M., Cronin S.C., Friel N., Bradley D.G., Hardiman O., Evans A., Wilson J.F., Loftus B.J. (2010) "Sequencing and analysis of an Irish human genome." *Genome Biology* **11**, R91 doi:10.1186/gb-2010-11-9-r91.

Truberg B., Hühn M. (2000) "Contribution to the analysis of genotype by environment interactions: Comparison of different parametric and non-parametric tests for interactions with emphasis on crossover interactions." *Journal of Agronomy and Crop Science* **185**, 267–274.

Tukey J.W. (1977) "Exploratory data analysis." Addison-Wesley, Reading, MA, USA.

Unnevehr L.J., Duff B., Juliano B.O. (1992) "Consumer demand for rice grain quality." International Rice Research Institute, Manila, and International Development Research Center, Ottawa.

Venu R.C., Jia Y., Gowda M., Jia M., Jantasuriyarat C., Stahlberg E., Li H., Rhineheart A., Boddhireddy P., Singh P., Rutger N., Kudrna D., Wing R., Nelson J.C., Wang G.L. (2007) "RL-SAGE and microarray analysis of the rice transcriptome after *Rhizoctonia solani* infection." *Molecular Genetics and Genomics* **278**, 421–431.

Walsh T., Shahin H., Elkan-Miller T., Lee M.K., Thornton A.M., Roeb W., Abu Rayyan A., Loulus S., Avraham K.B., King M.C., Kanaan M. (2010) "Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82." *The American Journal of Human Genetics* **87**, 90–94.

Wan X.Y., Wan J.M., Su C.C., Wang C.M., Shen W.B., Li J.M., Wang H.L., Jiang L., Liu S.J., Chen L.M., Yasui H., Yoshimura A. (2004) *Theoretical and Applied Genetics* **110**, 71–79.

Wang H., Waller L., Tripathy S., St. Martin S.K., Zhou L., Krampis K., Tucker D.M., Mao Y., Hoeschele I., Maroof M.A.S., Tyler B.M., Dorrance A.E. (2010) **"**Analysis of genes underlying soybean quantitative trait loci conferring partial resistance to *Phytophthora sojae*." *The Plant Genome* **3**, 23–40.

Xie W., Feng Q., Huihui Y., Xuehui H., ZhaoQ., Xing Y., Yu S., Han B., Zhang Q. (2010) "Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing." *Proceedings of the National Academy of Sciences of the United States of America* **107**, 10578–10583.

Xu J., Zhao Q., Du P., Xu C., Wang B., Feng Q., Liu Q., Tang S., Gu M., Han B., Liang G. (2010) "Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome re-sequencing in rice (*Oryza sativa* L.)." *BMC Genomics* **11**, 656.

Yan W., Falk D.E. (2002) "Biplot analysis of host-by-pathogen data." *Plant Disease* **86**, 1396–1401.

Yan W., Holland J.B. (2010) "A heritability-adjusted GGE biplot for test environment evaluation." *Euphytica* **171**, 355–369.

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y., Nagamura Y., Sasaki T. (2000) "Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS." *Plant Cell* **12**, 2473–2484.

Yi X., Liang Y., Huerta-Sanchez E., Jin X., Cuo Z.X., Pool J.E., Xu X., Jiang H., Vinckenbosch N., Korneliussen T.S., Zheng H., Liu T., He W., Li K., Luo R., Nie X, Wu H., Zhao M., Cao H., Zou J., Shan Y., Li S., Yang Q., Asan., Ni P., Tian G., Xu J., Liu X., Jiang T., Wu R., Zhou G., Tang M., Qin J., Wang T., Feng S., Li G., Huasang, Luosang J., Wang W., Chen F., Wang Y., Zheng X., Li Z., Bianba Z., Yang G., Wang X., Tang S., Gao G., Chen Y., Luo Z., Gusang L., Cao Z., Zhang Q., Ouyang W., Ren X., Liang H., Zheng H., Huang Y., Li J., Bolund L., Kristiansen K., Li Y., Zhang Y., Zhang X., Li R., Li S., Yang H., Nielsen R., Wang J., Wang J. (2010) "Sequencing of 50 human exomes reveals adaptation to high altitude." *Science* **329**, 75–78.

Zhao C.J., Wang A.R., Shi Y.J., Wang L.Q., Liu W.D., Wang Z.H., Lu G.D. (2008) "Identification of defense-related genes in rice responding to challenge by *Rhizoctonia solani*." *Theoretical and Applied Genetics* **116**, 501–516.

Zuo S., Zhang L., Wang H., Yin Y., Zhang Y., Chen Z., Ma Y., Pan X. (2008) "Prospect of the QTL-qSB-9Tq utilized in molecular breeding program of *japonica* rice against sheath blight." *Journal of Genetics and Genomics* **35**, 499–505.

CHAPTER 2 GGE BIPLOT EXPLORATION OF RESISTANCE TO SHEATH BLIGHT DISEASE IN DOUBLED-HAPLOID LINES OF RICE

2.1 Introduction

Development of sheath blight-resistant commercial rice cultivars is a high priority for the U.S. rice industry and for other regions of the world (Marshall and Rush, 1980; Savary et al. 2000; Slaton et al. 2003). However, progress has been slow in transferring stable resistance to commercial cultivars due to complex inheritance and few good sources of stable resistance in exotic or adapted germplasm (Eizenga et al., 2002). Another major challenge is low repeatability in field-plot and greenhouse ratings due to variation and potential interactions among temperature, humidity and other factors across years and locations (Oard and Groth, unpublished observations).

Interpretation of genotype-by-environment (GE) effects that impact performance and stability of agronomic and other traits can be obtained by inspection of a biplot graphical display (Gabriel, 1971; Yan and Tinker, 2006; Yan and Holland, 2010). The GGE biplot was reported to provide insights into patterns of lines and environments that contribute to potential interactions (Samonte et al., 2005). Biplot analyses have been used primarily for GE interactions of yield and related traits in multi-location trials (Yan and Kang, 2003). This methodology has also been utilized recently to characterize and identify breeding lines and cultivars that are resistant to various diseases. Examples include identification of wheat (*Triticum aestivum* L.) lines with stable resistance to the powdery mildew pathogen *Blumeria graminis* f. sp. *tritici* (Lillemo et al., 2010). Similarly, biplots were used to select elite wheat lines resistant to Fusariam head blight caused by *Fusarium graminearum* (Kadariya et al., 2008). GGE biplot evaluation was also

[1] Reprinted by permission of "*Crop Science*"

conducted to identify barley (*Hordeum vulgare L.)* lines showing resistance to net blotch caused by *Drechslera teres f. sp. maculata* (Yan and Falk, 2002).

Identification of outliers and interactions between lines and environments can also be accomplished by the "median polish" method (Tukey, 1977) that was recently implemented to identify stable resistance in wheat against *B. graminis* f. sp. *tritici* and *Septoria tritici* Roberge in Desmaz (Arraiano and Brown, 2006; Lillemo et al., 2010). The same approach was used in sunflower (*Helianthus annuus* L.) to identify lines resistant to *Phoma macdonaldii* Boerema (Darvishzadeh et al., 2007). Hühn's nonparametric methods (Nassar and Hühn, 1987; Truberg and Hühn, 2000) have been exploited primarily for stability analyses of yield and associated traits in various crops, including potato (*Solanum tuberosum* L.) (Hassanpanah and Chakherchaman, 2010), lentils (*Lens culinaris* L.) (Sabaghnia et al., 2006), and maize (*Zea mays* L.) (Scapim et al., 2000). Recently, this approach was utilized to identify powdery mildew resistance in wheat (Lillemo et al., 2010). Similar to GGE biplot, the additive main effects and multiplicative interaction (AMMI) method (Gauch, 1992) uses two-dimensional graphical displays to evaluate GE. AMMI has been used to evaluate host-pathogen interactions of rice-*Xanthomonas oryzae pv. oryzae* (Nayak et al., 2008), of tulip (*Tulipa* sp.)-*R. solani* interactions (Schneider and Van den Boogert, 1999), and stability of faba beans (*Vicia faba* L.) for resistance to *Orobanche* sp. (Flores et al., 1996).

The SB2 mapping population of 322 doubled-haploid (DH) lines (Chu et al., 2006) developed during the RiceCAP project (www.uark.edu/ua/ricecap) was derived from a cross between the resistant parent MCR10277 (Rush et al., 2006) and the susceptible parent Cocodrie (Linscombe et al., 2000). Given the challenges described above for selecting stable genotypes with high levels of resistance, we were interested in exploring the potential of different

approaches to identify candidate rice lines with high and stable levels of sheath blight resistance. The specific objective of this research was to explore performance of and stability for sheath blight resistance among DH lines of the SB2 population using GGE biplot and other methods.

2.2 Materials and Methods

2.2.1 SB2 Mapping Population and Field Plot Trials

The SB2 population was developed as a genetic mapping resource to identify lines containing molecular markers associated with sheath blight resistance (Chu et al., 2006). SB2 consists of 322 DH lines derived from a cross between the susceptible parent Cocodrie (CCDR) (Linscombe et al., 2000) and the resistant parent MCR10277 (MCR) (Rush et al., 2006). The SB2 lines and parents were planted at the Rice Research Station, Crowley, Louisiana and the Rice Research and Extension Center, Stuttgart, Arkansas in 2006 and 2007 for a total of four test environments. A randomized complete-block design with three replications was used, with plots consisting of a single row 1.8 m long and 0.17 m spacing between rows. Standard agronomic practices were carried out to maximize growth and to control pests. During the late tillering stage, plants were inoculated with mycelia of isolate LR72 from the fungal pathogen *R. solani* grown on a sterile rice hull: grain medium. DH lines were rated at the soft dough stage of grain maturity for sheath blight (SB) severity on a 0–9 scale, with 0 = no disease and 9 = dead plants.

2.2.2 Variance Components and Estimation of Broad-Sense Heritabilities

A combined ANOVA for randomized complete-block designs, as described in Table 2.3 of McIntosh (1983), was carried out using SAS (Release 9.1.3; SAS Institute, Cary, NC, 2009a). For each location-year, the genotypic variance (σ_g^2) and error variance (σ_e^2) were estimated

using Proc Varcomp of SAS (Release 9.1.3; SAS Institute, Cary, NC, 2009a). Broad sense heritabilities were calculated as

$$H = 1 - \frac{\sigma_e^2}{b\sigma_g^2 + \sigma_e^2}$$

where *b* is the number of replications. For this study, *H* was considered a measure of the usefulness of the trial in genotype evaluation where *H*=0 indicates the differences among genotypic means in the trial are completely due to random error, while *H*=1 indicates differences are entirely due to genetic effects (Yan and Holland, 2010).

2.2.3 Heritability-Adjusted Genotype plus Genotype-by-environment (HA-GGE) Interaction Biplot Analysis

SB mean scores from the GE table of means were transformed by subtracting each mean score from nine. Using this transformation, new scores exhibited the same general interpretability principles as yield and other similar data where high values are preferred. Transformed SB mean scores were stored in a 324 genotypes (322 DH lines plus their two parents) *x* 4 environments matrix ***M*** and heritability-adjusted scaling (Yan and Holland, 2010) was performed in SAS/IML (Release 9.1.3; SAS Institute, Cary, NC, 2008) using the following expression

$$m_{ij} = \frac{\bar{y}_{ij} - \bar{y}_{.j}}{s_j}\sqrt{H_j}$$

where i=1 to 324, j=1 to 4, m_{ij} is the entry for the i^{th} row and j^{th} column of the scaled matrix, $\bar{y}_{ij}$ is the transformed SB mean score for the i^{th} genotype in the j^{th} environment, $\bar{y}_{.j}$ is the overall mean of the transformed SB mean scores from environment j, s_j is the standard deviation for the j^{th} environment, and H_j is the estimated broad-sense heritability for the j^{th} environment.

2.2.4 Environmental and Genotypic Principal Component Scores

Two different biplots can be constructed depending on how the singular values (characteristic roots of a matrix) are partitioned: the "environment-focused" biplot and the "genotype-focused" biplot (Yan and Tinker, 2006). When study of the relationships among environments is desired, the singular values must be entirely partitioned into the environment principal components (PC). Therefore, the set of PC for environments and genotypes are given by the first two columns of the matrix product $\boldsymbol{E}_{4x4}\, Diag(\boldsymbol{L}_{4x1})$ and the first two columns of the matrix $\boldsymbol{G}_{324x4}$, respectively. The matrices $\boldsymbol{E}$, $\boldsymbol{L}$, and $\boldsymbol{G}$ were obtained from the singular value decomposition of the matrix $\boldsymbol{M}$. When the goodness-of-fit of the approximation using the first two PCs is close to 1, the cosine of the angle between two environmental vectors is approximately equal to the genetic correlation between them (Gabriel, 1971; Kroonenberg, 1995; Yan and Tinker, 2006). Inspection of the angle formed by two environmental vectors in an environment-focused GGE biplot visually conveys the following information: An acute angle (close to 0^{o}) implies a high positive genetic correlation. A right angle (90^{o}) implies no genetic correlation between the two environments. An obtuse angle (close to 180^{o}) implies a high negative genetic correlation.

To study relationships among genotypes, the singular values must be entirely partitioned into genotype PCs. In this case, the set of PCs for environments and genotypes is given by the first two columns of the matrix $\boldsymbol{E}_{4x4}$ and the first two columns of the matrix product $\boldsymbol{G}_{324x4}\, Diag(\boldsymbol{L}_{4x1})$, respectively**.** The Euclidean distance between any pair of genotypes is a measure of the overall dissimilarity between them (Yan and Tinker 2006). To assess HA-GGE biplot displays, balanced bootstrap confidence intervals for PC scores of individual DH lines across the four test environments were calculated as described by Lebart (2007) using R statistical software

(http://www.r-project.org). Results were displayed as confidence interval regions or ellipses using JMP 8.0 (SAS Institute, Cary, NC, 2009b).

2.2.5 Additional Methods to Study GE

To complement the HA-GGE biplot analyses, three additional methods were also used to study GE: Hühn's nonparametric stability analysis, median polish, and additive main effects and multiplicative interaction (AMMI). Based on rankings across environments, two nonparametric stability statistics were computed as proposed by Hühn (1990a, 1990b): $S_i^{(1)}$ for measuring the mean absolute rank difference of genotype *i*, and $S_i^{(2)}$ which provides the variance of the ranks. The mean absolute residual from a median polish analysis (Tukey, 1977) was used as another stability indicator where high mean absolute residual values identify lines with high phenotypic stability. The residuals from an ANOVA using the GE table of means were used to compute PC scores to construct an AMMI2 biplot (Gauch, 1992). Results from the additional methods were compared against the mean and variance of the raw SB scores, and against the HA-GGE biplot statistics (PC1 and absolute value of PC2) using correlation analysis.

2.3 Results

2.3.1 Sheath Blight Severity Scores and Estimation of Broad-sense Heritabilities

Mean values, standard deviations, and the frequency distribution for SB scores within and across years and locations suggest that the DH lines exhibited similar levels of disease severity at both locations in 2006 and 2007 (Figure 2.1). Conversely, differences in both mean and variances are evident from the histograms for Louisiana 2006 and 2007 (LA06 and LA07, respectively). The average SB score in Arkansas 2006 (AR06) was slightly smaller than the

average scores for Louisiana, but the shape of the distribution was flatter, showing a higher level of variability ($s \approx 1.7$). Even though Arkansas 2007 (AR07) showed the smallest variability ($s \approx 1.1$), the distribution was skewed with SB scores considerably higher than the remaining three environments. Perhaps the environmental conditions for AR07 were more favorable for the development of higher disease pressure than in the other environments.

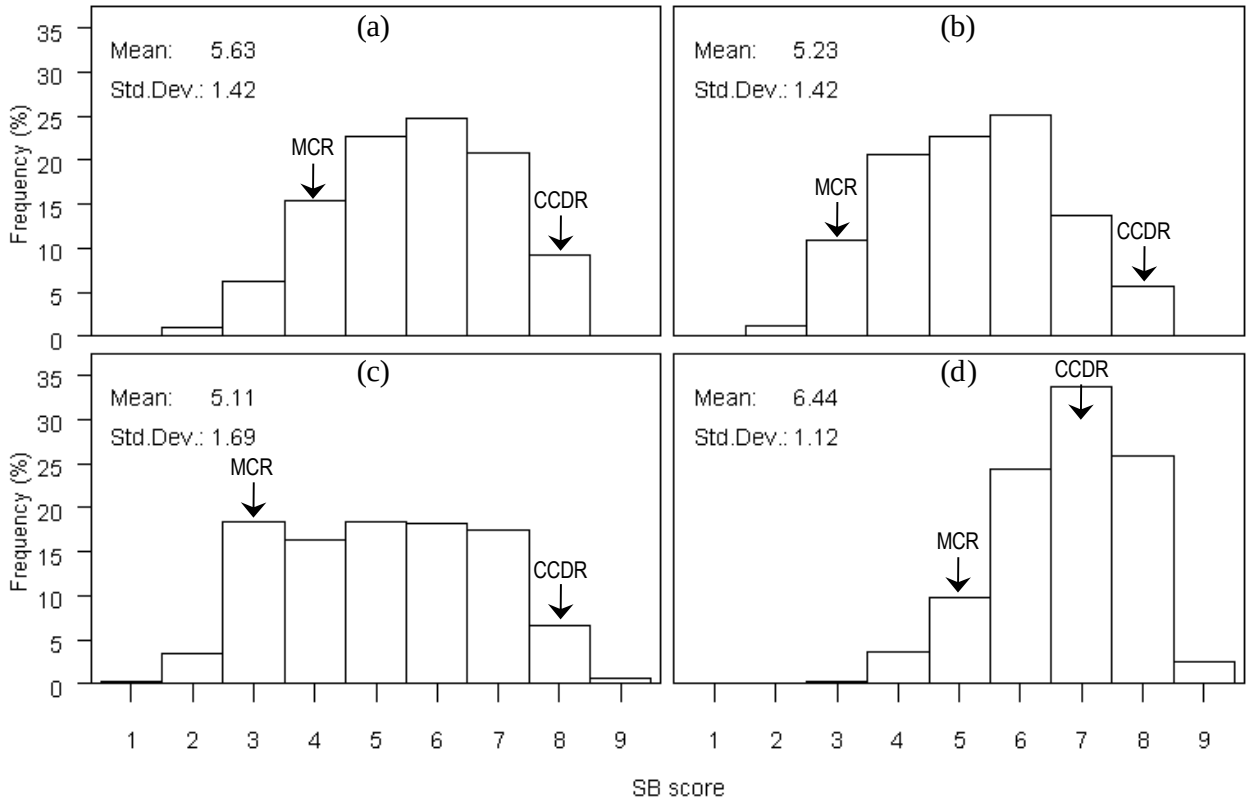

Figure 2.1 Frequency distribution for SB scores in Louisiana 2006 (a), Louisiana 2007 (b), Arkansas 2006 (c), and Arkansas 2007 (d). The arrows indicate scores for the resistant (MCR) and susceptible (CCDR) parents.

The analysis of variance, combining data from all four environments, is shown in Table 2.1. The genotype-by-environment interaction (genotype-by-location-by-year in this study) was highly significant, as were all pair-wise interaction effects. Because of the availability of "large" sample sizes (> 300 DH lines in three blocks for every location-year), the tests for location and year main effects were statistically significant. However, small differences between average SB

scores for locations and years (5.4–5.7 for Louisiana-Arkansas, 5.3–5.8 for 2006–2007) suggested that the statistically significant tests for location and year main effects did not have practical importance.

Table 2.1 Analysis of variance for SB scores of 322 DH lines plus SB2 parental cultivars Cocodrie and MCR10277, Louisiana, Arkansas, 2006, and 2007.

Source of variation	Degrees of freedom	Sum of squares	Mean square	F value
Location	1	115.1	115.1	177.8 **
Year	1	214.8	214.8	331.9 **
Location x Year	1	729.1	729.1	1126.6 **
Block (Location x Year)	8	50.4	6.3	
Genotype	323	4393.6	13.6	21.0 **
Genotype x Location	323	614.2	1.9	2.9 **
Genotype x Year	323	703.3	2.2	3.4 **
Genotype x Location x Year	323	445.1	1.4	2.1 **
Error	2583	1671.7	0.65	
Total	3886	8940.3		

** Statistically significant differences (P-value < 0.01).

The results for the within-environment ANOVA, variance components, and heritabilities are shown in Table 2.2. Parameter estimates for 2006 and 2007 were, in general, more similar for SB scores obtained in Louisiana than in Arkansas. In 2007 the genotypic and phenotypic variances were smaller than in 2006, with more striking differences in Arkansas. Moreover, the genotypic variance in 2006 was almost three times the genotypic variance in 2007 for Arkansas, whereas the values for Louisiana were similar. Nevertheless, the estimated heritability values were relatively high in all four environments (between 80% and 90%) with slightly higher estimates for 2006.

Table 2.2 Broad-sense heritability (H) and parameter estimates for SB severity among 322 DH lines plus SB2 parental cultivars Cocodrie and MCR10277, Louisiana, Arkansas, 2006, and 2007.

Parameters	LA 2006	LA 2007	AR 2006	AR 2007
Blocks (b)	3	3	3	3
Genotypes	324	324	324	324
Blocks mean square	1.51	3.00	19.97	0.73
Genotypes mean square	4.96	4.54	6.85	2.73
Mean Square Error (σ^2_e)	0.53	0.74	0.80	0.52
σ^2_{blk}	0.003	0.007	0.059	0.001
σ^2_g	1.48	1.27	2.02	0.74
$\sigma^2_p = \sigma^2_g + \sigma^2_e/b$	1.65	1.51	2.28	0.91
SB Mean (μ) (0-9 scale)	5.63	5.23	5.11	6.44
SE (σ_e) (0-9 scale)	0.73	0.86	0.89	0.72
SD (σ_p) (0-9 scale)	1.29	1.23	1.51	0.95
CV% = SE/Mean x 100	12.93	16.44	17.53	11.17
$H = 1 - (\sigma^2_e/\sigma^2_p) / b$	0.89	0.84	0.88	0.81

2.3.2 Mega-environment Identification

To identify and characterize potential GE interactions, three biplots were constructed (Figures 2.2, 2.3, and 2.4). The environment-focused HA-GGE biplot shown in Figure 2.2 was used to investigate relationships among environments and to identify a potential "mega-environment" - defined as meaningful subsets of similar environments (Yan and Tinker, 2006). Approximately 84% of the total variability was accounted for by the first two PCs. The lengths of the displayed environmental vectors (distance from the biplot origin to the environment marker point) were proportional to the square root of the heritability estimates, and as indicated in Figure 2.2, the four vectors exhibited similar lengths. The cosine of the angle between two environmental vectors provided an estimate of their correlation coefficient. Although all four environments were positively correlated (acute angles), some of the correlations with AR06 were only moderate.

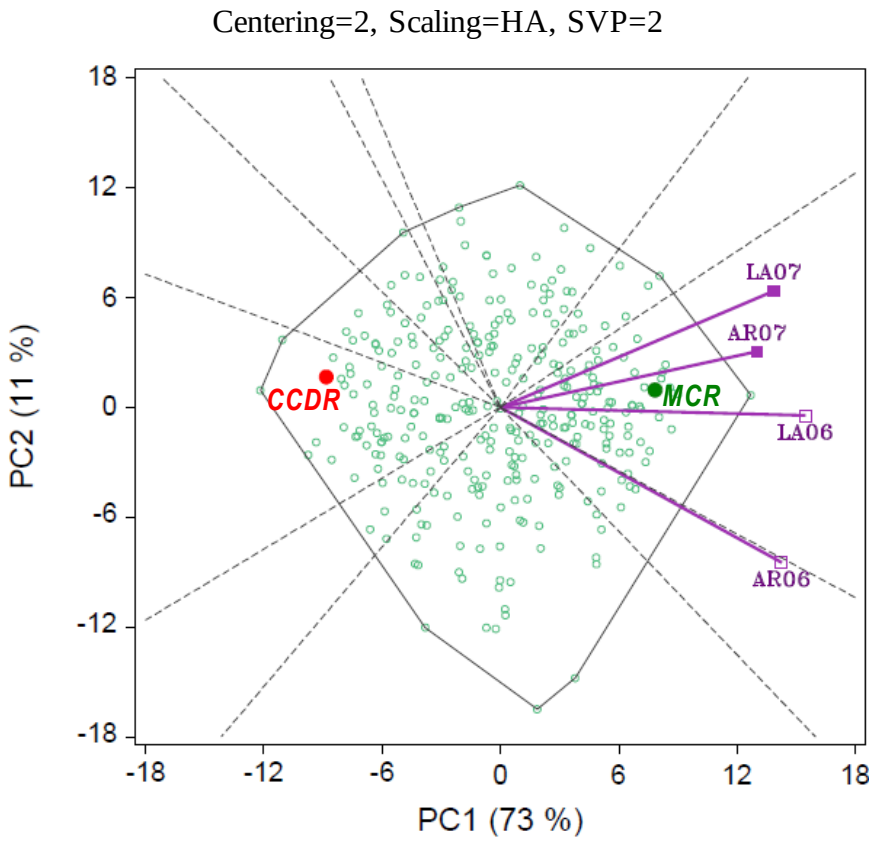

Figure 2.2 Which-won-where display of the environment-focused HA-GGE biplot for mega-environment identification, Louisiana and Arkansas, 2006 and 2007. "Centering=2" means the data were centered by the means of environments. "Scaling=HA" means heritability adjusted where the environment standardized data were multiplied by the heritability in each environment. "SVP=2" means the singular values were partitioned into the environment eigenvectors for visualizing the correlation among environments.

For example, the correlation between AR06 and AR07 was only 0.57, whereas the correlation between LA06 and LA07 was approximately 0.90. As shown in Figure 2.2, the AR07 environmental vector was located within the angle formed by the environmental vectors for Louisiana. Consequently, the correlation of AR07 with either LA06 (0.98) or LA07 (0.96) was even higher than the correlation between LA06 and LA07. A polygon that encloses all marker points is shown in Figure 2.2. The lines perpendicular to its edges divided the plot into sectors. Three out of the four environments (LA06, LA07, and AR07) fell into the same sector, whereas

AR06 fell outside but very close to the edge of that sector and its correlation with LA06 was very high (0.87). Therefore, a single mega-environment consisting of the four sub-environments was identified. The susceptible parent CCDR (red-filled circle), and the resistant parent MCR (green-filled circle) were located on opposite sides, which was consistent with their mean performance.

2.3.3 Test Environment Evaluation

A second biplot shown in Figure 2.3 was created to conduct test-environment evaluation. Approximately 84% of the total variability was accounted for by the first two PCs.

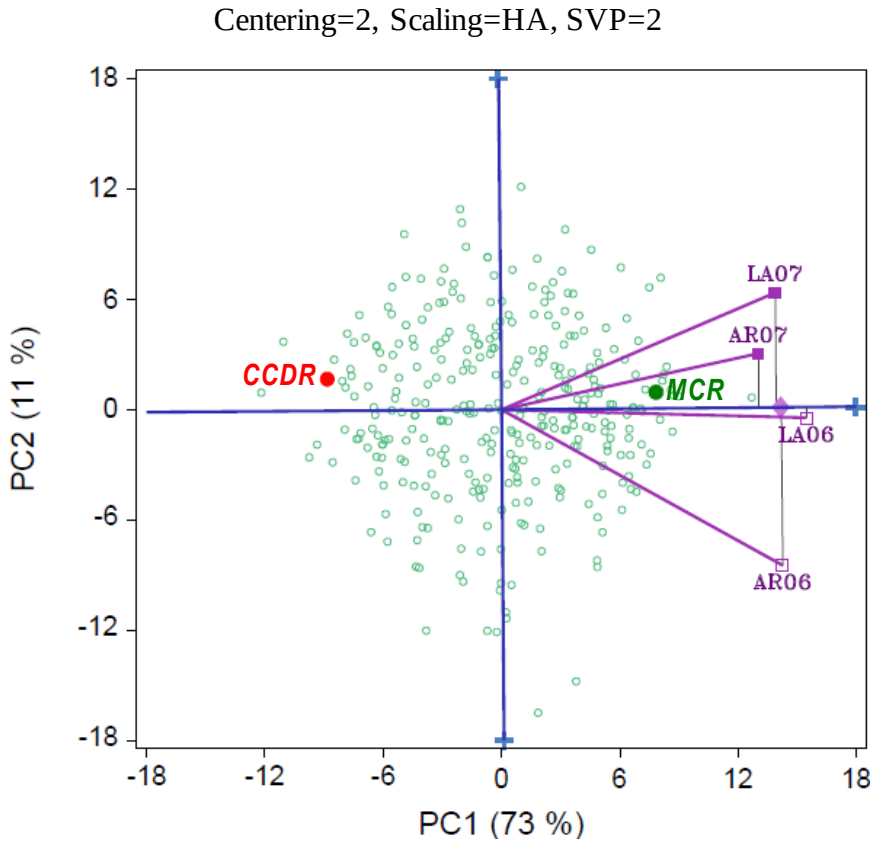

Figure 2.3 Test environment evaluation display of the environment-focused HA-GGE biplot, Louisiana and Arkansas, 2006 and 2007. "Centering=2" means the data were centered by the means of environments. "Scaling=HA" means heritability adjusted where the environment standardized data were multiplied by the heritability in each environment. "SVP=2" means the singular values were partitioned into the environment eigenvectors for visualizing the correlation among environments.

The light-purple diamond in Figure 2.3 represents coordinates equal to the average coordinates of the four marker points for environments, referred to as the "Average Environment Coordinates" (AEC; Yan and Holland, 2010). The blue axis that passed through the origin of the biplot and in the direction of the AEC was labeled the "Average Environment Axis" (AEA) and the plus sign over the AEA pointed in the direction of high transformed SB scores. According to Yan and Holland, (2010), usefulness of the four environments is determined by the projection of every environment onto the AEA, which allowed the four environments to be ranked as: AR07 < LA07 < AR06 < LA06. The blue axis that passed through the origin and was perpendicular to the AEA showed two plus signs that pointed away from stability, regardless of direction (Yan and Holland, 2010).

2.3.4 Genotype Evaluation

A third GGE biplot, designated the "Genotype-focused biplot" (SVP=1), is shown in Figure 2.4 that was used to study relationships among genotypes (DH lines). Only genotypes are shown that were either better or worse than the resistant/susceptible parents or were highly unstable (high PC2 values regardless of direction). It is worth noting that the correlation between the PC1 scores and the mean performance for DH lines was almost perfect (r = 0.99). A similar result was obtained by Lillemo et al. (2010) who used GGE biplot to identify stable resistance to powdery mildew disease in wheat. The correlation between the absolute value of PC2 (stability indicator) and the variance of genotypes across environments in our study was moderate (r = 0.54). To assess the HA-GGE biplot display in Figure 2.4, confidence interval regions or ellipses for PC scores were determined for all 324 genotypes across locations (Louisiana, Arkansas), and years (2006, 2007). Clear separation or non-overlap of the 95% confidence intervals was observed for the subsets of susceptible and resistant DH lines (results not shown). Because of

software limitations, representative confidence ellipses for good (left portion of plot), intermediate (central portion), and poor (right portion) DH performers are displayed in Figure 2.5. We interpreted all confidence interval results to be consistent with and provide support for HA-GGE biplot displays shown in Figure 2.4.

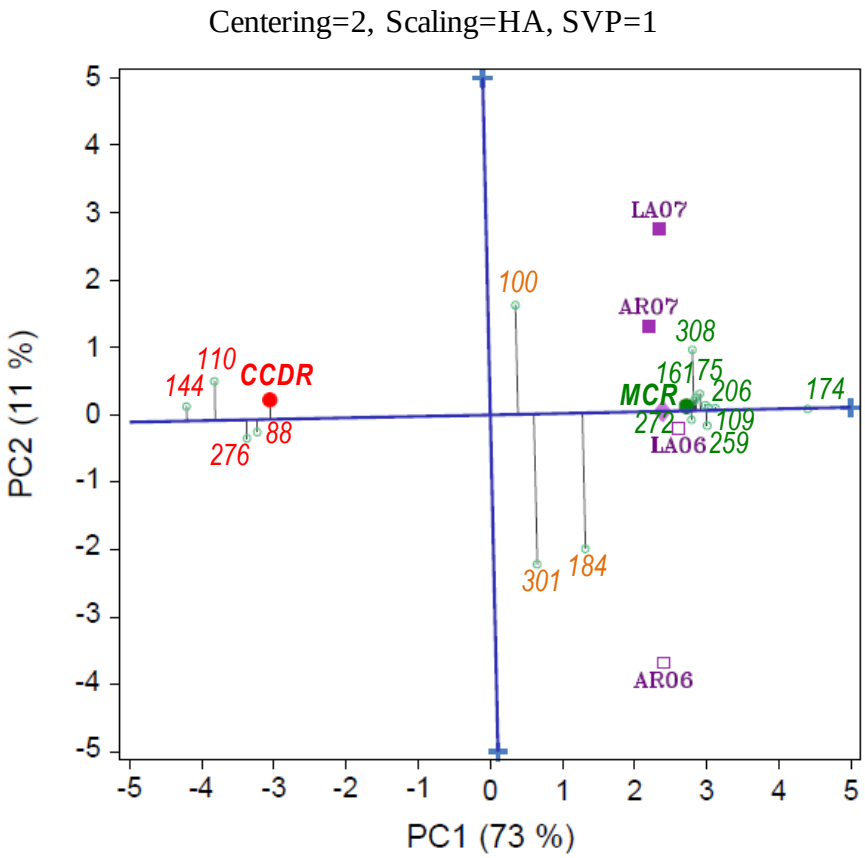

Figure 2.4 Genotype evaluation display of the environment-focused, HA-GGE biplot, Louisiana and Arkansas, 2006 and 2007. "Centering=2" means the data were centered by the means of environments. "Scaling=HA" means heritability adjusted where the environment standardized data were multiplied by the heritability in each environment. "SVP=1" means the singular values were partitioned into the genotype eigenvectors for visualizing the correlation among genotypes.

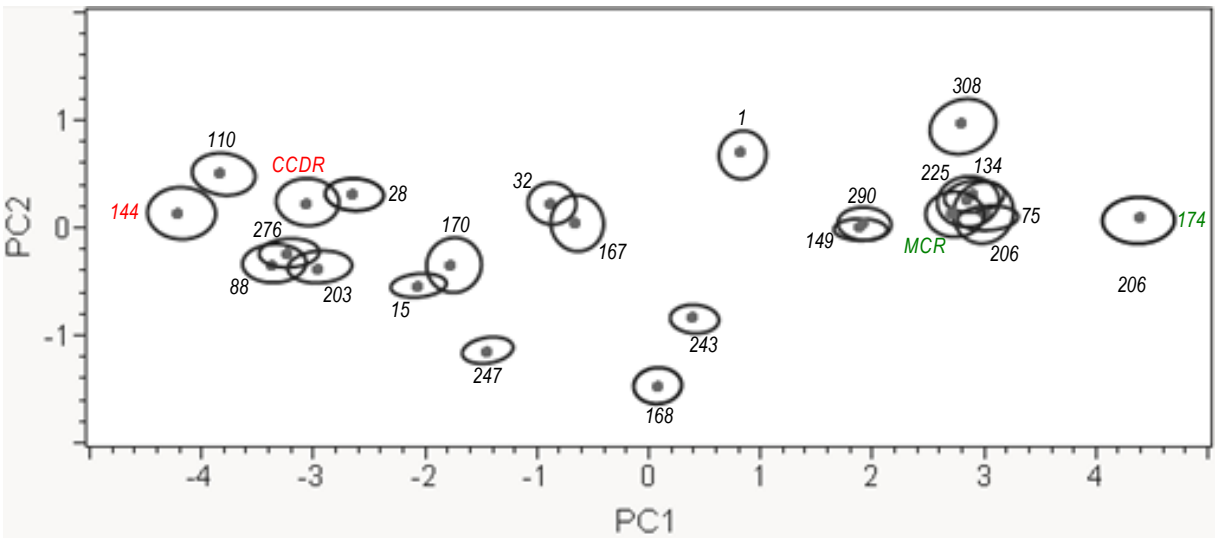

Figure 2.5 Confidence regions (95% coverage ellipses) for PC scores of 23 DH lines and parents from a genotype-focused HA-GGE for SB transformed scores across Louisiana and Arkansas, 2006 and 2007. Software limitations precluded display of confidence interval regions for all 322 DH lines. Regions depicted are representative of DH lines with good, poor, and intermediate levels of sheath blight resistance.

2.3.5 Comparison Among Methods

For each DH genotype, means and variances were computed using the transformed SB scores across environments. High correlations between mean transformed SB scores and various estimates were detected: PC1 (0.999; $P < 0.001$) and Mean Rank (0.991; $P < 0.001$). Similarly, intermediate to high correlations between variances of transformed SB and various estimates were detected: absolute value of PC2 (0.542; $P < 0.001$), Hühn's $S_i^{(1)}$ (0.683; $P < 0.001$), $S_i^{(2)}$ (0.772; $P < 0.001$), and mean absolute residual from median polish (0.732; $P < 0.001$). High to intermediate correlations were also observed between HA-GGE biplot statistics and the other methods used in the study. For example, Mean Rank was highly correlated with PC1 (0.992; $P < 0.001$); intermediate correlations were found between the absolute value of PC2 and $S_i^{(1)}$ (0.701; $P < 0.001$), $S_i^{(2)}$ (0.702; $P < 0.001$), and mean absolute residual from median polish (0.711; $P < 0.001$).

An AMMI2 biplot analysis was also carried out with the same dataset, where the susceptible and resistant parents, and good and poor DH performers were located near the origin and close to each other (results not shown), suggesting that winner and loser genotypes were close to each other, which is counter-intuitive. This result is consistent with one of the AMMI's potential drawbacks as pointed out in the review paper by Yan et al. (2007, p. 649). Therefore, the AMMI2 and AMMI1 results were not considered for further analysis.

2.4 Discussion

Breeding elite cultivars resistant to sheath blight disease is a high priority not only for the U.S. rice industry, but also for Asia and other rice-growing regions of the world (Marshall and Rush, 1980; Savary et al., 2000; Slaton et al., 2003). A major challenge has been to identify high and stable levels of resistance in exotic and unadapted germplasm. Accurate and repeatable field-plot selection for resistance in breeding material is hampered by complex inheritance and location/year variations in environmental factors, such as temperature and humidity that prolong development of elite germplasm. Several recent studies have reported success in the use of graphical display methods, such as GGE and AMMI biplots, to identify sources of stable resistance against different crop pathogens (Flores et al., 1996; Yan and Falk, 2002; Kadariya et al., 2008; Nayak et al., 2008; Lillemo et al., 2010). We were, therefore, interested in exploring the potential of GGE biplot and other approaches to identify candidate rice lines with high and stable levels of sheath blight resistance.

The ANOVA for sheath-blight resistance showed that nearly all sources of variation were significant with the location-by-year effect contributing the most to the observed variation. The contribution of the DH genotypes was relatively small, but larger than its interactions with

location and year. While the ANOVA can provide a general overview of variation and detect potential GE interactions, it generates no useful information about trends that may arise by interactions or create a viable method to select stable genotypes (Samonte et al., 2005). In contrast, GGE analyses combine ANOVA and PC approaches to graphically display G and GE interactions that serve as a first step to identify candidate genotypes with both desirable and consistent performance across years and locations.

Even though location-by-year interactions were a major component of the observed variation in our study, estimates of broad-sense heritabilities were high within each test environment. These estimates were used to create the HA-GGE biplots and rank the four test environments for selecting superior DH lines and to improve efficiency of genotype evaluation, as described by Yan and Holland (2010). The biplot results were consistent with mean SB scores obtained across years and locations for the SB2 parents CCDR and MCR and the 322 DH lines. The GGE methods are considered graphical tools for exploring relationships among genotypes and environments. The utility of biplot displays needs to be tested further to make inferences and critical decisions. We attempted to satisfy this requirement by two different methods. The first was creation of confidence interval regions of PC values corresponding to the HA-GGE biplot in Figure 2.4 by balanced bootstrapping suggested by Yang et al. (2009) and carried out as described by Lebart (2007). The confidence ellipses depicted in Figure 2.5 indicated that the identified (top five) resistant DH lines exhibited significantly better mean performance than the group (bottom five) of susceptible DH lines (the confidence regions of tolerant DH lines do not overlap with the confidence regions of susceptible DH lines). The second method involved calculating median polish values and Hühn's nonparametric stability statistics $S_i^{(1)}$ and $S_i^{(2)}$. Results from both methods complemented and provided support for the HA-GGE biplot analysis

of performance and stability of DH lines from the SB2 mapping population. Specifically, the HA-GGE biplot analysis indicated that 11 DH lines exhibited high and stable levels of sheath-blight resistance across all test environments. Moreover, five susceptible DH lines were identified as potential checks with greater stability than the susceptible parents. Performance and stability of all selected material warrant additional testing in southern U.S. rice-growing regions for future development of elite, sheath blight-resistant cultivars.

2.5 References

Arraiano L.S., Brown J.K.L. (2006) "Identification of isolate-specific and partial resistance to *Septoria tritici* blotch in 238 European wheat cultivars and breeding lines." *Plant Pathology* **55**, 726–738**.**

Chu Q.R., Linscombe S.D., Rush M.C., Groth D.E., Oard J., Sha X., Utomo H.S. (2006) "Registration of a C/M Doubled Haploid Mapping Population of Rice." *Crop Science* **46**, 1416.

Darvishzadeh R., Dechamp-Guillaume G., Hewezi T., Sarrafi A. (2007) "Genotype-isolate interaction for resistance to black stem in sunflower (*Helianthus annuus*)." *Plant Pathology* **56**, 654–660.

Eizenga G.C., Lee F.N., Rutger J.N. (2002) "Screening *Oryza* species plants for rice sheath blight resistance." *Plant Dis*ease **86**, 808–812.

Flores F., Moreno M.T., Martinez A., Cubero J.I. (1996) "Genotype-environment interaction in faba bean: comparison of AMMI and principal coordinate model." *Field Crops Research* **47**, 117–127.

Gabriel K.R. (1971) "The biplot graphic display of matrices with application to principal component analysis." *Biometrika* **58**, 453–467.

Gauch H.G. (1992) "Statistical analysis of regional yield trials: AMMI analysis of factorial designs." Eksevuerm Amsterdam, the Netherlands.

Hassanpanah D., Chakherchaman S.H.A. (2010) "Estimating rank and stability of cultivars by non-parametric analysis." *Research Journal of Environmental Sciences* **2**,173–179.

Hühn M. (1990a) "Nonparametric measures of phenotypic stability. I. Theory." *Euphytica* **47**,189–194.

Hühn M. (1990b) “Nonparametric measures of phenotypic stability. II. Applications.” *Euphytica* **47**, 195–201.

Kadariya M., Glover K.D., Mergoum M., Osborne L.E. (2008) “Biplot analysis of agronomic and *Fusarium* head blight resistance traits in spring wheat.” *Journal of Crop Improvement* **22**, 147–170.

Kroonenberg P.M. (1995) “Introduction to biplots for GxE tables.” Department of Mathematics, Research Report 51. University of Queensland, Australia.

Lebart L. (2007) “Which bootstrap for principal axes method?” In Brito P., Cucumel G., Bertrand P., de Carvalho F. (Eds.). Selected contributions in data analysis and classification, Part VII. Springer, Berlin, Heidelberg. p. 581–588.

Lillemo M., Singh R.P., van Ginkel M. (2010) “Identification of stable resistance to powdery mildew in wheat based on parametric and non-parametric methods.” *Crop Science* **50**, 478–485.

Linscombe S.D., Jodari F., Bollich P.K., Groth D.E., White L.M., Chu Q.R., Dunand R.T., Sanders D.E. (2000) “Registration of ‘Cocodrie’ rice.” *Crop Science* **40**, 294.

Marshall D.S., Rush M.C. (1980) “Infection cushion formation on rice sheaths by *Rhizoctonia solani*.” *Phytopathology* **70**, 947–950.

McIntosh M.S. (1983) “Analysis of combined experiments.” *Agronomy Journal* **75**, 153–155.

Nassar R., Hühn M. (1987) “Studies on estimation of phenotypic stability: Tests of significance for non parametric measures of phenotypic stability.” *Biometric* **43**, 45–53.

Nayak D., Bose L.K., Reddy P.R., Nayak P. (2008) “Host-pathogen interaction in rice-bacterial blight pathosystem.” *Journal of Plant Protection Research* **48**, 371–384.

Rush M.C., Nandakumar R., Zhang S., Groth D.E. (2006) “Status of our research on developing high levels of partial resistance to sheath blight in lines with high yield potential and good agronomic characteristics.” *In: Proceedings Thirty-first Rice Technical Working Group*. The Woodlands, Texas. Louisiana State University Agricultural Center, Louisiana Agricultural Experiment Station Southwest Region, Rice Research Station, Crowley, Louisiana. p. 91.

Sabaghnia N., Dehghani H., Sabaghpour S.H. (2006) “Nonparametric methods for interpreting genotype x environment interaction of lentil genotypes.” *Crop Science* **46**, 1100–1106.

Samonte S.O.P.B., Wilson L.T., McClung A.M., Medley J.C. (2005) “Targeting cultivars onto rice growing environments using AMMI and SREG GGE biplot analyses.” *Crop Science* **45**, 2414–2424.

SAS Institute Inc. 2008. SAS/IML® 9.2 User's Guide. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2009a. SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2009b. JMP® 8 Introductory Guide, Second Edition. Cary, NC: SAS Institute Inc.

Savary S., Willocquet L., Elazegui F.A., Castilla N., Teng P.S. (2000) "Rice pest constraints in tropical Asia: Quantification of yield losses due to rice pests in a range of production situations." *Plant Disease* **84**, 357–369.

Scapim C.A., Oliveira V.R., Braceini A.L., Cruz C.D., Andrade C.A., Vidigal M.C.G. (2000) "Yield stability in Maize (*Zea mays* L.) and correlation among the parameters of the Eberhart and Russell, Lin and Binns and Huehn models." *Genetics and Molecular Biology* **23**, 387–393.

Schneider J.H.M., Van den Boogert P.H.J.F. (1999) **"**Exploring Differential Interactions between *Rhizoctonia solani* AG 2-t Isolates and Tulip Cultivars." *Plant Disease* **83**, 474–481.

Slaton N.A., Cartwright R.D., Meng J., Gbur E.E, Norman R.J. (2003) "Sheath blight severity and rice yield as affected by nitrogen fertilizer rate, application method, and fungicide." *Agronomy Journal* **95**, 1489–1496.

Truberg B., Hühn M. (2000) "Contribution to the analysis of genotype by environment interactions: Comparison of different parametric and non-parametric tests for interactions with emphasis on crossover interactions." *Journal of Agronomy and Crop Science* **185**, 267–274.

Tukey J.W. (1977) "Exploratory data analysis." Addison-Wesley, Reading, MA, USA.

Yan W., Falk D.E. (2002) "Biplot analysis of host-by-pathogen data." *Plant Disease* **86**, 1396–1401.

Yan W., Holland J.B. (2010) "A heritability-adjusted GGE biplot for test environment evaluation." *Euphytica* **171**, 355–369.

Yan W., Kang M.S. (2003) *GGE biplot analysis: a graphical tool for breeders, geneticists, and agronomists*. CRC Press, Boca Raton, FL.

Yan W., Kang M.S., Ma B., Woods S., Cornelius P.L. (2007) "GGE biplot vs. AMMI analysis of genotype-by-environment data." *Crop Science* **47**, 643–655.

Yan W., Tinker N.A. (2006) "Biplot analysis of multi-environment trial data: principles and applications." *Canadian Journal of Plant Science* **86**, 623–645.

Yang R.C., Crossa J., Cornelius P.L., Burgueno J. (2009) “Biplot analysis of genotype x environment interaction: proceed with caution.” *Crop Science* **49**, 1564–1576.

CHAPTER 3 IDENTIFICATION OF CANDIDATE GENES IN RICE FOR RESISTANCE TO SHEATH BLIGHT DISEASE BY WHOLE GENOME SEQUENCING

3.1 Introduction

Sheath blight (SB), caused by the fungal pathogen *Rhizoctonia solani* Kuhn, causes significant yield loss and reduction in grain quality for rice (*Oryza sativa* L.) in the southern U.S. and other regions of the world (Lee and Rush, 1983; Rush and Lindberg, 1996). All current U.S. rice cultivars are susceptible to *R. solani* with costly fungicide applications as the primary means of control. Various studies have shown that response of different rice lines to infection by *R. solani* is expressed as partial resistance (Liu et al., 2009), also referred to as incomplete, quantitative, field, or horizontal resistance (Wang et al., 2010). Numerous genetic and QTL mapping studies have reported partial resistance, hereafter referred to simply as "resistance", to *R. solani* is controlled by multiple regions in the genome each with small or moderate effect (www.gramene.org). As part of the RiceCAP research efforts (www.uark.edu/ua/ricecap), a recombinant inbred line mapping population was used to identify a "major" QTL for SB resistance on chromosome 9 (Liu et al., 2009). This same region was also reported in previous studies to impact SB resistance (Pinson et al., 2005; Tan et al., 2005; Zuo et al., 2008), including the RiceCAP SB2 mapping population evaluated in Louisiana and Arkansas (Nelson et al., unpublished data). A recent study reported a QTL of large effect for resistance on chromosome 11 containing 154 genes of which 11 were tandem repeats of xylanase inhibitor (chitinase) genes (Channamallikarjuna et al., 2010). Zhao et al. (2008) found 50 genes of diverse function that were transcriptionally activated in rice after challenge by *R. solani*. Venu et al. (2007) detected numerous up and down-regulated rice genes after infection by *R. solani* using SAGE and microarray analysis. Increased resistance was observed in transgenic rice containing an engineered ribosome inactivating protein (Kim et al., 2003), thaumatin and chitinase genes from

[1] Reprinted by permission of "*Theoretical and Applied Genetics*"

rice (Maruthasalam et al., 2007), and chitinase genes from *Trichoderma atroviride* and *T. virens* (Liu et al., 2004; Shah et al., 2009). *O. nivara* accessions IRGC 104443 and IRGC 100898 were shown recently to exhibit SB resistance under greenhouse, growth chamber, or laboratory conditions (Prasad and Eizenga, 2008). In spite of the research efforts described above, the routine use of marker-assisted selection to enhance SB resistance in commercial rice cultivars has not been reported.

The advent of next-generation sequencing has been proposed as a rapid, cost effective alternative to Sanger sequencing for identification of candidate genes and variants underlying simple and even complex traits (Hobert, 2010; Teer and Mullikin, 2010). Whole genome sequencing (WGS) of one or a few individuals has recently identified single or multiple variants associated with different Mendelian disorders in humans (Rios et al., 2010; Roach et al., 2010; Sobreira et al., 2010; Tong et al., 2010; Lupski et al., 2010). Similar progress has been made with whole-exome sequencing to uncover rare or recessive variants in humans causing different diseases or adaptations to different environments (Bilgüvar et al., 2010; Krawitz et al., 2010; Ng et al., 2010a, b; Walsh et al., 2010; Yi et al., 2010). Xie et al. (2010) recently used WGS of recombinant inbred lines of rice (*Oryza sativa* L.) at low coverage to construct a linkage map of about 209 K SNPs that successfully identified a known QTL associated with grain width. A similar WGS strategy for chromosome segment substitution lines allowed identification of a QTL containing the *sd1* locus for plant height (Xu et al., 2010). A genomic DNA library enriched for genic sequences in rice was recently constructed followed by deep sequencing that revealed approximately 2,600 SNPs between an *indica* and a *tropical japonica* line (Deschamps et al., 2010).

In addition to QTL mapping for SB resistance by Liu et al. (2009), RiceCAP completed WGS of 13 rice lines using the Illumina GA IIx platform in cooperation with the National Center for Genome Resources (Scheffler et al., unpublished data). The objective of our research was to use sequence data of 13 lines to identify nsSNPs and corresponding candidate genes for SB resistance. We chose to focus on nsSNPs in our study because this class of variants was reported to play a role in the function and evolution of plant resistance (Fu et al., 2010; Ling et al., 2009; McNally et al., 2009; Song et al., 1995) that may complement microarray or other gene expression studies.

3.2 Materials and Methods

3.2.1 Plant material, DNA Isolation, and Variant Selection Strategies

To identify nsSNPs and candidate genes by the "common variant" (CV) selection strategy (see below), three SB resistant lines [Jasmine 85 (PI 595927), MCR010277 (GSOR 200327), and TeQing (PI 536047)] and three susceptible lines [Cocodrie (PI 606331), Cypress (PI 561734), and Lemont (PI 475833)] were used. To further evaluate the initial CV-selected nsSNPs, the following 11 highly/moderately resistant lines were used: Shu Feng 121-1655 (mutant of PI 615015), Rondo (mutant of PI 615022), Taducan (PI 280681), Oryzica Llanos 5 (GSOR 301111), 09DN/Rush072 (M.C. Rush, D.E. Groth, unpublished,), CIAT 4 (F. Correra, unpublished), IR64 (GSOR 301401), Jhona 349 (GSOR 301071), Jouiku 393G (GSOR 301072), *O. nivara* (IRGC 100898), and *O. nivara* (IRGC 10443). In addition, the following nine highly/moderately susceptible lines were used: Azucena (GSOR 301665), Bengal (PI 561735), Bowman (RU0404191), Francis (PI 632447), L-201 (CIor 9971), LaGrue (PI 568891), Leah (GSOR 310045), Nipponbare (GSOR 301164), and Wells (PI 612439).

For the "Principal Component-Biplot" (PB) variant selection strategy (see below), the following 13 lines were used: Bengal, Bowman, Cocodrie, Cypress, Francis, Jasmine 85, LaGrue, Lemont, L-201, MCR010277, TeQing, Shu Feng 121-1655, and Wells. Seedlings of these lines were grown in the dark for approximately 14 days to minimize presence of chloroplasts in the leaves collected for total DNA isolation using the DNeasy 96 Plant kit (Qiagen, Inc., Hilden, Germany). Genomic DNA from each of the 13 lines was used for whole genome sequencing described below. For the remaining lines, DNA was isolated from leaves grown in light using the method described by Li et al. (2010).

3.2.2 WGS and SNP Calling

Genomic DNA isolated from each line as described above was sheared by a Covaris S2 sonicator, and Illumina paired-end genomic libraries were built according to standard protocols. Cluster generation was performed on an Illumina cluster station using a version 2 cluster generation kit and 54 bp paired-end sequencing was carried out on an Illumina Genome Analyzer IIx. Base calling and quality filtering were performed with Illumina Pipeline version 1.4.0 with default parameters. Paired reads were aligned to version 6.0 of the MSU rice genome assembly using GSNAP (Wu and Nacu, 2010) with trimming enabled and allowing up to six mismatches with indels scored as equivalent to three mismatches. Alignments were filtered and variants called and characterized for changes to coding potential via the Alpheus pipeline (Miller et al., 2008). Alignments were required to have at least 50 bp matched for a read aligned singly or 100 bp matched for a paired alignment. Reads mapping equivalently to more than five locations were discarded. Variants were called from alignments meeting these criteria where in at least one of the sequenced lines, the variant allele was detected in at least two uniquely aligning reads, with the bases calling the variant having an phred-equivalent average quality at least 20, and that at

least 20% of the reads aligned to the site in that variety called the variant allele. For each variant meeting these criteria, evidence for the genotype at that site was reported for each of the lines.

3.2.3 Identification of Non-synonymous SNPs in Candidate Resistance Genes

The following steps were carried out for the CV filtering strategy using the six lines described above: (1) Select all variants, except those identified in transposable elements, from output of Alpheus analysis pipeline with quality score at least 25 (2) Select variants from Step 1 with coverage 5 or more (3) Select common variants from step 2 with 3 or more reads in susceptible Cocodrie, Cypress, and Lemont (4) Select common variants from step 2 with 3 or more reads in resistant Jasmine 85, TeQing, and MCR010277 (5) Given that the reference Nipponbare is SB susceptible, select variants that have 100% frequency in the resistant lines and 0% frequency in the susceptible lines (6) Select nsSNPs from Step 5 and identify corresponding candidate genes.

The PB variant selection strategy was carried out using the 13 lines described above in the following steps: (1) Select all variants, except those in transposable elements, from output of Alpheus analysis pipeline with quality scores at least 25 (2) Complete remaining steps using SAS software (Release 9.1.3; SAS Institute, Cary, NC) (3) Select variants from Step 1 with coverage 5 or more (4) Remove common variants selected in Step 3 across all 13 lines with 3 or more reads (5) Perform Principal Component Analysis (PCA) using standardized variant frequencies of the 13 lines (6) Perform Ward's minimum variance clustering (Everitt et al., 2001) using PC1 and PC2 scores obtained in Step 5 (7) For each cluster identified in Step 6, compute average variant frequencies for the 13 lines. Given that the reference Nipponbare is SB susceptible, identify a single cluster with highest average variant frequency in resistant lines and lowest

average variant frequency in susceptible lines (8) Create GGE biplot display (Yan and Tinker, 2006) using PC scores from Step 5. (9) Select nsSNPs from Step 8 and identify corresponding candidate genes.

3.2.4 nsSNP-specific PCR

Primers approximately 25 nt long were designed to amplilfy about 350 bases flanking each nsSNP using the SNAP Program (http://ausubellab.mgh.harvard.edu) based on sequences of Cocodrie and MCR010277 generated by the Illumina GA IIx platform and the reference Nipponbare sequence posted at the Gramene website (www.gramene.org). A 10 uL PCR reaction consisted of the following: 0.5 uL 10 ng DNA template, 1 uL 10X buffer solution (containing 1mM $MgCl_2$)(Applied Biosystems), 7.22 uL of dH_2O, 0.8 uL of 10mM dNTPs mix (Applied Biosystems), 0.2 uL each of 20 uM forward and reverse primers, and 0.08 uL of 5U uL^{-1} of TAQ polymerase (Applied Biosystems). PCR reactions were carried out on the BioRad ICycler consisting of the following steps: 95° C, 3 min; 95° C, 20 sec; 60° C, 20 sec; 72° C, 20 sec; repeat 30X previous three steps; 72° C, 5 min. Amplified PCR products were visualized by running on a 2% agarose gel and stained with ethidium bromide.

3.2.5 Sanger Sequencing and SNP Validation

Using the Nipponbare reference sequence from Gramene (www.gramene.org), 24 primer pairs were designed using the software Primer 3 (http://frodo.wi.mit.edu/primer3) to flank putative nsSNPs located within 23 CV-selected genes identified with Locus ID in Table B.3a (Appendix B). Sequences of the primers designed and evaluated are shown in Table B.3b (Appendix B). The 24 primer pairs were used to amplify PCR fragments of approximately 500 bp from the SB susceptible line Cocodrie and resistant line CIAT4. Both strands of amplified

fragments were directly sequenced at the Pennington Biomedical Research Institute, Louisiana State University, Baton Rouge, Louisiana. The nsSNPs were identified from sequence data using the ClustalW software (http://www.ebi.ac.uk/Tools/msa/clustalw2). To detect predicted amino acid changes in "resistant" and "susceptible" alleles, nsSNP-containing codons from Sanger sequence data were compared manually with corresponding codons posted at the Gramene website.

To detect candidate nsSNPs in *O. nivara* accessions IRGC 104443 and IRGC 100898, primers were designed and evaluated as described above for 12 CV-selected nsSNPs in genes identified with locus ID given in Table B.4a (Appendix B). Sequences of the primers designed for amplification of PCR fragments containing the nsSNPs are shown in Table B.4b (Appendix B).

3.3 Results

3.3.1 Genomic Variants, Reads, and Coverage

The total number of sequenced read counts, aligned reads %, total number of reads generated, filtered reads, high quality (HQ) reads, and variants detected from HQ reads are shown in Table 3.1. The total number of sequence reads produced across chromosomes before filtering was around 520,000,000 with a range from about 21,000,000 for Bengal to about 92,000,000 for Cocodrie. The average percentage of reads generated across chromosomes that aligned to the Nipponbare reference genome was approximately 72% with a range of about 60% for LaGrue to 77% for Bengal. Moderate variation in the total number of reads and filtered reads was observed for the 13 lines except for the relatively high values of Cocodrie and the low values for Bengal. Variation in the number of HQ reads and variants detected with HQ reads for

this study was found to be consistent with next-generation sequencing of whole genomes in other plant species (Farmer and Woodward, unpublished results). The average coverage across lines of 5.3x generated in this study was nearly identical to that reported for WGS of six maize (*Zea mays* L.) inbred lines (Lai et al., 2010).

Table 3.1 Sequenced read counts, aligned reads %, total reads generated, filtered reads, high quality (HQ) reads, and number of variants with HQ reads for each of 13 rice lines.

Variety	Sequenced read counts[a]	Aligned reads %[b]	Total reads generated[c]	Filtered total reads[d]	Total HQ reads[e]	No. of HQ variants[e]
01. Bengal	20,969,202	76.7	1,660,179	1,595,010	958,078	203,597
02. Bowman	51,256,956	75.8	6,396,817	6,127,689	5,149,205	805,497
03. Cocodrie	92,260,896	75.4	11,430,664	10,966,097	10,062,538	1,091,783
04. Cypress	55,911,024	75.7	5,773,192	5,657,835	4,870,823	793,807
05. Francis	24,893,286	72.3	2,998,710	2,853,346	1,928,852	368,011
06. Jasmine 85	28,749,946	73.2	7,654,829	7,282,676	5,468,966	1,055,316
07. L-201	29,647,596	72.4	3,842,709	3,700,517	2,602,934	506,932
08. LaGrue	52,059,398	61.1	5,051,049	4,933,165	3,975,002	711,918
09. Lemont	37,742,062	74.1	3,598,482	3,494,686	2,568,242	477,985
10. MCR 010277	26,803,094	74.8	5,203,829	4,959,707	3,524,884	693,606
11. Shu-Feng 121-1655	27,684,224	68.4	6,286,567	5,863,943	4,014,601	789,462
12. TeQing	37,330,856	63.8	8,277,487	7,770,577	6,048,175	1,091,937
13. Wells	36,463,096	68.2	4,040,260	3,871,333	2,857,966	530,755
TOTAL	521,771,636	--	72,214,774	69,076,581	54,030,266	9,120,606
AVERAGE	40,136,280	71.7	5,554,983	5,313,583	4,156,174	701,585

[a] Defined as total amount of read counts after initial base calling
[b] Defined as % reads aligned with the Nipponbare reference genome
[c] Defined as reads from sites at which variants were called
[d] Defined as total reads generated having an average quality greater than or equal to 25
[e] HQ=high quality, defined as having an average quality greater than or equal to 25, and reads count greater than or equal to 3

The number of variants with minimum quality scores of 25 and coverage of 5 for each chromosome, and the number of selected nsSNPs and genes are shown in Table 3.2. Across all chromosomes, the total number of unselected variants before filtering relative to the reference Nipponbare varied from about 300 K to 500 K that represented approximately 10% of the total unfiltered reads produced by the Illumina procedure. Only a small reduction in variants with minimum quality was observed, but an approximate 10-fold reduction in those with minimum

coverage relative to the original variants was also found. Wide variation in the number of selected nsSNPs across chromosomes was observed with 1.3 nsSNPs detected on average for each gene.

Table 3.2 Total number of variants for each chromosome with minimum quality, minimum coverage, present in susceptible line, absent in resistant, and present in resistant line, absent in susceptible.

Chromosome	Total number of variants	With minimum quality	With Minimum coverage	Present in susceptible line, absent in resistant[a]	Present in resistant line, absent in susceptible
01	510,984	479,208	34,267	55 (44)	260 (169)
02	436,827	412,025	31,147	80 (38)	1,874 (857)
03	406,883	379,702	29,804	16 (13)	402 (273)
04	435,265	409,133	35,770	67 (51)	344 (200)
05	324,967	303,357	32,730	8 (8)	130 (100)
06	380,628	357,149	32,065	203 (129)	807 (465)
07	353,948	330,366	31,373	3 (3)	22 (8)
08	407,818	384,005	32,323	119 (85)	401 (271)
09	290,054	272,574	20,730	153 (109)	633 (408)
10	341,340	322,393	29,417	0	0
11	434,484	410,715	28,448	101 (67)	1,082 (475)
12	346,137	325,546	23,559	53 (38)	266 (170)

[a] The number in parenthesis is the number of genes

Figure 3.1 shows the percentage of reads and coverage in different combinations for the 13 lines. Variation was observed across lines for depth of coverage and reads where Cocodrie produced the highest percentage of reads ≥ 3 and coverage ≥ 5 while Bengal generated the lowest percentage of all lines. The combination of reads ≥ 3 and coverage ≥ 5 comprised the highest percentage for all lines except for Bengal.

The percentage of all variants within intergenic, untranslated 5', untranslated 3', coding sequence, and intron regions for each chromosome across all 13 lines is shown in Figure 3.2. A large majority of variants (about 60%) were detected within intergenic regions across lines that consisted of both *indica* and *tropical japonica* sub populations.

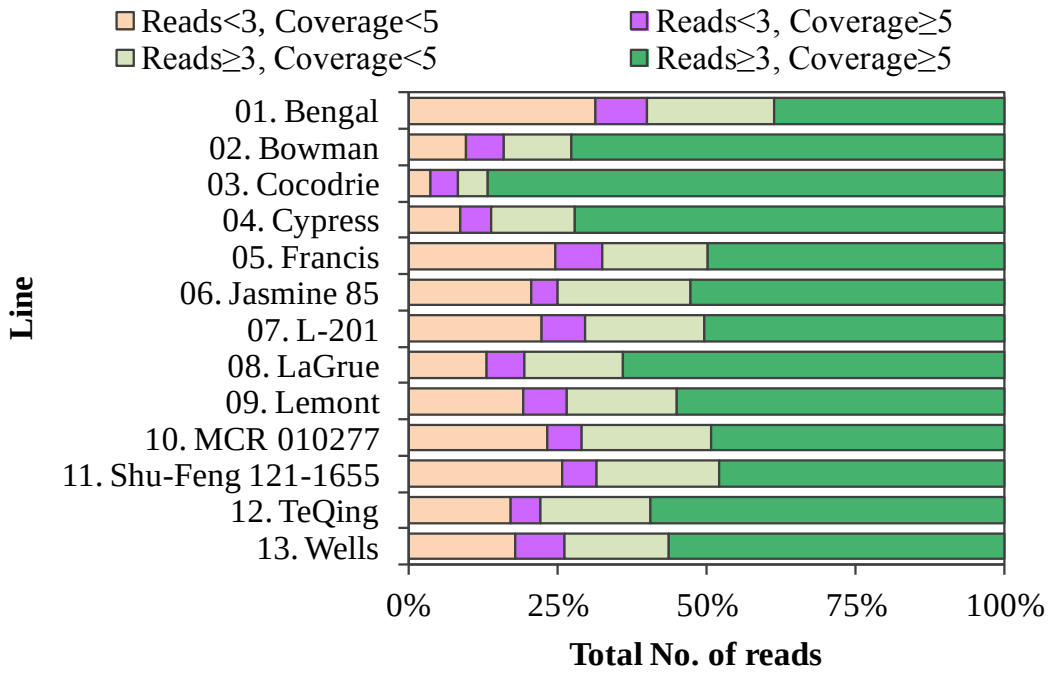

Figure 3.1 Frequency of reads < 3, coverage < 5 (*tan bar*); reads < 3, coverage ≥ 5 (*purple*); reads ≥ 3, coverage < 5 (*light green*); reads ≥ 3, coverage ≥ 5 (*dark green*) from WGS of 13 rice lines.

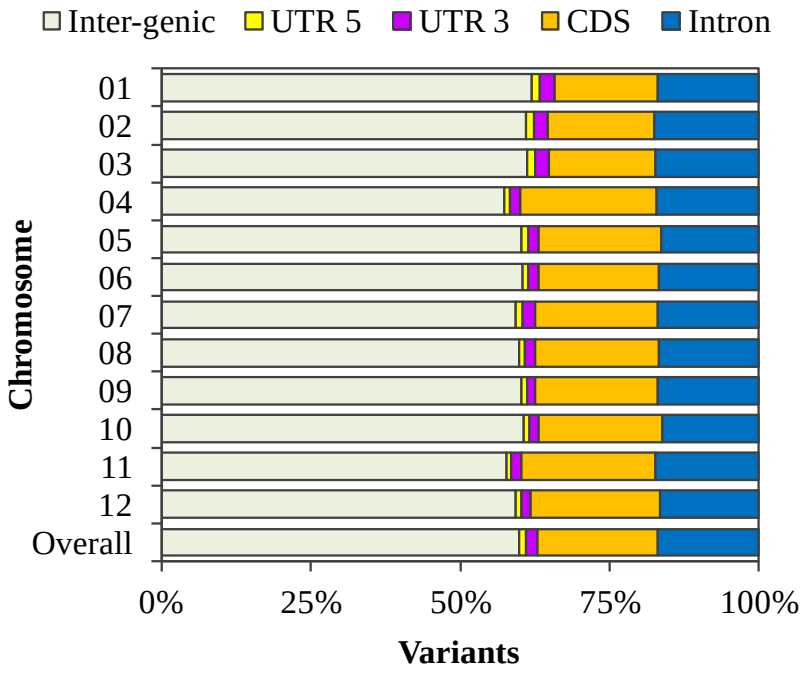

Figure 3.2 Frequency of all variants detected by Alpheus pipeline analysis for intergenic (*light green bar*), untranslated 5'(UTR 5)(*yellow*), untranslated 3' (UTR 3)(*purple*), coding sequencing (CDS)(*orange*), and intron (*blue*) regions within and across 13 rice lines.

Coding sequences and introns shared similar proportions across all lines (about 15% to 20%) while untranslated 5'and untranslated 3' regions comprised a small percentage of the total variants (approximately 1% and 2%, respectively).

Figure 3.3 shows the percentage of variants identified as insertions, deletions, non-synonymous SNPs, and synonymous SNPs for each chromosome across all 13 lines. The most striking result was the large percentage (about 80%) of variants that consisted of synonymous SNPs for both *indica* and *tropical japonica* lines. The second largest class was comprised of nsSNPs, although at a much smaller percentage at approximately 12%, while the remaining variants were made up of insertions and deletions at very low frequencies (about 3% and 4%, respectively).

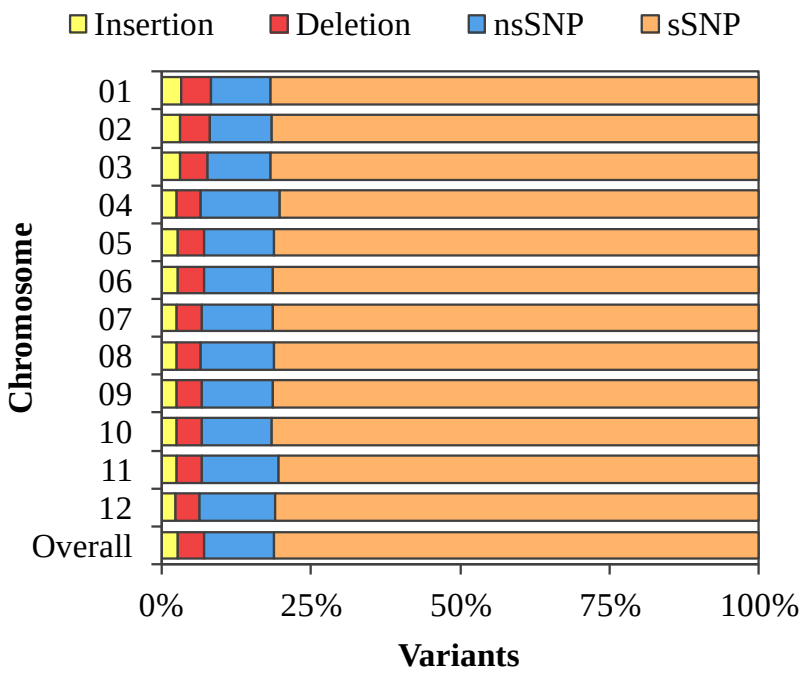

Figure 3.3 Frequency of all insertions (*yellow bar*), deletions (*red*), non-synonymous SNPs (nsSNPs)(*blue*), and synonymous SNPs (sSNPs)(*orange*) detected by Alpheus pipeline analysis for each chromosome across all 13 lines.

3.3.2 Detection of nsSNPs and Candidate Genes inside SB QTL *qShB9-2* on Chromosome 9

qShB9-2, a QTL for sheath blight, was mapped to a region at the bottom of chromosome 9 consisting of approximately 1.2 M bp flanked by SSR markers RM215 and RM245 (Liu et al., 2009). Before the CV selection procedure was carried out, a total of 155 variants were detected within qShB9-2 with 3 or more reads and coverage 5 or more from resistant Jasmine 85, TeQing, and MCR010277. The majority of variants in *qShB9-2* were classified as sSNPs (73%), a substantially smaller percentage as nsSNPs (26%), and the smallest fractions identified were insertions (1.0 %) or deletions (0%). When the CV selection procedure was carried out to identify candidate nsSNPs for SB resistance within *qShB9-2*, relatively few selected nsSNPs (10) were found that mapped throughout most of the QTL (approximately 1.1 M bp). The nsSNPs were detected in a total of 10 genes that were placed into seven groups based on gene ontology/gene function. The physical location of selected nsSNPs within *qShB9-2* along with corresponding genes are shown in Table B.1 (Appendix B).

3.3.3 Detection of nsSNPs, Candidate Genes, and New QTLs outside *qShB9-2*

QTL *qShB9-2* explained approximately 25% of the observed variation for SB resistance when Jasmine 85 was used as the resistant parent (Liu et. al., 2009). Because the majority of variation was detected outside of *qShB9-2,* we scanned all remaining regions of the genome other than this QTL using the CV selection strategy. The selected regions also showed that sSNPs were the most common variant at 78 % while insertions and deletions were rare at 0.35%. As shown in Table B.2 (Appendix B), the distribution of selected nsSNPs and corresponding genes across chromosomes was not uniform. For example, a maximum of 70 nsSNPs and 49 genes were

found on chromosome 2 whereas 0, 7, and 9 nsSNPs and 0, 7, and 2 genes were detected on chromosomes 10, 5, and 7, respectively.

Table B.2 (Appendix B) includes three new candidate QTL regions for SB resistance that have not been reported in the literature. These regions include the top of chromosome 2 (975,892 bp to 6,210,412 bp), the bottom of chromosome 3 (30,523,344 bp to 35,667,086 bp), and the bottom of chromosome 5 (21,585,027 bp to 28,979,361bp). The homologues of certain selected genes within these QTLs have been implicated in stress and disease response of plants and humans. Examples include phytosulfokine receptors (LOC_Os02g06200, LOC_Os02g06210)(Motose et al., 2009), cytokinin-O-glucosyltransferase (LOC_Os02g11130)(Havlova et al., 2008), U5 small nuclear ribonucleoprotein helicase (LOC_Os03g53220)(Hahn and Beggs, 2010), and CCR4-NOT transcription factor (LOC_Os05g40790)(Sarowar et al., 2007). The following three genes on chromosome 5 are reported here for the first time as candidates associated with SB resistance in plants: VHS and GAT domain containing protein (LOC_Os05g39760), kri1 protein (LOC_Os05g41100), and PX domain containing protein (LOC_Os05g50660).

3.3.4 Principal Component-Biplot (PB) Display of Variants on Chromosome 9

The PB selection strategy was conducted across all chromosomes for the 13 lines described above. An important step in this procedure was the construction of a biplot that simultaneously displayed the relationships among variants, relationships among lines, and the underlying interactions between variants and lines (Yan and Tinker, 2006). For ease of visualization, Figure 3.4 shows the biplot of variants on chromosome 9 from the PB selection among four SB resistant (MCR010277, Jasmine 85, TeQing, and Shu-Feng 121-1655), one moderately susceptible (Bengal), and eight highly susceptible lines (Cocodrie, Cypress, Lemont, Bowman, LaGrue,

Francis, L-201, and Wells). Vectors (solid lines) in the biplot showed a clear separation between the four resistant lines and the nine remaining lines. TeQing (TQNG) and MCR010277 (MCR) were found in the same region which was reasonable given that TeQing was one SB resistant parent of MCR010277. Resistant Shu-Feng 121-1655 was found in the same region as TeQing and MCR010277, but its pedigree was not known because the parental line is an undescribed accession from China (www.ars-grin.gov). All nine remaining lines occurred in one large region including Bengal that generated a relatively short vector length compared to the susceptible lines in that region.

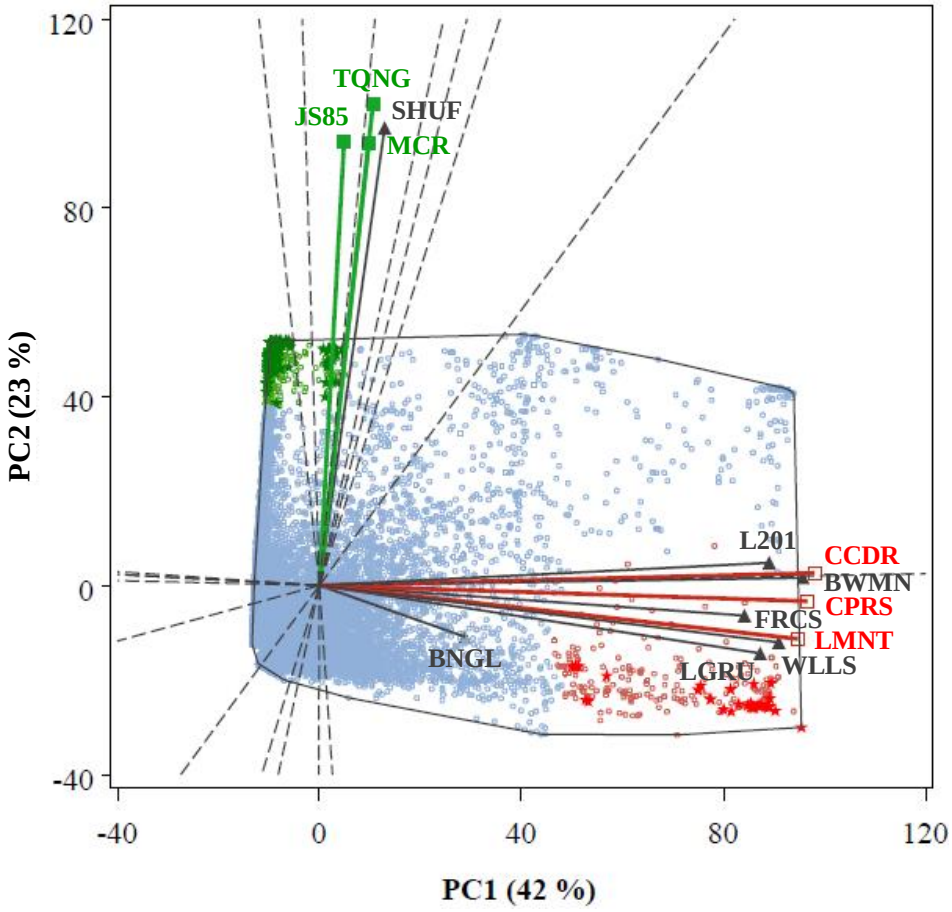

Figure 3.4 Biplot display of all variants on chromosome 9 in four sheath blight resistant lines [MCR010277 (MCR), Jasmine 85 (J85), TeQing (TQNG), and Shu Feng 121-1655 (SHUF)] and nine highly/moderately susceptible lines [Cocodrie (CCDR), Cypress (CPRS), Lemont (LMNT), Bengal (BNGL), Bowman (BWMN), LaGrue (LGRU), Francis (FRCS), L-201 (L201), and Wells (WLLS)].

3.3.5 Grouping of CV-Selected Candidate Genes Based on Gene Family/Gene Ontology

Figure 3.5 shows the groupings of CV-selected candidate genes across all lines and chromosomes based on gene family/gene ontology. A total of 240 genes were assigned to 42 diverse groups with kinase, nucleotide binding, and peptide repeat as the top three with the greatest number of candidate genes. One-half (22/42) of the groups contained only one or two candidate genes.

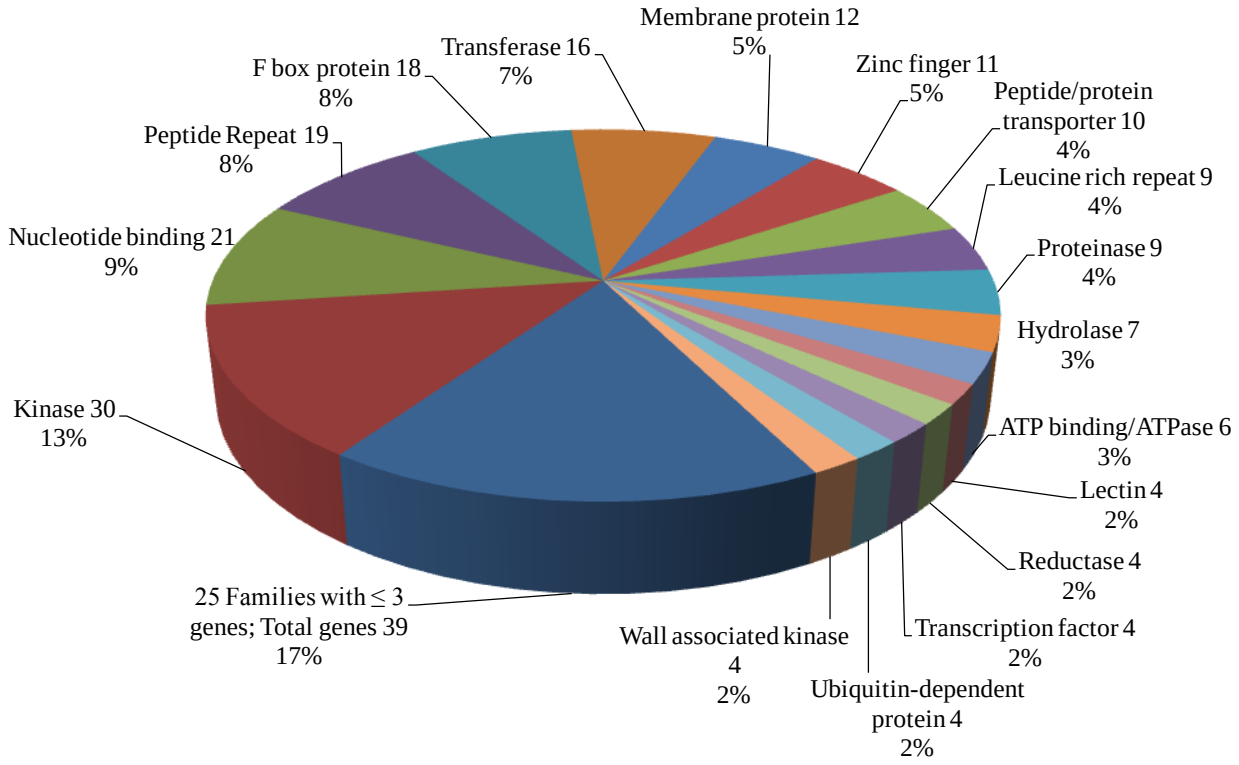

Figure 3.5 Grouping of CV-selected candidate genes based on gene family/gene ontology.

3.3.6 Genotypes of Selected Candidate nsSNPs Evaluated in Different Resistant and Susceptible Lines

Based on selected nsSNPs from the six lines used in the CV selection strategy, we examined nsSNP profiles of the remaining seven lines sequenced by the Illumina method. The

susceptible lines Bowman, Francis, L-201, LaGrue, Leah, and Wells were found with "susceptible" alleles at all loci consistent with susceptible Cocodrie, Cypress, and Lemont. The moderately susceptible Bengal displayed a combination of "susceptible" and "resistant" nsSNPs (results not shown).

A subset of 24 nsSNPs found in 23 randomly-selected candidate genes was selected for further study (Table B.3a, Appendix B). All nsSNPs in this subset were found in dbSNP (posted on Gramene website). All PCR-generated SNP specific alleles for susceptible Nipponbare were consistent with those from the Illumina WGS results and the published Nipponbare reference genome sequence. Susceptible Azucena and Leah produced the same allele profiles as those of Nipponbare. "Resistant" SNP genotypes generated from PCR amplification in resistant MCR010277 and TeQing were in complete agreement for all 23 genes and were consistent with all corresponding genotypes produced by the Illumina GA IIx platform. Profiles for the remaining seven moderately resistant lines varied when compared to MCR010277 and TeQing, ranging from one allele difference in IR64 and Shu Feng 121-1655 to five in Oryzica Llanos 5 and Jhona 349.

The two *O. nivara* accessions, IRGC 104443 and IRGC 100898, along with resistant CIAT 4 and susceptible Catahoula, were screened with 12 random CV-selected nsSNPs (Table B.4a, Appendix B). CIAT 4 produced "resistant" alleles from nine genes (LOC_Os02g19200, LOC_-Os02g54330, LOC_Os02g54500, LOC_Os03g37720, LOC_04g59540, LOC_Os06g28124, LOC_Os06g29700, LOC_Os06g32350, LOC_Os09g37880). Susceptible Catahoula carried only "susceptible" alleles. IRGC 104443 produced "resistant" and "susceptible" alleles at heterozygous loci from two genes on chromosome 2 (LOC_Os02g54330, LOC_Os02g54500) while IRGC 100898 produced one resistant allele on chromosome 4 (LOC_Os04g59540).

We also genotyped eight individuals derived from the moderately resistant Louisiana inbred (F_6) line 09DN/Rush072 with 11 CV-selected nsSNPs chosen at random (data not shown). No individual possessed all 11 "resistant" alleles, although five individuals contained seven resistant alleles from LOC_Os01g52880, LOC_Os02g56380, LOC_Os04g20680, LOC_Os04-g55760, LOC_12g06740, LOC_Os12g09710, and LOC_Os12g10180. All eight individuals carried "susceptible" alleles from four genes on chromosome 9 (LOC_Os09g36900, LOC_Os09-g37590, LOC_Os09g37800, LOC_Os09g37880).

3.3.7 Selection of Variants Using the CV vs. the PB Selection Strategies

The CV strategy for selection of variants in this study was developed as a modification of the approaches used to identify variants for rare human disorders. As shown here, the CV method appears to successfully select candidates associated with SB resistance, but the procedure is somewhat tedious. We therefore developed the PB approach that does involve more steps, but is actually less time consuming and more systematic than the CV method. As part of the PB strategy, the biplot display allows rapid and informative inspection of variant information not possible by other statistical methods. After the CV and PB procedures were completed for QTL *qShB9-2* and the remaining portions of the genome, we found that if a low number of clusters was identified for an individual chromosome, the PB approach selected slightly greater number of variants than the CV method. With high numbers of clusters, both methods were virtually indistinguishable in terms of selected variants.

3.3.8 Sanger Sequencing of Fragments Containing Candidate nsSNPs

Both strands of 12 putative nsSNP-containing fragments from Cocodrie and CIAT4 were sequenced by the Sanger method for the following CV-selected genes: NBS-LRR type disease

resistance protein Rps1-k-2 (LOC_Os12g10180), receptor-like protein kinase 2 (LOC_Os09-g17630), resistance protein (LOC_Os02g35210), OsFBDUF47-F box and DUF domain containing protein (LOC_Os09g37590), receptor protein kinase TMK1 precursor (LOC_Os04-g58910), OsFBDUF14-F-box and DUF domain containing protein (LOC_Os02g54330), leucine-rich repeat family protein (LOC_Os01g52880), NBS-LRR type disease resistance protein Rps1-k-1 (LOC_Os03g37720), phosphatidylinositol-4-phosphate 5-Kinase (LOC_Os04g59540), THION21 - Plant thionin family protein precursor (LOC_Os02g02650), OsFBD11-F-box and FBD domain containing protein (LOC_Os06g29700), and glycosyltransferase (LOC_Os06-g28124). Sanger sequencing results confirmed presence of nsSNPs within all 12 genes (results not shown). In addition, predicted amino acid changes of all nsSNPs were consistent between Sanger and GA IIx sequencing results.

3.4 Discussion

A major rice breeding goal for the southern U.S. is the development of high-yielding cultivars that are resistant to sheath blight, a disease that causes substantial reductions in grain yield and quality in the southern U.S., South America, and Asia. There is currently no resistant U.S. commercial cultivar, primarily due to challenges in selection for quantitative resistance and inconsistencies in phenotyping across years and locations. Several QTLs of small effect have been reported over the years from different studies using Jasmine 85, TeQing, MCR010277, and other lines as sources of resistance. Liu et al. (2009) crossed Jasmine 85 with Lemont to generate a mapping population that showed a QTL at the bottom of chromosome 9 with a “large” effect ($R^2 \approx 0.25$). However, only a modest increase in resistance was observed using three markers within this region for selection in a backcross population (Zuo et al., 2008). This result highlights

the quantitative nature of SB resistance and the need to identify additional markers across the entire genome to assist in development of new cultivars with high levels of resistance.

Several recent studies in humans have shown the potential of WGS to identify variants and genes responsible for rare Mendelian disorders (Rios et al., 2010; Roach et al., 2010; Sobreira et al., 2010; Tong et al., 2010; Lupski et al., 2010). Based on initial success of the human sequencing efforts, we initiated a study to evaluate WGS of rice by the Illumina GA technology to identify candidate nsSNPs that are associated with resistance to sheath blight. An important component of the RiceCAP efforts was to complete WGS of 13 inbred rice lines that have been used in applied breeding of elite U.S. southern cultivars. As shown in Table 3.1, the number of total and high quality variants produced by the Illumina platform differed across the 13 lines, a result that is consistent with other plant species using the Illumina GA IIx technology. The average coverage across lines in our study was nearly identical to that reported for WGS of six maize (*Zea mays* L.) inbred lines (Lai et al., 2010).

Figure 3.2 shows that the majority of variants detected by the Alpheus pipeline for the 13 lines occurred not in the coding sequences, but in the intergenic regions. Therefore, only a small portion of the rice genome from the coding sequences was actually evaluated in this study for candidate variants associated with SB resistance. It is therefore likely that more variants other than the nsSNPs detected in this study could play a role in resistance. Similar conclusions can be drawn from inspection of variant distributions as shown in Figure 3.3 of insertions, deletions, nsSNPs, and sSNPs.

The display in Figure 3.4 shows that Bengal produced a relatively short biplot vector length compared to the remaining eight susceptible lines. One interpretation to account for the

difference is that Bengal, classified as a line of medium grain-length, possesses a different genetic makeup compared to the remaining susceptible long-grain types. A second possibility is that U.S. southern medium grain-length lines such as Bengal generally exhibit slightly higher, and therefore slightly different, levels of SB resistance than most long-grain lines. However, the most likely explanation for the short vector length of Bengal is that it represents a reduced ability to discriminate among variants compared to the remaining susceptible lines. This reduction is consistent with the number of HQ variants generated for Bengal which was the smallest for all 13 lines (Table 3.1). The removal of Bengal variant data should therefore be considered for initial identification of candidate nsSNPs and corresponding genes. This conclusion might not have been possible using only PCA, cluster or other similar statistical approaches, and demonstrates the potential value of biplot display during the variant selection process.

When the CV selection strategy was applied across all lines and chromosomes, a wide array of gene families was identified based on gene ontology/gene function as shown in Figure 3.5. Other than the 25 families grouped together, each with less than four genes, the kinase, nucleotide binding, peptide repeat, and F-box protein categories were the top four that have been detected in several previous investigations of rice and *A. thaliana* resistance (Jwa et al., 2006; Venu et al., 2007; Zhao et al., 2008). Fourteen families, including calcium binding, heat shock, and polygalacturonase, consisted of a single candidate gene. Similar high levels of gene family diversity were also found in previous studies of resistance to the rice blast pathogen *Magnaporthe oryzae* (Vergne et al., 2010), to soybean *Glycine max* L. pathogen *Phytophthora sojae* (Wang et al., 2010), and response of *A. thaliana* to a plant defense elicitor (Libault et al., 2007). Although many of the same gene families were shared, none of the candidate SB resistance genes isolated by our CV or PB selection strategies was identified by suppression

subtractive hybridization or MPSS/SAGE methods in rice (Venu et al., 2007; Zhao et al., 2008). This discrepancy may be explained by the possibility that genes identified in the RNA-based methods contained variants other than nsSNPs.

The QTL *qShB9-2* reported by Liu et al. (2009) represents a region of potential importance for SB resistance breeding because of the relatively large stable effect detected across different greenhouse and field conditions. We therefore decided to identify candidate markers and genes within this region as shown in Table B.1 (Appendix B). Genes homologous to four selected candidates within *qShB9-2* have been implicated previously in resistance to different pathogens. For example, serine/threonine kinases such as those at LOC_Os09g37800 and LOC_Os09-g37880 in the current study have been shown previously to play a role in disease resistance (Afzal et al., 2008). F Box proteins such as OsFBDUF47 at LOC_Os09g37590 were reported to improve disease resistance in tobacco (Cao et al., 2008). Zinc finger proteins like that at LOC_Os09g38970 were reported to be important in resistance signaling in barley (Shirasu et al., 1999). The wall-associated kinase OsWAK91 at LOC_Os09g38850 represents a category found to be associated with resistance to *Pseudomonas syringae* in *A. thaliana* (He et al., 1998). The following five selected candidates have not been reported in the literature to be associated with biotic stress in rice, and therefore represent potential new factors contributing to SB resistance: aspartic proteinase nepenthesin (LOC_Os09g38380), WD domain, G-beta repeat domain containing protein (LOC_Os09g36900), STRUBBELIG-RECEPTOR FAMILY 5 precursor (LOC_Os09g38700), HEAT repeat family protein (LOC_Os09g38710), and potassium transporter (LOC_Os09g38960).

Table B.2 (Appendix B) shows candidate nsSNPs and genes identified by the CV selection strategy outside of *qShB9-2*. A review of the candidates shows that many selections belong to

different plant resistance pathways. For example, various kinases and corresponding receptors like those detected on chromosomes 1 4, 6, 8, 9, and 12 in this study have been reported to play a role in disease resistance. Different F-Box and ubiquitin proteins like those found on chromosomes 2 6, 9, and 12 presumably assist in regulation of the salicylic acid pathway (Llorente et al., 2008). Thionins similar to those on chromosomes 2 and 6 have been shown to accumulate after jasmonic acid induction (Anderson et al., 1992). NB-ARC and leucine rich repeat proteins help modulate R gene-based resistance (Zhang et al., 2003) with homologues in this study detected on chromosomes 1 4, 8, and 12. The pathogenesis-related (PR) protein glucan endo-1,3-beta-glucosidase (glucanase) was found at two loci on chromosomes 8 and 9. Others of interest include a GTPase on chromosome 2, heat shock protein on chromosome 4, a MYB family transcription factor on chromosome 5, a "cell death" protein on chromosome 11, and a RING-H2 finger protein on chromosome 12. Certain selected candidates outside of *qShB9-2* not reported in the literature represent potential new resistance factors. Examples include the rapid alkalinization factor protein (LOC_Os01g10470), cystathionine β-synthase (CBS) domain-containing protein (LOC_Os02g42640), multidrug resistance protein (LOC_Os02g46680), mitochondrial transcription termination factor (LOC_Os02g54200), KIP1 (LOC_Os03g43684), amidase (LOC_Os04g10460), and cadmium tolerance factor (LOC_Os06g19110).

Table B.3a (Appendix B) shows PCR-based SNP allele genotypes from 23 candidate genes of MCR010277 and TeQing originally used in the CV selection process and 10 additional resistant and susceptible lines. Complete agreement observed between PCR and Illumina-generated alleles for all 23 genes of MCR010277 and TeQing suggests that the Illumina platform is suitable for accurate genotyping of rice breeding material. Similarly, all genotypes found for Azucena, Leah, Nipponbare are consistent with their known susceptibility to *R. solani*. The

remaining seven lines showed different combinations of "resistant" and "susceptible" alleles that are in accord with their moderate level of resistance compared to MCR010277 and TeQing. Similar results were obtained for the Louisiana breeding line 09DN/Rush072 (results not shown). Taken together the PCR-based profiles of the 12 lines described here are consistent with corresponding nsSNPs identified from Illumina sequencing of Jasmine 85, TeQing, and MCR010277.

We also examined the possibility that 12 selected nsSNPs were present in two resistant accessions of *O. nivara* and a SB resistant *indica* line from South America. Two *O. nivara* accessions contained three resistant nsSNPs that were common with the nine resistant nsSNPs from the *indica* resistant line. These results suggest that sources of SB resistance do occur in related species other than *O. sativa* consistent with previous work of Prasad and Eizenga (2008). Channamallikarjuna et al. (2010) identified a stable SB QTL on chromosome 11 from the cultivar Tetep that contained 11 xylanase inhibitor genes presumably functioning as class III chitinases. No such genes were found on chromosome 11 in our study which suggests that additional candidates may occur in other SB resistant sources. It is interesting that a xylanase inhibitor gene was identified in this study on chromosome 9.

Our study was the first to use WGS to identify candidate rice genes associated with SB resistance. The outcome from this investigation suggests that WGS may be a useful strategy to identify candidate variants associated with other rice diseases that can complement QTL mapping and microarray/transcriptome approaches. Several new candidate QTLs and genes were identified in our study that warrant further investigation. Moreover, SNP profiles detected in the original three resistant lines were found to be consistent with additional resistant/tolerant material. This information may prove valuable in development of marker assisted breeding for

SB resistance. Proof that these selected genes actually play a role in resistance will require transgene over-expression and/or knock-out experiments.

The candidate markers and genes identified in this study appear promising, but it is important to state that the WGS approach used in our research very likely did not detect all genes associated with SB resistance for the following reasons: (1) The majority of variants detected in the initial screening were not nsSNPs. Additional research will be required to determine any potential role(s) of sSNPs and other variants in SB resistance (2) The Illumina GA IIx and other similar sequencing platforms generate data from short DNA fragments that cannot readily identify large deletions, insertions, or copy number variants (3) Additional accessions, lines or *Oryza sp*. may contain different alleles or genes not evaluated in this study. Nevertheless, the output generated from this study should provide new information for future basic and applied research of SB resistance in rice.

3.6 References

Afzal A.J., Wood A.J., Lightfoot D.A. (2008) "Plant receptor-like serine threonine kinases: roles in signaling and plant defense." *Molecular Plant Microbe Interactions* **21**, 507–517.

Andresen I., Becker W., Schlüter K., Burges J,, Parthier B., Apel K. (1992) "The identification of leaf thionin as one of the main jasmonate-induced proteins of barley (*Hordeum vulgare*)." *Plant Molecular Biology* **19**, 193–204.

Bilgüvar K., Oztürk A.K., Louvi A., Kwan K.Y., Choi M., Tatli B., Yalnizoğlu D., Tüysüz B., Cağlayan A.O., Gökben S., Kaymakçalan H., Barak T., Bakircioğlu M., Yasuno K., Ho W., Sanders S., Zhu Y., Yilmaz S., Dinçer A., Johnson M.H., Bronen R.A., Koçer N., Per H., Mane S., Pamir M.N., Yalçinkaya C., Kumandaş S., Topçu M., Ozmen M., Sestan N., Lifton R.P., State M.W., Günel M. (2010) "Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations." *Nature* **467**, 207–210.

Cao Y., Li Z., Chen T., Zhang Z., Zhang J., Chen S. (2008) "Overexpression of a tobacco small G protein gene NtRop1 causes salt sensitivity and hydrogen peroxide production in transgenic plants." *Science in China Series C: Life Sciences* **51**, 383–390.

Channamallikarjuna V., Sonah H., Prasad M., Rao G.J.N., Chand S., Upreti H.C., Singh N.K., Sharma T.R. (2010) "Identification of major quantitative trait loci qSBR11–1 for sheath blight resistance in rice." *Molecular Breeding* **25**, 155–166.

Deschamps S., La Rota M., Ratashak J.P., Biddle P., Thureen D.A., Luck S., Beatty M., Nagasawam N., Michael L., Llaca V., Sakai H., May G., Lightner J., Campbell M.A. (2010) "Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina Genome Analyzer." *The Plant Genome* **3**, 53–68.

Everitt B.S., Landau S., Leese M. (2001) *Cluster Analysis, 4th Edition*. Oxford: Oxford University Press.

Fu Y.B., Peterson G.W., McCallum B.D., Huang L. (2010) "Population-based resequencing analysis of improved wheat germplasm at wheat leaf rust resistance locus Lr21." *Theoretical and Applied Genetics* **121**, 271–281.

Hahn D., Beggs J.D. (2010) "Brr2p RNA helicase with a split personality: insights into structure and function." *Biochemical Society Transactions* **38**, 1105–1109.

Havlova M., Dobrev P.I., Motyka V., Storchová H., Libus J., Dobrá J., Malbeck J., Gaudinová A., Vanková R. (2008) "The role of cytokinins in responses to water deficit in tobacco plants over-expressing trans-zeatin O-glucosyltransferase gene under 35S or SAG12 promoters." *Plant, Cell and Environment* **31**, 341–353.

He H.H., He D., Kohorn B.D. (1998) "Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response." *The Plant Journal* **14**, 55–63.

Hobert O. (2010) "The impact of whole genome sequencing on model system genetics: get ready for the ride." *Genetics* **184**, 317–319.

Jwa N.S., Agrawal G.K., Tamogami S., Yonekura M., Han O., Iwahashi H., Rakwal R. (2006) "Role of defense/stress-related marker genes, proteins and secondary metabolites in defining rice self-defense mechanisms." *Plant Physiology and Biochemistry* **44**, 261–273.

Kim J.K., Jang I.C., Wu R., Zuo W.N., Boston R.S., Lee Y.H., Ahn I.P., Nahm B.H. (2003) "Co-expression of a modified maize ribosome-inactivating protein and a rice basic chitinase gene in transgenic rice plants confers enhanced resistance to sheath blight." *Transgenic Research* **12**, 475–484.

Krawitz P.M., Schweiger M.R., Rödelsperger C., Marcelis C., Kölsch U., Meisel C., Stephani F., Kinoshita T., Murakami Y., Bauer S., Isau M., Fischer A., Dahl A., Kerick M., Hecht J., Köhler S., Jäger M., Grünhagen J., de Condor B.J., Doelken S., Brunner H.G., Meinecke P., Passarge E., Thompson M.D., Cole D.E., Horn D., Roscioli T., Mundlos S., Robinson P.N. (2010) "Identity-by-descent filtering of exome sequence data identifies PIGV

mutations in hyperphosphatasia mental retardation syndrome." *Nature Genetics* **2**, 827–829.

Lai J., Li R., Xu X., Jin W., Xu M., Zhao H., Xiang Z., Song W., Ying K., Zhang M., Jiao Y., Ni P., Zhang J., Li D., Guo X., Ye K., Jian M., Wang B., Zheng H., Liang H., Zhang X., Wang S., Chen S., Li J., Fu Y., Springer N.M., Yang H., Wang J., Dai J., Schnable P.S., Wang J. (2010) "Genome-wide patterns of genetic variation among elite maize inbred lines." *Nature Genetics* **42**, 1027–1030.

Lee F.N., Rush M.C. (1983) "Rice sheath blight: a major rice disease." *Plant Disease* **67**, 829–832.

Li J.F., Li L., Sheen J. (2010) "Protocol: a rapid and economical procedure for purification of plasmid or plant DNA with diverse applications in plant biology." *Plant Methods* **6**, 1 doi:10.1186/1746-4811-6-1.

Libault M., Wan J., Czechowski T., Udvardi M., Stacey G. (2007) "Identification of 118 *Arabidopsis* transcription factors and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor." *Molecualr Plant Microbe Interactions* **20**, 900–911.

Ling K.S., Harris K.R., Meyer J.D., Levi A., Guner N., Wehner T.C., Bendahmane A., Havey M.J. (2009) "Non-synonymous single nucleotide polymorphisms in the watermelon eIF4E gene are closely associated with resistance to zucchini yellow mosaic virus." *Theoretical and Applied Genetics* **120**, 191–200.

Liu M., Sun Z.X., Zhu J., Xu T., Harman G.E., Lorito M.. (2004) "Enhancing rice resistance to fungal pathogens by transformation with cell wall degrading enzyme genes from *Trichoderma atroviride.*" *Journal of Zhejian University SCIENCE* **5**, 133–136.

Liu G., Jia Y., Correa-Victoria F.J., Prado G.A., Yeater K.M., McClung A., Correll J.C. (2009) "Mapping quantitative trait loci responsible for resistance to sheath blight in rice." *Phytopathology* **99**, 1078–1084.

Llorente F., Muskett P., Sánchez-Vallet A., López G., Ramos B., Sánchez-Rodríguez C., Jordá L., Parker J., Molina A. (2008) "Repression of the auxin response pathway increases *Arabidopsis* susceptibility to necrotrophic fungi." *Molecular Plant* **1**, 496–509.

Lupski J.R., Reid J.G., Gonzaga-Jauregui C., Deiros D.R., Chen D.C.Y., Nazareth L., Bainbridge M., Dinh H., Jing C., Wheeler D.A., McGuire A.L., Zhang F., Stankiewicz P., Halperin J.J., Yang C., Gehman C., Guo D., Irikat R.K., Tom W., Fantin N.J., Muzny D.M., Gibbs R.A. (2010) "Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy." *The New England Journal of Medicine* **362**, 1181–1191.

Maruthasalam S., Kalpana K., Kumar K.K., Loganathan M., Poovannan K., Raja J.A., Kokiladevi E., Samiyappan R., Sudhakar D., Balasubramanian P. (2007) "Pyramiding

transgenic resistance in elite *indica* rice cultivars against the sheath blight and bacterial blight." *Plant Cell Reports* **26**, 791–804.

McNally K.L., Childs K.L., Bohnert R., Davidson R.M., Zhao K., Ulata V.J., Zeller G., Clark R.M., Hoen D.R., Bureau T.E., Stokowski R., Ballinger D.G., Frazer K.A., Cox D.R., Padhukasahasram B., Bustamante C.D., Weigel D., Mackill D.J., Bruskiewich R.M., Ratsch G., Buell C.R., Leung H., Leach J.E. (2009) "Genomewide SNP variation reveals relationships among landraces and modern varieties of rice." *Proceedings of the National Academy of Sciences of the United States of America* **28**, 12273–12278.

Miller N.A., Kingsmore S.F., Farmer A., Langley R.J., Mudge J., Crow J.A., Gonzalez A.J., Schilkey F.D., Kim R.J., van Velkinburgh J., May G.D., Black C..F, Myers M.K., Utsey J.P., Frost N.S., Sugarbaker D.J., Bueno R., Gullans S.R., Baxter S.M., Day S.W., Retzel E.F. (2008) "Management of high-throughput DNA sequencing projects: Alpheus." *Journal of Computer Science and Systems Biology* **1**, 132–148.

Motose H., Iwamoto K., Endo S., Demura T., Sakagami Y., Matsubayashi Y., Moore K.L., Fukuda H. (2009) "Involvement of phytosulfokine in the attenuation of stress response during the transdifferentiation of Zinnia mesophyll cells into tracheary elements." *Plant Physiology* **150**, 437–447.

Ng S.B., Buckingham K.J., Lee C., Bigham A.W., Tabor H.K., Dent K.M., Huff C.D., Shannon P.T., Jabs E.W., Nickerson D.A., Shendure J., Bamshad M.J. (2010a) "Exome sequencing identifies the cause of a Mendelian disorder." *Nature Genetics* **42**, 30–35.

Ng S.B., Bigham A.W., Buckingham K.J., Hannibal M.C., McMillin M.J., Gildersleeve H.I., Beck A.E., Tabor H.K., Cooper G.M., Mefford H.C., Lee C., Turner E.H., Smith J.D., Rieder M.J., Yoshiura K., Matsumoto N., Ohta T., Niikawa N., Nickerson D.A., Bamshad M.J., Shendure J. (2010b) "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome." *Nature Genetics* **42**,790–793.

Pinson S.R.M., Capdevielle F.M., Oard J.H. (2005) "Confirming QTLs and finding additional loci conditioning sheath blight resistance in rice using recombinant inbred lines." *Crop Science* **45**, 503–510.

Prasad B., Eizenga G.C. (2008) "Rice sheath blight disease identified in *Oryza* spp. accessions." *Plant Disease* **92**, 1503–1509.

Rios J., Stein E., Shendure J., Hobbs H.H., Cohen J.C. (2010) "Identification by whole genome resequencing of gene defect responsible for severe hypercholesterolemia." *Human Molecular Genetics* **19**, 4313–4318.

Roach J.C., Glusman G., Smit A.F.A., Huff C.D., Hubley R., Shannon R.T., Rowen L., Pant K.P., Goodman N., Bamshad M., Shendure J., Drmanac R., Jorde L.B., Hood L., Galas D.J. (2010) **"**Analysis of genetic inheritance in a family quartet by whole-genome sequencing." *Science* **328**, 636–639.

Rush M.C., Lindberg G.D. (1996) “Rice disease research.” *Rice Journal* **77**, 49–52.

Sarowar S., Oh H.W., Cho H.S., Baek K.H., Seong E.S., Joung Y.H., Choi G.J., Lee S., Choi D. (2007) “Capsicum annuum CCR4-associated factor CaCAF1 is necessary for plant development and defense response.” *Plant Journal* **51**, 792–802.

Shah J.M., Raghupathy V., Veluthambi K. (2009) “Enhanced sheath blight resistance in transgenic rice expressing an endochitinase gene from Trichoderma virens.” *Biotechnology Letters* **31**, 239–244.

Shirasu K., Lahaye T., Tan M.W., Zhou F., Azevedo C., Schulze-Lefert P. (1999) “A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*.” *Cell* **99**, 355–366.

Sobreira N.L.M., Cirulli E.T., Avramopoulos D., Wohler E., Oswald G.L., Stevens E.L., Ge D., Shianna K.V., Smith J.P., Maia J.M., Gumbs C.E., Pevsner J., Thomas G., Valle D., Hoover-Fong J.E., Goldstein D.B. (2010) “Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene.” *PLoS Genetics* **6**(6), e1000991. doi:10.1371/journal.pgen.1000991.

Song W.Y., Wang G.L., Chen L.L., Kim H.S., Pi L.Y., Holsten T., Gardner J., Wang B., Zhai W.X., Zhu L.H., Fauquet C., Ronald P. (1995) “A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21.” *Science* **270**, 1804–1806.

Tan C.X., Ji X.M., Yang Y., Pan X.Y., Zuo S.M., Zhang Y.F., Zou J.H., Chen Z.X., Zhu L.H., Pan X.B. (2005) “Identification and marker-assisted selection of two major quantitative genes controlling rice sheath blight resistance in backcross generations.” *Chinese Journal of Genetics and Genomics* **32**, 399–405.

Teer J.K., Mullikin J.C. (2010) “Exome sequencing: the sweet spot before whole genomes.” *Human Molecular Genetics* **19**, 145–151.

Tong P., Prendergast J.G.D., Lohan A.J., Farrington S.M., Cronin S.C., Friel N., Bradley D.G., Hardiman O., Evans A., Wilson J.F., Loftus B.J. (2010) “Sequencing and analysis of an Irish human genome.” *Genome Biology* **11**, R91 doi:10.1186/gb-2010-11-9-r91.

Venu R.C., Jia Y., Gowda M., Jia M., Jantasuriyarat C., Stahlberg E., Li H., Rhineheart A., Boddhireddy P., Singh P., Rutger N., Kudrna D., Wing R., Nelson J.C., Wang G.L. (2007) “RL-SAGE and microarray analysis of the rice transcriptome after *Rhizoctonia solani* infection.” *Molecular Genetics and Genomics* **278**, 421–431.

Vergne E., Grand X., Ballini E., Chalvon V., Saindrenan P., Tharreau D., Nottéghem J.L., Morel J.B. (2010) “Preformed expression of defense is a hallmark of partial resistance to rice blast fungal pathogen *Magnaporthe oryzae*.” *BMC Plant Biology* **10**, 206 doi:10.1186/1471-2229-10-206.

Walsh T., Shahin H., Elkan-Miller T., Lee M.K., Thornton A.M., Roeb W., Abu Rayyan A., Loulus S., Avraham K.B., King M.C., Kanaan M. (2010) "Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82." *The American Journal of Human Genetics* **87**, 90–94.

Wang H., Waller L., Tripathy S., St. Martin S.K., Zhou L., Krampis K., Tucker D.M., Mao Y., Hoeschele I., Maroof M.A.S., Tyler B.M., Dorrance A.E. (2010) **"**Analysis of genes underlying soybean quantitative trait loci conferring partial resistance to *Phytophthora sojae*." *The Plant Genome* **3**, 23–40.

Xie W., Feng Q., Huihui Y., Xuehui H., ZhaoQ., Xing Y., Yu S., Han B., Zhang Q. (2010) "Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing." *Proceedings of the National Academy of Sciences of the United States of America* **107**, 10578–10583.

Xu J., Zhao Q., Du P., Xu C., Wang B., Feng Q., Liu Q., Tang S., Gu M., Han B., Liang G. (2010) "Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome re-sequencing in rice (*Oryza sativa* L.)." *BMC Genomics* **11**, 656.

Yan W., Tinker N.A. (2006) "Biplot analysis of multi-environment trial data: principles and applications." *Canadian Journal of Plant Science* **86**, 623–645.

Yi X., Liang Y., Huerta-Sanchez E., Jin X., Cuo Z.X., Pool J.E., Xu X., Jiang H., Vinckenbosch N., Korneliussen T.S., Zheng H., Liu T., He W., Li K., Luo R., Nie X, Wu H., Zhao M., Cao H., Zou J., Shan Y., Li S., Yang Q., Asan., Ni P., Tian G., Xu J., Liu X., Jiang T., Wu R., Zhou G., Tang M., Qin J., Wang T., Feng S., Li G., Huasang, Luosang J., Wang W., Chen F., Wang Y., Zheng X., Li Z., Bianba Z., Yang G., Wang X., Tang S., Gao G., Chen Y., Luo Z., Gusang L., Cao Z., Zhang Q., Ouyang W., Ren X., Liang H., Zheng H., Huang Y., Li J., Bolund L., Kristiansen K., Li Y., Zhang Y., Zhang X., Li R., Li S., Yang H., Nielsen R., Wang J., Wang J. (2010) "Sequencing of 50 human exomes reveals adaptation to high altitude." *Science* **329**, 75–78.

Zhang Y., Goritschnig S., Dong X., Li X. (2003) "A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1." Plant Cell **15**, 2636–2646.

Zhao C.J., Wang A.R., Shi Y.J., Wang L.Q., Liu W.D., Wang Z.H., Lu G.D. (2008) "Identification of defense-related genes in rice responding to challenge by *Rhizoctonia solani*." *Theoretical and Applied Genetics* **116**, 501–516.

Zuo S., Zhang L., Wang H., Yin Y., Zhang Y., Chen Z., Ma Y., Pan X. (2008) "Prospect of the QTL-qSB-9Tq utilized in molecular breeding program of *japonica* rice against sheath blight." *Journal of Genetics and Genomics* **35**, 499–505.

CHAPTER 4 SUMMARY AND CONCLUSIONS

4.1 Association Mapping of Grain Quality and Flowering Time

The level of variation among rice lines across all locations and traits evaluated in the study summarized in Appendix A demonstrated that using elite inbred tropical *japonica* germplasm for association mapping can generate sizeable phenotypic variation with "acceptable" commercial levels such that important contributions toward the success of commercial rice markets can be achieved. Although for all three traits mean values approached acceptable commercial levels, relatively wide ranges were observed at each site, and consequently the estimated low values for heritabilities indicated that non-genetic sources (e.g., location, environment, and location-by-environment) contributed to the expression of the traits. Population structure analyses confirmed that the inbred lines used represented a single genetic collection, while association mapping revealed up to 30 effects (mainly epistatic) significantly contributing to the expression of each trait. A small proportion of the total variation was accounted for by selected effects, a result that was consistent with the low broad-sense heritability estimates and the complex nature of these traits.

Results suggest that association mapping analysis for complex agronomic traits should consider gene-gene interactions. Although several selected alleles for each trait mapped either within or near previously reported QTL, several loci were reported here for the first time, representing new genetic regions associated with these three important agronomic characters. Selected loci (e.g., allele RM190_122 inside the *Waxy* locus) were also found to be associated with more than one trait, suggesting pleiotropic effects. New information on the genetic components of grain quality and flowering time in *japonica* rice has been provided, contributing

to the development of effective breeding strategies for the improvement of cooking quality and whole-grain rice yields.

4.2 GGE Biplot Exploration of Resistance to Sheath Blight

Sheath blight (SB), caused by the fungus *Rhizoctonia solani*, is a major foliar disease of southern U.S. rice that is difficult to evaluate under field conditions due to large GxE effects. The observed SB scores within and across years and locations suggest that the DH lines used in this study exhibited similar levels of disease severity, but the GE effect was highly significant. The estimated heritability values, however, were relatively high in all four environments in LA and AR.

Three HA-GGE biplot analyses were used to identify and characterize GE interactions. The "which-won-were" of the environment-focused HA-GGE biplot analysis (Figure 2.2) explained ~ 84% of the total variability and was used to identify a single meaningful mega-environment with high genetic correlations among environments. An alternative view of the environment-focused HA-GGE biplot analysis, the "test environment evaluation" (Figure 2.3), allowed the four environments to be ranked as: AR07 < LA07 < AR06 < LA06. A third HA-GGE biplot (Figure 2.4) was constructed to conduct genotype evaluation and allowed the identification of stable-resistant and stable-susceptible DH lines.

Assessment of the genotype evaluation HA-GGE biplot using 95% confidence regions for PC scores showed clear separation and non-overlapping between subsets of susceptible and resistant DH lines, providing consistency for our results. Support for the HA-GGE biplot analysis of performance and stability of DH lines was also provided by median polish values and Hühn's nonparametric stability statistics. My results indentified 11 stable-resistance DH lines

that can be used as potential parents for breeding purposes. Five susceptible DH lines were identified as potential checks with greater stability than the susceptible parents. Performance and stability of all selected material warrant additional testing in southern U.S. rice-growing regions for future development of elite, sheath blight-resistant cultivars.

4.3 Identification of Candidate Genes for Resistance to Sheath Blight

An important contribution of the USDA-funded RiceCAP efforts was the generation of WGS of 13 inbred rice lines used in applied breeding of elite U.S. Inspired by the initial success of the human sequencing efforts, we used WGS of the 13 rice lines by the Illumina GA technology to identify candidate nsSNPs that are associated with resistance to sheath blight. In agreement with other plant species using the Illumina GA IIx technology, the number of total and high quality variants produced by the Illumina platform differed across the 13 lines. Most of the variants detected by the Alpheus pipeline occurred inside the intergenic regions indicating that only a small portion of the rice genome from the coding sequences was actually evaluated in this study. Therefore, it is likely that other variants than the nsSNPs detected in this study could play a role in resistance to SB.

The two strategies applied to perform identification of variants associated with SB resistance showed consistent results and allow the confirmation of the role of several previously reported genes in disease resistance as well as the identification of new candidate genes. Although a wide array of gene families was identified based on gene ontology/gene function, we decided to identify candidate markers for genes inside the *qShB9-2* on chromosome 9 because its relatively large-stable effect has been validated across different locations and environments.

This study was the first to use WGS to identify candidate rice genes for SB resistance, but WGS may be a useful strategy to identify candidate variants associated with other traits. Although proof that the selected genes are actually involved in SB resistance requires additional experimentation, we provided valuable information for the development of marker assisted breeding not only for SB resistance, but for future basic and applied research for other traits of recognized economical importance for worldwide rice markets.

4.4 Significance and Impact of the Results

Under the constraints established by the limited financial resources of small-sized rice breeding programs, the number of selected SSR molecular markers (in Appendix A) may seem too large for future practical applications. Inclusion of additional SSR molecular markers and relevant epistatic effects, however, will improve the statistical power of the fitted models, which translates into a more accurate and effective implementation of marker-assisted selection strategies.

The use of GGE biplots (in Chapter 2) has proven to be a simple and practical way for identifying SB-resistant commercial rice lines with high levels of stability, significantly contributing to the breeding priorities not only for the U.S., but for other regions of the world. The usefulness of GGE biplot representation was also demonstrated in Chapter 3 where summary and interpretation of interactions between rice varieties and genomic variants allowed selection of candidate genes for a complex and important disease of the U.S. rice industry.

Analogous to the research achievements from studies to identify disease susceptibility in humans, exploitation of WGS of rice inbred lines was extremely useful to identify candidate genes for SB resistance. The outcomes described in Chapter 3 suggest that the use of WGS

information may overcome the resolution limitations of other technologies (e.g., SSR markers). Implementing our filtering strategies using WGS may be useful for identifying candidate genomic variants associated with other traits of similar importance for world-wide rice markets, significantly impacting research efforts of rice breeding programs around the world.

APPENDIX A ASSOCIATION MAPPING OF GRAIN QUALITY AND FLOWERING TIME IN ELITE *japonica* RICE GERMPLASM

A.1 Introduction

Rice grain quality traits command worldwide attention not only from consumers, but they also play a crucial economic role for millers, manufacturers of conventional and convenience foods, and exporters to international markets. Appearance, eating, cooking, and milling qualities comprise the primary components of rice grain quality. Factors such as grain shape and translucency of the endosperm have been shown to impact quality of appearance (Unnevehr et al., 1992; Juliano and Villareal, 1993).

One of the most important determinants of eating and cooking quality is apparent amylose content reported in several studies to be governed primarily by the *Waxy* (Wx) locus on chromosome 6 (Hao et al., 2009; Kepiro et al., 2008; Aluko et al., 2004; Septiningsih et al., 2003; Tan et al. 1999). However, other studies have shown that amylose content is a quantitative trait governed by additional QTL of minor effect at various chromosomal locations (Aluko et al., 2004; Wan et al., 2004; Tan et al., 2001). Separate QTL studies by Fan et al. (2005) and Wan et al. (2004) demonstrated that environment was a major source of variation for amylose content while epistasis played a minor role. Zhou et al. (2003) showed that marker-based strategies can be exploited to enhance eating characteristics of hybrid rice.

Milling quality is typically assessed as brown rice percentage, milled rice percentage, and head-milled rice. Grain from which the bran has not been removed is considered brown rice, and the combined whole and broken rice grains with the bran removed constitute milled rice. Whole grain or head rice is defined as the proportion of whole kernels that also includes broken kernels

[1] Reprinted by permission of "*Theoretical and Applied Genetics*"

75% to 80% of the whole rice grain. It is well known that head rice is an important component for establishing market value, and the most important characteristic of overall milling quality.

Several QTL mapping studies for head rice yield have been reported within the last ten years (Tan et al., 2001; Septiningsih et al., 2003; Aluko et al., 2004; Zheng et al., 2007; Kepiro et al., 2008; Hao et al., 2009; Lou et al., 2009). Most studies were carried out at a single location in a single year due to the time-consuming task of evaluating lines with multi-step procedures in replicated field plots and laboratory analyses. Numerous QTL were identified each with small effects across different chromosomal regions. A few investigations also detected QTL with major effects when evaluated at individual locations (Aluko et al., 2004; Kepiro et al., 2008; Lou et al., 2009). Four studies identified two-way QTL interactions on six chromosomes (Tan et al 2001; Septiningsih et al. 2003; Aluko et al. 2004; Lou et al., 2009). In a multi-environment trial by Lou et al. (2009) genotype-by-environment interaction was not significant for head rice and two-way interactions (epistasis) produced only a minor effect.

The time of flowering or heading date is considered a crucial factor not only for quantity, but also for quality of rice grain production (Fan et al., 2005). Tabien et al. (2009) found that the rate and duration of flowering influenced grain quality among elite *japonica* inbred lines. More than 100 QTLs associated with heading date have been identified primarily from *O. sativa* x *O. indica* crosses (www.gramene.org). Certain QTLs were recently shown to be directly involved in complex interactions for heading date and/or photoperiodic responses. For example, a major QTL at the top of chromosome 6 designated Hd1 was shown to promote flowering under short-day conditions and inhibit heading in long-day environments (Yano et al., 2000). Hd1 was later found to regulate the action of a second major QTL under short-day conditions on chromosome 6

designated Hd3a (Kojima et al., 2002). A recent study suggested that expression of Hd3a was also impacted by variation in temperature and day-length (Luan et al., 2009).

Nearly all QTL studies for grain quality to date have utilized predominantly *indica*, temperate *japonica, O. rufipogon*, or *O. glaberrima* sources as parents to develop various mapping populations. Kepiro et al. (2008) recently detected QTLs on chromosome 6 for apparent amylose content and head rice in a tropical *japonica* long-grain cross evaluated at a single location. However, information is lacking on candidate DNA markers and their potential interactions associated with grain quality traits at multiple locations for lowland tropical *japonica* germplasm.

QTL mapping that evaluates progeny from bi-parental crosses has been the conventional approach to identify chromosomal regions associated with grain quality. Association mapping is an alternative approach that captures multiple historical recombination events among a selected panel or population of unrelated inbred individuals (Myles et al., 2009). The principal advantage of this strategy is that use of unrelated inbred lines is amenable to rapid evaluation at multiple locations since development of specific mapping populations is not required, thus saving time, money, and labor. Various candidate markers associated with agronomic traits in rice, and other cereals have been reported recently by this method (Breseghello et el., 2006; Wilson et al., 2004; Zhang et al., 2005). The objective of our research was to identify candidate marker effects associated with two grain quality and one flowering trait by association mapping in a collection of elite tropical *japonica* lines evaluated at five different locations.

A.2 Materials and Methods

A.2.1 Plant Material and Field Evaluation of Traits

Phenotypic data for this study were obtained from the University of Arkansas Rice Research and Extension Center, Stuttgart, Arkansas. A collection of 192 elite rice breeding lines and varieties representing a narrow tropical *japonica* germplasm base were evaluated by public rice breeders in replicated field plot trials in 2000 at Crowley (Louisiana), Beaumont (Texas), Stuttgart (Arkansas), Stoneville (Mississippi), and Cape Girardeau (Missouri). The germplasm consisted of 52 lines from Arkansas, one from California, 55 from Louisiana, 25 from Mississippi, and 58 lines from Texas. Based on grain length, 161 were long grain types, 26 were medium grain, and five were short grain. All 192 inbred lines were planted from March to April, 2000 in each of the five states listed 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 carried out to minimize weed and insect damage for maximum grain yield. The center four rows of each plot were used to collect data for heading date (days from seedling emergence to panicle emergence from swollen stem or boot), and percent head rice (whole grains/(whole grains + broken grains) x 100). Grain samples for each line were dried to about 12% moisture in Texas and Arkansas and sent to the USDA-Beaumont grain quality laboratory to determine apparent amylose content. Phenotypic data expressed as trait means across replications at each location were obtained to compute means and variances using Proc Tabulate (SAS Institute Inc., 2006). Phenotypic outliers were identified by Proc Univariate and replaced with imputed values using the TASSEL software program (www.maizegenetics.net, v. 2.1). TASSEL was also used to impute missing values for apparent amylose content (5 out of 192 for Arkansas, and 2 out of 192 for Texas).

A.2.2 Marker Genotyping

Microsatellite (SSR) marker data for the 192 lines were obtained from Dr. Thomas Tai, USDA-ARS, UC-Davis, Davis, California. A total of 97 SSR markers, evenly spaced over the 12 chromosomes at approximately 20 cM intervals, generated a total of 579 alleles with an average of six alleles/locus. Rare alleles at less than seven percent frequency were removed to provide 194 marker alleles at 97 bi-allelic loci for the final analysis. Missing marker data (1.8% of total) were estimated using the SAS Multiple Imputation Procedure (SAS Institute Inc., 2009).

A.2.3 Statistical Analyses and Association Mapping

Data for the three traits were averaged across replications within each location to compute means and variances using PROC TABULATE (SAS Institute Inc., 2006). Data for trait variation at each location were not available. Therefore, an Analysis of Variance (ANOVA) in the form of a Tukey test for non-additivity (Tukey, 1949) was conducted to test for genotype-by-location interaction of the form $(\alpha\beta_{ij}) = D\alpha_i\beta_j$ in the model

$$y_{ij} = \mu + \alpha_i + \beta_j + D\alpha_i\beta_j + \varepsilon_{ij}$$

where μ is the overall mean effect, α_i is the main fixed effect of rice line i, β_j is the main fixed effect of location j, and D is a constant fitted to the data. The test for $H_0 : D = 0$ is equivalent to the test for interaction, so if the hypothesis is accepted, the additive model can be assumed to be reasonable, and lines would respond in a similar manner across different locations (SAS Institute Inc., 1991).

The Pearson's product-moment correlation coefficients among phenotypic traits were obtained for each location using PROC CORR (SAS Institute Inc., 2006). This analysis was also

repeated according to the rice lines classified as long, medium, or short grain length. Broad-sense heritability estimates (h^2) were calculated using the TASSEL software. Detection of potential population structure was carried out with the model-based "Structure" software program, v. 2 (http://pritch.bsd.uchicago.edu/structure.html). A burn in of 5,000,000 and a run length of 2,000,000 were carried out. A total of 2 to 7 *K* clusters were evaluated with LnP(*D*) probabilities used to detect putative subpopulations. In addition, a genetic distance-based procedure based on Ward's hierarchical clustering of the 192 lines with all 579 marker alleles was performed in PROC CLUSTER (SAS Institute Inc., 2009). Estimates of kinship relationships between lines were carried out with the TASSEL software program using the *K* model with phenotypic data and marker effects selected at $p = 0.15$. Note that the "*Q*" population structure component was not included in the TASSEL analysis since neither the STRUCTURE nor the Ward's program described above found evidence of sub-population clustering in this elite germplasm.

For each trait-location combination, the selected marker effects from the TASSEL analysis were used to fit a multiple linear regression model with all possible pair-wise interaction terms using the procedure GLMSELECT of SAS. To reduce multicollinearity issues, a variable selection procedure was incorporated using the STEPWISE option in the model statement, with the PRESS sub-option (equivalent to the leave-one-out cross validation procedure) used as a stopping criterion. The final specification of the multiple linear regression model having *k* selected regressors was as follows

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon_i; \text{ with } \varepsilon_i \sim Normal(0, \sigma^2), i = 1,2,...,192$$

where y_i was the phenotypic response of the line *i*, β_0 was the intercept, $(\beta_1, \beta_2, ..., \beta_k)$ were the regression coefficients associated with the selected regressors $(x_1, x_2, ..., x_k)$, and ε_i was a random error term. Note that the x_m's were indicator variables where

a selected main marker effect $x_m = \begin{Bmatrix} 1 \text{if line } i \text{ has the marker} \\ 0 \text{ otherwise} \end{Bmatrix}$

a selected epistatic effect $x_m = \begin{Bmatrix} 1 \text{if line } i \text{ has both markers} \\ 0 \text{ otherwise} \end{Bmatrix}$

The procedure calculated the least square estimates of the regression coefficients corresponding to each selected regressor, as well as a *P*-value associated with the null hypothesis that the regression coefficient was zero. The False Discovery Rate (FDR) was calculated for the selected effects using SAS PROC MULTTEST with *P*-value < 0.05 as threshold for marker-trait associations. Marker class values of lines carrying alleles of main and interaction effects were calculated from SAS GLMSLECT. For the top five selected effects, marker class values of lines carrying alleles of main and interaction effects were compared using SAS PROC GLM. SAS PROC ALLELE was used to estimate polymorphism information content (PIC) and allelic diversity (SAS Institute Inc., 2008). SSR markers were placed on the genetic map in Figure A.1 of this study based on the Cornell 2001 mapping population that consisted of 96 doubled-haploid progeny from the *indica* (IR64) x tropical *japonica* (Azucena) cross (www.gramene.org).

A.3 Results

A.3.1 Trait Means, Correlations, and Heritabilities

The imputed mean, range, and heritability estimates for the three traits investigated in this study are shown in Table A.1. Substantial variation was observed among the rice lines evaluated for all three characters at each location. For example, while apparent amylose content mean values approached “acceptable” commercial levels within states, a relatively wide range in values about 15% was also observed at each site. These values demonstrated that sizeable phenotypic variation can be generated with elite inbred tropical *japonica* germplasm for

association mapping and other studies. The mean values of apparent amylose content for long and medium grain lines within each location were not statistically different as judged by *t*-tests, so values across grain type were combined. Heritability estimates for apparent amylose content were consistent within locations, but were surprisingly low in this study, given that a major QTL at the *Waxy* locus is considered to play an important role in expression of this character. The heritability values suggest that non-genetic sources such as location and/or environment contributed to expression and production of apparent amylose content in this *japonica* germplasm.

Table A.1 Mean, range, and heritability estimates for apparent amylose content (AC), heading date (HD), and head rice (HR) among 192 lines evaluated in Arkansas (AR), Louisiana (LA), Missouri (MO), Mississippi (MS), and Texas (TX).

Trait	Location	Mean ± SD[a]	Range	Heritability[b]
AC	AR	20.03 ± 2.61	14.00–25.50	0.42
	TX	20.15 ± 1.15	15.70–23.80	0.47
HD	AR	83.62 ± 4.10	73.00–95.00	0.18
	LA	87.01 ± 3.70	76.50–97.50	0.35
	MO	91.47 ± 3.80	82.50–105.50	0.15
	MS	83.51 ± 3.72	74.00–93.50	0.29
	TX	80.04 ± 3.68	70.00–95.00	0.28
HR	AR	47.12 ± 10.76	17.68–66.86	0.27
	LA	65.50 ± 3.28	56.75–72.35	0.23
	MS	48.72 ± 6.46	31.30–63.10	0.28
	TX	53.63 ± 4.50	42.10–62.20	0.34

[a] Standard deviation of the mean
[b] Broad-sense heritability calculated from the TASSEL software program

Mean values for days to heading exhibited moderate variation of 11 days that was expected given the geographical location between the most northern location in Missouri and those of Louisiana and Texas. The Missouri location required the longest average heading time, but the range in heading was extensive within locations at nearly three weeks. Heritability for heading date was low in all states, especially for the Arkansas and Mississippi locations.

The mean values of head rice for long and medium grain lines were not statistically different within each location, so values across grain type were combined. Overall mean head rice values fell essentially into two groups. The first consisted of Louisiana head rice production that was within "acceptable" values 0.60 or more for commercial standards. The second group consisted of head rice produced from the remaining locations that was substantially lower than that of Louisiana. The range of head rice values was extensive, with the greatest observed for the Arkansas site, and the smallest detected for Louisiana. Heritability values were low at each location that underscored complex genetic control and large environmental influences that affect this important grain quality trait.

When values were averaged across the 192 lines within each location, correlations between the traits were either weak or nonexistent. For example, apparent amylose content was weakly correlated with heading date in Arkansas ($r = 0.226$, $P < 0.005$) and Texas ($r = 0.175$, $P < 0.05$). Heading date and head rice showed a weak negative correlation in Louisiana ($r = –0.176$, $P < 0.05$). All other correlations within each location were not statistically significant. Similar results were obtained when accounting for grain length.

A.3.2 Analysis of Variance of Traits

The analysis of variance for apparent amylose content, heading date, and head rice is shown in Table A.2. The results indicated that genotype and location were significant sources of observed variation except for location of apparent amylose content. The non-additivity or genotype-by-location source was also significant in all cases except for heading date, even though location was the predominant source of variation for this trait. Location was the most important source of variation observed for head rice. These results, consistent with the low heritability values shown in Table A.1, indicate that location and its interaction with genotypes should be considered when conducting association mapping of grain quality and flowering traits.

Table A.2 Analysis of variance of apparent amylose content, heading date and head rice based on adjusted mean values averaged within each location using a fixed effect, general linear model.

Trait	Source	df	Sum of squares	Mean square	*F* value	*P*-value
Apparent amylose content	Genotype	191	1161.99	6.08	7.35	< 0.0001
	Location	1	2.13	2.13	2.57	0.1104
	Non-additivity[a]	1	238.7	238.7	288.18	< 0.0001
	Error	190	157.37	0.83		
	Total	383	1560.19			
Heading date	Genotype	191	11362.8	59.49	18.60	< 0.0001
	Location	4	14314.3	3578.6	1118.75	<0.0001
	Non-additivity	1	6.33	6.33	1.98	0.1599
	Error	763	2440.64	3.2		
	Total	959	28124.1			
Head rice	Genotype	191	17435.7	91.29	3.75	< 0.0001
	Location	3	39802	13267	544.86	< 0.0001
	Non-additivity	1	4646.03	4646	190.8	< 0.0001
	Error	572	13928.2	24.35		
	Total	767	75811.9			

[a] Equivalent to genotype-by-location interaction as defined by Tukey's test for non-additivity

A.3.3 Marker Analysis and Population Structure

The average PIC value across the bi-allelic dataset used for this study was 0.37 while allelic diversity varied moderately from 0.37 to 0.50. When the model-based "Structure" program was implemented, no population stratification was detected. Specifically, no peak values for the K = 2 to 7 LnP(D) probabilities were observed during the analysis. This result is in agreement with the known pedigrees of the 192 lines that consisted almost exclusively of tropical *japonica*, one of the five major subpopulations previously identified in rice (Garris et al., 2005). Moreover, the Ward's clustering results confirmed that the inbred lines chosen for this study represented a single genetic group or collection (results not shown). We also accounted for pair-wise kinship relationships of the inbred lines using the TASSEL program. Wide variation was observed in the percentage of lines with detected kinship relationships (results not shown).

A.3.4 Association Mapping

Association mapping carried out in this study revealed the identification of up to 30 effects associated for each trait at FDR values less than 0.05 (results not shown). For simplicity the top effects for each trait in terms of explained variation, FDR values, and significant marker class differences by Tukey's test are shown in Table A.3. A striking result was that the marker variables associated with all three traits were comprised in almost all cases of two-way interactions. These interactions consisted in certain instances of one allele that mapped within previously reported QTLs and a second allele reported here for the first time (Table A.3, Figure A.1). One example involved the RM437_274 allele for heading date in Arkansas that mapped within a reported QTL on chromosome 5 (www.gramene.org) and its interaction with RM317_161 on chromosome 4 identified in this study.

Table A.3 Summary statistics for top main and two-way interaction alleles identified by association mapping for amylose content, heading date, and head rice at Arkansas (AR), Louisiana (LA), Missouri (MO), Mississippi (MS), and Texas (TX) locations.

Traits and Locations	Allele 1	Chr.	Allele 2	Chr.	*P*- value	FDR[b]	Adj. R^2	LS Estimate[c]	Marker Class 0[d]	Marker Class 1[e]
Amylose content										
AR	RM190 122	6	RM5752 12	7	0.0001	0.0003	0.434	-3.05	20.32	15.63
	RM459 060	5	RM202 176	11	0.0094	0.0134	0.187	-1.63	20.43	16.98
	RM1167 17	1	RM116 279	11	0.0001	0.0003	0.049	-5.20	20.45	18.07
	RM459 060	5	RM3430 21	6	0.0001	0.0003	0.021	-4.50	20.11	14.73
	RM5752 12	7	RM435 163	6	0.0001	0.0003	0.017	-3.28	20.10	16.29
TX	RM190 122	6			0.0110	0.0128	0.132	-0.51	20.35	19.27
	RM149 241	8	RM316 212	9	0.0002	0.0007	0.034	0.72	19.99	20.99
	RM459 060	5	RM5 114	1	0.0029	0.0068	0.025	-1.32	20.18	19.11
Heading date										
AR	RM317 161	4	RM437 274	5	0.0079	0.0153	0.056	2.38	83.24	86.57
	RM279 164	2	RM132 080	3	0.0001	0.0006	0.053	9.04	83.49	91.87
	RM5 114	1	RM459 064	5	0.0001	0.0006	0.053	2.79	82.37	84.32
	RM459 064	5	RM144 253	11	0.0001	0.0006	0.044	1.61	83.17	86.26
LA	RM190 122	6	RM144 256	11	0.0001	0.0002	0.201	-3.61	87.71	83.04
	RM279 164	2	RM132 080	3	0.0001	0.0002	0.098	9.85	86.85	97.00
	RM293 198	3	RM408 127	8	0.0455	0.0455	0.052	-1.16	87.37	84.33
	RM478 212	7	RM3912 19	9	0.0001	0.0002	0.049	-2.84	87.35	84.38
	RM486 097	1	RM433 221	8	0.0001	0.0002	0.042	-2.47	87.44	84.91
MO	RM190 122	6			0.0002	0.0005	0.127	-2.44	92.13	88.61
	RM144 253	11	RM437 274	5	0.0001	0.0005	0.068	4.23	91.17	95.60
	RM475 185	2	RM55 227	3	0.0004	0.0008	0.028	-10.62	91.52	82.50
MS	RM184 204	10	RM420 186	7	0.0001	0.0002	0.139	5.67	80.77	84.23
	RM420 186	7	RM190 122	6	0.0001	0.0002	0.104	-5.32	83.93	80.11
	RM132 080	3	RM431 254	1	0.0001	0.0002	0.054	9.84	83.38	91.67
	RM317 161	4	RM232 157	3	0.0003	0.0006	0.041	1.72	83.05	86.08
	RM184 215	10	RM2 164	7	0.0157	0.0196	0.039	-3.44	83.64	75.25
TX	RM3430 21	6	RM433 221	8	0.0001	0.0004	0.087	-1.10	80.37	76.13
	RM184 215		RM408 127	8	0.0001	0.0004	0.079	-3.67	80.29	75.44
	RM478 212	7	RM231 181	3	0.003	0.0048	0.052	-1.54	80.40	78.22
Head rice										
AR	RM315 132	1	OSR13 098	3	0.0001	0.0003	0.129	-13.12	50.00	41.76
	RM498 211	2	RM435 167	6	0.0001	0.0003	0.074	12.29	46.31	53.68
	RM475 199	2	RM408 127	8	0.0001	0.0003	0.047	4.86	43.78	48.20
	RM333 165	10	RM338 179	3	0.0001	0.0003	0.048	-7.22	48.65	45.19
LA	RM312 094	1	RM190 122	6	0.0049	0.011	0.121	1.66	64.96	68.00
	RM109 095	2	RM5 114	1	0.0001	0.0005	0.092	2.32	65.15	67.06
	RM431 250	1	RM5 114	1	0.0001	0.0005	0.068	-1.98	65.88	64.21
	RM1189 18	9	RM5 114	1	0.0001	0.0005	0.045	2.53	65.13	66.82
	RM341 142	2	RM312 094	1	0.0002	0.0007	0.043	1.48	64.38	66.35
MS	RM181 239	11	RM475 199	2	0.0001	0.0002	0.13	3.67	47.31	52.65
	RM437 274	5	RM104 222	1	0.0001	0.0002	0.082	-4.06	50.27	45.57
	RM341 136	2	RM106 293	2	0.0001	0.0002	0.062	-3.48	49.86	44.89
	RM234 141	7	RM315 137	1	0.0004	0.0008	0.021	6.62	48.43	54.70
	RM403 239	1	RM340 114	6	0.0001	0.0002	0.021	9.02	48.48	55.06
TX	RM418 298	7	RM296 119	9	0.0001	0.0003	0.138	3.58	51.77	55.17
	RM1167 17	1	RM206 131	11	0.0001	0.0003	0.064	2.68	51.63	54.89
	RM315 137	1	RM481 156	7	0.0005	0.0011	0.053	1.82	52.64	56.37
	RM408 127	8	OSR13 098	3	0.0001	0.0003	0.012	2.32	51.86	54.10

[a] Allele designation in SSR marker and allele size in base pairs

[b] False Discovery Rate with *P*- value < 0.05 as threshold for marker-trait association

[c] Least square estimate of phenotypic value obtained from GLMSELECT multiple regression model used in this study

[d] Marker class 0 value = phenotypic value of lines not carrying allele of main effect or both alleles of interaction effect

[e] Marker class 1 value = phenotypic value of lines carrying allele of main effect or both alleles of interaction effect

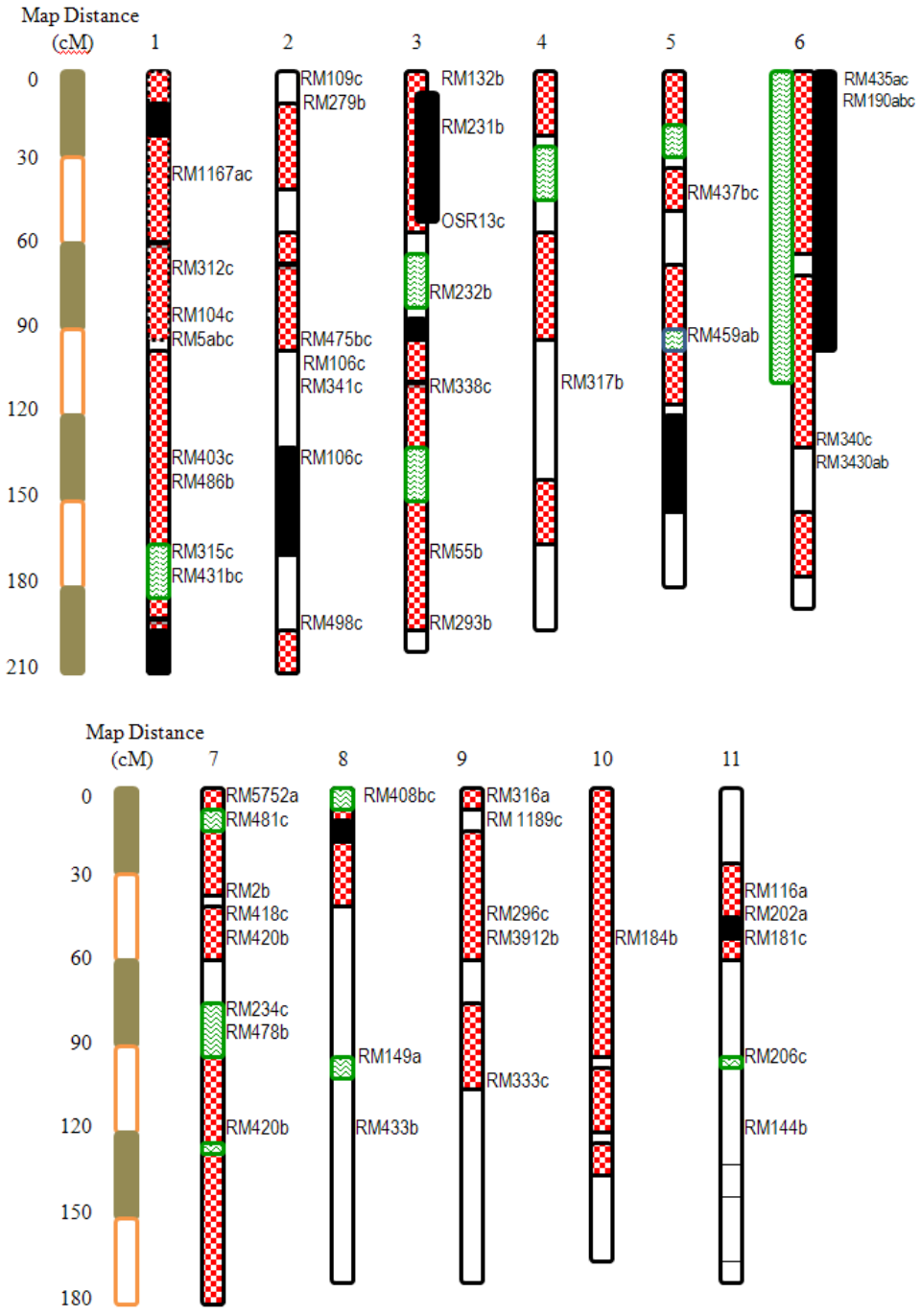

Figure A.1 Chromosomal locations (top: 1–6, bottom: 7–11) of selected loci as main or components of interaction effects associated with apparent amylose content (a), heading date (b), and head rice (c). Green stippled, red checkered, and black boxes represent QTLs detected in previous research for apparent amylose content, heading date, and head rice, respectively. Markers placed on map were based on Cornell 2001 SSR mapping population (www.gramene.org).

In other instances such as RM418_298 * RM296_119 for head rice in Texas, both alleles represent new interacting candidate markers for this important trait. The RM190_122 allele was found as a single main effect associated with apparent amylose content in Texas and heading date in Missouri and Mississippi.

The vast majority of selected effects explained only a small proportion of observed variation based on adjusted R^2 values (Table A.3). These results were in accordance with low broad-sense heritability estimates of the traits shown in Table A.2. Given the complex nature of these traits, the results were not unexpected. The sole exception was the RM190_122 * RM5752_176 interaction that explained 43% of variation for apparent amylose content at the Arkansas location.

The selected effects generally consisted of two-way interactions formed by unique combinations of alleles specific for each location (Table A.3). One exception was RM279_164 * RM132_080 observed for heading date at both the Arkansas and Louisiana locations. In addition, certain individual loci, as components of two-way interactions, were found associated with a trait at more than one location. For example, the RM190 and RM459 loci were common at both Arkansas and Texas for apparent amylose content, and a total of 11 loci associated with heading date were found at more than one location. These loci included RM317, RM437, RM279, RM132, RM190, RM144, RM408, RM478, RM433, RM184, and RM420. The following five loci were detected for head rice at more than one location: RM315, OSR13, RM408, RM475, and RM341. In some cases, different alleles were associated with a given trait at two or three locations. The RM144_253 allele was identified for heading date in Arkansas and Missouri while the RM144_256 allele was detected in Louisiana. The RM341_142 and RM341_136 alleles were associated with head rice in Louisiana and Mississippi, respectively.

We found that a few selected alleles were associated with more than one trait (Table A.3, Figure A.1). The RM190_122 allele was associated with apparent amylose content in Arkansas, and Texas, for heading date in Louisiana, Missouri, Mississippi, and with head rice in Louisiana. Another example was RM5_114 selected for apparent amylose content in Texas, heading date in Arkansas and head rice in Louisiana.

Least squares estimates of the regression of identified main or epistatic marker classes were calculated and shown in Table A.3. Both positive and negative values were found for estimates of a given trait. In four cases though, estimates were either all positive or all negative. Each estimate for apparent amylose content in Arkansas and heading date in Texas was negative while values for heading date in Arkansas and head rice in Texas were positive.

The selected candidate markers and two-way interactions shown in Table A.3 were further evaluated for significant phenotypic differences between marker classes. Marker class zero defined in this study consisted of the phenotypic value of lines not carrying an allele of a main effect or not carrying one or both alleles of a selected interaction effect. Marker class one was defined as the phenotypic value of lines carrying an allele of a main effect or both alleles of a selected interaction effect. The difference between marker classes was consistent in both direction and size when compared with the least square estimate for each effect. The presence of both alleles in each two-way interaction was associated with a reduction in apparent amylose content in both Arkansas and Texas. The reduction was substantially greater in Arkansas (2% to 5%) compared to selected effects in Texas of only one percent.

For heading date the majority of selected effects were observed with either an increase or decrease of approximately two to four days as shown in Table A.3. Certain interactions were also

associated with large increases or decreases in heading date of corresponding marker classes. For example the RM279_164 * RM132_080 interaction was associated with 8 to 10-day increase in heading date in both Arkansas and Louisiana. In addition the RM132_080 * RM431_254 effect was associated with an 8-day increase in heading date at the Mississippi location. On the other hand, RM475_185 *RM55_227 was found at the Missouri location with a reduction in heading date of 9 days.

Table A.3 and Figure A.1 show that the majority (13/18, 72%) of selected effects for head rice mapped across several chromosomes and were associated with an increase in value for this important quality trait. While modest increases were observed for effects at the Louisiana and Texas locations, relatively large increases of about five to seven percent were found with RM498_211 * RM435_167 in Arkansas and RM181_239 * RM475_199 in Missouri. The largest reduction in head rice of eight percent was detected with the RM315_132 * OSR_09 in Arkansas. The effects identified for Louisiana were somewhat unique in that the majority of alleles mapped to chromosome 1 with three different interactions consisting of the RM5_114 allele.

A.4 Discussion

Relatively large amounts of phenotypic variability were measured for all three traits within each of the five locations. The high level of observed variation suggests that this collection of elite inbred *japonica* lines was appropriate to conduct association studies for grain quality and flowering traits. Non-genetic factors such as location and/or environment were most likely important in contributing to the low broad-sense heritability estimates obtained in our study. Indeed, the ANOVA indicated that location and/or its interaction with genotype were important

sources of variability for the three characters. Similar results were found for amylose content in a mapping study of *indica* rice (Fan et al., 2005) and with a field evaluation of 171 accessions of diverse origin (Chen et al., 2008). Location was found to play a major role for heading date among doubled-haploid progeny from an *indica* x *japonica* cross (Li et al., 2003).

Both the model-based and genetic distance-based measures of population structure (STRUCURE, Ward's) indicated that the panel of elite inbred lines belonged to the same population cluster or group. This result is not unexpected as the lines were derived from a narrow elite germplasm base of tropical *japonica*. On the other hand, kinship relationships were detected between the lines, so the "*K* model" in the TASSEL software program was implemented for association mapping to reduce spurious associations and Type I error.

The vast majority of effects associated with the traits in this study consisted of two-way interactions. This suggests that genetic factors affected the action of other loci to generate most of the variation observed in this study. Epistasis has been previously reported to impact grain quality and flowering in rice (Tan et al 2001; Septiningsih et al. 2003; Aluko et al. 2004; Fan et al., 2005; Lou et al., 2009). Our results, similar to these studies, suggest that gene-gene interactions should be considered for association mapping and even QTL studies for complex agronomic traits as discussed by Li et al. (2003). Several selected alleles for each trait, as components of two-way interactions, mapped either within or near previously reported QTL as shown in Figure A.1 and at the Gramene website. This suggests that the selected effects in this study may be also associated with grain quality and flowering in other populations, although separate validation is required. Our study also identified various loci reported here for the first time that represent new genetic regions associated with these three important agronomic characters.

The small size of the selected effects in our study emphasizes the quantitative nature of genomic regions associated with grain quality as reported in previous research (Tan et al., 2001; Septiningsih et al., 2003; Aluko et al., 2004; Fan et al., 2005; Kepiro et al., 2008; Lou et al., 2009). Our study detected specificity of selected interactions, but common loci at more than one location interacting with specific alleles was also observed. Selected loci were also found to be associated with more than one trait. The most prominent example is the RM190_122 allele associated with all three traits. RM190 is known to occur in the *Waxy* locus that affects amylose content, but this marker also mapped approximately 5 cM from the Hd3a locus reported as a major factor in flowering time (Tamaki et al., 2007). These results suggest that genetic factors for heading date at or near the Hd3a region interact with loci identified in this study to affect cooking and milling quality at two (Arkansas, Louisiana) of the four locations. In summary, our study has confirmed previous reports and provided new information on the genetic components of grain quality in *japonica* rice and their mode of interaction with the environment. This information should also help develop effective breeding strategies for the improvement of cooking quality and whole-grain rice yields.

A.5 References

Aluko G., Martinez C., Tohme J., Castano C., Bergman C., Oard J.H. (2004). "QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* x *O. glaberrima*." *Theoretical and Applied Genetics* **109**, 630–639.

Breseghello F., Sorrells M.E. (2006). "Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars." *Genetics* **172**, 1165–1177.

Chen M.H., Bergman C., Pinson S., Fjellstrom R. (2008). "*Waxy* gene haplotypes: associations with apparent amylose content and the effect by the environment in an international rice germplasm collection." *Journal of Cereal Science* **47**, 536–545.

Fan C.C., Yu X.Q., Xing Y.Z., Xu C.G., Luo L.J., 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 population. *Theoretical and Applied Genetics* **110**, 1445–1452.

Garris A.J., Tai T.H., Coburn J., Kresovich S., McCouch S. (2005). "Genetic structure and diversity in *Oryza sativa* L." *Genetics* **169**, 1631–1638.

Hao W., Zhu M-Z., Gao J-P., Dun S-Y., Lin H-X. (2009). "Identification of quantitative trait loci for rice quality in a population of chromosome segment substitution lines." *Journal of Integrative Plant Biology* **51**, 500–512.

Iwata H., Ebana K., Fukuoka S., Jannink J.L. Hayashi T. (2009). "Bayesian multilocus mapping on ordinal and censored traits and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms." *Theoretical and Applied Genetics* **118**, 865–880.

Juliano B.O., Villareal C.P. (1993). "Grain quality evaluation of world rices." International Rice Research Institute, Manila.

Kepiro J.L., McClung A.M., Chen M.H., Yeater K.M., Fjellstrom R.G. (2008). "Mapping QTLs for milling yield and grain characteristics in a tropical *japonica* long grain cross." *Journal of Cereal Science* **48**, 477–485.

Kojima S., Takahashi Y., Kobayashi Y., Monna L., Sasaki T., Araki T., Yano M. (2002). "Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions." *Plant Cell Physiology* **43**, 1096–1105.

Li Z.K., Yu S.B., Lafitte H.R., Huang N., Courtois B., Hittalmani S., Vijayakumar C.H., Liu G.F., Wang G.C., Shashidhar H.E., Zhuang J.Y., Zheng K.L., Singh V.P., Sidhu J.S., Srivantaneeyakul S., Khush G.S. (2003). "QTL x environment interactions in rice. I. heading date and plant height." *Theoretical and Applied Genetics* **108**, 141–153.

Lou J., Chen L., Yue G.Y., Lou Q., Mei H., Xiong L. (2009) "QTL mapping of grain quality traits in rice. *Journal of Cereal Science* **50**, 145–151.

Luan W., Chen H., Fu Y., Si H., Peng W., Song S., Liu W., Hu G., Sun Z., Xie D., Sun C. (2009). "The effect of the crosstalk between photoperiod and temperature on the heading-date in rice." *PLos ONE* **4(6)**: e5891. Doi;10.1371/journal.pone0005891.

Myles S., Peiffer J., Brown P., Ersoz E., Zhang Z., Costich D.E., Buckler E.S. (2009). "Association mapping: critical considerations shift from genotyping to experimental design." *The Plant Cell* **21**, 2194–2202.

SAS Institute Inc. (1991). *SAS® System for Statistical Graphics, First Edition.* Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2006). *Base SAS® 9.1.3 Procedures Guide, Second Edition, Volumes 1, 2, 3, and 4.* Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2008). *SAS/Genetics*^{TM} *9.2User's Guide.* Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2009). *SAS/STAT ® 9.2 User's Guide, Second Edition*. Cary, NC: SAS Institute Inc.

Septiningsih E.M., Trijatmiko K.R., Moeljopawiro S., McCouch S.R. (2003). "Identification of quantitative trait loci for grain quality in and advanced backcross population derived from the Oryza sativa variety IR64 and the wild relative *O. rufipogon*." *Theoretical and Applied Genetics* **107**, 1433–1441.

Tabien R.E., Samonte S.O.P.B., Tiongco E.R. (2009). "Relationship of milled grain percentages and flowering-related traits in rice." *Journal of Cereal Science* **49**, 122–127.

Tamaki S., Matsuo S., Wong H.L., Yokoi S., Shimamoto K. (2007). "Hd3a protein is a mobile flowering signal in rice." *Science* **316**, 1033–1036.

Tan Y.F., Li J.X., Yu S.B., Xing Y.Z., Xu C.G., Zhang Q. (1999) "Three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63." *Theoretical and Applied Genetics* **99**, 642–648.

Tan Y.F., Sun M., Xing Y.Z., Hua J.P., Sun X.L., Zhang Q.F. 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." *Theoretical and Applied Genetics* **103**, 1037–1045.

Tukey J.W. (1949). "One degree of freedom for non-additivity." *Biometrics* **5**, 232–42.

Unnevehr L.J., Duff B., Juliano B.O. (1992). "Consumer demand for rice grain quality." International Rice Research Institute, Manila, and International Development Research Center, Ottawa.

Wan X.Y., Wan J.M., Su C.C., Wang C.M., Shen W.B., Li J.M., Wang H.L., Jiang L., Liu S.J., Chen L.M., Yasui H., Yoshimura A. (2004). *Theoretical and Applied Genetics* **110**, 71–79.

Wilson L.M., Whitt S.R., Ibanez A.M., Rocheford T.R., Goodman M.M., Buckler E.S. (2004). "Dissection of maize kernel composition and starch production by candidate gene association." *Plant Cell* **16**, 2719–2733.

Yano M., Katayose Y., Ashikari M., Yamanouchi U., Monna L., Fuse T., Baba T., Yamamoto K., Umehara Y., Nagamura Y., Sasaki T. (2000). "Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS." *Plant Cell* **12**, 2473–2484.

Zhang N., Xu Y., Akash M., McCouch S., Oard J.H. (2005). "Identification of candidate markers associated with agronomic traits in rice using discriminant analysis." *Theoretical and Applied Genetics* **110**, 721–729.

Zheng T.Q., Xu J.L., Li Z.K., Zhai H.Q., Wan J.M. (2007). "Genomic regions associated with milling quality and grain shape identified in a set of random introgression lines of rice (*Oryza sativa* L.)." *Plant Breeding* **126**, 158–163.

APPENDIX B SUPPLEMENTARY TABLES

Table B.1 Selected nsSNP positions, Locus ID, and corresponding genes identified within QTL *qShB9-2* (Liu et al., 2009) from resistant Jasmine 85, Teqing and MCR010277.

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Nipponbare Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
21279866	C	t	H	Y	LOC_Os09g36900	WD domain, G-beta repeat domain containing protein, expressed	21275966	21280139
21666818	T	c	M	T	LOC_Os09g37590	OsFBDUF47 - F-box and DUF domain containing protein, expressed	21666226	21673416
21781200	T	c	H	R	LOC_Os09g37800	serine/threonine kinase, putative	21778729	21782005
21841580	G	c	V	L	LOC_Os09g37880	serine/threonine-protein kinase receptor precursor, putative	21840875	21844761
22096465	C	t	S	N	LOC_Os09g38380	aspartic proteinase nepenthesin, putative	22095603	22096898
22245913	C	t	D	V	LOC_Os09g38700	STRUBBELIG-RECEPTOR FAMILY 5 precursor, putative,	22243412	22248821
22252463	C	t	G	D	LOC_Os09g38710	HEAT repeat family protein, putative, expressed	22247967	22253928
22317968	T	c	*	Q	LOC_Os09g38850	OsWAK91 - OsWAK receptor-like protein kinase, expressed	22315245	22318384
22367742	C	t	D	N	LOC_Os09g38960	potassium transporter, putative, expressed	22366890	22373139
22381404	T	a	S	A	LOC_Os09g38970	zinc finger family protein, putative, expressed	22376407	22382011

[a] Base pair position at which the nsSNP occurs based on Nippbare MSU6 reference genome sequence

[b] Allele based on Nipponbare MSU6 reference genome sequence

[c] Variant allele based on Illumina GA IIx sequencing

[d] Predicted amino acid based on Nipponbare MSU6 reference genome sequence; A=alanine, D=aspartic acid, G=glycine, H=hisitidine, M=methionine, N=asparagine, Q=glutamine, R=arginine, S=serine, T=threonine, Y=tyrosine, V=valine

[e] Predicted variant amino acid based on Illumina GA IIx sequencing

Table B.2 Selected nsSNP positions and corresponding Nipponbare reference allele, variant allele, Nipponbare reference amino acid, variant amino acid, locus identification, and candidate genes located outside of QTL *qShB9-2*.

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
Chromosome 1								
5540388	A	c	L	R	LOC_Os 01g10470	RALFL17 - Rapid ALkalinization Factor RALF family protein precursor, expressed	5539629	5540520
7420797	A	c	I	L	LOC_Os 01g13300	B3 DNA binding domain containing protein, expressed	7416303	7421624
30075242	G	a	E	K	LOC_Os 01g52330	NB-ARC domain containing protein, expressed	30074579	30076081
30406859	G	a	C	Y	LOC_Os 01g52880	leucine-rich repeat family protein, putative, expressed	30405992	30407510
30675476	A	g	T	A	LOC_Os 01g53390	glucosyltransferase, putative, expressed	30674429	30676377
30675791	T	a	S	T	LOC_Os 01g53390	glucosyltransferase, putative, expressed	30674429	30676377
30689063	T	c	R	G	LOC_Os 01g53420	anthocyanidin 5,3-O-glucosyltransferase, putative, expressed	30688803	30690554
30897396	G	a	V	M	LOC_Os 01g53750	anthocyanidin 5,3-O-glucosyltransferase, putative, expressed	30896929	30899576
31005520	C	t	T	I	LOC_Os 01g53920	receptor-like protein kinase 5 precursor, putative, expressed	31004982	31009379
31005532	G	a	S	N	LOC_Os 01g53920	receptor-like protein kinase 5 precursor, putative, expressed	31004982	31009379
31005889	T	a	F	Y	LOC_Os 01g53920	receptor-like protein kinase 5 precursor, putative, expressed	31004982	31009379
33065796	T	g	E	A	LOC_Os 01g57230	BTBN1 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain, expressed	33063342	33066790
33479245	G	a	R	C	LOC_Os 01g57900	PPR repeat domain containing protein, putative, expressed	33477940	33480299
33659492	G	c	R	T	LOC_Os 01g58240	OsSub6 - Putative Subtilisin homologue, expressed	33658738	33671494
Chromosome 2								
975892	T	g	N	T	LOC_Os 02g02650	THION21 - Plant thionin family protein precursor	975623	976077
3098887	A	g	F	S	LOC_Os 02g06200	phytosulfokine receptor precursor, putative, expressed	3097245	3099377
3104097	G	a	T	M	LOC_Os 02g06210	phytosulfokine receptor precursor, putative, expressed	3102629	3105505
5065045	A	g	T	A	LOC_Os 02g09820	zinc finger, C3HC4 type domain containing protein, expressed	5063134	5066802
5277344	T	g	K	N	LOC_Os 02g10120	lipoxygenase, putative, expressed	5276617	5282623
5760143	A	g	M	T	LOC_Os 02g10860	lipoxygenase, putative, expressed	5760034	5763645
5786160	G	a	A	V	LOC_Os 02g10900	NB-ARC domain containing protein, expressed	5785295	5788769
5788240	T	g	N	T	LOC_Os 02g10900	NB-ARC domain containing protein, expressed	5785295	5788769
5967533	C	g	G	A	LOC_Os 02g11130	cytokinin-O-glucosyltransferase 3, putative, expressed	5966029	5967668

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
6114451	T	a	V	E	LOC_Os 02g11820	GTPase-activating protein, putative, expressed	6110092	6118738
6209027	G	a	T	M	LOC_Os 02g11980	receptor-like protein kinase precursor, putative, expressed	6208684	6212210
6210341	T	c	T	A	LOC_Os 02g11980	receptor-like protein kinase precursor, putative, expressed	6208684	6212210
6210412	T	c	E	G	LOC_Os 02g11980	receptor-like protein kinase precursor, putative, expressed	6208684	6212210
11193184	C	t	A	T	LOC_Os 02g19200	OsFBX46 - F-box domain containing protein, expressed	11188427	11193872
20661950	G	c	W	S	LOC_Os 02g34490	Leucine Rich Repeat family protein, expressed	20657590	20662944
20798058	G	a	A	T	LOC_Os 02g34680	ZOS2-09 - C2H2 zinc finger protein, expressed	20795085	20798606
20899450	T	a	V	D	LOC_Os 02g34850	histone-lysine N-methyltransferase ASHH2, putative, expressed	20893940	20901307
21160861	G	a	D	N	LOC_Os 02g35210	resistance protein, putative	21160810	21164084
21466875	A	t	T	S	LOC_Os 02g35750	pentatricopeptide repeat domain containing protein, putative, expressed	21465862	21469764
21658261	A	g	I	V	LOC_Os 02g36030	cytochrome P450, putative, expressed	21656739	21658454
23887432	T	a	K	M	LOC_Os 02g39590	GDSL-like lipase/acylhydrolase, putative	23887389	23888600
25509520	G	t	L	M	LOC_Os 02g42412	F-box/LRR-repeat protein 2, putative, expressed	25508639	25509778
25633683	C	t	G	D	LOC_Os 02g42620	protein kinase, putative, expressed	25633218	25635286
25642200	A	c	D	A	LOC_Os 02g42640	CBS domain-containing protein, putative, expressed	25640756	25642222
26032397	A	c	N	K	LOC_Os 02g43194	aldehyde dehydrogenase, putative, expressed	26028451	26035553
26228789	C	g	A	G	LOC_Os 02g43460	required to maintain repression 1, putative	26227039	26232421
26229122	A	g	Q	R	LOC_Os 02g43460	required to maintain repression 1, putative	26227039	26232421
26388585	G	a	R	Q	LOC_Os 02g43740	AGC_PVPK_like_kin82y.6 - ACG kinases include homologs to PKA, PKG	26387094	26390851
26624790	A	g	F	S	LOC_Os 02g44104	F-box family protein, putative, expressed	26622195	26626241
26663691	T	a	S	T	LOC_Os 02g44120	ZOS2-13 - C2H2 zinc finger protein, expressed	26662927	26665275
27099654	T	a	M	K	LOC_Os 02g44730	tetracycline transporter protein, putative, expressed	27097597	27100509
27113311	A	c	I	L	LOC_Os 02g44770	uncharacterized mscS family protein, putative, expressed	27112122	27116571
27115988	A	g	I	V	LOC_Os 02g44770	uncharacterized mscS family protein, putative, expressed	27112122	27116571
27387949	A	g	S	P	LOC_Os 02g45160	aluminum-activated malate transporter, putative, expressed	27384978	27388319
27732042	C	a	R	L	LOC_Os 02g45590	PPR repeat domain containing protein, putative, expressed	27729978	27735070

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
28014024	C	t	T	M	LOC_Os 02g45980	ZR1 protein, putative, expressed	28008953	28014945
28471433	A	t	N	I	LOC_Os 02g46650	ubiquitin carboxyl-terminal hydrolase domain containing protein, expressed	28466956	28472140
28483772	C	t	A	T	LOC_Os 02g46680	multidrug resistance protein, putative, expressed	28481680	28492811
31659253	A	c	M	L	LOC_Os 02g51680	uncharacterized glycosyl hydrolase Rv2006/MT2062, putative, expressed	31657211	31659949
31782040	T	c	S	G	LOC_Os 02g51900	cytokinin-O-glucosyltransferase 2, putative, expressed	31780932	31783162
31782340	G	c	H	D	LOC_Os 02g51900	cytokinin-O-glucosyltransferase 2, putative, expressed	31780932	31783162
31859549	G	a	T	M	LOC_Os 02g52060	peptide transporter like protein, putative	31859051	31860015
31859870	C	t	R	H	LOC_Os 02g52060	peptide transporter like protein, putative	31859051	31860015
31859888	A	c	L	R	LOC_Os 02g52060	peptide transporter like protein, putative	31859051	31860015
32173206	A	g	I	V	LOC_Os 02g52590	xyloglucan fucosyltransferase, putative, expressed	32171248	32173255
32180391	T	a	L	Q	LOC_Os 02g52610	xyloglucan fucosyltransferase, putative, expressed	32173486	32181562
32180424	A	g	Q	R	LOC_Os 02g52610	xyloglucan fucosyltransferase, putative, expressed	32173486	32181562
32180439	C	t	S	F	LOC_Os 02g52610	xyloglucan fucosyltransferase, putative, expressed	32173486	32181562
32827717	T	g	N	T	LOC_Os 02g53680	RPA1A - Putative single-stranded DNA binding complex subunit 1, expressed	32826963	32830612
32976770	A	t	N	I	LOC_Os 02g53850	OsSub21 - Putative Subtilisin homologue	32975405	32977751
33002679	G	a	T	M	LOC_Os 02g53910	OsSub23 - Putative Subtilisin homologue	33002645	33004717
33002854	T	c	M	V	LOC_Os 02g53910	OsSub23 - Putative Subtilisin homologue	33002645	33004717
33003286	T	c	M	V	LOC_Os 02g53910	OsSub23 - Putative Subtilisin homologue	33002645	33004717
33004216	T	c	I	V	LOC_Os 02g53910	OsSub23 - Putative Subtilisin homologue	33002645	33004717
33040089	T	c	S	G	LOC_Os 02g53970	OsSub24 - Putative Subtilisin homologue, expressed	33039511	33042134
33220680	G	a	R	*	LOC_Os 02g54200	mitochondrial transcription termination factor family protein, putative, expressed	33218155	33221808
33220700	A	g	L	P	LOC_Os 02g54200	mitochondrial transcription termination factor family protein, putative, expressed	33218155	33221808
33220868	A	g	L	P	LOC_Os 02g54200	mitochondrial transcription termination factor family protein, putative, expressed	33218155	33221808
33220883	T	c	K	R	LOC_Os 02g54200	mitochondrial transcription termination factor family protein, putative, expressed	33218155	33221808
33221331	G	t	L	I	LOC_Os 02g54200	mitochondrial transcription termination factor family protein, putative, expressed	33218155	33221808
33307448	C	g	R	T	LOC_Os 02g54330	OsFBDUF14 - F-box and DUF domain containing protein	33306895	33308136

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
33794880	C	t	T	I	LOC_Os 02g55180	ubiquitin carboxyl-terminal hydrolase domain containing protein, expressed	33791497	33797995
33981713	C	g	V	L	LOC_Os 02g55510	CXXXC2 - Cysteine-rich protein with paired CXXXC motifs precursor, putative,	33981155	33981845
34434822	C	t	V	M	LOC_Os 02g56280	zinc finger family protein, putative, expressed	34433892	34438893
34511349	C	a	A	S	LOC_Os 02g56380	OsWAK21 - OsWAK receptor-like cytoplasmic kinase OsWAK-RLCK,	34510753	34513605
34568863	T	c	H	R	LOC_Os 02g5648	PB1 domain containing protein, expressed	34568083	34570304
35103382	G	c	P	A	LOC_Os 02g57305	disease resistance protein, putative	35102960	35104238
35103568	A	g	S	P	LOC_Os 02g57305	disease resistance protein, putative	35102960	35104238
35104101	A	t	F	L	LOC_Os 02g57305	disease resistance protein, putative	35102960	35104238
35109755	G	a	Q	*	LOC_Os 02g57310	pib, putative, expressed	35107768	35112900
35118371	T	g	C	G	LOC_Os 02g57340	tetratricopeptide-like helical, putative, expressed	35116970	35120947
35425301	G	c	Q	E	LOC_Os 02g57860	OsFBX71 - F-box domain containing protein	35425163	35426971
35425348	G	a	P	L	LOC_Os 02g57860	OsFBX71 - F-box domain containing protein	35425163	35426971
35425615	A	g	V	A	LOC_Os 02g57860	OsFBX71 - F-box domain containing protein	35425163	35426971
35426140	A	g	V	A	LOC_Os 02g57860	OsFBX71 - F-box domain containing protein	35425163	35426971
35426333	G	a	L	F	LOC_Os 02g57860	OsFBX71 - F-box domain containing protein	35425163	35426971
35495080	G	a	P	S	LOC_Os 02g57860	Leucine Rich Repeat family protein	35492353	35495167
35528303	G	c	I	M	LOC_Os 02g58040	OsFBX75 - F-box domain containing protein	35528270	35530261
35528649	C	t	R	H	LOC_Os 02g58040	OsFBX75 - F-box domain containing protein	35528270	35530261
35657337	A	g	S	P	LOC_Os 02g58260	metallo-beta-lactamase family protein, putative, expressed	35655094	35657614
35774119	C	t	P	S	LOC_Os 02g58530	transporter family protein, putative, expressed	35773561	35775519
35778055	G	a	A	V	LOC_Os 02g58540	RING-H2 finger protein, putative, expressed	35777415	35778277
35817716	A	g	I	V	LOC_Os 02g58620	pentatricopeptide, putative, expressed	35817127	35823762
35844321	G	a	R	H	LOC_Os 02g58660	ATCHX15, putative, expressed	35841745	35844754
Chromosome 3								
390749	A	g	S	G	LOC_Os 03g01630	expansin precursor, putative	390569	391372
17206912	C	t	R	K	LOC_Os 03g30130	phospholipase C, putative, expressed	17206103	17209401

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
17207137	T	g	L	F	LOC_Os 03g30130	phospholipase C, putative, expressed	17206103	17209401
18052635	G	a	V	M	LOC_Os 03g31630	OsSub29 - Putative Subtilisin homologue	18051792	18054158
20914617	A	g	L	P	LOC_Os 03g37720	NBS-LRR type disease resistance protein Rps1-k-1, putative	20912120	20915920
21745084	A	c	M	L	LOC_Os 03g39150	protein kinase domain containing protein	21744559	21745614
22369241	C	t	R	K	LOC_Os 03g40250	Leucine Rich Repeat family protein, expressed	22367808	22369834
24429583	T	c	I	V	LOC_Os 03g43684	KIP1, putative, expressed	24424802	24433681
30523344	A	t	Q	L	LOC_Os 03g53220	U5 small nuclear ribonucleoprotein 200 kDa helicase, putative	30515962	30523987
31819527	A	g	Y	H	LOC_Os 03g55890	ternary complex factor MIP1, putative, expressed	31818383	31822562
31868577	A	c	N	T	LOC_Os 03g56000	PHLOEM 2-LIKE A10, putative, expressed	31867852	31869448
31868747	G	a	A	T	LOC_Os 03g56000	PHLOEM 2-LIKE A10, putative, expressed	31867852	31869448
32007857	T	c	S	G	LOC_Os 03g56180	legume lectins beta domain containing protein, expressed	32006093	32008127
32144849	T	c	Q	R	LOC_Os 03g56400	pentatricopeptide, putative, expressed	32143681	32146259
32145467	T	c	D	G	LOC_Os 03g56400	pentatricopeptide, putative, expressed	32143681	32146259
32163618	A	g	D	G	LOC_Os 03g56450	OsFBX111 - F-box domain containing protein	32163401	32164804
32586703	C	t	G	D	LOC_Os 03g57160	zinc ion binding protein, putative, expressed	32584566	32591899
33260375	A	g	T	A	LOC_Os 03g58390	zinc finger, C3HC4 type domain containing protein, expressed	33260062	33265172
34335480	A	g	K	R	LOC_Os 03g60380	cinnamoyl CoA reductase, putative, expressed	34333157	34336630
35667086	A	g	V	A	LOC_Os 03g63110	prefoldin, putative, expressed	35666152	35669732
Chromosome 4								
2441294	G	a	D	N	LOC_Os 04g05030	serine-rich 25 kDa antigen protein, putative, expressed	2437107	2443384
5684447	C	g	H	D	LOC_Os 04g10460	amidase, putative, expressed	5681891	5684930
6377725	A	g	Q	R	LOC_Os 04g11640	methyl-CpG binding domain containing protein	6374200	6377753
6560546	G	a	A	T	LOC_Os 04g11970	O-methyltransferase, putative, expressed	6560068	6562330
8505140	G	t	G	C	LOC_Os 04g15650	Leucine Rich Repeat family protein, expressed	8503235	8506337
10443450	A	g	S	P	LOC_Os 04g18790	OsFBX126 - F-box domain containing protein, expressed	10442964	10444424
11560624	A	g	Y	H	LOC_Os 04g20680	wall-associated receptor kinase 3 precursor, putative, expressed	11560043	11565349

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
12387967	A	c	Q	P	LOC_Os 04g21890	disease resistance protein RPM1, putative, expressed	12386832	12389630
12388621	A	g	K	R	LOC_Os 04g21890	disease resistance protein RPM1, putative, expressed	12386832	12389630
13514379	A	c	S	A	LOC_Os 04g23620	D-mannose binding lectin family protein	13512075	13515129
13640560	T	c	Q	R	LOC_Os 04g23890	AGC_PVPK_like_kin82y.10 - ACG kinases include homologs to PKA, PKG	13632299	13645569
33008803	G	a	E	K	LOC_Os 04g55760	OsWAK55 - OsWAK receptor-like protein kinase	33007511	33011019
33349688	G	a	T	I	LOC_Os 04g56250	OsFBX152 - F-box domain containing protein, expressed	33349274	33350954
34150615	C	t	V	M	LOC_Os 04g57670	pentatricopeptide, putative, expressed	34148357	34151562
34731835	A	g	I	V	LOC_Os 04g58720	anthranilate phosphoribosyltransferase, putative, expressed	34730681	34734979
34732078	A	g	T	A	LOC_Os 04g58720	anthranilate phosphoribosyltransferase, putative, expressed	34730681	34734979
34732090	G	a	G	S	LOC_Os 04g58720	anthranilate phosphoribosyltransferase, putative, expressed	34730681	34734979
34732235	A	c	Q	P	LOC_Os 04g58720	anthranilate phosphoribosyltransferase, putative, expressed	34730681	34734979
34804587	G	a	R	K	LOC_Os 04g58820	ATOFP18/OFP18, putative, expressed	34803723	34805015
34826838	T	g	H	P	LOC_Os 04g58860	harpin-induced protein 1 domain containing protein, expressed	34825828	34827340
34856814	T	c	N	D	LOC_Os 04g58910	receptor protein kinase TMK1 precursor, putative, expressed	34854803	34858678
34943898	T	g	I	L	LOC_Os 04g59060	heat shock protein DnaJ, putative, expressed	34943467	34947490
35113021	G	t	A	S	LOC_Os 04g59380	ZOS4-14 - C2H2 zinc finger protein, expressed	35112479	35115230
35230058	C	g	Q	E	LOC_Os 04g59540	phosphatidylinositol-4-phosphate 5-Kinase, putative, expressed	35228123	35233722
Chromosome 5								
21585027	A	g	S	P	LOC_Os 05g37040	MYB family transcription factor, putative	21584362	21585144
23293209	G	a	S	N	LOC_Os 05g39760	VHS and GAT domain containing protein, expressed	23289137	23293955
23860975	A	g	D	G	LOC_Os 05g40790	CCR4-NOT transcription factor, putative, expressed	23853793	23860985
24014563	T	c	N	D	LOC_Os 05g41100	protein kri1, putative, expressed	24013013	24015304
24027934	C	a	G	C	LOC_Os 05g41130	OsFBX168 - F-box domain containing protein, expressed	24027165	24029164
24122910	T	g	N	K	LOC_Os 05g41290	disease resistance RPP13-like protein 1, putative, expressed	24121866	24126622
28979361	A	g	N	D	LOC_Os 05g50660	PX domain containing protein, putative, expressed	28975904	28983664

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
Chronmosome 6								
3056773	C	t	R	K	LOC_Os 06g06520	GDSL-like lipase/acylhydrolase, putative, expressed	3056017	3058018
3056780	C	t	E	K	LOC_Os 06g06520	GDSL-like lipase/acylhydrolase, putative, expressed	3056017	3058018
7208678	C	a	G	C	LOC_Os 06g13140	WD domain, G-beta repeat domain containing protein, expressed	7208517	7211491
8598272	T	c	I	V	LOC_Os 06g15170	3-ketoacyl-CoA synthase, putative, expressed	8596871	8598913
10871554	T	c	N	D	LOC_Os 06g19110	cadmium tolerance factor, putative, expressed	10870810	10874145
10871580	C	t	R	K	LOC_Os 06g19110	cadmium tolerance factor, putative, expressed	10870810	10874145
10871589	G	t	T	K	LOC_Os 06g19110	cadmium tolerance factor, putative, expressed	10870810	10874145
10886942	G	a	A	T	LOC_Os 06g19130	cadmium tolerance factor, putative	10880038	10888582
11535373	A	g	H	R	LOC_Os 06g20120	CND41, chloroplast nucleoid DNA binding protein, putative	11534324	11536445
12056838	G	c	L	V	LOC_Os 06g20870	pentatricopeptide repeat protein PPR1106-17, putative, expressed	12053983	12057675
12750932	G	a	D	N	LOC_Os 06g22020	cytochrome P450, putative	12750746	12752287
12751136	C	t	R	W	LOC_Os 06g22020	cytochrome P450, putative	12750746	12752287
12751175	A	g	M	V	LOC_Os 06g22020	cytochrome P450, putative	12750746	12752287
12751231	G	a	M	I	LOC_Os 06g22020	cytochrome P450, putative	12750746	12752287
12751263	T	a	V	E	LOC_Os 06g22020	cytochrome P450, putative	12750746	12752287
12751320	C	t	A	V	LOC_Os 06g22020	cytochrome P450, putative	12750746	12752287
12751686	T	c	M	T	LOC_Os 06g22020	cytochrome P450, putative	12750746	12752287
13056419	T	c	S	G	LOC_Os 06g22460	disease resistance protein RPM1, putative, expressed	13054163	13057028
13601739	T	a	K	M	LOC_Os 06g23290	phosphatidylinositol 3- and 4-kinase family protein, putative, expressed	13600952	13603918
13651123	G	a	D	N	LOC_Os 06g23390	IQ calmodulin-binding motif family protein, putative	13648911	13651230
13725000	A	c	D	E	LOC_Os 06g23530	pre-mRNA-splicing factor ATP-dependent RNA helicase, putative, expressed	13722595	13726111
15930212	T	c	K	R	LOC_Os 06g28060	ATP-binding region, ATPase-like domain containing protein, expressed	15921987	15932060
15968674	T	c	D	G	LOC_Os 06g28124	glycosyltransferase, putative, expressed	15967623	15973051
16329889	G	t	V	F	LOC_Os 06g28670	polygalacturonase, putative, expressed	16328397	16330466
16596715	C	g	Q	E	LOC_Os 06g29110	MLO domain containing protein, putative	16592799	16598302

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
17044919	A	g	H	R	LOC_Os 06g29700	OsFBD11 - F-box and FBD domain containing protein, expressed	17043173	17045354
17066203	C	g	R	T	LOC_Os 06g29730	RALFL28 - Rapid ALkalinization Factor RALF family protein precursor, expressed	17065922	17067621
17066224	T	c	D	G	LOC_Os 06g29730	RALFL28 - Rapid ALkalinization Factor RALF family protein precursor, expressed	17065922	17067621
17195755	T	g	S	A	LOC_Os 06g29844	MATE efflux family protein, putative, expressed	17193077	17199586
17209038	G	a	W	*	LOC_Os 06g29870	far1-like, putative	17208897	17210710
17772543	A	g	D	G	LOC_Os 06g30680	WD domain, G-beta repeat domain containing protein	17770594	17776884
18071409	T	a	K	N	LOC_Os 06g31070	PROLM24 - Prolamin precursor, expressed	18071225	18071907
18827854	A	c	N	K	LOC_Os 06g32350	THION12 - Plant thionin family protein precursor	18827824	18828457
19357228	A	g	E	G	LOC_Os 06g33250	crooked neck, putative, expressed	19356235	19359928
19401755	G	a	R	H	LOC_Os 06g33320	extra-large G-protein-related, putative, expressed	19400019	19405763
20768668	C	g	R	T	LOC_Os 06g35590	reticuline oxidase-like protein precursor, putative, expressed	20766518	20768688
20916895	G	c	R	T	LOC_Os 06g35850	lectin protein kinase family protein, putative, expressed	20914617	20917500
20960032	G	a	A	T	LOC_Os 06g35930	aquaporin protein, putative	20959709	20960822
22193618	C	t	V	I	LOC_Os 06g37500	cytokinin dehydrogenase precursor, putative	22191036	22193966
22397264	G	t	N	K	LOC_Os 06g37500	cytokinin dehydrogenase precursor, putative	22395865	22397671
22863200	A	g	I	V	LOC_Os 06g38590	receptor-like protein kinase precursor, putative, expressed	22862177	22865875
22863207	C	a	P	Q	LOC_Os 06g38590	receptor-like protein kinase precursor, putative, expressed	22862177	22865875
25712817	A	t	D	E	LOC_Os 06g42770	type II intron maturase protein, putative, expressed	25710871	25713684
27075561	G	a	E	K	LOC_Os 06g44820	PPR repeat domain containing protein, putative	27074173	27075641
Chromosome 7								
27048590	C	t	H	Y	LOC_Os 07g45340	hypothetical protein	27047795	27049099
27048630	A	g	N	S	LOC_Os 07g45340	hypothetical protein	27047795	27049099
27048761	C	t	P	S	LOC_Os 07g45340	hypothetical protein	27047795	27049099
27141597	C	t	S	L	LOC_Os 07g45490	conserved hypothetical protein	27140252	27145948
27141600	T	c	V	A	LOC_Os 07g45490	conserved hypothetical protein	27140252	27145948
27142510	T	g	F	L	LOC_Os 07g45490	conserved hypothetical protein	27140252	27145948

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
27142832	G	a	A	T	LOC_Os 07g45490	conserved hypothetical protein	27140252	27145948
27143119	T	g	I	M	LOC_Os 07g45490	conserved hypothetical protein	27140252	27145948
27143226	C	t	T	I	LOC_Os 07g45490	conserved hypothetical protein	27140252	27145948
Chromosome 8								
6003503	G	a	E	K	LOC_Os 08g10300	SHR5-receptor-like kinase, putative, expressed	5997024	6004223
6216207	A	t	I	N	LOC_Os 08g10560	histone-like transcription factor and archaeal histone family protein	6211078	6217280
7587176	T	c	V	A	LOC_Os 08g12800	glucan endo-1,3-beta-glucosidase precursor, putative, expressed	7582451	7587424
7760443	G	c	P	R	LOC_Os 08g13070	MBTB23 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain, expressed	7759524	7761246
7878475	T	c	I	V	LOC_Os 08g13250	MBTB23 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain, expressed	7878233	7880681
8282016	T	g	K	Q	LOC_Os 08g13870	S-locus lectin protein kinase family protein, putative	8281966	8284338
8282546	T	c	N	S	LOC_Os 08g13870	S-locus lectin protein kinase family protein, putative	8281966	8284338
8282993	C	g	G	A	LOC_Os 08g13870	S-locus lectin protein kinase family protein, putative	8281966	8284338
8283225	C	t	V	I	LOC_Os 08g13870	S-locus lectin protein kinase family protein, putative	8281966	8284338
8283922	A	t	N	K	LOC_Os 08g13870	S-locus lectin protein kinase family protein, putative	8281966	8284338
11786501	A	c	D	E	LOC_Os 08g19694	NB-ARC domain containing protein, expressed	11784534	11797472
11796984	A	t	L	Q	LOC_Os 08g19694	NB-ARC domain containing protein, expressed	11784534	11797472
11987684	C	t	G	R	LOC_Os 08g20020	octicosapeptide/Phox/Bem1p, putative, expressed	11983657	11988804
12381205	G	a	A	T	LOC_Os 08g20610	pentatricopeptide containing protein, putative	12380719	12382926
17198760	A	g	N	D	LOC_Os 08g28180	PPR repeat domain containing protein, putative, expressed	17196596	17200370
19042526	G	a	G	D	LOC_Os 08g30850	YDG/SRA domain containing protein, expressed	19041037	19044346
19085103	T	c	I	T	LOC_Os 08g30910	YDG/SRA domain containing protein, expressed	19082792	19086784
19213472	G	t	R	S	LOC_Os 08g31110	PPR repeat domain containing protein, putative, expressed	19213116	19215096
21088401	C	g	G	A	LOC_Os 08g33750	myb-like DNA-binding domain containing protein, expressed	21087229	21089395
22084235	T	c	D	G	LOC_Os 08g35050	ARID/BRIGHT DNA-binding domain containing protein, expressed	22079859	22091044
22277158	C	a	G	C	LOC_Os 08g35310	O-methyltransferase, putative	22275143	22277242
22876630	T	c	D	G	LOC_Os 08g36320	decarboxylase, putative, expressed	22875955	22880235

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
23212262	A	c	Y	S	LOC_Os 08g36760	remorin C-terminal domain containing protein, putative, expressed	23210433	23214913
26966254	A	c	L	R	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
26966516	C	g	E	Q	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
26966634	C	g	R	S	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
26966645	T	c	T	A	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
26966657	T	c	K	E	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
26967222	G	c	I	M	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
26967368	C	g	E	Q	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
26967395	C	t	V	I	LOC_Os 08g42670	resistance protein, putative, expressed	26965828	26975527
27330584	T	g	K	Q	LOC_Os 08g43240	LTPL97 - Protease inhibitor/seed storage/ LTP family protein precursor, expressed	27330142	27330915
27633238	C	t	A	V	LOC_Os 08g43730	DUF630/DUF632 domains containing protein, putative, expressed	27632257	27636933
27651157	A	g	H	R	LOC_Os 08g43800	carrier, putative	27650784	27651488
27665742	C	g	H	D	LOC_Os 08g43860	carrier, putative	27665622	27666329
27683714	G	a	R	H	LOC_Os 08g43950	carrier, putative	27683704	27684870
27684146	G	a	R	K	LOC_Os 08g43950	carrier, putative	27683704	27684870
Chromosome 9								
3804855	G	a	A	V	LOC_Os 09g07590	RGH1A, putative	3802863	3804874
5086798	T	g	N	T	LOC_Os 09g09450	PPR repeat domain containing protein, putative, expressed	5086530	5086994
5120970	C	t	E	K	LOC_Os 09g09500	lectin-like receptor kinase, putative	5119567	5121573
5121056	G	a	A	V	LOC_Os 09g09500	lectin-like receptor kinase, putative	5119567	5121573
5124435	G	a	A	T	LOC_Os 09g09510	legume lectins beta domain containing protein	5124150	5126327
5125811	A	c	Q	H	LOC_Os 09g09510	legume lectins beta domain containing protein	5124150	5126327
7759028	T	c	S	G	LOC_Os 09g13420	mucin, putative	7758918	7761038
10153331	A	g	R	G	LOC_Os 09g16540	protein kinase, putative, expressed	10151936	10156507
10153340	T	c	C	R	LOC_Os 09g16540	protein kinase, putative, expressed	10151936	10156507
10740864	T	a	Q	H	LOC_Os 09g17560	O-methyltransferase, putative, expressed	10739339	10740910

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
10766714	G	a	R	K	LOC_Os 09g17600	membrane protein, putative, expressed	10762249	10767267
10792494	T	c	I	T	LOC_Os 09g17630	receptor-like protein kinase 2, putative, expressed	10784209	10792868
14666043	A	g	T	A	LOC_Os 09g24640	pentatricopeptide, putative, expressed	14665228	14666808
14971252	C	a	Q	K	LOC_Os 09g25050	PPR repeat domain containing protein, putative, expressed	14970347	14974385
15385777	A	g	L	S	LOC_Os 09g25620	CPuORF8 - conserved peptide uORF-containing transcript, expressed	15384600	15388095
15422862	A	g	N	D	LOC_Os 09g25700	TsetseEP precursor, putative	15421387	15423389
15532799	T	a	F	I	LOC_Os 09g25890	trehalose-6-phosphate synthase, putative, expressed	15531902	15535139
15558634	G	a	V	M	LOC_Os 09g25910	xylanase inhibitor, putative	15558430	15559764
15593365	T	c	R	G	LOC_Os 09g25960	glutamate receptor, putative	15587269	15593377
15769422	C	g	T	R	LOC_Os 09g26160	glutamate receptor, putative, expressed	15764944	15769860
15891490	A	g	V	A	LOC_Os 09g26300	hypro1, putative, expressed	15890728	15891986
16413734	A	g	Y	C	LOC_Os 09g26999	keratin-associated protein 5-4, putative, expressed	16410150	16414861
16748987	A	g	F	S	LOC_Os 09g27570	OsFBA3 - F-box and FBA domain containing protein, expressed	16747246	16750290
16758017	A	g	K	E	LOC_Os 09g27580	potassium transporter, putative, expressed	16753448	16758733
16885819	G	t	R	S	LOC_Os 09g27750	1-aminocyclopropane-1-carboxylate oxidase 1, putative, expressed	16885018	16886616
17289330	G	a	V	I	LOC_Os 09g28400	alpha-amylase precursor, putative, expressed	17287992	17290294
17309928	A	g	T	A	LOC_Os 09g28450	paramyosin, putative, expressed	17308965	17311559
19117102	C	t	V	I	LOC_Os 09g32020	ubiquitin fusion degradation protein, putative, expressed	19114579	19117606
19122509	A	c	L	W	LOC_Os 09g32040	no apical meristem protein, putative, expressed	19121415	19124621
19591594	C	t	L	F	LOC_Os 09g32860	OsFBX336 - F-box domain containing protein, expressed	19589132	19592371
19913544	T	g	N	H	LOC_Os 09g33710	Os9bglu33 - beta-glucosidase homologue, expressed	19908161	19914000
20182171	T	g	L	R	LOC_Os 09g34180	formin, putative, expressed	20179233	20182554
20239255	A	g	N	S	LOC_Os 09g34280	ankyrin repeat-containing protein, putative, expressed	20235938	20240138
22736162	G	a	A	T	LOC_Os 09g39620	protein kinase family protein, putative, expressed	22736078	22740681
Chromosome 11								
1807366	A	g	H	R	LOC_Os 11g04350	cell death associated protein, putative, expressed	1807130	1808538

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
1807803	G	a	A	T	LOC_Os 11g04350	cell death associated protein, putative, expressed	1807130	1808538
1807945	C	t	A	V	LOC_Os 11g04350	cell death associated protein, putative, expressed	1807130	1808538
1808002	T	c	V	A	LOC_Os 11g04350	cell death associated protein, putative, expressed	1807130	1808538
2026004	G	t	G	V	LOC_Os 11g04770	EF hand family protein, putative	2025739	2026179
7469515	G	t	P	H	LOC_Os 11g13650	cellulose synthase, putative, expressed	7469190	7469942
9764292	G	c	G	R	LOC_Os 11g17530	pentatricopeptide, putative, expressed	9762646	9766930
9765173	T	a	D	E	LOC_Os 11g17530	pentatricopeptide, putative, expressed	9762646	9766930
10772091	A	g	I	V	LOC_Os 11g18940	WW domain containing protein, expressed	10771040	10772747
10841544	A	g	D	G	LOC_Os 11g19030	FAR1 family protein	10839850	10849990
10842305	T	c	M	T	LOC_Os 11g19030	FAR1 family protein	10839850	10849990
10848563	C	t	R	*	LOC_Os 11g19030	FAR1 family protein	10839850	10849990
11342380	C	a	N	K	LOC_Os 11g19700	cycloeucalenol cycloisomerase, putative, expressed	11336541	11342887
11645547	G	a	L	F	LOC_Os 11g20160	O-methyltransferase, putative, expressed	11644431	11646204
13199356	T	c	V	A	LOC_Os 11g24060	permease domain containing protein, putative, expressed	13197876	13199888
13321629	A	t	V	E	LOC_Os 11g24180	OsSCP50 - Putative Serine Carboxypeptidase homologue, expressed	13306746	13321765
13648166	T	a	S	C	LOC_Os 11g24770	ankyrin repeat domain containing protein	13645885	13648247
15283625	G	a	V	I	LOC_Os 11g27370	UDP-glucoronosyl and UDP-glucosyl transferase domain containing protein,	15282902	15284487
15665554	A	g	*	Q	LOC_Os 11g28065	cytochrome P450, putative, expressed	15665552	15667318
16287232	T	c	E	G	LOC_Os 11g28950	pollen signalling protein with adenylyl cyclase activity, putative, expressed	16286575	16294594
16287233	C	a	E	*	LOC_Os 11g28950	pollen signalling protein with adenylyl cyclase activity, putative, expressed	16286575	16294594
16293088	C	t	S	N	LOC_Os 11g28950	pollen signalling protein with adenylyl cyclase activity, putative, expressed	16286575	16294594
16293091	C	a	R	L	LOC_Os 11g28950	pollen signalling protein with adenylyl cyclase activity, putative, expressed	16286575	16294594
16293106	G	a	T	I	LOC_Os 11g28950	pollen signalling protein with adenylyl cyclase activity, putative, expressed	16286575	16294594
16293571	A	g	L	S	LOC_Os 11g28950	pollen signalling protein with adenylyl cyclase activity, putative, expressed	16286575	16294594
16567411	A	c	S	A	LOC_Os 11g29360	pentatricopeptide, putative, expressed	16566605	16568088
28284687	C	g	A	G	LOC_Os 11g47760	DnaK family protein, putative, expressed	28281027	28285329

(Table continued)

nsSNP Position[a]	Nipponbare Reference Allele[b]	Variant Allele[c]	Reference Amino Acid[d]	Variant Amino Acid[e]	Locus ID	Gene	Gene Start Position	Gene Stop Position
28284702	T	c	M	T	LOC_Os 11g47760	DnaK family protein, putative, expressed	28281027	28285329
28284753	T	c	M	T	LOC_Os 11g47760	DnaK family protein, putative, expressed	28281027	28285329
Chromosome 12								
1411478	C	t	R	W	LOC_Os 12g03554	zinc finger C-x8-C-x5-C-x3-H type family protein	1411001	1420875
1973059	G	c	T	R	LOC_Os 12g04660	zinc finger, C3HC4 type domain containing protein, expressed	1972804	1973670
2028637	A	g	K	E	LOC_Os 12g04660	zinc finger, C3HC4 type domain containing protein, expressed	2027537	2029787
2349667	C	t	P	L	LOC_Os 12g05280	zinc finger, C3HC4 type domain containing protein	2349648	2350376
2425834	A	g	F	S	LOC_Os 12g05370	RING-H2 finger protein, putative, expressed	2425130	2428788
3280174	A	g	I	V	LOC_Os 12g06740	F-box domain containing protein, expressed	3279334	3282712
3410207	G	t	Q	K	LOC_Os 12g06980	SAP domain containing protein, expressed	3401730	3412835
3743984	G	a	G	D	LOC_Os 12g07530	splicing factor, putative, expressed	3742732	3748503
3744523	A	g	T	A	LOC_Os 12g07530	splicing factor, putative, expressed	3742732	3748503
3744551	A	g	Y	C	LOC_Os 12g07530	splicing factor, putative, expressed	3742732	3748503
3941715	T	c	M	T	LOC_Os 12g07800	S-locus-like receptor protein kinase, putative, expressed	3937881	3942935
3942174	G	a	G	E	LOC_Os 12g07800	S-locus-like receptor protein kinase, putative, expressed	3937881	3942935
4033132	C	t	R	H	LOC_Os 12g07950	transcriptional regulator Sir2 family protein, putative, expressed	4031200	4035956
4709578	T	c	L	S	LOC_Os 12g09000	phosphomethylpyrimidine kinase/thiamin-phosphate pyrophosphorylase, putative,	4705832	4710321
5128266	T	a	I	N	LOC_Os 12g09710	NBS-LRR disease resistance protein, putative	5124188	5128660
5378630	T	g	M	L	LOC_Os 12g10180	NBS-LRR type disease resistance protein Rps1-k-2, putative, expressed	5375852	5382028
5468607	A	g	L	S	LOC_Os 12g10330	NB-ARC domain containing protein, expressed	5468030	5470355
5508921	G	c	A	G	LOC_Os 12g10410	NB-ARC domain containing protein, expressed	5507548	5514002
7284433	C	t	R	C	LOC_Os 12g13100	WW domain containing protein, expressed	7283319	7284990
26185651	A	g	L	S	LOC_Os 12g42260	initiation factor 2 subunit family domain containing protein, expressed	26180352	26188203

[a] Base pair position at which the nsSNP occurs based on Nippbare MSU6 reference genome sequence

[b] Allele based on Nipponbare MSU6 reference genome sequence

[c] Variant allele based on Illumina GA IIx sequencing

[d] Predicted amino acid based on Nipponbare MSU6 reference genome sequence; A=alanine, C=cysteine, D=aspartic acid, H=hisitidine, I=isoleucine, K=lysine, L=leucine, N=asparagine, M=methionine, P=proline, Q=glutamine, R=arginine, S=serine, T=threonine, W=tryptophan, Y=tyrosine, V=valine

[e] Predicted variant amino acid based on Illumina GA IIx sequencing

Table B.3a SNP alleles detected by PCR in 23 candidate SB resistance genes from nine resistant/tolerant and three susceptible lines.

Locus ID	MCR010277[a]	TeQing[a]	Taducan[a]	Rondo[a]	Shu Feng 121-1655 [a]	IR64[a]	O. Llanos 5[a]	Jouiku 393G[a]	Jhona 349[a]	Leah[b]	Nipponbare[b]	Azucena[b]
LOC_Os01g52880	G	G	G	G	G	G	A	G	G	A	A	A
LOC_Os02g34490	C	C	C	C	C	C	C	G	G	G	G	G
LOC_Os02g35210	A	A	A	A	A	A	A	A	A	G	G	G
LOC_Os02g54500	G	G	G	G	G	G	G	G	G	A	A	A
LOC_Os02g56380	A	A	A	A	A	A	A	A	A	C	C	C
LOC_Os02g57960 (1)	A	A	A	A	A	A	A	A	A	G	G	G
LOC_Os02g57960 (2)	A	A	A	A	A	A	A	A	A	G	G	G
LOC_Os03g37720	G	G	G	A	A	G	A	G	A	A	A	A
LOC_Os04g15650	T	T	T	T	T	T	G	T	T	G	G	G
LOC_Os04g20680	G	G	G	A	G	G	A	G	G	A	A	A
LOC_Os04g55760	A	A	G	A	A	A	G	A	A	G	G	G
LOC_Os04g58910	C	C	C	C	C	C	C	C	C	T	T	T
LOC_Os08g10300	A	A	A	A	A	A	A	A	A	G	G	G
LOC_Os09g17630	C	C	C	C	C	C	C	C	C	T	T	T
LOC_Os09g36900	T	T	T	T	T	C	T	C	T	C	C	C
LOC_Os09g37590	C	C	C	C	C	C	C	C	C	T	T	T
LOC_Os09g37800	C	C	C	C	C	C	C	C	T	T	T	T
LOC_Os09g37880	C	C	C	C	C	C	C	C	C	G	G	G
LOC_Os09g38850	C	C	C	C	C	C	C	C	C	T	T	T
LOC_Os09g39620	A	A	A	A	A	A	A	A	A	G	G	G
LOC_Os12g06740	T	T	T	T	T	T	T	T	T	A	A	A
LOC_Os12g09240	A	A	A	A	A	A	A	A	A	G	G	G
LOC_Os12g09710	A	A	A	A	A	A	A	A	A	T	T	T
LOC_Os12g10180	G	G	G	G	G	G	G	G	G	T	T	T

[a] SB resistant/tolerant line

[b] SB susceptible line

Table B.3b Primer sequences for resistant and susceptible allele PCR fragments containing nsSNPs given in Table B.3a.

Locus ID	Forward Primer Sequence	Reverse Primer Sequence	Allele
LOC_Os-	GAACACCAGCGCCATTGTCTTCC	TGCACGGCCAAGAAGCCGTC	Resistant
09g37880	CGTCGGTGTCGATGATCGCGTC	ATGAACACCGGCAACCTCGTCG	Susceptible
LOC_Os-	TCCCCGGCCACGAAAGACGTA	CCATGTATCCAATACCTGCGGAAAATCA	Susceptible
12g06740	CTCCCCGGCCACGAAAGACAAT	CCATGTATCCAATACCTGCGGAAAATCA	Resistant
LOC_Os -	CCGGAGTCGCTCAACAGGCAAT	TGGCAGAGCTTTAGCCAGCCGA	Susceptible
09g37800	CCGGAGTCGCTCAACAGGGAAC	TGGCAGAGCTTTAGCCAGCCGA	Resistant
LOC_Os-	GGCACGAGTCATCATCATTGTCACG	GCCCAACTGAAACTAAAGCCTGCATTCT	Susceptible
09g36900	GGGCACGAGTCATCATCATTGTCAAA	CCCACTGACATGATAGATTGATAGATTCCTGC	Resistant
LOC_Os-	AGTGACTTCCACGACGCCTCGC	CTCTGTGAACTGGATATTAACTTCCAAAAGCTCC	Susceptible
09g37590	GACGTAAGTGACTTCCACGACGCCTACT	CTCTGTGAACTGGATATTAACTTCCAAAAGCTCC	Resistant
LOC_Os-	CACCCTGCTGCACAGGGAATTACA	CAACTATCCACCAGAAAATTAGGCAGTAACAGCTAT	Susceptible
02g54500	CCTGCTGCACAGGGAATTCGG	CAACTATCCACCAGAAAATTAGGCAGTAACAGCTAT	Resistant
LOC_Os-	TGGTTAGCTCACCGAGGCACTCGATATAG	GAGAGAAGTGATGGACCTGACCGGC	Suceptible
12g09240	TGGTTAGCTCACCGAGGCACTCGATATAA	GAGAGAAGTGATGGACCTGACCGGC	Resistant
LOC_Os-	GGCCTCCGAAACCTCCAGCG	CCATCCGGTCATCCAGGCACA	Susceptible
01g52880	CCGGCCTCCGAAACCTCCACTA	CCATCCGGTCATCCAGGCACA	Resistant
LOC_Os-	TTGAAGCTCTGAGAGGGAGGTGATCTCTC	ATGTGTATCGGCTCCCATATTGCTTGTTATC	Susceptible
02g34490	AGCTCTGAGAGGGAGGTGATCTGCG	ATGTGTATCGGCTCCCATATTGCTTGTTATC	Resistant
LOC_Os-	GATGACAAGCTCAACGCCAAAGTCG	CATGAGGAGGTCTGCAATCTCTGTTGC	Susceptible
02g56380	TTGATGACAAGCTCAACGCCAAAGTCT	CATGAGGAGGTCTGCAATCTCTGTTGC	Resistant
LOC_Os-	GCCAAGAAGATGGGCGGCGT	AACCAAATCTTCAAAGAACTTGCTTCCAATGT	Susceptible
03g37720	CTAGCCAAGAAGATGGGCGGACC	AACCAAATCTTCAAAGAACTTGCTTCCAATGT	Resistant
LOC_Os-	AAGAAATACTACATGAGGATAACATGGAACTGCTGT	CATAGAAGCCAAATGTAGCTCAGACAAAAACTTTC	Susceptible
04g20680	AAGAAATACTACATGAGGATAACATGGAACTGCTTC	CATAGAAGCCAAATGTAGCTCAGACAAAAACTTTC	Resistant
LOC_Os-	CATCCATCACGGATGTAAGGATTGCCTAC	CCAGGTCACGTCTCTGATAGACCGAAATT	Susceptible
04g55760	CCATCACGGATGTAAGGATTGCGTTT	CCAGGTCACGTCTCTGATAGACCGAAATT	Resistant
LOC_Os-	TGGGTCGAACTACTGTTGCCATCATTTTT	GTGTGAAGGTGAATGTGACCGGCA	Susceptible
04g58910	GGGTCGAACTACTGTTGCCATCATTCTC	GTGTGAAGGTGAATGTGACCGGCA	Resistant
LOC_Os-	CAGTGGCATGCCCAGTATGCCTG	GGTTTTCGTGGTCCAATGTTGAGCATAG	Susceptible
04g15650	GCAGTGGCATGCCCAGTATGCTCT	GGTTTTCGTGGTCCAATGTTGAGCATAG	Resistant
LOC_Os-	GGACTTGCCAAGCTCTATGATGAAACGG	TCATACGGACAACGTGTTGATTGTGAGAA	Susceptible
08g10300	GGACTTGCCAAGCTCTATGATGAAACGA	TCATACGGACAACGTGTTGATTGTGAGAA	Resistant
LOC_Os-	CCCTTGTCTCCTCAGCCGGTAGTACTTG	ATGGAAATACAACCGTTGTTGCCTGCT	Susceptible
09g39620	CCCTTGTCTCCTCAGCCGGTAGTACATA	ATGGAAATACAACCGTTGTTGCCTGCT	Resistant
LOC_Os-	GAACACTTTCGAGTGTCATCTCCACCAA	CATTCCAGCTGAACAAACTGGGATAACAAC	Susceptible
09g38850	ACACTTTCGAGTGTCATCTCCACCCG	CATTCCAGCTGAACAAACTGGGATAACAAC	Resistant
LOC_Os-	GACTTCTCCCACAAGCCTAGTGAAGCTATGA	GCGCAAGAGCAAAGATGTGGCTG	Susceptible
12g09710	TCCCACAAGCCTAGTGAAGCTGGGT	GCGCAAGAGCAAAGATGTGGCTG	Resistant
LOC_Os-	GCCACATGCAAACGGCTAGAGTATCTTC	AAAGTAATTACCTTTTCGCTCAAGAAATTGAGGTG	Susceptible
02g57960(1)	GCCACATGCAAACGGCTAGAGTATGTGT	AAAGTAATTACCTTTTCGCTCAAGAAATTGAGGTG	Resistant
LOC_Os-	CGCAACTTAAAGCTTGCTGAAACTGACATAC	TGGTGGGGGCACTAGAAAGGAACTG	Susceptible
02g57960(2)	CGCAACTTAAAGCTTGCTGAAACTGACACTT	TGGTGGGGGCACTAGAAAGGAACTG	Resistant
LOC_Os-	GGACTCTGTCCTCAGCAAGCTCATCG	CATCTCCTTGGCAATTTGGTAGTGATTCC	Susceptible
02g35210	ATGGACTCTGTCCTCAGCAAGCTCAACA	CATCTCCTTGGCAATTTGGTAGTGATTCC	Resistant
LOC_Os-	TTGAGCCTGCTTGAGGGGCAGAT	TCACTATCCTAAAGATTTAAGCAGAGTGTCCATCTT	Susceptible
09g17630	TTGAGCCTGCTTGAGGGGCAAAC	TCACTATCCTAAAGATTTAAGCAGAGTGTCCATCTT	Resistant
LOC_Os-	CCTCGAGACCAAGTCATCCAGGGTG	CTTCTCCAACACCAGCTCAGAAAGATGC	Susceptible
12g10180	TCGAGACCAAGTCATCCAGGCCC	CTTCTCCAACACCAGCTCAGAAAGATGC	Resistant

Table B.4a SNP alleles detected by PCR in 12 candidate SB resistance genes from three resistant/tolerant (CIAT4, *O. nivara* 100898, 104443) and one susceptible line (CTHL).

Locus ID	CIAT4	*O. nivara* 100898	*O. nivara* 104443	CTHL
LOC_01g52880	G	G	G	G
LOC_02g54500	G	A	G/A[a]	A
LOC_02g34490	G	G	G	G
LOC_02g19200	T	C	C	C
LOC_02g44104	A	A	A/G[a]	A
LOC_02g54330	G	C	G/C[a]	C
LOC_03g37720	G	A	A	A
LOC_04g59540	G	G	C	C
LOC_06g28124	C	T	T	T
LOC_06g29700	G	A	A	A
LOC_06g32350	C	A	A	A
LOC_09g37880	C	G	G	G

[a] Heterozygous at this SNP locus

Table B.4b Primer sequences for resistant and susceptible allele PCR fragments containing nsSNPs given in Table B.4a.

Locus ID	Forward primer sequence	Reverse primer sequence	Allele
LOC_Os-	GGCCTCCGAAACCTCCAGCG	CCATCCGGTCATCCAGGCACA	Susceptible
01g52880	CCGGCCTCCGAAACCTCCACTA	CCATCCGGTCATCCAGGCACA	Resistant
LOC_Os-	CACCCTGCTGCACAGGGAATTACA	CAACTATCCACCAGAAAATTAGGCAGTAACAGCTAT	Susceptible
02g54500	CCTGCTGCACAGGGAATTCGG	CAACTATCCACCAGAAAATTAGGCAGTAACAGCTAT	Resistant
LOC_Os-	TTGAAGCTCTGAGAGGGAGGTGATCTCTC	ATGTGTATCGGCTCCCATATTGCTTGTTATC	Susceptible
02g34490	AGCTCTGAGAGGGAGGTGATCTGCG	ATGTGTATCGGCTCCCATATTGCTTGTTATC	Resistant
LOC_Os-	TGGCGATGGCGATGGCAATG	CCACATGGATCAGATAAAGCCCAGATTTC	Susceptible
02g19200	GGTGGCGATGGCGATGGCGTTA	CCACATGGATCAGATAAAGCCCAGATTTC	Resistant
LOC_Os-	GCGATCATTGTAATATATCAACAACCTAGATTCAAA	CTTGAGGAGCTCACCATCGCCAAC	Susceptible
02g44104	GCGATCATTGTAATATATCAACAACCTAGATTAAAG	CTTGAGGAGCTCACCATCGCCAAC	Resistant
LOC_Os-	GCCAAGAAGATGGGCGGCGT	AACCAAATCTTCAAAGAACTTGCTTCCAATGT	Susceptible
03g37720	CTAGCCAAGAAGATGGGCGGACC	AACCAAATCTTCAAAGAACTTGCTTCCAATGT	Resistant
LOC_Os-	CGTCTTCAGCTGATCGTCCGCA	GGCTTTCGCATGACAAATAACACAGCTAAATA	Susceptible
06g29700	CGTCTTCAGCTGATCGTCCGCG	GGCTTTCGCATGACAAATAACACAGCTAAATA	Resistant
LOC_Os-	CATCGTCGACTTCAACCAGGACAGCTA	ACCACCCGGGAGAACTCCTCGA	Susceptible
06g28124	TCGTCGACTTCAACCAGGACAGAGG	ACCACCCGGGAGAACTCCTCGA	Resistant
LOC_Os-	GGATACAGGTGACGAGGAATCCCCTTC	CACGCCATGATCAACCTCCGGT	Susceptible
02g54330	TACAGGTGACGAGGAATCCCCACG	CACGCCATGATCAACCTCCGGT	Resistant
LOC_Os-	GGACACAACGGTGACAGTCTGAGCTACA	CAATATTTCTGGCTCAATCATTCTTGCCTG	Susceptible
06g32350	CACAACGGTGACAGTCTGAGCTGCC	CAATATTTCTGGCTCAATCATTCTTGCCTG	Resistant
LOC_Os-	CCGAAAGGATCAGGCTGTGACATTTTATG	TCATTACTGGAATACCATGATGGGGATCAC	Susceptible
04g59540	CGAAAGGATCAGGCTGTGACATTTTCTC	TCATTACTGGAATACCATGATGGGGATCAC	Resistant
LOC_Os-	CGTCGGTGTCGATGATCGCGC	ATGAACACCGGCAACCTCGTCG	Susceptible
09g37880	GAACACCAGCGCCATTGTCTTCC	TGCACGGCCAAGAAGCCGTC	Resistant

APPENDIX C PERMISSION LETTERS

James Silva Garcia <jsilva9@tigers.lsu.edu>

---

student authorship on manuscript

1 message

Oard, James H. <JOard@agcenter.lsu.edu> Wed, Jun 13, 2012 at 11:10 AM

To: "graddeanoffice@lsu.edu" <graddeanoffice@lsu.edu>

Cc: James Silva Garcia <jsilva9@tigers.lsu.edu>

Dear Graduate School Dean:

In February 2010 a manuscript entitled “Association mapping of grain quality and flowering time in elite *japonica* rice germplasm” was accepted for publication by the *Journal of Cereal Science*. Although my graduate student James Silva-Garcia was not listed as first author in that publication, James actually performed the statistical analysis and contributed to the written manuscript we submitted. Therefore, all content of that publication can be credited to James and be included as part of his Dissertation.

Regards,

James Oard

Professor
School of Plant, Environmental, and Soil Sciences
104 Sturgis Hall
Louisiana State University
Baton Rouge, LA 70803
Phone: 225-578-1301
Fax: 225-578-1403

ELSEVIER LICENSE
TERMS AND CONDITIONS

Mar 13, 2012

This is a License Agreement between James Silva ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane. Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	James Silva
Customer address	104 MB Sturgis Hall Baton Rouge, LA 70803
License number	2867220115013
License date	Mar 13, 2012
Licensed content publisher	Elsevier
Licensed content publication	Journal of Cereal Science
Licensed content title	Association mapping of grain quality and flowering time in elite*japonica* rice germplasm
Licensed content author	Samuel A. Ordonez,James Silva,James H. Oard
Licensed content date	May 2010
Licensed content volume number	51
Licensed content issue number	3
Number of pages	7
Start Page	337
End Page	343
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Application of Genetic and Statistical Tools for Improvement of Louisiana Rice
Expected completion date	Aug 2012
Estimated size (number of pages)	120
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD

From: James Silva Garcia jsilva9@tigers.lsu.edu
Date: Tue, Mar 13, 2012 at 2:09 PM
Subject: Permission request to reprint publish paper as part of my dissertation
To: egebhardt@sciencesocieties.org

104 M.B. Sturgis Hall
LSU Campus
Baton Rouge, LA 70803

March 13, 2012

Liz Gebhardt
Managing Editor
Crop Science
5585 Guilford Rd.
Madison, WI 53711-5801

Dear Dr. Gebhardt:

I am completing a doctoral dissertation at Louisiana State University entitled "Application of Genetic and Statistical Tools for Improvement of Louisiana Rice." I would like your permission to reprint below cited research paper as part of my Ph.D. dissertation.

Silva J, Groth DE, Moldenhauer KA, Oard JH (2011) "GGE biplot exploration of resistance to Sheath Blight disease in doubled-haploid lines of rice." Crop Science 51:1028–1035. doi: 10.2135/cropsci2010.10.0612.

Thank you very much.

Sincerely,

James Silva
Ph.D. Candidate
School of Plant Environmental and Soil Sciences
Louisiana State University

13 Mar. 2012

Dear Mr. Silva,

Permission for re-use of the following article:

"GGE biplot exploration of resistance to Sheath Blight disease in doubled-haploid lines of rice" from *Crop Science* Volume 51 (p. 1028-1035),

is granted for use in your dissertation entitled "Application of Genetic and Statistical Tools for Improvement of Louisiana Rice."

Sincerely,

Elizabeth Gebhardt

Managing Editor, *Crop Science*

ASA-SSSA-CSSA

SPRINGER LICENSE
TERMS AND CONDITIONS

Mar 15, 2012

This is a License Agreement between James Silva ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2870230850113
License date	Mar 15, 2012
Licensed content publisher	Springer
Licensed content publication	Theoretical and Applied Genetics
Licensed content title	Identification of candidate genes in rice for resistance to sheath blight disease by whole genome sequencing
Licensed content author	James Silva
Licensed content date	Jan 1, 2011
Volume number	124
Issue number	1
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are the editor of the new work
Order reference number	
Title of your thesis / dissertation	Application of Genetic and Statistical Tools for Improvement of Louisiana Rice
Expected completion date	Aug 2012
Estimated size(pages)	120
Total	0.00 USD

VITA

James Silva-Garcia was born in Santiago de Cali, Valle del Cauca, Colombia, in 1970. He is the first child among five siblings. He begun his formal education at República del Paraguay elementary school, graduating later from República de Israel high school. He was admitted to Universidad del Valle in 1988 and completed a Bachelor of Science (BS) degree in Statistics, graduation in 1994.

Two years before getting his BS degree he started to work as a Research Visitor in the Centro de Investigación de la Caña de Azúcar de Colombia, CENICAÑA. Upon graduation he transferred to the International Center for Tropical Agriculture (CIAT) and worked as a Programmer Statistician contributed to multi-disciplinary teams and collaborated with researchers from several areas (agronomy, genetics, pathology, among others) in designing appropriate experimental procedures, fitting associated statistical models, and presenting research findings.

In Fall 2008 he was accepted in the doctoral program at Louisiana State University and was awarded a research assistantship by the School of Plant, Environmental and Soil Sciences through Dr. James H. Oard, his Major Professor. After four years of hard and diligent work he is finally graduating in Summer 2012.

He is married to Millis D. Arismendi and they have three wonderful daughters Viviana (24 years old), Isabella (10 years old), and Ashley Sophia (1 year old).