DEVELOPMENT OF INTROGRESSION LINES AND VALIDATION OF QUANTITATIVE TRAIT LOCI (QTL) CONFERRING TRANSGRESSIVE VARIATION FOR YIELD IN RICE

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Ize Imai

May 2012

ABSTRACT

A set of introgression lines (ILs) were developed in an advanced backcross population from a cross between the donor, *O. rufipogon* (IRGC 105491) in the genetic background of an elite U.S. variety, cv Jefferson, to confirm the performance of six yield-enhancing QTLs identified from a previous study. At the BC3F4 generation, 50 lines containing the donor alleles at the target QTL regions and the least number of background introgressions were selected for evaluation in standard yield plots at four locations in southern U.S. rice growing regions. The top performing lines were genotyped with fixed SNP assays to provide a higher resolution definition of the size and location of donor introgressions in each IL. Based only on genotypic data, 1-2 plants/family were simultaneously selected for an additional backcrossing to Jefferson and screened with SSRs and InDel markers for selecting individuals with minimum background, and for recombination within the target QTL regions as the basis for fine mapping. NIL status was confirmed using the 384 SNP assay and additional genotyping conducted with SNP markers.

BIOGRAPHICAL SKETCH

Ize Imai is from Tokyo, Japan. She attended the British School in Tokyo and graduated from the American School in Japan in 2004. While attending Macalester College in St. Paul, Minnesota, she was a struggling chemistry student until the end of junior year when Professor Paul Overvoorde offered her a summer research position. After spending a year and a half investigating SAUR (small auxin up-regulated) genes in *Arabidopsis thaliana*, and thoroughly immersed in the elegance and complex reactions of plants, she applied to graduate schools in the plant sciences. By extraordinary luck, she crossed paths with Professor Susan McCouch at a seminar in the University of Minnesota and learned about the approach to increase the yield and genetic diversity of rice by crossing elite cultivars with wild progenitor species. Since joining the McCouch lab in the fall of 2008, there have been many lessons learned in classrooms, laboratories, meetings, and in the greenhouse. Ize believes that these wonderful experiences have prepared her towards an exciting career.

This thesis is dedicated to my family (Kumiko, Seiji, Zeo and my grandparents) for wholeheartedly supporting my desire to study abroad and shouldering the financial and emotional burden of a nonconventional education; my husband, Jacob, for standing by me, and his family for accepting me without reservations; the friends, medical professionals, spiritual guides and mentors that kept me in one body and mind; and finally, my Advisors, Susan and Paul, who gave me wonderful opportunities and taught me how a scientific mind can be cultivated.

ACKNOWLEDGMENTS

First and foremost, I want to express my deepest gratitude to Professor Susan McCouch for her boundless patience and inspirational presence as a mentor, role model and scientist.

I would also like to acknowledge the co-principal investigator in this project, Dr. Anna McClung, for providing phenotypic data of the lines in the field over multiple years, and Dr. Kathleen Yeater for extensive statistical analysis of the data. Also my sincere thanks to my committee members, Professor Mark Sorrells and Professor Rebecca Nelson, for their time and guidance in developing this thesis.

I am also deeply appreciative of Dr. Mark "Koni" Wright, who has provided data for developing SNP markers and taught me many concepts. Dr. Chih-Wei Tung also has my utmost respect for her unrivaled laboratory skills and insights in how to design experiments. Thank you for all the suggestions and help.

We thank Tony Beaty, Laduska Simpson, Sarah Hendrix, Bill Luebke, Curtis Kerns for yield trials at Stuttgart; Jodie Cammack, Kip Landry, Carl Henry, Piper Roberts, and Jason Bonnette for yield trials at Beaumont; Shannon Moon and RiceTec, Inc., for collection of field data in Jonesboro and Alvin; WenGui Yan and Tiffany Sookaserm for sheath blight and straighthead evaluations; and Paul Doran, Hema Liyanage and Bob Shiels from Sequenom, Inc., for technical assistance in developing the iPLEX MassARRAY assays.

Many colleagues, visiting scientists, former and current students in the McCouch lab have contributed to the 'Yield NIL' project over the years (in alphabetical order): Raja Guru Bohar, Joshua Cobb, Benjamin Conway, David Drew, Dr. Adam Famoso, Hyunjung Kim, Jennifer Kimball, Dr. Michael Kovach, Dr. Luis Maas, Dr. Michael Thomson, and John Waldon. Special thanks to Diane Wang for the

KASP optimization protocol and plink instruction tutorial.

Also, none of this would be possible without the supportive staff; Sandra Harrington, Kazi Manna Akther, Gen Fumio Onishi, Cheryl Utter, Lois Swales, and maintenance crew at the Guterman greenhouses; Jean Koski and Janet Myrick.

This project was funded by the National Science Foundation (Grant number 0606461: “Exploring the genetic basis of transgressive variation in rice”).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AMYLOSE = Amylose content

ASV = Alkali spreading value (standard assay to estimate gelatinization temperature and evaluate cooking quality by measuring speed at which grain dissolves)

AVPANL = Average panicle length

AVPANWT = Average panicle weight

AVSDPAN = Average seeds per panicle

AVTILL = Average number of tillers

BC = Backcross

CHKPCT = Chalk percent (percent of grain that shows chalkiness)

CL = culm length (cm) taken from base of plant to the neck node of the first panicle to flower

CW = culm width (mm) taken at 20cm above the base of the plant

D2HD = days to heading (germination time to the date on which the first panicle emerged from its leaf sheath or from a fixed date, 30 days before the first plant in the experiment headed).

GL = Grain length (mm)

GW = Grain width (mm)

GY = grain yield (g) of plant

GxE = Genotype by environment

GxG = Genotype by genotype

HIF = Heterogenous inbred families

KSDWT = thousand-seed weight

IL = Introgression lines

InDel = Insertion Deletion

LD = Linkage disequilibrium

MAB = marker assisted breeding

MAS = marker assisted selection

Mb = Megabase (1,000,000 bp)

MY = Milling yield

NIL = Near Isogenic Line

OPA = oligo pool assay

ORF = open reading frame

PCTSB = percent sheath blight

PL = panicle length (cm); average length of five panicles at harvest, or length of first panicle at heading time or 80% ripening stage

PLTHT = plant height (cm)

PLSQM = plant per square meter

QTL = Quantitative trait loci

RFLP = Random Fragment Length Polymorphism

SIL = Sub-introgression line

SNP = Single Nucleotide Polymorphism

SSR = Simple Sequence Repeats

TotalMY = Total milling yield (filled and unfilled seed)

URRN = Uniform rice regional nursery

WholeMY= Whole milling yield (filled seed only)

WS = water stressed conditions

CHAPTER 1

Literature Review

1. Diversity of rice

The genus *Oryza* is comprised of 23 species, including domesticated Asian rice, *O. sativa,* and domesticated African rice, *O. glaberrima* (Vaughan *et al.,* 2003). These species comprise nine different genome groups, as outlined in Table 1. The AA genome has the largest number of species, including both domesticates, and is characterized by a large amount of eco-geographic differentiation. While most *Oryza* species are diploid (2n=24), nine are allotetraploid (Table 1). The CC and BB genomes (but not the AA genome) have been repeatedly involved in polyploid events, though the reason for this is not well understood. The use of molecular markers has helped to clarify the way genetic variation is partitioned both between and within *Oryza* species. In the case of *O. sativa,* five well-differentiated subpopulations have been identified using SSR and SNP markers (Garris *et al*., 2005; Zhao *et al*., 2010; 2011), and three subpopulations were identified within *O. glaberrima* (Mande *et al*., 2005).

***Oryza* species: their chromosome number, DNA content, genome group and usual habitat.**

Section Complex *Species*	Chromosome number (DNA content [pg/2C])*	Genome group	Usual habitat
Oryza			
***Oryza sativa* complex**			
Oryza sativa L.	24 (0.91–0.93)	AA	Upland to deepwater; open
O. rufipogon sensu lacto† (syn: *O. nivara* for the annual form *O. rufipogon sensu stricto* for the perennial form)	24 (0.95)	AA	(Annual) Seasonally dry; open (Perennial) Seasonally deepwater and wet year round; open
O. glaberrima Steud.	24 (0.87)	AA	Upland to deepwater; open
O. barthii A. Chev.	24	AA	Seasonally dry; open
O. longistaminata Chev. et Roehr.	24 (0.81)	AA	Seasonally dry to deepwater; open
O. meridionalis Ng	24 (1.02)	AA	Seasonally dry; open
O. glumaepatula Steud.‡	24 (0.99)	AA	Inundated areas that become seasonally dry; open
***O. officinalis* complex**			
O. officinalis Wall ex Watt	24 (1.45)	CC	Seasonally dry; open
O. minuta JS Presl. ex CB Presl.	48 (2.33)	BBCC	Stream sides; semi shade
O. rhizomatis Vaughan	24	CC	Seasonally dry; open
O. eichingeri Peter§	24 (1.47)	CC	Stream sides, forest floor; semi shade
O. malapuzhaensis Krishnaswamy and Chandrasakaran	48	BBCC	Seasonally dry forest pools; shade
O. punctata Kotschy ex Steud.	24 (1.11), 48	BB, BBCC	(Diploid) seasonally dry; open (Tetraploid) forest floor; semi shade
O. latifolia Desv.#	48 (2.32)	CCDD	Seasonally dry; open
O. alta Swallen	48	CCDD	Seasonally inundated; open
O. grandiglumis (Doell.) Prod.	48 (1.99)	CCDD	Seasonally inundated; open
O. australiensis Domin	24 (1.96)	EE¥	Seasonally dry; open
Ridleyanae Tateoka			
O. schlechteri Pilger	48	Unknown**	River banks; open
***O. ridleyi* complex**			
O. ridleyi Hook.	48 (1.31–1.93)	HHJJ	Seasonally inundated forest floor; shade
O. longiglumis Jansen	48	HHJJ	Seasonally inundated forest floor; shade
Granulata Roschev.			
***O. granulata* complex**•			
O. granulata Nees et Arn ex Watt	24	GG	Forest floor; shade
O. meyeriana (Zoll. et Mor. ex Steud.) Baill.	24	GG	Forest floor; shade
Brachyantha B.R. Lu			
O. brachyantha Chev. Et Roehr.	24 (0.72)	FF	Rock pools; open

*Data for diploid species from [71] and tetraploid species from [72]. †Many workers have considered that the annual and perennial wild relatives of *O. sativa* should be considered separate species. However, crop complexes consisting of perennial and annual wild relatives together with the cultigen have generally been given sub-specific ranking [11]. Research results suggest that for rice and its relatives, the evolution of annual forms from perennial forms is a local phenomenon, morphologically intermediate types are abundant and no major crossing barriers exist between rice and its close relatives [12]. ‡We refer to the Latin American AA genome as *O. glumaepatula* because this name is widely used in the literature despite the fact that the taxonomy and nomenclature of this species is in a state of flux. No key characters have been found to distinguish this species from perennial *O. rufipogon* [52]. §There have recently been several reports of tetraploid *O. eichingeri*. However, all correctly identified germplasm of *O. eichingeri* that has had chromosome numbers checked by collectors has been diploid ([73]; DA Vaughan unpublished data). #A diploid population of *O. latifolia* from Paraguay has been reported but attempts to confirm this have failed. This report is thus discounted. •Two other species have recently been named within this complex: *Oryza indandamanica* Ellis is restricted to Rutland Island, the Andamans, India, whereas *Oryza neocaledonica* Morat is from the Pouembout region of New Caledonia. The former is a diminutive variant of *O. granulata* and the latter was distinguished primarily on the basis of microscopic epidermal characters. Both species probably warrant sub-specific status only, but further studies of these two taxa are needed. ¥Recently, it has been suggested that the EE genome is the same as the DD genome [24]. However, this has been shown not to be the case [21•,39]. **It has been suggested on the basis of molecular studies of part of the genome that *O. schlechteri* has the HHKK genome [24]. However, recent data indicate that this may not correct [39]. Here, *O. schlechteri* is tentatively placed in section Ridleyanae. Further information is necessary to determine the sectional status of this species.

Table 1-1: *Oryza* species: their chromosome number, DNA content, genome group and usual habitat. Table from Vaughan *et al*. 2003.

2. Genotype

Isozymes and DNA markers have been used as tools for identifying *Oryza* genomes and species since the early 1980's (Second, 1982; Glaszmann, 1987; Wang and Tanksley, 1989). The use of molecular markers for genetic characterization dramatically changed the approach to phylogenetic and evolutionary inference, and made it possible to map genes and quantitative trait loci (QTL) conferring phenotypes of interest based on co-segregation of molecular polymorphisms and phenotypic variation. The ability to identify molecular markers closely linked to genes and QTLs also opened the door to the use of marker-assisted selection in plant improvement. To facilitate mapping studies, phylogenetic reconstruction and molecular breeding applications, a wide assortment of both dominant and codominant marker systems have been developed for rice over the last 30 years and used as tools to dissect traits, identify genes and molecular mechanisms responsible for traits of interest, and to enhance the efficiency of selection in the context of applied plant breeding.

Isozymes

Research on rice isozyme variability began in the mid 1960's in Japan (Shahi *et al.,* 1969; reviewed by Endo and Morishima, 1983). This work was extended by Second and Trouslot (1980), Second (1982; 1985), Glaszmann (1987) and Khush *et al.* (2003). The study by Second (1982) utilized 40 isozyme loci to evaluate 1,948 accessions representing different AA species of *Oryza* and the data were interpreted to support independent domestications of *O. sativa* (in Asia) and *O. glaberrima* (in Africa), as well as independent domestications of the *indica* and *japonica* subgroups within *O. sativa* (Second*,* 1982). Glaszmann (1987) later used 15 isozyme loci coding for eight different enzymes to examine the genetic structure of a diverse panel of 1,688 *Oryza*

sativa accessions and identified six discrete varietal groups or subpopulations (Glaszmann, 1987). Khush *et al.* (2003) evaluated 20,562 accessions at 20 isozyme loci. These studies provided greatly enhanced resolution for evolutionary and phylogenetic evaluation compared to previous work based on morphological characters, but the small number of isozymes remained a limiting factor for plant breeding purposes.

RFLP (Restriction Fragment Length Polymorphism)

RFLPs represented a leap forward in terms of resolution for mapping and potential for plant breeding application (Tanksley *et al.,* 1989). RFLPs are far more numerous in eukaryotic genomes than isozymes and since most are neutral markers, the genetic variation they detect is not generally associated with any phenotypic consequences, unlike morphological markers (Botstein *et al.,* 1980). During the 1980's, RFLPs were used to construct genetic maps, classify germplasm, detect major genes, and identify QTLs. The first 'saturated molecular map' of the rice genome was made through the use of 135 RFLP markers (McCouch *et al.,* 1988). Subsequently, molecular maps of rice with a higher density of RFLP markers were published (Saito *et al.,* 1991; Causse *et al.,* 1994; Harushima *et al.,* 1998). The first studies showing that RFLPs could be used to detect major genes and QTLs were focused on disease resistance and drought-related traits (Wang *et al.,* 1994; Yoshimura *et al.,* 1995; Champoux *et al.,* 1995). RFLP markers were also used to examine genetic variation within and between species of *Oryza* (Wang and Tanksley, 1989; Wang *et al*., 1992, Zhang *et al.,* 1992).

Although RFLPs made many contributions to genetic studies, they were technically cumbersome; turnaround time was slow, and required large amounts of DNA and the use of hazardous radioactive labels to generate marker data. The fact that RFLPs were co-dominant

made them valuable, but a marker system that was faster, easier to handle and would not require as much tissue was highly sought after for breeding applications.

SSR markers (microsatellite, SSLP)

The discovery of *Taq* polymerase and PCR, coupled with the development of new sequencing technologies, allowed for the use of smaller amounts of DNA for genetic analysis and provided the basis for a wide array of new marker systems. The discovery of highly mutable, repetitive sequence motifs called SSR (simple sequence repeats), and the ubiquitous nature of these sequences in eukaryotic genomes paved the way for targeting SSR loci as genetic markers. By designing PCR primers in regions of unique sequence flanking the repetitive SSR motifs, these highly mutable loci could be targeted for use as codominant markers and used for genetic analysis. The fact that SSR loci mutate at a higher rate than unique sequence DNA means that they typically harbor many more alleles per locus than RFLPs or isozymes. As a result, SSRs have significant power to differentiate even closely related varieties, where RFLPs and isozymes were unable to do so (Li *et al.,* 2002). SSR markers could be readily developed by sequencing, and because they were PCR-based, they were technically much easier and faster to use than RFLPs, and they could be multi-plexed in moderately high throughput systems. Thus, SSRs quickly replaced RFLPs for most mapping and genetic analysis in rice during the 1990's and early 2000's. Because of their power to discriminate closely related varieties, fewer markers were required for cultivar identification and they proved useful for many plant breeding applications (McCouch *et al.* 1997; 2002; Ni *et al.,* 2002).

The first SSR markers were identified in rice based on screening of genomic or cDNA libraries to identify clones carrying SSR motifs (Wu and Tanksley*,* 1993; Panaud *et al.,* 1996;

Chen *et al.,* 1997; McCouch *et al.,* 1997; Temnykh *et al.,* 2000). With the sequencing of the rice genome (Goff *et al.,* 2002; Yu *et al.,* 2002), many more simple sequence repeat (SSR) markers were developed and became available to the plant breeding community (McCouch *et al.,* 2002; IRGSP, 2005). Capillary multiplexing of fluorescence-based SSRs markers allowed for higher multiplex levels and accuracy, but required optimization and heralded the introduction of subsequent technologies that were designed for increasingly high throughput applications (Mansfield *et al.,* 1994). In plants, SSR markers proved to be more informative, accessible, efficient and cost-effective than RFLPs, and were widely used for mapping and genetic evaluation for almost two decades (Antoni and Tingey*,* 1993).

In rice, SSR markers shed new light upon the nature of genetic variation and population structure and added a new dimension to previous studies based on isozymes and RFLPs. In a study involving 234 diverse *O. sativa* accessions, 169 SSRs were used to evaluate subpopulation structure and identified five subgroups; *indica*, *aus, temperate japonica, tropical japonica* and *aromatic* (Garris *et al.,* 2005). This was largely consistent with the isozyme groups reported by Glaszmann (1987) but the SSRs clearly differentiated the *tropical* and *temperate japonica* subpopulations, which had not been distinguishable with isozymes. The SSRs also clarified the nature of shared ancestry between *indica* and *aus,* and between *temperate japonica, tropical japonica* and *aromatic,* demonstrating clearly for the first time that the *aromatic* or *basmatic* group shared a more recent common ancestor with *japonica* than with *indica* (Garris *et al*., 2005). Further, the minor isozyme groups identified as 'Group III' and 'Group IV' were shown to cluster with the *aus* subpopulation when SSRs were used. Despite the minor differences, isozyme and SSRs identified groups that were largely consistent within domesticated Asian rice.

SNPs

In some species, such as humans, the level of genetic variation among individuals is low and the number and distribution of SSRs does not suffice for differentiation of haplotypes. SNPs are the most abundant form of variation in eukaryotic organisms, and their high density allows for high resolution, genome-wide coverage. SNPs have been used extensively to understand genetic variation among human populations (Sachidanandam *et al.,* 2001; Haga *et al.,* 2002). As sequencing became more routine, the rice scientific community also started to use sequence information to discover SNPs. Comparisons between the *indica* variety GLA4 and the *temperate japonica* variety Nipponbare were used to identify SNPs in a 2.3 Mb region on chromosome 4 (Han and Xue, 2003). In another study, the 9311 (*indica*) sequences was compared to the Nipponbare genome and provided the basis for detecting 384,341 SNPs (Feltus *et al.,* 2004). SNPs were ever more used to increase the resolution of genetic and evolutionary studies and breeding applications (Caicedo *et al.,* 2007; Huang *et al.,* 2009; McNally *et al.,* 2009; Huang *et al.,* 2010; McCouch *et al.,* 2010; Xu *et al.,* 2010; Yu *et al.,* 2011). In Japan, SNP discovery made through sequence comparisons between the closely related *temperate japonica* cultivars, Nipponbare and Koshihikari, enabled the genomic evaluation of 151 closely related Japanese cultivars (Yamamoto *et al.,* 2009). Previously, there were not enough polymorphisms to do genome-wide marker dissection in this narrow gene pool, but SNPs were abundant enough to dissect complex traits in *temperate japonica* populations.

Automated, high-throughput SNP detection methods allowed researchers to undertake highly parallel experiments involving the genotyping of multiple accessions at the same time. Fixed SNP arrays at various levels of resolution have now been developed for rice (Thomson *et al*., 2011). These include a custom-designed Affymetrix array consisting of 44K SNPs (Tung et

al., 2010), a set of 4357 SNPs discovered from sequencing 1578 genes (Ebana *et al.,* 2010), Illumina GoldenGate assays consisting of 1536 SNPs (Zhao *et al.,* 2010) and 768 SNPs (Yamamoto *et al.,* 2010; Nagasaki *et al.,* 2010), and lower resolution 384 'breeder's chips' based on the Illumina BeadXpress platform (Thomson *et al.,* 2011). A SNP-calling algorithm, ALCHEMY, tailored for highly homozygous inbreeding populations such as rice, was developed to increase the call rate on fixed SNP assays (Wright *et al.,* 2010). Fixed arrays have been particularly attractive for breeders and smaller research groups that do not have the means or access to computational staff, computing facilities or storage capabilities to handle large sequencing datasets. The use of a set of SNPs selected for a particular objective can greatly reduce the time required for analysis and the dependence on computationally sophisticated people and facilities (McCouch *et al.,* 2010).

Sequencing

Improvements in sequencing technology over the past decade have made resequencing of plant genomes cheaper and more accessible to those with the computational power to wield large datasets (Delseny *et al.,* 2010). First-generation sequencing began with whole-genome Sanger sequencing, where long contiguous reads were made using fluorescently dye-labeled terminators. With the introduction of several 'next-generation sequencing' technologies, costs have been significantly reduced, and throughput has been significantly increased by eliminating the tedious growth of *E.coli* colonies, adjusting the read length and improving the efficiency of tiling of the copies (Shendure and Ji*,* 2008). These sequencing systems include Roche 454 pryosequencing which involves pyrophosphate signaling base incorporation by light (Ronaghi *et al.,* 1998), Illumina GAII (Genome Analyzer II, Illumina) and AB SOLiD™ (Sequencing by

Oligonucleotide Ligation and Detection, Applied Biosystems, Beverly, MA) with varied read lengths and run times. PacBio SMRT sequencing (*Single Molecule, Real-Time*) may be considered a "third-generation" sequencing technology; it utilizes modified DNA polymerases, allowing read lengths of several kilobases and shorter run times. These next-generation sequencing technologies and various template and library preparation methods and algorithms for assembly have been extensively reviewed (Metzker*,* 2010). Some limitations of sequencing remain, such as the tedious process of creating fragment libraries, the detection of sequencing errors or biases, depending on the platform used, and the requirements for vast storage capacity due to the shear volume of the datasets (Delseny *et al.,* 2010).

Sequencing technology also provides a form of direct genotyping, based on the 'genotyping-by-sequencing' (GBS) strategy. The earliest form of GBS is known as RAD (restriction-site associated DNA) sequencing; as the name implies, this method targets polymorphisms in DNA directly flanking restriction sites. The strategy uses enzyme-digested DNA that has linkers ligated to the sticky ends, and the linkers contain molecular barcodes and primer annealing sites (Miller *et al.,* 2007). Methylation-sensitive restriction enzymes can be used to reduce complexity because methylated regions are not digested and the large fragments representing methylated genomic regions are removed from the analysis. The use of RAD with methylation-sensitive enzymes is ideal for larger genomes. This reduces the size and complexity of the genome and allows faster SNP discovery and genetic marker development compared to sequencing without complexity reduction (Baird *et al.,* 2008). A variety of next-generation sequencing methods for marker discovery are reviewed extensively by Davey *et al*. (2011**).**

For organisms with high-quality reference genome sequences, such as humans or rice, resequencing short fragments to provide low genome coverage can be combined with imputation

as a viable alternative to complexity reduction. Low coverage resequencing (selective sequencing) of 150 RIL of rice was conducted using imputation from 0.02X coverage and this strategy was used to identify previously-discovered QTLs and genes, such as *sd1*, providing evidence that it could be used effectively (Huang *et al.,* 2009). Compared to 287 PCR-based SSR markers for genotyping a rice population, the study claimed that sequencing-based genotyping was roughly 20 times faster in data collection and 35 times more precise in recombination breakpoint determination. A simpler and more cost effective approach to GBS was developed by Elshire *et al.* (2011), where the restriction enzyme *Ape*K1 was used for complexity reduction and a novel method was used for attaching barcodes, such that either 96-plex or 384-plex sequencing could be done very efficiently. Resequencing and subsequent alignment however, failed to detect SNPs in repetitive regions and in regions that were misaligned or highly diverged from the reference genome (McCouch *et al.,* 2010). An additional constraint of resequencing is the time and cost required for management and storage of large amounts of data, and the computational requirements to draw out what is useful for breeding purposes (Batley *et al.,* 2009).

With so many genotyping options available, the emphasis shifts to the requirement for appropriate mapping populations and types of germplasm and genetic stocks to take advantage of the new potential for rapidly discovering genes of interest for trait improvement and beginning to understand how they interact to determine complex phenotypes. Further advances in genotyping methods are on the horizon, and these will further increase efficiency, accuracy and power of detecting polymorphisms that capture the materials' potential for trait improvement.

3. QTLs: Dissecting phenotypic variation using genetic maps and markers

Information about molecular variability does not lead to crop improvement without linking it to traits of interest. Single-gene traits are phenotypes associated with variation at a single locus, while quantitative or complex traits are associated with variation at multiple loci. Quantitative trait loci (QTL) are identified as correlations between molecular markers and a phenotype of interest detected in a segregating population (Tanksley, 1993; Mackay, 2001). Markers have made it easier to track complex traits by associating portions of the phenotypic variation with individual loci in the genome (Collard *et al.,* 2005). The use of molecular markers has become indispensible to plant breeding in recent years (Bernardo *et al.,* 2008). Marker-assisted breeding (MAB) or marker-assisted selection (MAS) in rice utilizes genetic markers linked to agronomically important genes to improve crops (Collard *et al.,* 2008). MAB utilizing naturally-occurring genetic variation has lead to the release of several varieties with enhanced and novel characteristics. For example, the major QTL *sub1* was cloned from the *aus* cultivar FR13A (Xu *et al.,* 2000) and subsequently bred into commercial elite varieties using MAB to combat yield-loss in flood prone regions (Neeraja *et al.,* 2007; Septiningsih *et al.,* 2009). Various lines containing the *sub1* gene have been tested in farmers' fields in multiple regions and two submergence tolerant varieties were released in 2009: Swarna-*sub1* in India and IR64-*sub1* in Indonesia and Philippines (Manzanilla *et al.,* 2011).

Although the *sub1* example has demonstrated the potential of targeting a major gene or QTL, there have been constraints to incorporating complex traits using similar schemes because multiple genetic loci must be transferred to confer the required benefit (Finkers *et al.,* 2007). Transgenic methods facilitate the rapid transfer of genes that have previously been cloned and characterized, but few agronomically valuable genes are currently available and there are

enormous costs and delays involved in obtaining permits to evaluate transgenic varieties in the field. For rice, marker-assisted breeding utilizing natural variation is running ahead of transgenics in terms of deliverables to farmers. For example, "Golden rice", with high-carotenoid levels imparted through transgenics, is not yet in the hands of farmers over a decade after its first publication (Enserink *et al.,* 2008). In fact, no commercial transgenic rice is yet grown anywhere in the world. Marker-assisted backcrossing to capture favorable alleles present as naturally existing genetic variation within the *Oryza* genus is currently the most practical approach for targeted improvement of rice.

Association mapping

In addition to QTL mapping, two other methods are widely used to identify genetic components underlying complex traits: Association Mapping and Selection Screening as reviewed by Takeda and Matsuoka (2008). Early attempts at genome-wide association mapping in rice were made using RAPDs (Virk *et al.,* 1998) and SSRs (Agrama *et al.,* 2007) but these were largely inconclusive due to lack of marker resolution and control for population structure. With the increased genome coverage provided by SNP markers, association genetics can be more effectively utilized for crop improvement (Mather *et al.,* 2007; Flint-Garcia *et al.,* 2005). The first targeted association mapping study in rice focused on the *xa5* region on chromosome 5, with SNPs every 0.5 kilobase (Garris *et al.,* 2003), but the LD extended for approximately 90 kb and the resolution of the study made it impossible to identify which of the nine genes in the region was "*xa5*" (Blair *et al.,* 2003). Subsequently, a candidate gene-based association study was successfully used to identify the *xa5* gene and the functional polymorphism associated with the resistant phenotype (Iyer and McCouch, 2004).

Genome-wide association studies (GWAS) have been undertaken at various levels of resolution. In a study by Zhao *et al.* (2010), 395 rice varieties were genotyped with 1536 SNPs and phenotyped for grain morphology and grain quality. GWAS demonstrated the ability to identify known major genes. Ebana *et al.* (2010) genotyped 140 diverse varieties with 4,357 SNPs discovered in PCR amplicons of predicted gene regions and evaluated population structure, demonstrating that the *indica* subpopulation was comprised of several sub-groups. In a study by Huang *et al.* (2010), 517 landraces were genotyped with several hundred thousand SNPs and 14 agronomic traits were targeted for GWAS. This investment in low-coverage sequencing did not detect as many novel associations as expected due, in part, to the extent of LD in rice and in part to the imputation methods used. The results largely coincided with the peaks already identified through previous QTL analyses. For example, genes for grain width, *GW3,* and grain size, *GS5,* were detected, but the LOD score peaks were not any sharper (more significant) than those from the GWAS study by Zhao *et al.* (2011) using 44K SNPs, or from previous QTL studies. Association mapping is also constrained by the allele frequencies in the panel of germplasm being evaluated; alleles that occur at low frequency (<10%) are generally masked in the analysis, so QTL discovered in a specific cross are not always detectable in an association mapping panel (Famoso *et al*. 2011).

It is well documented that large blocks of the rice genome are either fixed, or do not undergo recombination in certain crosses (Chen *et al*. 2008; Ouyang *et al*. 2010). For inbred species with extensive LD, such as rice, enhancing the resolution of mapping depends more on increasing the frequency and location of recombination breakpoints than increasing the density of markers (Famoso *et al*., 2011).

Although marker-assisted breeding (MAB) is faster and more economical than conventional breeding (Alpuerto *et al*., 2009), the rice breeding community has been slow to incorporate MAB on a global scale. A review article on marker-assisted selection highlights some of the reasons why marker-assisted selection appears to have made such a small impact in the rice community (Collard *et al*., 2008). This is due to: (1) the fact that it is often not clear whether markers have been used because MAB is not always documented when a variety is released, particularly if the private sector is involved, (2) the restricted scope of current QTL mapping studies means that traits of interest may have no mapping information, (3) the effects of genotype by genotype (GxG) interaction make it difficult to reliably transfer a QTL from one genetic background to another, (4) insufficient understanding of environmental and GxE (genotype by environment) effects, and (5) lack of access to required technology due to financial and/or practical reasons.

Selection for multiple loci in a genetic background of interest

Although many large-effect QTLs have been utilized in the development of varieties with improved qualities, not many studies have tried to utilize the many small-effect QTLs that have been identified. GWAS aims to map multiple loci that contribute to a complex trait. In rice, many key QTLs are subpopulation-specific and many quantitative traits can be enhanced by moving introgressions from one subpopulation to another. This makes GWAS a particularly interesting approach for identifying interacting components that contribute to complex phenotypic variation, particularly when they have never been combined into a single genetic background before (Famoso *et al*., 2011).

Genomic selection (GS) represents a marker-based approach to breeding that requires no prior information about which genes or specific regions of the genome are required to improve a variety's performance. It uses all marker information available to make predictions about performance, based on a model that is developed on a 'training population'. The training population has to be genotyped with a set of genome-wide markers and evaluated phenotypically in a set of target environments. The resulting dataset then serves as the basis for building a model that predicts how a new line derived from a population that is similar to the training population will perform in a similar set of environments, based solely on genotypic information (Lorenz *et al*., 2011). While this approach is still in its infancy in plants, it holds great promise for increasing the efficiency of breeding, as long as the variation that is being recombined in the population is largely additive and all members of the population are already well adapted (Bernardo and Yu, 2007; Heffner *et al*., 2009).

For complex traits that result from non-additive variation, it is difficult to predict with any accuracy how to favorably recombine alleles. QTL mapping has been successfully used to identify transgressive variation whereby selected offspring outperform the better parent. This phenomenon is widespread when divergent parents are used for crossing (Tanksley and McCouch, 1997; Rieseberg *et al.*, 2003; McCouch *et al*., 2007) and the QTLs that cause the transgressive variation can be backcrossed into elite genetic backgrounds or fine-mapped and cloned to determine what genes are involved and how they function (Paterson *et al*., 1991; Eshed and Zamir, 1995; Thomson *et al*., 2006; Maas *et al*., 2010).

4. Fine mapping

Once a QTL is mapped, the genes responsible for the QTL effect can sometimes be cloned based on a fine-mapping approach. Validation of QTL and subsequent fine mapping generally involves backcrossing the target QTL allele from the donor parent into the genetic background of the other parent to develop near isogenic lines (NILs). Backcrossing simplifies the genetic background, which allows the phenotypic difference to be associated with a single region of introgression. NILs have been instrumental in narrowing QTL regions into smaller and smaller chromosomal segments using recombination and marker-assisted selection to, identify the DNA polymorphism that is responsible for the phenotype of interest.

Successful fine mapping can separate a gene of interest from closely-linked donor alleles. When closely-linked loci confer a negative phenotype, the phenomenon is known as 'linkage drag'. A recent example of linkage drag involved a blast disease resistance gene, *PI21*, and a neighboring gene causing poor eating quality (Fukuoka *et al*., 2009). Once the linkage between the positive and negative allele was broken by identifying a recombinant, researchers could begin to utilize the new blast resistance gene in the context of plant breeding.

5. Yield: a composite phenotype

Unlike phenotypes that can be explained by a mutation in one gene, composite phenotypes are controlled by multiple gene interactions. Yield is a composite phenotype that is complicated by multiple interacting components. Virtually all the genes in a plant contribute in one way or another to defining yield potential. Solar radiation, water availability, and nutrient use efficiency, as well as the physiological architecture of the plant and a variety of quality traits are major players contributing to yield (Loomis and Amthor, 1999; Yamamoto *et al*., 2009; Xin-

Guang *et al*., 2010). Many yield-related QTL studies are documented in databases (Ni *et al*., 2009; Yonemaru *et al*., 2010), but relatively few have been utilized for crop improvement. Some important yield genes have been discovered using QTL fine mapping strategies, as reviewed in Chapter 2.

6. Transgressive variation in inbreds at gene level (GxG)

Transgressive variation refers to the scenario in which a cross between divergent parents gives rise to progeny having qualities that exceed the two parental lines. Transgressive variation in inbred crop species have been documented in tomato (Paterson *et al*., 1991; Eshed and Zamir, 1995), barley (Kuczynska *et al*., 2007), wheat (Krystkowiak *et al*., 2009), soybean (Mansur *et al*., 1996) and rice (Xiao *et al*., 1998; Moncada *et al*. 2001; Septininsih *et al*., 2003; Thomson *et al*., 2003; Sarla *et al*., 2005; McCouch *et al*., 2007). Since inbred crops tend to have less genetic variation compared to out-crossing species, it is important to expand the range of variation that is available for selection by breeders and this can be done by selecting divergent materials to serve as parents when making a cross.

The wild species *Oryza rufipogon* is known to be the progenitor of *Oryza sativa*, domesticated Asian rice. Introgressions from wild donors into several domesticated backgrounds have been shown to confer beneficial characteristics due to positive GxG interactions (Oka and Chang, 1961; Oka, 1988; Langevin *et al*., 1990; Tanksley and McCouch, 1997; Perera *et al*., 1998; Marri *et al.,* 2005; McCouch *et al.,* 2007). Utilizing QTLs derived from interspecific crosses is an exciting area for further study and holds great promise for making rapid genetic gain in rice improvement.

References

Agrama, H., Eizenga, G., & Yan, W. (2007). Association mapping of yield and its components in rice cultivars. *Molecular Breeding, 19*(4), 341-356.

Alpuerto, V. E. B., Norton, G. W., Alwang, J., & Ismail, A. M. (2009). Economic impact analysis of marker-assisted breeding for tolerance to salinity and phosphorous deficiency in rice. *Review of Agricultural Economics, 31*(4), 779-792.

Antoni Rafalski, J., & Tingey, S. V. (1993). Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends in Genetics, 9*(8), 275-280.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE, 3*(10), e3376.

Batley, J., & Edwards, D. (2009). Genome sequence data: Management, storage, and visualization. *BioTechniques, 46*(5), 333.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science, 48*(5), 1649-1664.

Bernardo, R., & Yu, J. (2007). Prospects for genome-wide selection for quantitative traits in maize. *Crop Science, 47*(3), 1082-1090.

Blair, M. W., Garris, A. J., Iyer, A. S., Chapman, B., Kresovich, S., & McCouch, S. R. (2003). High resolution genetic mapping and candidate gene identification at the *xa5* locus for bacterial blight resistance in rice (*Oryza* sativa L.). *Theoretical and Applied Genetics, 107*(1), 62-73.

Caicedo, A. L., Williamson, S. H., Hernandez, R. D., Boyko, A., Fledel-Alon, A., York, T. L., et al. (2007). Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet, 3*(9), e163.

Causse, M. A., Fulton, T. M., Cho, Y. G., Ahn, S. N., Chunwongse, J., Wu, K. S., et al. (1994). Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics, 138*(4), 1251-1274.

Champoux, M. C., Wang, G., Sarkarung, S., Mackill, D. J., O'Toole, J. C., Huang, N., et al. (1995). Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. *Theoretical and Applied Genetics, 90*(7), 969-981.

Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., et al. (2008). A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proceedings of the National Academy of Sciences of the United States of America, 105*(32), 11436-11441.

Chen, X., Temnykh, S., Xu, Y., Cho, Y. G., & McCouch, S. R. (1997). Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics, 95*(4), 553-567.

Collard, B., Jahufer, M., Brouwer, J., & Pang, E. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica, 142*(1), 169-196.

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 363*(1491), 557-572.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics, 12*(7), 499-510.

Delseny, M., Han, B., & Hsing, Y. I. (2010). High throughput DNA sequencing: The new sequencing revolution. *Plant Science, 179*(5), 407-422.

Ebana, K., Yonemaru, J., Fukuoka, S., Iwata, H., Kanamori, H., Namiki, N., et al. (2010). Genetic structure revealed by a whole-genome single-nucleotide polymorphism survey of diverse accessions of cultivated Asian rice (*Oryza* sativa L.). *Breeding Science, 60*(4), 390-397.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE, 6*(5), e19379.

Enserink, M. (2008). Tough lessons from golden rice. *Science, 320*(5875), 468-471.

Eshed, Y., & Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics, 141*(3), 1147-1162.

Famoso, A. N., Zhao, K., Clark, R. T., Tung, C., Wright, M. H., Bustamante, C., et al. (2011). Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet, 7*(8), e1002221.

Feltus, F. A., Wan, J., Schulze, S. R., Estill, J. C., Jiang, N., & Paterson, A. H. (2004). An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Research,* 14, 1812-1819.

Finkers, R., van Heusden, A., Meijer-Dekens, F., van Kan, J., Maris, P., & Lindhout, P. (2007). The construction of a *Solanum habrochaites LYC4* introgression line population and the identification of QTLs for resistance to *Botrytis cinerea*. *Theoretical and Applied Genetics, 114*(6), 1071-1080.

Flint-Garcia, S. A., Thuillet, A., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. *The Plant Journal, 44*(6), 1054-1064.

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., et al. (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science, 325*(5943), 998-1001.

Garris, A. J., McCouch, S. R., & Kresovich, S. (2003). Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). *Genetics, 165*(2), 759-769.

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., & McCouch, S. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics, 169*(3), 1631-1638.

Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics, 74*(1), 21-30.

Goff, S. A., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science, 296*(5565), 92-100.

Haga, H., Yamada, R., Ohnishi, Y., Nakamura, Y., & Tanaka, T. (2002). Gene-based SNP discovery as part of the Japanese millennium genome project: Identification of 190,562 genetic variations in the human genome. *Journal of Human Genetics, 47*(11), 605.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., et al. (1998). A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics, 148*(1), 479-494.

Heffner, E. L., Sorrells, M. E., & Jannink, J. (2009). Genomic selection for crop improvement. *Crop Science, 49*, 1-12.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Research, 19*(6), 1068-1076.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics, 42*(11), 961-967.

IRGSP (2005). The map-based sequence of the rice genome. Nature 436: 793-800.

Iyer, A. S., & McCouch, S. R. (2004). The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Molecular Plant-Microbe Interactions, 17*(12), 1348; 7-1354.

Khush, G. S., Brar, D. S., Virk, P. S., Tang, S. X., Malik, S. S., Busto, G. A., et al. (2003). Classifying rice germplasm by isozyme polymorphism and origin of cultivated rice. *IRRI Discussion Paper Series, 46*, 282.

Krystkowiak, K., Adamski, T., Surma, M., & Kaczmarek, Z. (2009). Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis effects in wheat (*Triticum aestivum* L.). *Euphytica, 165*(3), 419-434.

Kuczynska, A., Surma, M., Kaczmarek, Z., & Adamski, T. (2007). Relationship between phenotypic and genetic diversity of parental genotypes and the frequency of transgression effects in barley (*Hordeum vulgare* L.). *Plant Breeding, 126*(4), 361-368.

Langevin, S. A., Grace, J. B., & Clay, K. (1990). The incidence and effects of hybridization between cultivated rice and its related weed red rice (*Oryza* sativa L.). *Evolution, 44*(4), 1000-1008.

Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology, 11*(12), 2453-2465.

Loomis, R. S., & Amthor, J. S. (1999). Yield potential, plant assimilatory capacity, and metabolic efficiencies. *Crop Science, 39*(6), 1584-1596.

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., et al. (2011). Chapter two - genomic selection in plant breeding: Knowledge and prospects. *Advances in agronomy* (pp. 77-123) Academic Press.

Maas, L., McClung, A., & McCouch, S. (2010). Dissection of a QTL reveals an adaptive, interacting gene complex associated with transgressive variation for flowering time in rice. *Theoretical and Applied Genetics, 120*(5), 895-908.

Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics, 35*(1), 303-339.

Mansfield, D. C., Brown, A. F., Green, D. K., Carothers, A. D., Morris, S. W., Evans, H. J., et al. (1994). Automation of genetic linkage analysis using fluorescent microsatellite markers. *Genomics, 24*(2), 225-233.

Mansur, L. M., Orf, J. H., Chase, K., Jarvik, T., Cregan, P. B., & Lark, K. G. (1996). Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Science, 36*(5), 1327-1336.

Manzanilla, D. O., Paris, T. R., Vergara, G. V., Ismail, A. M., Pandey, S., Labios, R. V., et al. (2011). Submergence risks and farmers' preferences: Implications for breeding *Sub1* rice in Southeast Asia. *Agricultural Systems, 104*(4), 335-347.

Marri, P., N, S., Reddy, L., & Siddiq, E. A. (2005). Identification and mapping of yield and yield related QTLs from an Indian accession of *Oryza rufipogo*n. *BMC Genetics, 6*(1), 33.

Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S., & Purugganan, M. D. (2007). The extent of linkage disequilibrium in rice (*Oryza* sativa L.). *Genetics, 177*(4), 2223-2232.

McCouch, S. R., Kochert, G., Yu, Z. H., Wang, Z. Y., Khush, G. S., Coffman, W. R., et al. (1988). Molecular mapping of rice chromosomes. *Theoretical and Applied Genetics, 76*(6), 815-829.

McCouch, S. R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y. G., et al. (1997). Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Molecular Biology, 35*(1), 89-99.

McCouch, S. R., Teytelman, L., Xu, Y. B., Lobos, K. B., Clare, K., Walton, M., et al. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Research, 9*(6), 199-207.

McCouch, S., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septiningsih, E., et al. (2007). Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. *Euphytica, 154*(3), 317-339.

McCouch, S. R., Zhao, K., Wright, M., Tung, C., Ebana, K., Thomson, M., et al. (2010). Development of genome-wide SNP assays for rice. *Breeding Science, 60*(5), 524-535.

McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences, 106*(30), 12273-12278.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics, 11*(1), 31-46.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research, 17*, 248.

Moncada, P., Martínez, C. P., Borrero, J., Chatel, M., Gauch Jr, H., Guimaraes, E., et al. (2001). Quantitative trait loci for yield and yield components in an *Oryza sativa × Oryza rufipogon* BC2F2 population evaluated in an upland environment. *Theoretical and Applied Genetics, 102*(1), 41-52.

Nagasaki, H., Ebana, K., Shibaya, T., Yonemaru, J., & Yano, M. (2010). Core single-nucleotide polymorphisms—a tool for genetic analysis of the Japanese rice population. *Breeding Science, 60*(5), 648-655.

Neeraja, C., Maghirang-Rodriguez, R., Pamplona, A., Heuer, S., Collard, B., Septiningsih, E., et al. (2007). A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theoretical and Applied Genetics, 115*(6), 767-776.

Ni, J., Colowit, P. M., & Mackill, D. J. (2002). Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Science, 42*(2), 601-607.

Oka, H. (1988). *Origin of cultivated rice*. Tokyo; Amsterdam [Netherlands]; New York; New York, N.Y.: Japan Scientific Societies Press; Elsevier; Exclusive sales rights for the U.S.A. and Canada, Elsevier Science Pub. Co.

Oka, H., & Chang, W. (1961). Hybrid swarms between wild and cultivated rice species, *Oryza perennis* and *O. sativa*. *Evolution, 15*(4), pp. 418-430.

Ouyang, Y., Liu, Y., & Zhang, Q. (2010). Hybrid sterility in plant: Stories from rice. *Current Opinion in Plant Biology, 13*(2), 186-192.

Panaud, O., Chen, X., & McCouch, S. (1996). Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza* sativa L.). *Molecular and General Genetics, 252*(5), 597-607.

Paterson, A. H., Damon, S., Hewitt, J. D., Zamir, D., Rabinowitch, H. D., Lincoln, S. E., et al. (1991). Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. *Genetics, 127*(1), 181-197.

Perera, A. L. T., Fahim, M., Sriyoheswaran, S., Dhanapala, M. P., Senadhira, D., & Lawrence, M. J. (1998). Quantitative genetics of rice I. evidence of unexploited genetical variation for yield and other quantitative characters in modern indica cultivars. *Field Crops Research, 55*(3), 245-256.

Rieseberg, L. H., Widmer, A., Arntz, A. M., & Burke, B. (2003). The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences, 358*(1434), 1141-1147.

Ronaghi, M., Uhlén, M., & Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science, 281*(5375), 363-365.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*(6822), 928.

Saito A., Yano M., Kishimoto N., Nakagahara M., Yoshimura A., Saito K., et al. (1991). Linkage map of restriction fragment length polymorphism loci in rice. *Japanese Journal of Breeding, 41*(4), 665-670.

Sarla, N., & Mallikarjuna Swamy B.P. (2005). *Oryza glaberrima*: A source for the improvement of *Oryza sativa*. *Current Science, 89*(6), 955.

Second, G. (1982). Origin of the geneic diversity of cultivated rice (*Oryza* spp.): Study of the polymorphism scored at 40 isozyme loci. *Japanese Journal of Genetics, 57*(1), 25-57.

Second, G. (1985). Evolutionary relationships in the *sativa* group of *Oryza* based on isozyme data. *Genet. Sel. Evol., 17*(1), 89-114.

Second, G., & Trouslot, P. (1980). Electrophorèse d'enzyme de riz (*Oryza* sp.). *Travaux et documents de l'ORSTOM* (pp. 88). Paris: ORSTOM.

Septiningsih, E. M., Prasetiyono, J., Lubis, E., Tai, T. H., Tjubaryat, T., Moeljopawiro, S., et al. (2003). Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theoretical and Applied Genetics, 107*(8), 1419-1432.

Septiningsih, E. M., Pamplona, A. M., Sanchez, D. L., Neeraja, C. N., Vergara, G. V., Heuer, S., et al. (2009). Development of submergence-tolerant rice cultivars: The *Sub1* locus and beyond. *Annals of Botany, 103*(2), 151-160.

Shahi, B. B., Morishima, H., & Oka, H. I. (1969). A survey of variations in peroxidase, acid phosphatase and esterase isozymes of wild and cultivated *Oryza* species. *Japanese Journal of Genetics, 44*(5), 303-319.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology, 26*(10), 1135-1145.

Takeda, S., & Matsuoka, M. (2008). Genetic approaches to crop improvement: Responding to environmental and population changes. *Nature Reviews Genetics, 9*(6), 444-457.

Tanksley, S. D. (1993). Mapping polygenes. *Annual Review of Genetics, 27*(1), 205-233.

Tanksley, S. D., Young, N. D., Paterson, A. H., & Bonierbale, M. W. (1989). RFLP mapping in plant breeding: New tools for an old science. *Nature Biotechnology, 7*(3), 257-264.

Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science, 277*(5329), 1063-1066.

Temnykh, S., Park, W. D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., et al. (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics, 100*(5), 697-712.

Thomson, M. J., Tai, T. H., McClung, A. M., Lai, X., Hinga, M. E., Lobos, K. B., et al. (2003). Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and Applied Genetics, 107*(3), 479-493.

Thomson, M. J., Edwards, J. D., Septiningsih, E. M., Harrington, S. E., & McCouch, S. R. (2006). Substitution mapping of dth1.1, a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. *Genetics, 172*(4), 2501-2514.

Thomson, M., Zhao, K., Wright, M., McNally, K., Rey, J., Tung, C., et al. (2011). High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding,* , 12.

Tung, C., Zhao, K., Wright, M. H., Ali, M. L., Jung, J., Kimball, J., et al. (2010). Development of a research platform for dissecting phenotype-genotype associations in rice (*Oryza* spp.). *Rice, 3*(4), 205-217.

Vaughan, D. A., Morishima, H., & Kadowaki, K. (2003). Diversity in the *Oryza* genus. *Current Opinion in Plant Biology, 6*(2), 139-146.

Virk, P. S., Ford-Lloyd, B. V., Jackson, M. T., Pooni, H. S., Clemeno, T. P., & Newbury, J. H. (1998). Mapping AFLP markers associated with subspecific differentiation of oryza sativa (rice) and an investigation of segregation distortion. *Heredity, 81*(6), 613-620.

Wang, G. L., Mackill, D. J., Bonman, J. M., McCouch, S. R., Champoux, M. C., & Nelson, R. J. (1994). RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. *Genetics, 136*(4), 1421-1434.

Wang, Z. Y., & Tanksley, S. D. (1989). Restriction fragment length polymorphism in *Oryza sativa* L. *Genome, 32*, 1113-1118.

Wang, Z. Y., Second, G., & Tanksley, S. D. (1992). Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theoretical and Applied Genetics, 83*(5), 565-581.

Wright, M., Tung, C., Zhao, K., Reynolds, A., McCouch, S., & Bustamante, C. (2010). ALCHEMY: A reliable method for automated SNP genotype calling for small batch sizes and highly homozygous populations. *Bioinformatics, 26*(23), 2952-2960.

Wu, K., & Tanksley, S. D. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics, 241*(1), 225-235.

Xiao, J., Li, J., Grandillo, S., Ahn, S. N., Yuan, L., Tanksley, S. D., et al. (1998). Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics, 150*(2), 899-909.

Xin-Guang Zhu, Long, S. P., & Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annual Review of Plant Biology, 61*(1), 235-261.

Xu, J., Zhao, Q., Du, P., Xu, C., Wang, B., Feng, Q., et al. (2010). Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome re-sequencing in rice (oryza sativa L.). *BMC Genomics, 11*(1), 656.

Xu, K., Xu, X., Ronald, P. C., & Mackill, D. J. (2000). A high-resolution linkage map of the vicinity of the rice submergence tolerance locus *Sub1*. *Molecular and General Genetics, 263*(4), 681-689.

Yamamoto, T., Yonemaru, J., & Yano, M. (2009). Towards the understanding of complex traits in rice: Substantially or superficially? *DNA Research, 16*(3), 141-154.

Yamamoto, T., Nagasaki, H., Yonemaru, J., Kaworu, E., Nakajima, M., Shibaya, T., et al. (2010). Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics, 11*, 267.

Yonemaru, J., Yamamoto, T., Fukuoka, S., Uga, Y., Hori, K., & Yano, M. (2010). Q-TARO: QTL annotation rice online database. *Rice, 3*(2), 194-203.

Yoshimura, S., Yoshimura, A., Iwata, N., McCouch, S. R., Abenes, M. L., Baraoidan, M. R., et al. (1995). Tagging and combining bacterial blight resistance genes in rice using RAPD and RFLP markers. *Molecular Breeding, 1*(4), 375-387.

Yu, H., Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., et al. (2011). Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS ONE, 6*(3), e17595.

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science, 296*(5565), 79-92.

Zhang, Q., Maroof, M. A. S., Lu, T. Y., & Shen, B. Z. (1992). Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. *Theoretical and Applied Genetics, 83*(4), 495-499.

Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., et al. (2010). Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS ONE, 5*(5), e10780.

Zhao, K., Tung, C., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun, 2*, 467.

CHAPTER 2

Validation of yield-enhancing QTLs from a low-yielding wild ancestor of rice

Acknowledgements

This study represents a collaborative project between Dr. Susan McCouch at Cornell University and Dr. Anna McClung at USDA-ARS Dale Bumpers National Rice Research Center in Arkansas, funded by the NSF-Plant Genome Research Program.

Ize Imai (the author) was responsible for three generations of backcrossing, genotyping and selection to create the introgression lines described in this thesis. Her work was entirely conducted at Cornell University and she had responsibility for all genotyping and line selection on the project since spring 2009.

Field and disease evaluation data has been provided by Anna McClung. Yield trials at Stuttgart, AR were conducted by Tony Beaty, Laduska Simpson, Sarah Hendrix, Bill Luebke, and Curtis Kerns. Yield trials data from Beaumont, Texas were generated by Jodie Cammack, Kip Landry, Carl Henry, Piper Roberts, and Jason Bonnette. Shannon Moon and RiceTec, Inc., collected the field data in Jonesboro, AR and Alvin, TX. WenGui Yan and Tiffany Sookaserm evaluated sheath blight resistance and straighthead. Kathy Yeater at USDA-ARS-SPA analyzed the field data and disease evaluations using SAS models.

At Cornell, Jenny Kimball, David Drew, Adam Famoso, Luis Maas, Michael Kovach and John Waldon contributed to the genotyping of the third backcross generation (BC_3) using marker-assisted selection and/or selection of lines sent to the 2007 field trials. Jenny Kimball used the 1,536 SNP assay and initiated backcrossing to generate the BC_4 generation. Mark Wright provided the list of polymorphic SNPs between the parents of this population, using data

from the 44K SNP array developed by Keyan Zhao. Chih-Wei Tung was instrumental in optimizing the wet lab procedures required to generate data using the SNP arrays. Josh Cobb and Michael Kovach developed InDel markers for the population. Benjamin Conway generated material targeting *yld6.1* and Raja Guru Bohar aided with the selection of lines targeting *yld3.2* and *yld8.1*. Diane Wang and Adam Famoso optimized the KASP marker protocol and PLINK instruction, and with Micheal Kovach, conducted marker-assisted selection of family 85_2-8. The author been has involved in all phases of genotyping on this project since the spring of 2009, and has coordinated, analyzed and interpreted results from the entire team as the basis for writing this thesis.

Abstract

A set of introgression lines (ILs) containing chromosomal segments from *O. rufipogon* (IRGC 105491) in the genetic background of an elite U.S. variety, cv. Jefferson, was developed to confirm the performance of six yield-enhancing quantitative trait loci (QTLs). Fifty BC_3F_3 ILs containing homozygous *O. rufipogon* introgressions across each of the target QTL regions, and as few background introgressions as possible, were selected for evaluation of yield and 14 yield-related traits in standard yield plots over two years at four locations in the southern U.S. Performance of the IL families was compared with three commercial inbreds and one hybrid variety. IL families carrying introgressions from the low yielding wild parent at the QTLs, *yld2.1* and *yld6,* yielded 27.7% and 26.1% more than the Jefferson recurrent parent, and were among the top-yielding entries in consecutive years in the Uniform Rice Regional Nursery (URRN) trials. IL yld2A also performed well under water-stressed conditions in two field locations, and had greater levels of resistance to four diseases in replicated greenhouse experiments. After the

first year of field trials, 10 of the top-performing BC_3F_4 families, representing five of the QTL targets, were genotyped with an Illumina 1,536 assay to define the size and location of wild introgressions. Based on this genotypic data, BC_3F_4 families with the fewest background introgressions were backcrossed to Jefferson and self-pollinated. The resulting BC_4F_2 families were screened with targeted SNP assays to identify individuals carrying homozygous introgressions across the target QTLs. Twelve ILs, representing each of the six QTL targets, have been submitted to the Genetic Stocks *Oryza* Center as the basis for future studies to explore the basis of transgressive variation in these interspecific pre-breeding lines.

Introduction

The development of near isogenic lines (NILs) that incorporate introgressions from genetically divergent donors in the genetic background of elite, high yielding lines is an effective strategy for evaluating the genetic potential of wild and exotic alleles at candidate quantitative trait loci (QTL). This strategy is based on first identifying desirable QTLs in a segregating population derived from a bi-parental cross. Once desirable QTLs are detected, introgression lines (IL) or near isogenic lines (NILs) carrying a single target QTL can be developed through backcrossing and marker assisted selection (MAS) (Young and Tanksley, 1989; Eshed and Zamir, 1995; Yano and Sasaki, 1997). The performance of NILs can be used to evaluate the effect of target QTLs, and NILs can be sib-mated to construct desired combinations of introgressions in the background of an elite cultivar. This strategy has been used as the basis for introgressing exotic alleles into adapted breeding lines of tomato (Paterson *et al*., 1991; Eshed and Zamir, 1995; Monforte and Tanksley, 2000), barley (Feuerstein *et al*., 1990; Kandemir *et al*., 2000; Kuczynska *et al*., 2007), wheat (Briggle, 1968; Krystkowiak *et al*., 2009), sorghum (Cox

and Frey, 1984; Pedersen and Toy, 2001), soybean (Mansur *et al*., 1996) and rice (Tanksley and McCouch, 1997; Ashikari *et al.,* 2005; Sarla and Mallikarjuna Swamy, 2005; Tian *et al.,* 2006; McCouch *et al*., 2007; Ali *et al.,* 2010; Fukuoka *et al.,* 2010; Xu *et al.,* 2010).

A recent study reported a 50% yield increase based on pyramiding three independent introgressions from a wild tomato in the genetic background of a leading market variety (Gur *et al.,* 2011). In rice, yield QTLs introgressed from exotic landraces and/or wild species have been repeatedly demonstrated to confer large and highly significant effects on yield under both irrigated (Xiao *et al*., 1998; Septiningsih *et al*., 2003; Thomson *et al*., 2003) and water-limited conditions (Moncada *et al*. 2001; Bernier *et al.,* 2007; 2008; 2009a; 2009b; Venuprasad *et al.,* 2009; 2011; 2012; Kumar *et al.,* 2009). Under lowland drought stress, ILs out-yielded the susceptible recurrent parents by ~44% and under upland drought conditions by ~93% (Venuprasad *et al.,* 2011). Though none of the genes underlying these highly significant QTLs have yet been cloned, these reports demonstrate that a few genes or QTLs can have a dramatic effect on yield under stress.

ILs also provide the basis for fine mapping and gene isolation. In rice, several yield-related genes that were first identified as QTLs have been fine-mapped and cloned. These include *grain number 1* (*GN1)* (Ashikari *et al.,* 2005); *dense erect panicle 1* (*DEP1)* (Huang *et al.,* 2009); *flag leaf 1 (FL1)* (Wang *et al.,* 2011); *squamosa promoter binding protein-like 14 (OsSPL14)* (Miura *et al.,* 2010; Jiao *et al.,* 2010); *panicle branch number (PBN6)* (Ando *et al.,* 2008; Ohsumi *et al.,* 2011); *grain size 3 (GS3)* (Fan *et al.,* 2006; Takano-Kai *et al.,* 2009); *grain size 5 (GS5)* (Li *et al.,* 2011); *grain width 2 (GW2)* (Song *et al.,* 2007); *grain width 5 (GW5)* (Weng, *et al.,* 2008); *glucosamine-6-P acetyltransferase (OsGNA1*) (Jiang *et al.,* 2005); *spikelets per panicle 3 (SPP3)* and *thousand grain weight 3 (TGW3)* (Liu *et al.,* 2010). The breeding value

of *GN1* was demonstrated by pyramiding the favorable grain number allele from Habataki with the *sd1* allele conferring semi-dwarf stature and high harvest index into a single, elite NIL that performed well under irrigated conditions in Japan (Ashikari *et al.,* 2005). Similarly, QTLs for disease and insect resistance have been successfully cloned: (*Xa-21* (Song *et al.,* 1995), *xa-5* (Iyer & McCouch, 2004; Jiang *et al.,* 2006), *xa-13* (Yuan *et al.,* 2010), *Pi-ta* (Bryan *et al.,* 2000), *Pi-z* (Zhou *et al.,* 2006), *Pi-36* (Liu *et al.,* 2007), *Pi-km* (Ashikawa *et al.,* 2008), *RYMV* (Pinto *et al.,* 1999), etc.) and abiotic stress tolerance (*sub-1* (Xu *et al.,* 2006), *pup-1* (Heuer *et al.,* 2009)) and demonstrated the breeding value of exotic alleles when introduced into elite genetic backgrounds (Collard and Mackill, 2009, Septiningsih *et al.,* 2009).

Based on their experience in rice, Ashikari and Matsuoka (2006) argue that the development of NILs incorporating exotic introgressions into elite genetic backgrounds is a useful strategy for crop improvement. In rice, the relative ease of incorporating new alleles with large effect through marker-assisted backcrossing makes this method a viable approach to rice improvement (Collard and Mackill, 2009; Thomson *et al.,* 2011). However, in wheat, introgression of yield-related QTLs into elite backgrounds showed only moderate gains (Miedaner *et al.,* 2009; Kumar *et al.,* 2010; 2011), and in maize, no major-effect QTLs have been reported for yield or any yield-related trait. In fact, studies in maize indicate that complex traits such as yield, disease resistance and flowering time are typically controlled by many small-effect QTLs with additive effects (Buckler *et al.,* 2009; Poland *et al.,* 2011), requiring a very different strategy for incorporating favorable alleles into elite recurrent parent backgrounds. There are many reasons that can be invoked to explain the observed differences in genetic architecture of quantitative traits among crop species, including mating habit, strength of the domestication bottleneck, extent of linkage disequilibrium (LD), and degree of structural variation among

varieties or accessions within a species (Flint-Garcia *et al.,* 2005; Yu and Buckler*,* 2006; Eichten *et al.,* 2011). These biological and selection differences help explain the diversity of approaches used by plant breeders to improve the performance of their respective crops.

In a previous study on rice, Jefferson, a U.S. *tropical japonica* cultivar was crossed with its wild progenitor species, *Oryza rufipogon,* and QTL analysis was conducted on the BC_2F_2 generation evaluated in four field environments in Arkansas and Texas (Thomson *et al.,* 2003). In the current study, ILs carrying six yield QTLs identified by Thomson *et al.,* (2003) were developed and evaluated in multi-location yield trials in the southern U.S. The best performing BC_3F_3 IL families were selected for further backcrossing using marker-assisted selection. The subsequent BC_4 families retained the target QTLs, but have reduced number and size of background introgressions.

In the present study, we developed introgression lines (ILs) for six yield QTLs through marker-assisted backcross selection; tested ILs in replicated field trials in the southern U.S., at four locations and five years; selected material as genetic stocks and a germplasm release for use in rice breeding programs; and identified SNP and SSR markers to accelerate the introgression of QTLs from this study into new genetic backgrounds.

MATERIALS AND METHODS

Plant Material

The U.S. cultivar, Jefferson (*Oryza sativa* L. *tropical japonica*) (Reg. no. CV-103, PI 593892, RA4829, GS*OR*#301409) is an elite, long-grain, blast disease resistant, semi-dwarf *tropical japonica* variety developed for the U.S. market (McClung *et al.,* 1997) and was released in 1998, while *Oryza rufipogon* (IRGC105491, GS*OR*#311701) is an *aus*-like wild rice ancestor from Malaysia. It has no agronomic traits of interest but was used as the donor because it crosses readily with diverse *O. sativa* varieties (i.e. *indica* and *japonica*) (McCouch *et al.,* 2007).

Fourteen BC_3F_2 families carrying a favorable introgression from *O. rufipogon* (IRGC #105491) at each of the six yield-related QTLs identified by Thomson *et al.,* (2003) were selected as the starting point for this study (Table 2-1). Henceforth we will refer to QTLs in italics and without prefix or suffix (e.g. *yld2.1)*, while we will refer to lines containing the QTLs using the prefix 'IL' and no italics (e.g. IL yld2_A).

Individuals homozygous for the introgressions of interest were backcrossed to Jefferson and self-pollinated (Fig. 2-1; Appendix I: Backcross level of *Jefferson x O. rufipogon* population). During the course of development of the ILs (BC_1 through BC_3), phenotypic selection was conducted to eliminate progeny that were highly sterile and extremely late (Thomson *et al.,* 2003), and had dormancy, shattering, or red bran (A. McClung, USDA-ARS, Dale Bumpers National Rice Research Center, Stuttgart, AR, pers. comm.).

BC_3 introgression lines (ILs) and BC_4 IL derivatives developed for this project were evaluated for yield and yield components in the field at two sites in Arkansas and two sites in Texas during 2007-10, alongside the Jefferson recurrent parent and commercial checks. The BC_3

ILs were also evaluated for resistance to sheath blight disease (*Rhizoctonia solani*) and to the physiological disorder straighthead (see below) in replicated trials.

Backcrossing and marker-assisted selection procedure

For both the BC_3 and BC_4 populations, Jefferson, the recurrent parent, was used as the female during backcrossing and selected ILs were used as pollen parents, except for family 85 (targeting the yield QTL, *yld2.1*), for which asynchrony of flowering required that Jefferson be used as the male. Three Jefferson individuals were planted on a weekly basis starting two weeks before and ending four weeks after planting of ILs to ensure synchrony of flowering. Markers across the target QTL regions were used to screen BC populations each generation to eliminate 50% of plants that did not carry the *O. rufipogon* allele. Selected plants with the target introgression at the QTL were then screened for introgressions to identify lines carrying the least amount of *O. rufipogon* DNA in non-target regions.

Markers and Marker Assays

DNA was extracted from single plants using one of the following methods; (a) Qiagen DNAeasy Plant Tissue Kit following protocols provided by Qiagen Inc. (Valencia, CA), (b) CTAB Method as described by Dietrich *et al.,* (2002), or (c) Extract-N-Amp Plant Kit (Sigma-Aldrich, Saint Louis, MO).

The BC_3 material was genotyped with SSRs and fixed SNP arrays (1,536 SNP and 384 SNP), while the later BC_4 families were also genotyped with InDel and targeted SNP markers using KASP (KBiosciences Ltd, Hertfordshire, UK) and MassARRAY iPLEX® Gold (Sequenom, Inc., San Diego, USA) assays. Genotyping of ILs carrying target QTL

introgressions at *yld2.1* and *yld6.1* was conducted by competitive allele-specific PCR KASP chemistry. Based on SNP positions identified using the 44K SNP Affymetrix rice chip (Zhao *et al.,* 2011), markers were designed for every megabase (three primers per SNP) within the target regions of interest using the online tool, PrimerPicker (http://primer-picker.htm). The list of primers used for KASP marker detection is given in Table 2-2.

Additional genotyping of the *yld2.1* QTL and background introgressions in IL 43_1-2 and target QTL region of the *yld6.1* were conducted with MassARRAY iPLEX® Gold. Three multiplexes (26-28 SNP per assay) overlapping SNP targets as KASP and additional targets from the 44K chip as needed for SNP targets were designed at 0.5 Mb intervals in the target regions (Table 2-3). The protocol is outlined in the application notes found online (Oeth *et al.,* 2005; http://www.sequenom.com/sites/genetic-analysis/applications/snp-genotyping).

Different DNA extraction methods were used with different marker assays. For MassARRAY iPLEX® Gold assays, CTAB extractions included additional treatments of RNAse (1 ul of 10mg/ml RNAse with 100 μl of TE buffer for 1 hour in 37°C bath after resuspending DNA pellet) and dissolving DNA in 50 μl of AE buffer (Qiagen) to enhance the quality and reliability of the assays. For KASP assays, the Extract-N-Amp procedure was used with additional overnight dry-down step with 2 μl of extracted DNA diluted in 98 μl of ddH_20 placed in 384-well PCR plate (D. Wang, Cornell Univ., pers. comm.). For use with genome wide SNP assays, including the 44K SNP array (Zhao *et al.,* 2011), the 1,536-SNP OPA (oligo pool assay) (Zhao *et al*., 2010) or the 384 SNP OPA (VC0011530-OPA) (Thomson *et al*., 2011), we used DNA samples extracted using Qiagen DNAeasy Kits.

SSR markers used to detect the presence/absence of target introgressions were selected from the pool of SSRs reported for rice (IRGSP, 2005) and available on the Gramene database

(http://www.gramene.org/). InDel markers were identified based on a comparison of the Nipponbare and 9311 sequences across the target regions (http://www.ncbi.nlm.nih.gov/). A list of primers used for InDel marker detection is given in Table 2-4.

Phenotyping of Yield and Yield Components

For each of the six targeted introgressions, four to twelve BC_3 IL families were developed along with two to five sib-line families that lacked the target introgression. During 2007-2008, these 50 ILs and 20 control sib-lines were evaluated at four different locations within two major U.S. rice-growing states: Stuttgart (AR), Jonesboro (AR), Beaumont (TX), and Alvin (TX) in flooded paddies. In 2008, eight NILs were removed because they were observed to possess red bran (Table 2-5). The Jefferson parent and three commercial checks varieties were included: inbred cvs (Cocodrie (PI606331) and Trenasse (PI641796)) and the hybrid cultivar XL723 (RiceTec, Alvin, TX). The experiments were conducted using a randomized complete block design with three replications. Each plot was approximately 5.8 m^2 and was drill seeded at approximately 45 kg ha^{-1}, a low seeding rate commonly used for hybrids. Prior to planting, 112 kg ha^{-1} of fertilizer nitrogen was applied and incorporated. Plots were flush-irrigated until stand establishment and then maintained under a permanent flood. Common pesticides were used to manage weeds throughout the season. In addition, during 2008, the BC_3 ILs were evaluated at Stuttgart, AR and Beaumont, TX under water-stressed conditions. The same procedures were used as in the flooded plots except that, after irrigation to achieve stand establishment, the plots were subsequently flush-irrigated only after the ground had dried to the point of cracking. Plots were evaluated for days to heading, plant height, and lodging percentage. Plots were harvested at approximately 18-20% grain moisture with a combine harvester to determine grain yield (kg ha^{-1}

adjusted to 12% moisture). At early tillering, two plants in each plot were identified for use in yield component measurement and were hand harvested prior to plot yield determination.

Yield components included average panicle length (AVPANL), average seeds per panicle (AVSDPAN), average panicle weight (AVPANWT), average tiller number (AVTILL), average plant weight (AVPLTWT), and 1000 seed weight (KSDWT). Days to heading (D2HD), plant height (PLTHT), lodge percent, plants per square meter (PLSQM) and percent stand (STDPCT) were measured. Cleaned rough rice samples (125 g) were milled using a McGill No. 2 mill (Rapsilver Supply Co Inc., Brookshire, TX) according to a standard protocol for long grain rice. Total milled rice percentage (TotalMY) was determined using the weight of the whole plus broken kernels as a proportion of the rough rice sample. The milled rice was separated using a #12 screen (Seedburo Equipment Co., Chicago, IL) and the weight was used to determine the whole milling yield (WholeMY) as a percent of the rough rice. Grain length (GW), grain width (GW), and percent chalk (CHKPCT) were determined using a WinSeedle Pro 2005a™ image analysis system (Nelson *et al*., 2011) Apparent amylose content (AMYLOSE) was determined using modified iodine spectrophotometric method of Perez and Juliano (1978) with a continuous-flow analyzer (AutoAnalyzer 3 Seal Analytical, Mequon, WI, USA). Alkali spreading value (ASV), an indicator of starch gelatinization temperature, was determined on six milled kernels using sodium hydroxide digestion according to the methods of Little *et al.* (1958).

During 2009, selected BC_3 ILs were evaluated in six locations as part of the Uniform Regional Rice Nursery (URRN): Crowley, LA, Stuttgart, AR, Malden, MO, Stoneville, MS, Beaumont, TX, and Eagle Lake, TX. Among a set of breeding lines evaluated in the URRN were two of the cultivars that were used in the 2007-2008 field trials, Cocodrie (released in 2000) and Trenasse (released in 2006), along with Presidio (PI 636465), a 2008 release that is a derivative

of Jefferson. The study was conducted as a RCB with two to four replications. Yield plots were approximately 5.8 m^2 and were drill seeded using a 125 kg ha^{-1} seeding rate. Fertilizer and pesticides were used according to local recommendations to maximize yield. A wide range of agronomic, grain quality, and disease resistance traits were determined.

In addition, the URRN was screened with markers associated with major genes for blast resistance and for the grain quality traits amylose content, gelatinization temperature, and aroma. DNA was extracted from 20 mg of brown rice using the modified CTAB method described in Fjellstrom *et al*. (2004). Five markers were used to detect known blast resistance genes: AP5659-1 and AP5659-5 associated with the presence of *Pi-z* (Fjellstrom *et al.,* 2006), RM208 associated with the presence of *Pi-b*, RM224 associated with the presence of *Pi-k* (Fjellstrom *et al.,* 2004), and Pi indica, a marker indicative of the resistant *Pi-ta* allele (Wang *et al.,* 2009). RM190, an SSR marker associated with the granule-bound starch synthase gene, was used to screen for amylose content (Chen *et al.,* 2008, Bao *et al.,* 2006) along with a marker associated with the WxEx6 functional SNP in the Waxy gene (Chen *et al.,* 2010). Specific markers were developed to detect the 8 bp deletion that is associated with the presence of grain aroma (Kovach *et al.,* 2009) and the GC/TT SNP in the *Alk* gene, the most common SNP for gelatinization temperature (Bao *et al.,* 2006). For each marker, controls were included for each known allele. PCR reactions were performed in 25 μl reaction volumes consisting of 20 ng of DNA, 10 mM Tris–HCl pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 300 nM of each primer, 1 U of Taq DNA polymerase (Promega, Madison, WI, USA). Reverse primers were unlabeled in order to reduce the cost, and the forward primers were labeled with either 6FAM, Tamra, or Hex (Integrated DNA Technologies, Coralville, IA, USA). DNA was amplified with MJ Research Tetrad PCR machines (Waltham, MA, USA) under the following conditions: initial denaturation at 94°C for

5 minutes; then 30 cycles of 94°C for 30 seconds, 55–67°C (dependent on the marker) for 30 seconds, and 72°C for 1 minute; 5 minute final extension at 72°C. PCR products were pooled based on sizes of amplified fragments (typically three markers per run along with ROX-labeled size standard) to reduce the cost, and the DNA was denatured by heating samples at 94°C for 5 minutes. The samples were separated on an ABI Prism 3100 DNA analyzer using methods as described by the manufacturer (Applied Biosystems, Foster City, CA, USA). The sizes of SSR fragments were estimated using the software GeneMapper v 3.7 (Applied Biosystems, Foster City, CA, USA).

Sheath blight resistance was evaluated on 44 BC_3F_6 ILs representing each of the six targeted introgressions and 17 sib-lines (Table 2-5) along with the check cultivars that were very susceptible (Lemont), moderately susceptible (Jefferson, Cocodrie, Wells, and Spring), and moderately resistant (Jasmine 85). The study was conducted using RCB design with three replications using the micro-chamber method under controlled conditions in a greenhouse during 2009 (Jia *et al.,* 2007). Each experimental unit consisted of one pot with three seedlings, and eight pots arranged in a row-column order within one watering tray. Three-week old plants were inoculated with the sheath blight pathogen, *Rhizoctonia solani* (RR0140-1), and after about two weeks in the microchamber, plant height and lesion length were measured as described in Jia *et al.,* (2007). The percent infected area (PCTSB) was determined by lesion length divided by plant height. Disease index is calculated as lesion length divided by plant height and multiplied by nine; scores of less than four are considered moderately resistant.

Straighthead resistance was evaluated during 2010 in a RCB field trial conducted at Stuttgart, AR with three replications. Prior to planting, monosodium methanearsonate (MSMA) was applied at a rate of 6.7 kg ha^{-1} and incorporated into the soil. Each BC_3 IL was drill seeded

as a single row plot 0.6 m^2. The parent, Jefferson (moderately resistant), and Cocodrie (highly susceptible) were included as repeated checks. Plots were managed according to standard cultural practices for flooded rice. Approximately two weeks prior to maturity, plots were scored for straighthead severity using a 0 to 9 scale with (0= over 90% of spikelets filled and no panicle distortion, 5= 20-40% of spikelets filled, 16-20% of the panicles are erect with some panicle distortion, and 9= no panicles emerged) (Yan *et al.,* 2005).

Statistical Analysis (Conducted and written by Kathy Yeater, USDA-ARS-SPA, TX)

In the irrigated system, all variables were analyzed using the MIXED Procedure in SAS Version 9.2. The fixed effects were line (entry), location, and their interaction, and block effect and year effect and their interactions were treated as random effects. The water-stressed analysis did not include a year effect. Least-squares means and differences of the line means compared to Jefferson and other controls were calculated using the LSMEANS option and diff test with Dunnett's adjustment. Multiple comparison of line to line was achieved with the pdiff test option and family-wise error rate was controlled with the Tukey-Kramer adjustment for p-values.

Regression analyses were completed with the REG Procedure in SAS Version 9.2 utilizing the forward selection method. Pearson Correlation calculations were achieved with the CORR Procedure in SAS Version 9.2.

In assessing genotype x environment (GxE) interactions, the irrigated plots from 2008 in Beaumont and Stuttgart and water-stressed plots from the two locations were combined for analysis. The MIXED Procedure in SAS Version 9.2 was utilized, with location and irrigation method and their interaction defined as fixed effects, and block effect and its interactions as the random effects.

Huhn's Nonparametric Stability statistics, based on ranks of lines in each environment, were estimated using the MEANS Procedure and RANK Procedure in SAS Version 9.2, similar to SAS coding statements available in Lu (1995).

Sheath blight response data was analyzed using the GLIMMIX Procedure in SAS Version 9.2. The line effects were treated as a fixed effect, and the blocking effects (row and column layout) of the greenhouse study were random effects. Least-squares means and differences of the line means compared to Jefferson and other controls were calculated using the LSMEANS option and diff test with Dunnett's adjustment.

Straighthead rating data was also assessed using the GLIMMIX Procedure in SAS Version 9.2. The line effects were treated as a fixed effect, and the replicate block effect was the lone random effect. Least-squares means and differences of the line means compared to Jefferson and other controls were calculated using the LSMEANS option and diff test with Dunnett's adjustment.

Data from the 2009 URRN trial conducted at six locations was analyzed using the MIXED Procedure (SAS Version 9.2) with State and Entry, and their interaction considered as fixed effects, and the replication and its interaction with state and entry as random effects. Significant covariate effects were found for yield and height and for days to heading with whole and total milling yields. These covariates were used to determine adjusted least squares means.

RESULTS

Multi-location field trials (2007-8)

Fifty BC_3 and BC_4 ILs carrying six different yield QTLs (on chromosomes 1, 2, 3, 6, 8 and 9) from *O. rufipogon* were developed in the Jefferson background over the course of this project (Fig. 2-1). When evaluated over two years (2007-2008) in multi-location yield trials in the southern U.S., ILs carrying introgressions across *yld2.1* and *yld6.1* (previously reported by Thomson *et al.,* 2003), consistently outperformed the recurrent parent, Jefferson, in grain yield per plot (Table 2-6). The other four ILs, with introgressions on chromosomes 1, 3, 8, and 9, out-performed Jefferson in some environments but not others.

All 50 of the BC_3 ILs carried introgressions across the target QTL regions, and in addition, random background introgressions on other chromosomes that were identified using SNP panels (Fig. 2-2). The best line, IL yld2_A (family 43_1-2), yielded 27.7% better than Jefferson based on average grain yield across four locations during 2007-2008 (Table 2-6). This line contained an 11.7 Mb target introgression on chromosome 2 plus four background introgressions (Fig. 2-2b). The second best line, IL yld6_A (family 219_2-9), out-yielded Jefferson by 26.1% and contained a 14.2 Mb introgression on chromosome 6 plus two background introgressions (Fig. 2-2d). There was no significant reduction in yield performance based on the presence of spurious *O. rufipogon* introgressions in the genetic background of these families. For example, sister lines IL yld2_A and IL yld2_B (family 43_2-12) were genetically identical, except that IL yld2_B had fewer non-target introgressions (Fig. 2-2b). It lacked two of the background introgressions (a 9.2 Mb introgression on chromosome 5 and a 1.5 Mb introgression on chromosome 9) but carried the same *yld2.1* introgression across the target region. The yield of IL yld2_B was not significantly different than IL yld2_A in the multi-

location yield trials in 2007-8 (Table 2-6). Thus, we conclude that the additional background introgressions did not affect yield performance.

Under flooded conditions, there was significant genotype by location effects ($p < 0.0001$). ILs grown in the fields of Arkansas (Jonesboro and Stuttgart) out-yielded the ILs grown in Texas (Alvin and Beaumont), and the average performance of ILs grown in all locations during 2008 was cumulatively higher than in 2007 (Fig. 2-3).

To better evaluate the impact of the target QTL introgressions on yield performance, 20 'control' lines were selected from within each of the segregating BC_3 families. These control lines retained the random assortment of *O. rufipogon* background introgression but lacked the *O. rufipogon* introgression at the peak marker for each QTL (Table 2-5). This provided a way of testing the effect of only the target introgression for each yield QTL. With one exception, the yield performance of the controls was consistently lower than that of the corresponding ILs that carried the target introgression (Fig. 2-4). This confirmed that the superior yield performance of the ILs was due to the presence of *O. rufipogon* DNA across the target QTL regions. In the case of *yld9.1*, sib-lines from families 9, 13, and 16 provided an interesting set of contrasts. For family 9, controls yielded better than the QTL-containing ILs, while for family 16, the reverse was true (Fig. 2-4). Close inspection of these lines confirmed that the poor performance of lines carrying *yld9.1* in family 9 (9_2-9 and 9_2-10) was not indicative of the value of the *yld9.1* QTL. ILs 9_2-9 and 9_2-10 were both very tall (av. 127 cm) and had relatively high lodging (avg. 10%), compared to the controls (avg. 111 cm height and 4% lodging) while lines from family 16 with *yld9.1* (IL 16_1-1, 16_1-2 and 16_1-10) had an average height of 89 cm and 3% lodging. The excessive height and lodging of ILs carrying *yld9.1* in family 9 accounted for its poor yield performance in the field.

Evaluation under water-stressed conditions (2008)

To explore the phenotypic plasticity of the interspecific ILs, we evaluated them under both flooded and water-stressed (WS) conditions during 2008 in Stuttgart, AR and Beaumont, TX. Plants grown in water-stressed conditions yielded an average of 23.3% less than the entries in flooded conditions. Family 43 (*yld2.1*), however, performed better than other ILs and Jefferson under both irrigated and water-stressed conditions and performed as well as the inbred varieties, Trenasse and Cocodrie (p-value = 0.1044) (Fig. 2-5). The hybrid XL723 yielded significantly better than any of the inbreds under both irrigated and WS conditions. Our data suggests that the introgressions from *O. rufipogon* mitigate the effect of WS on grain yield in the Jefferson background, conferring enhanced yield potential under both well-watered and WS conditions. This introgression may have potential for enhancing yield in other genetic backgrounds.

Regression analysis was used to identify the traits that were most strongly correlated with yield among the 50 ILs and 20 controls in this study. Eight traits explained 38% of the variability for grain yield under irrigated conditions (Table 2-7A). The first five traits, panicle length, plant weight, apparent amylose content, seed weight, and grain length explained 35% of the variation, indicating that increased yield was associated with longer panicles, heavier and shorter seed, and lower amylose content (Table 2-8A). Under WS conditions, regression analysis identified six parameters that explained >58% of the variation for yield, and four of them (seed weight, panicle weight, plant height, and tiller number) accounted for the majority of the variation (Table 2-7B). Correlation analysis of lines under water-stress conditions revealed that fewer traits are significantly correlated with each other compared to irrigated conditions (Table 2-8B).

Yield evaluation in the Uniform Rice Regional Nursery (URRN) in 2009

The best performing BC_3 ILs, yld2_A (family 43_1-2) and yld6_A (family 219_2-9), were compared to a set of elite lines from the southern U.S. public rice breeding programs as part of the URRN during 2009. IL yld2_A ranked above Trenasse, Cocodrie, Presidio and yld6_A when yield performance was averaged across six locations. Despite its higher yield performance, IL yld2_A had poorer whole-milling yield than yld6_A (p-value = 0.00248) or any of the commercial checks (Table 2-9). IL yld6_A had similar whole-milling yield as the commercial varieties, Cocodrie and Presidio, though it yielded slightly less than Trenasse (p-value = 0.0654). It is noteworthy that the seeding rate of the yield trials in 2007-8 had been 45 kg ha^{-1}, while the URRN plots were 125 kg ha^{-1}, suggesting that wider spacing may favor yield performance of IL yld2_A. Amylose content (20-22%), alkali spreading values (4), and their associated markers (RM190 and *Alk*, respectively) indicated that both ILs had conventional long grain cooking quality similar to Jefferson.

Disease resistance evaluation

The URRN also provided an opportunity to evaluate the incidence of disease on ILs yld2_A and yld6_A in the field. Data was obtained for eight diseases using a visual disease index and artificial or natural inoculation depending on the disease as summarized in Table 2-10. IL yld2_A (family 43_1-2) demonstrated moderate levels of resistance to sheath blight, panicle blast, leaf smut, and narrow brown leaf spot, while IL yld6_A (family 219_2-9) was moderately resistant to leaf smut and straighthead (in MSMA treated soil), but was highly susceptible to bacterial panicle blight. Using markers, AP5659-1, AP5659-5 and RM224, IL yld2_A (family

43_1-2) was confirmed to carry the blast resistance genes, *Pi-z* and *Pi-km* (on chromosomes 6 and 12, respectively), inherited from Jefferson, while IL yld6_A (family 219_2-9) only carried the *Pi-km* gene.

In a separate test using the coke bottle evaluation method under greenhouse conditions, we evaluated the average percent sheath blight (PCTSB) for 44 ILs and 17 sib-lines. The disease pressure in this test was very high, and results identified two sib-lines, 16_2-3 (PCTSB = 48.8%, adj. p-value = 0.03) and 16_1-6 (PCTSB = 49.0%, adj. p-value = 0.03) that were significantly more resistant than Jefferson (Table 2-11). These lines do not carry any yield QTLs, but can be used to identify introgressions from *O. rufipogon* that confer sheath blight resistance and are likely to be useful as donors in future breeding efforts (PCTSB = 76.0%). Neither IL yld2_A (PCTSB = 64.5%) nor IL yld6_A (PCTSB = 75.3%) were highly resistant in the greenhouse evaluations, but they were moderately resistant in the URRN trials and performed as well or better than the commercial checks, Presidio, Trenasse and Cocodrie.

In 2010, straighthead was also evaluated based on the addition of MSMA in the field (see Materials and Methods). None of the ILs were more resistant than Jefferson, though all were more resistant or equally resistant than Cocodrie.

Grain quality evaluation

During 2007-8, we collected data on 17 agronomic and grain quality traits, in addition to yield. These data provided additional information about the BC_3 ILs (Table 2-12). Compared to Jefferson, IL yld2_A (line 43_1-2) showed no significant difference in milling yield, amylose content, alkali spreading value (ASV), chalk, grain morphology (length and width), panicle length, weight, number of seeds per panicle, plant height or heading date (Table 2-12).

Interestingly, a sib-line, IL yld2_B (line 43_2-12) had longer average panicle length ($p=0.0004$) and higher chalk ($p<0.0001$) compared to Jefferson, and narrower grains ($p< 0.002$) than yld2_A. IL yld6_A (line 219_2-9) had shorter grain length compared to Jefferson ($p< 0.0001$), shorter average panicle length ($p=0.02$), less panicle weight ($p<0.01$), lower milling yield, lower amylose ($p<0.0147$), lower alkali spreading value ($p<0.0123$), and was higher in chalk ($p< 0.0001$), but it showed no significant difference in plant height or flowering time.

The most serious problem associated with many of the other ILs was significantly higher levels of chalk and small differences in grain morphology and/or grain weight (Table 2-12). Interestingly, none of the selected ILs were significantly different in plant height or flowering time, and panicle characteristics remained stable, suggesting that the introgressions from *O. rufipogon* affected basic physiological components of yield performance without dramatically affecting plant architecture or morphology in the Jefferson background.

Submission of ILs to the Genetic Stocks *Oryza* Collection (GS*OR*)

Twelve ILs developed as part of this study were deposited in the Genetic Stocks *Oryza* (GS*OR)* Collection (http://www.ars.usda.gov/Main/docs.htm?docid=8318) for use in future genetic analysis and as parents in breeding programs (Table 2-6). These lines contain *O. rufipogon* introgressions across each of the six target QTL regions, and have minimal *O. rufipogon* DNA in the genetic background. Genotype data based on whole-genome SNP assays (see Materials and Methods) provided information about the size of the introgressed regions and the number of background introgressions in each of the ILs. This data was used to compute the percent recurrent parent in each of the lines (Table 2-6). When comparing the percent donor genome from the BC_2F_2 genotypic data generated by Thomson *et al.* (2003), and SNP genotypes

of BC_3F_4 and BC_4F_2 (using 1,536 and 384 SNP array, respectively), the decrease in percent genome content is comparable or faster than was expected. The percent donor genome in the BC_2F_2 generation averaged 18.7% (range = 8.1% to 30.8%), BC_3F_4 averaged 7.67% (ranging from 2.9% to 16.4%) and BC_4F_2 averaged 6.21% (with range from 4.9 to 9.7%). The relative decline from BC_3 to BC_4 level was less dramatic compared to the reduction of percent donor genome from BC_2 to BC_3 because of uneven density of polymorphic SNPs between the parental genotypes on the 1,536 SNP array, leading to underestimated values of donor genome content.

When subsequent backcrosses were made with selected BC_3 ILs to reduce the number of *O. rufipogon* background introgressions, the loss of background generally did not affect the yield performance of the lines (Fig. 2-6). Thus, the yield of BC_4 IL derivatives was similar to the corresponding BC_3 ILs.

The 12 ILs were at different stages of backcrossing (BC_3-BC_4). Some of the lines containing multiple background introgressions outperformed siblings that contained fewer donor fragments. The top performing line, IL yld2_A (family 43_1-2), contained a target introgression at *yld2.1* between 4.36-16.1 Mb on the short arm of chromosome 2 plus four background introgressions (Fig. 2-2b). This line out-yielded the Jefferson recurrent parent by 27.68% in the flooded yield trials during 2007-8 (Table 2-6). The second-best performing line, yld6.1_A (IL 219-2-9), contained a large target introgression at *yld6.1* between 6.3-20.5 Mb plus two additional background introgressions. This line out-yielded the Jefferson parent by 26.09% in these same trials.

Additional representatives of these high yielding families were selected for submission to the GS*OR*. These included IL yld2_B, yld2_C and yld2_D. IL yld2_B (family 43_2-12) was the second-best yielding line, out-yielding Jefferson by 21.2%. It was from the same family as IL

yld2_A and contained the same target introgression but contained only two of the four background introgressions (Fig. 2-2b). IL yld2_C (43_1-2_7) is a BC_4 derivative of yld2_A with fewer background introgressions; IL yld2_D (IL 85_2-8_16-8) was selected from family 85 and carried a target introgression that was smaller and slightly offset compared to family 43, in addition to two background introgressions on chromosomes 5 and 8. Surprisingly, this line yielded 11.4% less than Jefferson during 2010 (Fig. 2-6) and was observed to be highly susceptible to straighthead in non-MSMA treated plots (data not shown). Further work is necessary to determine whether the straighthead susceptibility in IL yld 2_D is due to the loss of an *O. rufipogon* introgression on either chromosome 9 or 12 (compared to lines from family 43), or to the gain of *O. rufipogon* DNA across an extended target region on chromosome 2 (Fig. 2-2).

For *yld6.1*, we included two additional representatives of family 219. IL yld6_B (family 219_1-5) contained the same size target introgression as IL yld6_A, along with two background introgressions, including a large introgression on chromosome 11. IL yld6.1_C (219_1-5_29-7) was a backcross derivative of IL yld6_B, for which genotypic selection allowed us to identify a line containing no detectable background introgressions. There was no difference in the performance of IL yld6_B and yld6_C, and both out-yielded Jefferson by more than 17% during 2010 (Fig. 2-6).

IL yld3_A (family 16_2-1), IL yld8_A (family 121_2-2), and IL yld9_A (family 13_1-1) outperformed Jefferson by between 15.4-17.5% (Table 2-6) when averaged across four locations and two years. These data confirmed the earlier findings of Thomson *et al.* (2003). IL yld1_A (family 158_2-7) performed well during 2007-8 but yielded less than Jefferson in recent field trials (Fig. 2-6). Of the 12 lines being deposited in the GS*OR,* eight were selected after the BC_3

multi-location yield trials to represent each of the six QTL targets, and the remaining four lines represent BC_4 derivatives from the same families.

SNP Assays to Facilitate Utilization of ILs in Breeding

A total of 42 KASP primer sets and three separate MassARRAY iPLEX assays (81 markers) were designed based on polymorphic SNP positions identified by the 44K SNP array (Zhao *et al.,* 2011; Fig. 2-7 and 2-8, Table 2-13). These SNP markers were chosen to facilitate selection for the *O. rufipogon* target introgressions for IL yld2_A and IL yld6_A, and against *O. rufipogon* introgressions elsewhere in the genome. Markers were selected at approximately 0.5 Mb intervals (Table 2-2 and 2-3). This marker set provides an economical and efficient selection platform so that breeders can readily utilize these ILs as parents in future plant improvement. We also tested the MassARRAY SNP genotyping platform using three sets of DNA samples: Qiagen Plant mini kit extracted DNA (set 1), chloroform extracted DNA (set 2), and various concentrations of lesser quality Extract-N-Amp DNA samples (set 3 ~ 5), and demonstrated that the higher quality DNA extractions gave the best performance (see Chapter 4).

Homogeneity of IL Families

We used targeted SNP markers and whole-genome SNP assays to measure the level of homozygosity of ILs that had undergone field-based seed propagation for four years (2007-2010) and to confirm the genetic identity and composition of the lines. In 2006, using SSR data from 159 markers, we identified the size and location of both target and background introgressions in the BC_3F_2 lines and in 2008, confirmed this information in the BC_3F_4 generation using a 1,536-SNP genome-wide assay. In 2009, we developed a 384 SNP genome-wide assay tailored for

tropical japonica x *O. rufipogon* populations (Thomson *et al.,* 2011) and used it to derive ILs with fewer background introgressions (Table 2-14). MassARRAY assays were designed to specifically target regions of introgression in the two top-performing ILs. For IL yld2_A (family 43_1-2), we positioned 24 markers along the *yld2.1* target QTL (~17 Mb) and 23 markers in background introgressions, and in IL yld6_A (family 219_2-9), we positioned 26 markers along the *yld6.1* QTL (~14 Mb) with none targeting background. From 12 different BC_3F_7 headrows of IL yld2_A (43_1-2) grown in the field in 2010, ten panicles were bulked for each headrow, and we took 4~8 seeds from each of these headrows to genotype with MassARRAY assays. Results confirmed that by the BC_3F_8 generation, IL yld2_A (line 43_1-2) had been fixed for *O. rufipogon* alleles in the *yld2.1* region as well as in background introgressions on chromosome 5, 9 and 12 (Fig. 2-7). On the other hand, in the BC_3F_8 generation, three IL yld6_A (family 219_2-9) samples out of 46 plants representing eight different headrows were still segregating across the target region on chromosome 6 (Fig. 2-8). As a result, only homozygous *O. rufipogon* individuals were selected and bulked for submission to the GS*OR*.

DISCUSSION

We used backcrossing with MAS to develop ILs carrying six different yield QTLs and evaluated them over two years in multi-location yield trials in the southern U.S. Two ILs carrying introgressions from the wild donor, *O. rufipogon* across yield QTLs on chromosomes 2 and 6 consistently outperformed the recurrent parent, Jefferson (6217 ± 810 kg ha^{-1}) and were similar to the newer commercial varieties, Trenasse (8097 ± 812 kg ha^{-1}) and Cocodrie (8244 ± 811 kg ha^{-1}). All inbred lines (both ILs and commercial varieties) yielded significantly less than the commercial hybrid, XL723 (10336 ± 811 kg ha^{-1}).

Of all the observed yield components, the most significant difference between the sister ILs yld2_A and yld2_B was grain width. IL yld2_B (43_2-12) had significantly narrower grain width than IL yld2_A (43_1-2) (Table 2-12), while IL yld2_A was not significantly different than Jefferson. Two genes determining grain morphology in rice, *grain size 5* (*GS5*) and *grain width 5* (*GW5*), reside at 3.45 Mb and 5.35 Mb, respectively, on rice chromosome 5 (Li *et al.,* 2011; Shomura *et al.*, 2008). Both genes fall within the *O. rufipogon* introgression that is present in IL yld2_A, but absent in IL yld2_B, and are implicated as a reason why the *O. rufipogon* introgression across this region of chromosome 5 is associated with increased grain weight, as reported by Thomson *et al.* (2003). *GW5* was shown to bind with ubiquitin, acting in the ubiquitin-proteasome pathway to regulate cell division during seed development, and a 1.2 kb deletion in the ORF region was associated with increased grain width (Shomura *et al*., 2008; Weng *et al.,* 2008). *GS5* was shown to positively regulate the plant cell cycle to increase cell number and, to some extent, cell size, leading to enhanced latitudinal growth in the grain (Li *et al.,* 2011). Three promoter haplotypes at *GS5* were predictive of grain width, grain weight and grain filling characteristics in 35 *O. sativa* cultivars (Li *et al*., 2011). It would be of interest to

investigate the relationship between the alleles from our *O. rufipogon* donor (IRGC 105491) at these loci and their interaction with Jefferson alleles at other genes in the pathway to determine whether they contribute to the variation for grain width and/or chalk observed in our material.

Another example involves a comparison between two families, both of which contain an *O. rufipogon* introgression at *yld2.1,* family 43 (ILs yld2_A and yld2_B) and family 141 (Table 2-5). We are currently pending the results from a new 384 SNP OPA that will identify the number and location of background introgressions for both families 43 and families 141. Through comparing the presence or absence of donor introgressions between these two pools of ILs, we can start to elucidate favorable GxG interactions between *O. rufipogon* introgressions in the target region and in the genetic background that are responsible for the yield advantage observed in the best ILs. A corollary of this observation is that when different families containing the same target QTLs are evaluated side-by-side in the field, the use of molecular markers makes it possible to identify genetic differences between them that can be hypothesized to account for differences in performance. Indeed, some of these differences may be quite surprising and if marker-assisted development of near-isogenic lines were allowed to proceed ahead of field-based phenotyping, many favorable background introgressions would simply be missed.

Yield QTLs in the Jefferson x *O. rufipogon* population were originally identified by Thomson *et al.* (2003) based on an evaluation of average seed weight per BC_2F_2 family estimated as the average weight of seeds harvested from a bulk of ten plants grown in small family plots. In the current study, yield performance of the ILs was evaluated using large-scale field trials and BC_3F_4-F_5 progeny. In these field trials (2007-2008), yield was measured in terms of grain weight per plot area (kilograms per hectare). Thus, there was a critical difference as to

how yield was estimated in the original QTL study by Thomson *et al.* (2003) and how yield was assessed in the yield trials during 2007-2011. We were unable to calculate the average grain weight per plant in the recent field trials because the field data lacked accurate information about the number of plants per plot in the drill-seeded field experiments.

Chalkiness is an important grain quality trait causing breakage during the milling process and decreasing the market value of rice. Certain environmental conditions influence the amount of chalk, and certain grain characteristics are associated with the amount of breakage (as reviewed in Nelson *et al.,* 2011). While none of the high-yielding ILs had significantly different ASV compared to Jefferson, 24 ILs and 6 sib-line controls suffered from significantly higher chalk than Jefferson. Most importantly, our highest yielding IL yld2_A (family 43_1-2) does not suffer from significantly higher chalk or higher amylose compared to the recurrent parent. Yet this is a general problem that deserves further attention because five of the lines that were chosen for submission to GS*OR* have higher chalk than Jefferson. The only exception among all ILs tested was IL 16_1-6, a line that was developed as a control line for *yld3.2* and *yld9.1*. This line did not yield as well as Jefferson, but chalk levels were significantly lower than in Jefferson (adjusted $p<0.015$) and similar to those of Cocodrie. With new 384 SNP OPA data, we can confirm the number and size of introgressions in IL 16_1-6, and possibly utilize those introgression(s) contributing to lower chalk into other lines. Among the four locations during 2007-2008, the highest chalk levels were recorded in flooded irrigation plots in Stuttgart during 2008. Chalk values are missing for Beaumont and Stuttgart in 2007 so we could not determine whether the chalk is due to the year or location.

IL 16_1-6 was also one of two ILs with sheath blight (SB) resistance rating significantly better than Jefferson based on the 2009 disease trials (Adj. p-value = 0.03; Table 2-11). In

previous studies, SB resistance QTLs were detected in a Jasmine 85 x Lemont population (Zou *et al.,* 2000; Liu *et al.,* 2009) using the same micro-chamber method as in our study (Jia *et al.,* 2007). Through upcoming genotyping results, we can test hypotheses to determine which background introgressions in IL 16_1-6 are responsible for the SB resistance as well as the lower levels of chalk.

Field observations from 2010-11 show that IL yld2_D is highly susceptible to straighthead under irrigated field conditions and suffers low yield (Fig. 2-6). Straighthead is an arsenic-induced physiological defect influencing the shape of the panicle and reducing yield through floret sterility (Abedin and Meharg, 2002). In rice, arsenic is taken up through various pathways, with at least three transporters implicated to date (Takahashi *et al.,* 2004; Ma *et al.,* 2008; Wu *et al.,* 2011). An association mapping study utilizing 75 SSR markers revealed five QTLs (at RM490 on Chr. 1, RM263 on Chr. 2, RM413 on Chr. 5, RM105 on Chr. 9 and RM277 on Chr. 12) associated with straighthead resistance in a global panel of germplasm, suggesting that resistance is quantitatively inherited (Agrama *et al.,* 2009). In our study, IL yld2_D differs from the Jefferson recurrent parent by a 17.2 Mb *O. rufipogon* introgression across the target region on chromosome 2 that does not colocalize with RM263, and smaller introgression on chromosome 10 (Fig. 2-2b), that does not co-locate with the QTLs identified by Agrama *et al.* (2009). Interestingly, three sister introgression lines, yld2_A, yld2_B and yld2_C, do carry an *O. rufipogon* introgression on chromosome 12 marked by RM277 at the QTL regions identified by Agrama *et al.* (2009), but these lines do not differ in straighthead susceptibility from the Jefferson recurrent parent, suggesting that this introgression is probably not associated with straighthead resistance or susceptibility, unless it is via a complex genotype by genotype (GxG)

interaction. Further work is needed to dissect the genetics underlying straighthead susceptibility in IL yld2_D.

Based on the Thomson *et al.,* (2003) study, *yld3.2* explained the greatest proportion of variance for yield (R^2 = 16%) of the six yield-QTL targets evaluated here. However, the ILs developed for *yld3.2* did not perform as well as other lines in our experiments, possibly due to the numerous background introgressions that remained in all the *yld3.2*-containing lines. IL yld3_A (family 16_2-1) contained 11.9% *O. rufipogon* SNP alleles (78/653). This family carried a 14 Mb introgression on chromosome 9 that contained the QTL-target, *yld9.1*. When the performance of IL yld3_A was compared to IL yld9_A (family 13_1-1), which contained a 5 Mb introgression across the target region on chromosome 9 and only 2.9% donor alleles across the entire genome, the two lines yielded similarly (and were not significantly different from Jefferson) across years and locations. It is interesting to note that while *yld3.2* does not coincide with previously reported yield-related QTLs, *yld9.1* co-locates with QTLs reported as *TGW9 (thousand grain weight)* detected in a Minghui 63 x Teqing RIL population (Liu *et al.,* 2010) and a QTL cluster including *gw9.1*, *hd9.1* and *ph9.1* detected in a Hwaseongbyeo x *O. rufipogon* BC population (Xie *et al.,* 2008). Further backcrossing to separate the effects of the target QTL regions in the various background introgressions will be needed to better understand the potential of *yld3.2* and *yld9.1*.

SNP Genotyping Platforms

Several genome-wide SNP assays were used in this study. Levels of polymorphism for the Jefferson *x O. rufipogon* population are provided for the 384-SNP Illumina BeadXpress OPA

(Thomson *et al.,* 2011), the 1,536-SNP Golden Gate assay (42.5%), and the 44,000-SNP Affymetrix array are provided in Table 2-13.

From the 1,536-SNP assay, 653 SNPs (42.5%) were informative between Jefferson and *O. rufipogon* but they were not always well distributed across the genome. As illustrated in Figure 2-2b, there was a paucity of SNPs across the *yld2.1* region when the 1,536 SNP assay was used, but we were able to design a targeted 384 SNP assay that was both efficient and economical for use in marker-assisted selection on this project by selecting additional polymorphic SNPs that had been identified based on sequencing or other assays. For example, a total of 18,760 SNPs (50.1%) were informative between Jefferson and *O. rufipogon* using the 44K chip. The OPA 6.0 was designed for *tropical japonica* x *O. rufipogon* and was able to capture 282 SNPs (73.4%) between Jefferson and *O. rufipogon* (IRGC105491) (Table 2-14).

We also developed targeted SNP assays (KASP and iPLEX MassARRAY) that will aid the swift integration of *O. rufipogon* introgressions into any *tropical japonica* background. These targeted assays were ideal for backcross conversion where the wild introgressions are either selected for retention or elimination, and where accurate detection of *O. rufipogon* alleles in either homozygous or heterozygous condition is important.

These collections of SNPs are immediately useful for further backcrossing of the ILs in a breeding program, for fine mapping experiments, or for moving target introgressions into other elite *tropical japonica* backgrounds. Further, approximately 90% of all SNP markers that were selected from the pool of polymorphic targets identified on the 44K SNP chip were immediately functional in KASP and iPLEX MassARRAY systems. Thus, the information provided by the 44K and 1,536 SNP assays (Zhao *et al.,* 2010; 2011) provide a valuable template for others to

design their own cost-effective strategies for SNP-based genotyping in rice (McCouch *et al.,* 2010).

Table 2-1. Six target QTLs with corresponding peak markers from Thomson *et al.* (2003).

QTL	QTLs reported by Thomson *et al.*	BC_2F_2 family ID	Chr	Peak Marker	Position (Mb) MSU v.6**	Flanking markers position (Mb) ***	LOD score	R_2
yld1.1	*gpp1.2**	89; 158; 185	1	RM5*	*23.9715*	23.9713-30.0275	6.35*	9.7*
yld2.1	*gpp2.1; spp2.1; yld2.1*	43; 141; 85	2	CDO718	*18.5343*	18.5342-18.5345	7.35	12.3
yld3.2	*gpp3.1; spp3.1; pss3.1; dth3.4; gw3.2; yld3.2*	16	3	RG1356	*35.1410*	33.3863-35.1411	11.56	16.6
yld6.1	*pss6.1; yld6.1*	219; 221	6	RM276	*6.2301*	6.2300-6.2302	3.74	6.5
yld8.1	*gpp8.1; ph8.1*	307; 338	8	RM210	*22.4719*	22.4718-22.4720	4.44	7.2
yld9.1	*pl9.1; tt9.1; yld9.1; gpp9.1; spp9.1*	9; 13; 16	9	RM215	*21.1892*	21.1891-21.1893	4.28	7.4

Abbreviations: yld= yield*; gpp = grains per panicle (average number of filled spikelets per primary panicle); spp = spikelets per panicle (average total number of spikelets per primary panicle); pss = percent seed set (gpp divided by spp); dth= days to heading; gw = grain weight; ph = plant height; pl = panicle length; tt = tiller type (erect or lazy tillering); RM = rice microsatellite; CDO = cDNA from oat used as RFLP marker; RG = rice genomic RFLP marker. *These QTLs and associated scores were detected in the BC_2F_2 generation (Thomson *et al.,* 2003). While the current study defines yield as kilograms per hectare, the original yield measurement indicated the average weight of bulked harvested grain per plant from at least ten plants. ** Rice pseudo-molecules assembled by Michigan State University Version 6 http://rice.plantbiology.msu.edu/ ***Marker IDs can be found in paper by Thomson *et al.* (2003).

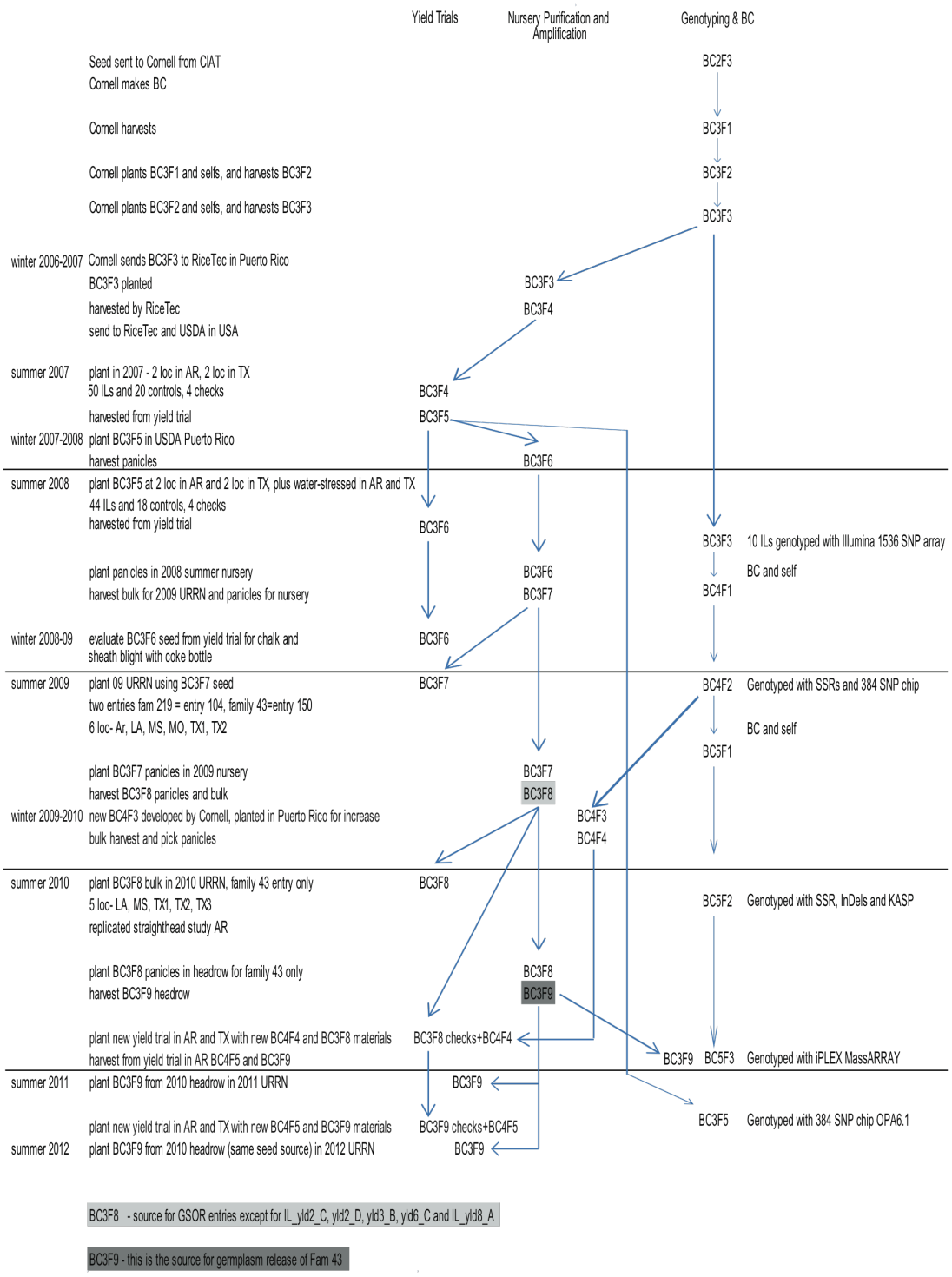


Figure 2-1. Pedigree diagram showing derivation of Jefferson x *O. rufipogon* ILs with timing and location of genotyping and phenotyping activities used during IL development. Author has participated in activities at Cornell since spring of 2009.

Table 2-2. KASP markers

SNP_id	iPLEX duplicate	Chr	MSU V.6 Position (Mb)	Jeff_allele	O. rufi_allele	Cycle #	Max Tm	Min Tm	A1	A2	C1
id2002637	*	2	4.989778	AA	GG	35	65.8	64	GAAGGTGACCAAGTTCATGCTCATCTGGCCTTGATGGGTATGCA	GAAGGTCGGAGTCAACGGATTATCTGGCCTTGATGGGTATGCG	AGGTGGACATCAGAGGAGTCGATTT
id2003133	*	2	5.936587	AA	GG	32	66.2	63.7	GAAGGTGACCAAGTTCATGCTCCAATATGTCGTCGAGTCTAAACTTT	GAAGGTCGGAGTCAACGGATTCCAATATGTCGTCGAGTCTAAACTTC	GTTGGCACAATATCGTGAGTTGGCTA
id2003563		2	6.87572	TT	AA	32	66.2	64.6	GAAGGTGACCAAGTTCATGCTCCCACACAAGCCTGTCCGGA	GAAGGTCGGAGTCAACGGATTCCCACACAAGCCTGTCCGGT	TTTCGGTTTGTACCTTGAGTGCAACAAA
id2004031	*	2	7.857478	CC	TT	32	65.9	64	GAAGGTGACCAAGTTCATGCTGCCTCGTCTTCACACCGCTC	GAAGGTCGGAGTCAACGGATTATGCCTCGTCTTCACACCGCTT	GGAGCCCTGTGTCGACGCTT
id2004457	*	2	8.980361	AA	TT	29	65.2	64.1	GAAGGTGACCAAGTTCATGCTGTTTGATGAGCACTTGCCCGTTAT	GAAGGTCGGAGTCAACGGATTGTTTGATGAGCACTTGCCCGTTAA	CATATAACCTTTGACTATCATCAGTAGCTT
id2004718	*	2	9.882938	TT	CC	32	66.3	63.7	GAAGGTGACCAAGTTCATGCTAATAGACAACTTTGCTGAAATGGGCA	GAAGGTCGGAGTCAACGGATTATAGACAACTTTGCTGAAATGGGCG	CAAGTTTATCCTATACATACCGTCGAGTT
wd2000589		2	10.89387	TT	GG	32	65.9	63.8	GAAGGTGACCAAGTTCATGCTGCGGTAGCGGGACGACAT	GAAGGTCGGAGTCAACGGATTGCGGTAGCGGGACGACAG	GCTCCCTGCCACAGCTGTGTA
ud2000665		2	11.896598	CC	TT	32	65.7	63.5	GAAGGTGACCAAGTTCATGCTGTCCGAAGTTGACGAGTCTG	GAAGGTCGGAGTCAACGGATTAACTGTCCGAAGTTGACGAGTCTA	GCTGTGCCATGTCTTCGTTTAGCTT
id2005728		2	12.8908	AA	GG	32	65.2	62.9	GAAGGTGACCAAGTTCATGCTAGACCACTTGATCGTATAAGCCCTT	GAAGGTCGGAGTCAACGGATTACCACTTGATCGTATAAGCCCTC	GCTGGAATTTATTGGACACTTTGTAATGAA
id2005731	*	2	12.895061	AA	CC	29	65.8	62.9	GAAGGTGACCAAGTTCATGCTGTAGTGGTAGCCTCCTAATTACGT	GAAGGTCGGAGTCAACGGATTAGTGGTAGCCTCCTAATTACGG	CGTCACGTGCCGGGCGTGA
wd2001061		2	13.3551	CC	TT	35	65.8	64.1	GAAGGTGACCAAGTTCATGCTGGACTCAGTTTTACCAGGACTCG	GAAGGTCGGAGTCAACGGATTATGGACTCAGTTTTACCAGGACTCA	CCGAATCAGACTCTGACTCTCCTAT
id2005915	*	2	13.899868	GG	TT	35	65.8	63.5	GAAGGTGACCAAGTTCATGCTATCAGGTTGCAATCCCTGATTTGC	GAAGGTCGGAGTCAACGGATTCATCAGGTTGCAATCCCTGATTTGA	GTGTAGTCGATGGGGCCCAGAA
ud2000810		2	14.995015	TT	CC	29	66.2	63.8	GAAGGTGACCAAGTTCATGCTCCATGCAGGCAGGGCACCAT	GAAGGTCGGAGTCAACGGATTCATGCAGGCAGGGCACCAC	CAAGCTATCATCATTCCTCTCTGATTCAT
id2006450	*	2	16.03187	TT	GG	35	65.9	63.3	GAAGGTGACCAAGTTCATGCTGTTGCTTGGAGTTGCCAGGGTA	GAAGGTCGGAGTCAACGGATTGCTTGGAGTTGCCAGGGTC	GCATTTGGCTTTGACCCCGCCAA
id2006627		2	16.530721	CC	TT	35	65.8	63.9	GAAGGTGACCAAGTTCATGCTTTCCTATCTACATTATTCCTCTGAAAATC	GAAGGTCGGAGTCAACGGATTATATTTCCTATCTACATTATTCCTCTGAAAATT	TTTTCCTGCCCTTTTAGGGCCTCTT
id2006798	*	2	17.041079	AA	GG	35	66.2	64.1	GAAGGTGACCAAGTTCATGCTGAAGGTCCTGAATCTTGATTGGTTCT	GAAGGTCGGAGTCAACGGATTAAGGTCCTGAATCTTGATTGGTTCC	GCGTACCTGAAACAGTTTAGGTAAGAAAT
id2007232		2	18.012987	TT	CC	35	65.2	62.9	GAAGGTGACCAAGTTCATGCTTCTTAGGCTAGGCTGTTTCCAA	GAAGGTCGGAGTCAACGGATTCTTAGGCTAGGCTGTTTCCAG	GAACCAAGGAGCTCCCGGCAT
id2007468	*	2	18.81379	AA	GG	29	65.8	63.5	GAAGGTGACCAAGTTCATGCTCGAAGTTAGTACAAGCCTCTGAT	GAAGGTCGGAGTCAACGGATTCGAAGTTAGTACAAGCCTCTGAC	TACACCGCATTATCCTCTCCAGCTT
id2008351	*	2	20.985363	CC	TT	32	66.4	62.9	GAAGGTGACCAAGTTCATGCTGGTGCAAGTTCTGAGCGATCTC	GAAGGTCGGAGTCAACGGATTGGTGCAAGTTCTGAGCGATCTT	ACCATCAGAGGTGATGTACGCCTT
id2008375		2	21.01131	GG	AA	35	65.8	63.5	GAAGGTGACCAAGTTCATGCTGAGGATGATGAGGACGAGCTTG	GAAGGTCGGAGTCAACGGATTATGAGGATGATGAGGACGAGCTTA	CTCCCTGAGGGTTCAAACCCCAA
id2009169		2	22.978727	GG	CC	32	66.3	64	GAAGGTGACCAAGTTCATGCTAGCAGGAGCCTCAACTGACAG	GAAGGTCGGAGTCAACGGATTAGCAGGAGCCTCAACTGACAC	CCAAATCATCTGGTGCTAGAGGTCTT
id6003270		6	4.497029	CC	AA	32	65.9	63.8	GAAGGTGACCAAGTTCATGCTGTGAGAGACACAGCCGCCG	GAAGGTCGGAGTCAACGGATTGAGTGAGAGACACAGCCGCCT	TGTCGCTGCCGCCACTGCAAAT
id6003397	*	6	4.867445	CC	TT	35	65.9	64	GAAGGTGACCAAGTTCATGCTGTGTGCGTTTTGCAGTGCTCTC	GAAGGTCGGAGTCAACGGATTCGTGTGCGTTTTGCAGTGCTCTT	GTCGTAGACGCCGACGTCGTA
id6003555	*	6	5.531392	GG	AA	32	66.2	64.1	GAAGGTGACCAAGTTCATGCTGGTTTCCTGATGCCCATTTGCAG	GAAGGTCGGAGTCAACGGATTGGTTTCCTGATGCCCATTTGCAA	CTAAACATTAAGCCACGTCCAGAAATCAA
id6003864		6	6.100589	CC	TT	32	67.2	63.8	GAAGGTGACCAAGTTCATGCTCATCACGGGAGCTGGCACC	GAAGGTCGGAGTCAACGGATTCCATCACGGGAGCTGGCACT	CTTGGAGTGCGCCAGTGAGCA
id6004180		6	6.513921	AA	GG	35	65.8	64.1	GAAGGTGACCAAGTTCATGCTCCTCTTAAGGCCTTTATAATGTGGCA	GAAGGTCGGAGTCAACGGATTCTCTTAAGGCCTTTATAATGTGGCG	CTGTGACGTGAAGAGATCGAGCTAA
id6004660		6	7.231978	GG	AA	35	65.9	64.1	GAAGGTGACCAAGTTCATGCTTAATAGGAGTTGCCTTAGTGGCAC	GAAGGTCGGAGTCAACGGATTGTTAATAGGAGTTGCCTTAGTGGCAT	AGTAGGGCACCCTTGTGCGCAA
id6005446		6	8.363542	CC	TT	32	66.4	63.5	GAAGGTGACCAAGTTCATGCTGTAAAGTGAGCCGTCACACGTC	GAAGGTCGGAGTCAACGGATTATGTAAAGTGAGCCGTCACACGTT	CATGGATCCGTGGACGGACACAT

id6005661		6	8.823427	TT	CC	32	66.2	63.5	GAAGGTGACCAAGTTCATGCTCAACAAGGAAAATAAAGAATTTTATGTTTGCA	GAAGGTCGGAGTCAACGGATTCAACAAGGAAAATAAAGAATTTTATGTTTGCG	AAATGAGGATTGCCACAAAGCCATTGAAA
id6006095	*	6	9.532265	TT	CC	35	63.7	65.2	GAAGGTGACCAAGTTCATGCTTCCCATCCTCCATAGAATGTACTTTT	GAAGGTCGGAGTCAACGGATTCCCATCCTCCATAGAATGTACTTTC	AAGGTACGTCAAGAGTAGCGGCTT
id6006538		6	10.5508	GG	CC	32	66.2	64	GAAGGTGACCAAGTTCATGCTTTGGAAGTGGTGGTCAGCAC	GAAGGTCGGAGTCAACGGATTCTTTGGAAGTGGTGGTCAGCAG	GGCATATGTCTTTCTTCCGCATTAATCTT
id6006541	*	6	10.552134	CC	TT	35	65.8	63.1	GAAGGTGACCAAGTTCATGCTAGATGACAAGATCGTTCCAACTCTC	GAAGGTCGGAGTCAACGGATTAAGATGACAAGATCGTTCCAACTCTT	CATCAAATCGAGCATGCATAAGAAGCATA
id6007245	*	6	11.518128	TT	GG	35	66.2	63.5	GAAGGTGACCAAGTTCATGCTATATATATGGCCATCACGCCCAGA	GAAGGTCGGAGTCAACGGATTATATATGGCCATCACGCCCAGC	CACAATGACATGCATCACTTGTACAGATA
wd6001025	*	6	12.537233	AA	GG	35	65.8	62.9	GAAGGTGACCAAGTTCATGCTAATTCACAGCACAGCCAATATGTAGTT	GAAGGTCGGAGTCAACGGATTCACAGCACAGCCAATATGTAGTC	CGCAAGCCTCGACGCAGATCAT
ud6000581		6	13.519496	TT	CC	32	65.8	64.3	GAAGGTGACCAAGTTCATGCTGGGCAGCATTTCAATTAATTTAGACGAAT	GAAGGTCGGAGTCAACGGATTGGCAGCATTTCAATTAATTTAGACGAAC	CAAACAGGGCCAAAAAGGCTGTGTA
id6008658		6	14.347413	AA	GG	35	65.8	62.9	GAAGGTGACCAAGTTCATGCTAGTAAGCCTCATAGCCTCATCTGTA	GAAGGTCGGAGTCAACGGATTAAGCCTCATAGCCTCATCTGTG	TTTGGTATTCGTGCCGGTACGCTAT
id6009425		6	16.5109	CC	GG	32	66.2	64	GAAGGTGACCAAGTTCATGCTCAGAGGCATGTTATGACGGGAG	GAAGGTCGGAGTCAACGGATTCAGAGGCATGTTATGACGGGAC	CTTGTCAAAAACTCCAAGCTGGAAACTTT
id6009428		6	16.512573	AA	GG	35	65.2	64	GAAGGTGACCAAGTTCATGCTGTGTTCTAATAGACCCGGTGGCA	GAAGGTCGGAGTCAACGGATTGTTCTAATAGACCCGGTGGCG	AACACCAGACCAGTCACCGCTATA
id6010146		6	18.831462	GG	AA	35	66.2	64	GAAGGTGACCAAGTTCATGCTCAGTACGGGTATTCCTCCAC	GAAGGTCGGAGTCAACGGATTGCTCAGTACGGGTATTCCTCCAT	CCAAGGAAATTGTATTTGGAATACATGCTA
id6010766	*	6	20.630627	TT	AA	35	65.7	63.3	GAAGGTGACCAAGTTCATGCTTTGGTTTTCTTTACCTTTATAGCTCTGT	GAAGGTCGGAGTCAACGGATTTGGTTTTCTTTACCTTTATAGCTCTGA	TTAGCAGATAGCATCAAACATGCCAGAA
wd6002805	*	6	21.031947	CC	AA	35	66.4	64	GAAGGTGACCAAGTTCATGCTGGTGAAGGGGTAAGTCCTACTG	GAAGGTCGGAGTCAACGGATTCGGTGAAGGGGTAAGTCCTACTT	GATACAACTGTCTGGGCCGGCAA
id6011272		6	21.500498	TT	CC	35	66.2	63.7	GAAGGTGACCAAGTTCATGCTACACTATCACGTCACATCAATTGCAT	GAAGGTCGGAGTCAACGGATTCACTATCACGTCACATCAATTGCAC	CCTAATTACTATAGTGCACATGGATGCTA

Abbreviations: SNP.id = SNP identification number; iPLEX duplicate = asterisks indicate that Sequenom MassARRAY iPLEX markers were developed targeting the same SNP; Chr = chromosome; MSU V.6 Position (Mb) = genetic position, Jeff allele = Jefferson recurrent parent allele; O. rufi allele = *O. rufipogon* (IRGC105491) allele; Cycle# = Number of cycles for PCR; Tm = melting temperatures; A1 = SNP specific marker for allele 1; A2 = SNP specific marker for allele 2; C1 = common reverse primer.

Table 2-3. MassARRAY iPLEX markers.

SNP.id	KASP duplicate	Chr	MSU V.6 Position (Mb)	Jeff allele	O. rufi allele	Multiplex set**	2W primer	1W primer	Extension primer
id2002162		2	4.033557	TT	AA	*Yld2.1*target	ACGTTGGATGTGGAGGAGGAGCCCGTCGT	ACGTTGGATGTTTCCACCGCATTCTTGGTC	gccaGTCGTCGCAGCTGACCGCTGAG
id2002637	*	2	4.989778	AA	GG	*Yld2.1*target	ACGTTGGATGAGGGATACAAGCCAGAGAAG	ACGTTGGATGATAACATCTGGCCTTGATGG	CATCAGAGGAGTCGATTT
id2003133	*	2	5.936587	AA	GG	*Yld2.1*target	ACGTTGGATGAGGTCCTCCAATATGTCGTC	ACGTTGGATGGGCACAATATCGTGAGTTGG	TATGTCGTCGAGTCTAAACTT
id2003553		2	6.871979	AA	GG	*Yld2.1*target	ACGTTGGATGCCACAACCATCTGTAGAGAG	ACGTTGGATGTAGGGTTTGTTCTTGGCAAC	taTGTAGAGAGAGCACCGG
id2003785		2	7.321268	CC	TT	*Yld2.1*target	ACGTTGGATGGGGAGTAGGTCTCATGAATC	ACGTTGGATGCCCATAATAGTTATGGCCCG	TGAATCATCCCTTGCGCGT
id2004031	*	2	7.857478	CC	TT	*Yld2.1*target	ACGTTGGATGTTAGCGAATGCCTCGTCTTC	ACGTTGGATGGTTTGTTTCGGTGAATCCCC	CCTCGTCTTCACACCGCT
id2004232		2	8.328053	GG	AA	*Yld2.1*target	ACGTTGGATGCAGGAACCTTCTCTACTTTG	ACGTTGGATGGCTATCGATGCTACCACAAC	gcTGTACTTCTTGATATCAACTTCAC
id2004457	*	2	8.980361	AA	TT	*Yld2.1*target	ACGTTGGATGCTCCGATGTTTGATGAGCAC	ACGTTGGATGCTCTGGCACATTAGCACAAG	ttttATGAGCACTTGCCCGTTA
id2004548		2	9.492964	GG	AA	*Yld2.1*target	ACGTTGGATGGGACAATGTAGATTCGTCAG	ACGTTGGATGCTGCATACCTTGAGGTGTTC	ccttgAGATTCGTCAGATAGCTTCCA
id2004718	*	2	9.882938	TT	CC	*Yld2.1*target	ACGTTGGATGTACCGTCGAGTTTTCGTGTC	ACGTTGGATGGCTAGCACTTTTTTTTGAGC	CGTGTCGATTCCTGTTA
id2005118		2	10.824535	TT	CC	*Yld2.1*target	ACGTTGGATGTGTCCTCACGAATTTCTGGG	ACGTTGGATGTAGGACAAAAGGGATCACAG	gctaACGAATTTCTGGGTAAGAATG
ud2000643		2	11.379299	AA	CC	*Yld2.1*target	ACGTTGGATGTGTGATTTCAGGCTTTCCGC	ACGTTGGATGAAAAGTCACAAAAGGCCAGC	gaggGGTATGCATGTAGTACATTGTA
id2005498		2	11.881302	GG	AA	*Yld2.1*target	ACGTTGGATGTATTCTACGTCGTAATGCGG	ACGTTGGATGTTTGTGGACGAATGGAATAC	cttCCTGTTCACCGTTCGCGTGTCC
id2005558		2	12.323073	GG	AA	*Yld2.1*target	ACGTTGGATGTTCCACTTGGGATTTATGGG	ACGTTGGATGTGGCTTTGTGAGAACCAGAG	ccagAGCAATACATTACTCTTGGA
id2005731	*	2	12.895061	AA	CC	*Yld2.1*target	ACGTTGGATGTAGTAGGGAGTAGGGTGTAG	ACGTTGGATGTTTTCGGCGGTGTCCTTTTC	gaggGGTAGCCTCCTAATTACG
wd2001042		2	13.334818	CC	TT	*Yld2.1*target	ACGTTGGATGGTGAAGCTATTCGGACATGG	ACGTTGGATGGCTCCATGTGATGAAGTGTC	aaATGGAAAACTTCGAAATAAATG
id2005915	*	2	13.899868	GG	TT	*Yld2.1*target	ACGTTGGATGTTCCCATCAGGTTGCAATCC	ACGTTGGATGAATGAACTGAGCTGGGTCTG	ccttTTGCAATCCCTGATTTG
id2005978		2	14.375133	TT	CC	*Yld2.1*target	ACGTTGGATGCATCTTACGCACTACTGCAC	ACGTTGGATGAGGGTCTCAATTTGCATTCG	tcccGCACCAACAGGGCCGTCTC
id2006124		2	14.933602	CC	TT	*Yld2.1*target	ACGTTGGATGGGTTTAGCATGTTTTGCAGG	ACGTTGGATGTCACTGCTAGTAGTGGGTTG	cGTTTTGCAGGTAAGGTC
id2006241		2	15.500726	CC	TT	*Yld2.1*target	ACGTTGGATGAGCATCATACACGTCAAGGG	ACGTTGGATGTGTTTAGTAGGTGTGGGAGG	gggTGGTTTGGGTGCATCAG
id2006450	*	2	16.03187	TT	GG	*Yld2.1*target	ACGTTGGATGTTGGAGTATGGAGGATCTGG	ACGTTGGATGAGAGCATTTGGCTTTGACCC	aggtCGTTGCTTGGAGTTGCCAGGGT
id2006621		2	16.499167	TT	CC	*Yld2.1*target	ACGTTGGATGGAACAACCTTAAGCAAAGAC	ACGTTGGATGAGCACTGGAAGTCTGAGTAG	ggCTCAACTGATTGGCACTA
id2006798	*	2	17.041079	AA	GG	*Yld2.1*target	ACGTTGGATGTGTAGTACGCACGTTCAAGG	ACGTTGGATGGACGAAGGTCCTGAATCTTG	cATCAAATCCAGCAACCGTA
id2007218		2	17.9721	CC	TT	*Yld2.1*target	ACGTTGGATGACTGATACTGCAACCACAGG	ACGTTGGATGTCAAGCTCATTACAGTCAAG	ggggaAGGCTTAAAAGTGTCCTT
id2007468	*	2	18.81379	AA	GG	*Yld2.1*target	ACGTTGGATGTACACCGCATTATCCTCTCC	ACGTTGGATGACTCAAAGCACATACCCCAG	CGCATTATCCTCTCCAGCTTC
id2007602		2	19.439264	GG	AA	*Yld2.1*target	ACGTTGGATGTGCATTGCCGTCTTGATTCC	ACGTTGGATGATCCTCAACTTTGGCACGAC	aacgCCTGACAAAATTCAGTTAGC
id2008351	*	2	20.985363	CC	TT	*Yld2.1*target	ACGTTGGATGACCATCAGAGGTGATGTACG	ACGTTGGATGAGAATGGCATCCTGTTGGAG	ggTTCGAGATGGCGAGATC
id2008815		2	22.113645	CC	TT	*Yld2.1*target	ACGTTGGATGCTGAAAAAGTCATGTCGGTC	ACGTTGGATGGGTCGTGACACAATTTTGCG	GTCGGTCTTTCGTCCTT
id6003397	*	6	4.867445	CC	TT	*Yld6.1*target	ACGTTGGATGCGTAGAGGAAGCGCTTGTTG	ACGTTGGATGCCTCCTTCTCCTTCTTCTTG	tgatGAAGCGCTTGTTGGAGTCG

id6003502		6	5.212799	CC	AA	*Yld6.1*target	ACGTTGGATGACCCTCGTATACCACAGATG	ACGTTGGATGTCTCGTACGGGTAGATGTTC	ggACAGATGGCGATCATTG
id6003555	*	6	5.531392	GG	AA	*Yld6.1*target	ACGTTGGATGTCTTGACCGTCGTGATGTTG	ACGTTGGATGTGATTGGTTTCCTGATGCCC	CCACGTCCAGAAATCAA
id6003812		6	5.976977	AA	TT	*Yld6.1*target	ACGTTGGATGACATACTGGAGTTATAAGGC	ACGTTGGATGGCCATGATCATCCGATAACG	AAGGCAACTAAACAAAGCAAG
id6004051		6	6.403848	AA	GG	*Yld6.1*target	ACGTTGGATGGGACACAATAACGGAGTTGG	ACGTTGGATGAGCCCTTTGCTGCAAGTTTG	CCCTCCTCTAATTCAAGTTCT
id6004356		6	6.792956	CC	TT	*Yld6.1*target	ACGTTGGATGCGAGTTTGTATTTAACAATGC	ACGTTGGATGAAACTTGCACGGCTGAATAC	ggtggATTTAACAATGCTTGGGACTG
id6004650		6	7.193359	CC	TT	*Yld6.1*target	ACGTTGGATGATCGTCCTCATCTCATCAGC	ACGTTGGATGACTTGCTTTGATCGAGGGTG	gataTCATCAGCGGCAACAATCA
id6004968		6	7.845712	TT	CC	*Yld6.1*target	ACGTTGGATGGGGCCGAATTCATTAATGGG	ACGTTGGATGAACAAATTCGATCCACGAGC	ggtAATTGAGTAGGCCGAATTTA
id6005428		6	8.358537	AA	GG	*Yld6.1*target	ACGTTGGATGGCTGAGCAAAAATAGCAACG	ACGTTGGATGGATTGAAACCGAGAGGACAG	caaacTAGCAACGTGCAATCCTTC
id6005761		6	8.977257	TT	CC	*Yld6.1*target	ACGTTGGATGAGAAGCCACTAAAGCCTTCG	ACGTTGGATGCCTCTTCAGCCTCATGTTTC	CTCCACTGATCCCAACC
id6006095	*	6	9.532265	TT	CC	*Yld6.1*target	ACGTTGGATGTGGAGACGACATTGTTTCCC	ACGTTGGATGCGATCACAAGGTACGTCAAG	tccccTCCTCCATAGAATGTACTTT
id6006235		6	10.008315	CC	GG	*Yld6.1*target	ACGTTGGATGCCGGTATACCAAGCAGTTTG	ACGTTGGATGAAGGACAGACAGAGACAGAG	TTTGTCAACTCCAGTTTATGAA
id6006541	*	6	10.552134	CC	TT	*Yld6.1*target	ACGTTGGATGCAAGATGACAAGATCGTTCC	ACGTTGGATGTAGGTGTCATCAAATCGAGC	CAAGATCGTTCCAACTCT
id6006838		6	11.010415	TT	CC	*Yld6.1*target	ACGTTGGATGAACCAAAGCTAGCCCCAATC	ACGTTGGATGACCATGGTGGAGATATACAG	gtaACAGGTAAGTAGGTGAACA
id6007245	*	6	11.518128	TT	GG	*Yld6.1*target	ACGTTGGATGGACATGCATCACTTGTACAG	ACGTTGGATGCCAGTTGTATATTGCAGCTC	gaaAGATAAGGGTCAAATTTTCCTA
id6007445		6	12.049783	CC	TT	*Yld6.1*target	ACGTTGGATGGAAATGCATCATCAATTCCG	ACGTTGGATGTGCAATAAGCGGATGTGTTG	AATTACAAGTCGCGTAAG
wd6001025	*	6	12.537233	AA	GG	*Yld6.1*target	ACGTTGGATGCGAGGATCACAATTCACAGC	ACGTTGGATGACGCAGATCATGGTTGTGAG	AGCACAGCCAATATGTAGT
id6007872		6	13.003783	AA	GG	*Yld6.1*target	ACGTTGGATGAGTTTGTGCCACTTTTAACC	ACGTTGGATGCTGGGTATGCACAGAAAGAG	ccccGTGTCACTTTGAACCAC
id6008426		6	14.010866	TT	CC	*Yld6.1*target	ACGTTGGATGGTGATTTTGCTACCCTGGAG	ACGTTGGATGTTGCCATCCCTAGGTAAGAG	CCCTGGAGAATGAGCGA
id6008871		6	15.037458	GG	TT	*Yld6.1*target	ACGTTGGATGCCCAACAAGCTAGTAGTGAG	ACGTTGGATGGATAACTAAGCGAATGGAGG	ccCCCATTGTCGATGCTAT
id6009213		6	16.021266	CC	TT	*Yld6.1*target	ACGTTGGATGTGTTTCCTCCTCCACATGTC	ACGTTGGATGGGGCTGTGTACTCTAACAAG	ATGTCTTCTTGATCTAATACTGAAAT
id6009639		6	17.030234	TT	GG	*Yld6.1*target	ACGTTGGATGTTGGTCCATGGACCATGAGG	ACGTTGGATGAATACAAGCAGACGCCGATG	ggggaATGGACCATGAGGTGACACGT
id6010081		6	18.072019	GG	AA	*Yld6.1*target	ACGTTGGATGGCTCCACCTAGCACAATATC	ACGTTGGATGGTTTCTTGTGTGCATCATCG	ATCATCATAAAAGGCATACAAATA
id6010178		6	19.022636	TT	CC	*Yld6.1*target	ACGTTGGATGTAGTGATCTGCTGCAGCGTG	ACGTTGGATGTCGCGCAGTTTGAGAAGTTG	ttttgGTTTTCAGCCTGCTGAGCAAA
id6010766	*	6	20.630627	TT	AA	*Yld6.1*target	ACGTTGGATGGTTAGCAGATAGCATCAAAC	ACGTTGGATGTCTCGTGTGTTTGGAAGCAG	gatCATCAAACATGCCAGAA
wd6002805	*	6	21.031947	CC	AA	*Yld6.1*target	ACGTTGGATGGAACAACACCAATGTAGCCG	ACGTTGGATGTCAGCTTTCGATCTTCGGTG	ctatcCAACTGTCTGGGCCGGCAAA
id1003932		1	4.816429	AA	CC	*Yld2.1*bkgd	ACGTTGGATGCGTGACTTGTTTTCATCAGG	ACGTTGGATGAACCCTCTCCTCCACGCCAT	ggtaATCAGGAGTACAGTTGTCAGT
id1004294		1	5.427556	TT	AA	*Yld2.1*bkgd	ACGTTGGATGCTCCTCTCGTAAGTAAATCTC	ACGTTGGATGTGCAGGGAAAGAATATTCAG	CTCGTAAGTAAATCTCACACAAT
id1004698		1	6.007068	TT	CC	*Yld2.1*bkgd	ACGTTGGATGGCCATTATTCTTCCGACCAG	ACGTTGGATGTGTTCACCGGAGTAGTAGAC	AGGAGCTTGAAGATTACA
id5003627		5	7.164489	TT	CC	*Yld2.1*bkgd	ACGTTGGATGTGGGTCTACCGGTCTAAATG	ACGTTGGATGATTTATGTCCGGTGGCACTC	gggagAAATGAAAGTGCCACG
id5004393		5	8.602435	TT	GG	*Yld2.1*bkgd	ACGTTGGATGCCTGTGTCTCAAACACCAAG	ACGTTGGATGCCTGTGTTGGAACATATGAG	cccaAAGTTTCTATGCTTTCTCATCT
wd5001098		5	10.117129	GG	CC	*Yld2.1*bkgd	ACGTTGGATGACTTGATCATATGAAGCACC	ACGTTGGATGCGGAGGGAGTGATAAAAAAG	gagCATATGAAGCACCACATGA
wd5001400		5	11.617315	AA	GG	*Yld2.1*bkgd	ACGTTGGATGCAAGTGGACATCACAAACCG	ACGTTGGATGTAATCTTTTTGCGGTCGGGC	gTTCCGCAGAGAGGCA
id5005425		5	13.128247	GG	CC	*Yld2.1*bkgd	ACGTTGGATGGACGTAAAGGAGATTGCGAC	ACGTTGGATGCTGAGGAAGAGATGTTACAG	ggggaTAGACGCGTAGCACCGA

wd5002158		5	14.630526	AA	GG	*Yld2.1*bkgd	ACGTTGGATGAACACTCATGGGCAATACCG	ACGTTGGATGGCTGCTCCGTTTTTCCAAAG	cctccTGGGCAATACCGTTTTCC
id5006506		5	16.228592	TT	GG	*Yld2.1*bkgd	ACGTTGGATGACGACTCTTACTTGCGATGG	ACGTTGGATGCGAACTACTGTTGAGAGAGC	aCCTAGTCATAAGGACTTTGAATC
id5007105		5	17.61974	TT	CC	*Yld2.1*bkgd	ACGTTGGATGTTCGGCCTCAGATTTTTGAC	ACGTTGGATGAGTCAAATCAGGAGCAAGAG	CTACTTTTGACTCTCTACCA
wd5002587		5	19.135999	TT	CC	*Yld2.1*bkgd	ACGTTGGATGGTGTTAGGGTTGGGTTTTGC	ACGTTGGATGATCAAGGACTCAAAGTTGGG	CCTTTGCAGTTTCGCT
id5008654		5	20.628688	GG	AA	*Yld2.1*bkgd	ACGTTGGATGTATCATAAAGCAGCCGCCAG	ACGTTGGATGCAGACCAGTGAAAATACACC	CGAAAAGGCATCTTCG
id9007407		9	21.839835	GG	TT	*Yld2.1*bkgd	ACGTTGGATGGACCGTTGGTCCAAAATGTG	ACGTTGGATGTGCTGTACCGTTGGTTGAAG	gggaTTGGTCCAAAATGTGGACTATT
id9007821		9	22.700945	GG	TT	*Yld2.1*bkgd	ACGTTGGATGTGAAACCAGGTTCCTGATGC	ACGTTGGATGTGTGTGCTGTTTATCGCGAG	gagggGCAAGACGATTTTCGTGT
id10005037		10	17.534004	GG	AA	*Yld2.1*bkgd	ACGTTGGATGCTCAAGTTAGATGATGTGAGC	ACGTTGGATGCACAAACATCATACTTGACC	tagcTGATGTGAGCGTTGTT
id10005801		10	18.846347	AA	CC	*Yld2.1*bkgd	ACGTTGGATGAAATGTGGAATTGTGGAGGG	ACGTTGGATGGTTTGATTTCCACACCTTCAG	ggacAATCGTTTGAACCTGCAGTGA
id10006386		10	20.104983	AA	GG	*Yld2.1*bkgd	ACGTTGGATGCCAAATCATGATTGGTTCTCG	ACGTTGGATGTTCACAAACACCGTACGTGC	ccccTGATTGGTTCTCGTATAATTT
id10006761		10	21.415521	TT	CC	*Yld2.1*bkgd	ACGTTGGATGTTGTGACCCTTTTCTGATGT	ACGTTGGATGCTGGGGAGGAAAAATATATGG	cccaTTTTCTGATGTGACGCA
wd10003936		10	22.727812	TT	GG	*Yld2.1*bkgd	ACGTTGGATGGACAGCAGAAATTCTACCAG	ACGTTGGATGTGCGCAAAGAGTCAGTTTCC	CCAGGATCTTACAAGCTTA
id12000380		12	0.910754	GG	AA	*Yld2.1*bkgd	ACGTTGGATGATGTTGCCAAGCTTCACGTC	ACGTTGGATGACGATGGTGATCCAGGAGAC	aaGCCTCCTGCATCATCA
id12000824		12	1.71126	GG	AA	*Yld2.1*bkgd	ACGTTGGATGGATGGGTATTGCTAACACTG	ACGTTGGATGCTGCTGATTGGGCATCAGAG	gttgTTTTATTTTGTCAACACTAGTAC
wd12000212		12	2.525627	GG	AA	*Yld2.1*bkgd	ACGTTGGATGACATCCCTGAACCGCAATAC	ACGTTGGATGCCACGTCAATAAAACCAACC	cCCAGAAATCTTGACCCA
id12001409		12	3.484066	TT	CC	*Yld2.1*bkgd	ACGTTGGATGGCGGCATGTTGACATAATCC	ACGTTGGATGGGTTGAAACTCACTGGGAAG	gtacTCCTTGCAGGACTGCTCAAACA
ud12001503		12	25.595115	TT	CC	*Yld2.1*bkgd	ACGTTGGATGCCTCCTCTTTGGTCGTCATC	ACGTTGGATGTGGCATCAAAGTGAACAGGC	GCTGCCATGTCCCGT
id12009772		12	26.348436	TT	CC	*Yld2.1*bkgd	ACGTTGGATGGACATCACCATCATCGATGC	ACGTTGGATGAAAGCGTACCGTCTCAGCGA	GCACGTTGACCATGTTCAGG
id12009955		12	27.214173	AA	GG	*Yld2.1*bkgd	ACGTTGGATGATGTCAAGTGCGACCACTCC	ACGTTGGATGGTTCACATGCTGAGCATGAG	CGACCACTCCAATTGTATCTG

Abbreviations: SNP.id = SNP identification number; iPLEX duplicate = asterisks indicate that KBioscience KASP markers were developed targeting the same SNP; Chr = chromosome; MSU V.6 Position (Mb) = genetic position, Jeff allele = Jefferson recurrent parent allele; O.rufi allele = *O. rufipogon* (IRGC105491) alleles.

**There were three different multiplex sets: *Yld2.1target* was used to genotype the target introgression on chromosome 2; *Yld6.1target* for the chromosome 6 target introgression; and *Yld2.1bkgd* for genotyping the background introgressions for family 43 and 85.

Table 2-4. InDel markers designed for *yld2.1, yld3.2*, *yld8.1*, and *yld9.1* introgressions.

InDel Marker Name	For YldQTL	Chr	Position MSUV6 (Mb)	Design by*	Forward primer sequence	Reverse primer sequence	Jeff allele	O. rufi allele	Approx. polym size
Chr1 6.2	*Yld2.1* 85_2-8	1	6.257638	MJK	GCACCGAATGTTGTAGG	TTGTTGCTTCGTCTGTTTCG	lower	higher	NA
Chr3 10.3	*Yld2.1* 85_2-8	3	10.357441	MJK	TCCGGTGTTTCGACGATAAT	ATCAGTCGCGGTTGCTGT	lower	higher	NA
Chr3 14.5	*Yld2.1* 85_2-8	3	14.493296	MJK	CATAACACCCAAATGCCTTC	AATGTTCACGGTCCAATAAGAA	lower	higher	NA
Chr10 22.2	*Yld2.1* 85_2-8	10	22.496274	MJK	GCGTCAGTCACAAAGGCAAT	GGAGATTGCGGGGAGAGAG	higher	lower	NA
R5M29	*Yld2.1* 85_2-8	5	13.861882	MJK	CTCGCTGTTTACTGACTGG	TTTGATGTACTGCCTGCTCT	higher	lower	NA
R5M30	*Yld2.1* 85_2-8	5	21.466926	MJK	CTCAATTTCACCCATCCC	CGCTCCGTCTCCAACCTC	higher	lower	NA
2.1_43_Tg_L-1	*Yld2.1* 43_1-2	2	4.94836	JNC	CGTTTTCTCGTCCTCGAAAG	CACAAGTTTTAGCGGTCTGGA	192	222	30
2.1_43_Tg_M-1	*Yld2.1* 43_1-2	2	9.926502	JNC	AAGAGGGCGCAGATATAGGG	TTTAACTGGTAACGTGAATGACA	179	159	20
2.1_43_Bkg2_M-1	*Yld2.1* 43_1-2	9	21.928231	JNC	AGAGGCGGTACAAGGAGGAG	CAGAGCAGGATCACCCACAC	187	163	24
2.1_43_Bkg2_R-1	*Yld2.1* 43_1-2	9	22.658931	JNC	TTGCTGTGTTGACGATTATGC	CGAAAATGAGCGGCAAATAC	151	130	21
2.1_43_Bkg4_M-1	*Yld2.1* 43_1-2	12	26.374105	JNC	TGCGACACGAAAATCCATAA	GCTCGAAAAGACCGATAACC	182	210	28
2.1_43_Bkg4_R-1	*Yld2.1* 43_1-2	12	27.276298	JNC	ATAATTGGCCTGCAAACTGC	AATCGTGCCAATCAATACCA	195	169	26
3.2_16_Bkg1_M-1a	*Yld3.2* fam16	3	9.023442	JNC	TGCGTCACACACCTAAATGC	AGCCGGAATATTGCTGACC	185	204	19
3.2_16_Bkg4_R-1	*Yld3.2* fam16	10	22.531296	JNC	CCTCCTAGTGCCAAGTGTCC	GCTAATTTATACCCCTTCCGTTT	191	162	29
8.1_121_Bkg3_R-1	*Yld8.1* fam121	5	22.904895	JNC	CCCAAATAGTCCACAGTTTCTGA	ATGGGCCTATTTGTGGAAGTT	165	135	30
8.1_121_Bkg4_L-1	*Yld8.1* fam121	7	17.797305	JNC	TTGCACAAAATCGAATATAGGG	ACCAACGCTACTGAAATTTGG	119	99	20
8.1_121_Tg_M-1	*Yld8.1* fam121	8	21.074265	JNC	CCTGCATGTTGCACATCTTT	TGTGTGTATGACAGGTTGATGTG	129	108	21
9.1_13_Tg_R-1	*Yld9.1* fam13	9	22.658931	JNC	TTGCTGTGTTGACGATTATGC	CGAAAATGAGCGGCAAATAC	151	130	21

Abbreviations: For YldQTL = Yield QTL target introgression and intended family for use; Chr = chromosome; Position MSU V6 (Mb) = genetic position, Jeff allele = Jefferson recurrent parent allele; *O. rufi* allele = *O. rufipogon* (IRGC105491) alleles; Approx. polym size = approximate polymorphism size between the alleles. * InDel markers were developed by MJK= Michael J. Kovach, and JNC = Joshua N. Cobb. All markers are visible on 1.5% agarose gels.

Table 2-5. Yield of 50 introgression lines (ILs) encompassing six yield QTL targets and 20 control sib-lines evaluated over eight flooded field environments.

Lines tested	*Yld1.1*	Lsm	*Yld2.1*	Lsm	*Yld3.2*	Lsm	*Yld6.1*	Lsm	*Yld8.1*	Lsm	*Yld9.1*	Lsm
1	158_1-5	6919.4	43_1-2	7937.7	16_1-4	7151.1	219_1-5	7762.6	307_1-5	6778.1	9_2-9	6275.0
2	158_1-7	7014.8	43_1-4	7800.3	16_1-12	7044.5	219_2-9	7838.6	307_1-6	6885.5	9_2-10	6363.9
3	158_2-6	7146.9	43_1-5	7645.0	16_2-1	7306.2	221_1-11	6863.8	307_1-12	6913.4	13_1-1	7176.3
4	158_2-7	7428.0	43_2-1	7543.1	16_2-11	6745.4	221_2-4	6401.1	338_1-3	6749.1	13_1-3	6878.6
5	158_2-12	7178.9	43_2-8	7779.0					338_1-4	7053.1	13_2-9	6941.0
6	185_2-4	6747.3	43_2-12	7534.9					338_2-11	7299.6	13_2-10	7153.6
7	185_2-11	6857.6	141_1-1	6963.0					121_2-2*	7242.9	16_1-1	6388.6
8	89_1-5	6684.3	141_2-3	6927.9					121_2-8*	7169.4	16_1-2	6990.4
9	89_1-7	7066.5	141_2-6	6752.2					121_2-12*	7239.9	16_1-10	6813.8
10	89_2-5	6278.9	141_2-11	7175.2							13_1-12*	6014.3
11	89_2-6	6841.7									13_2-2*	6752.5
											13_2-5*	6119.1
Controls												
1	185_2-2	6571.3	141_1-8	5836.0	16_1-6**	6025.8	219_1-12	6818.2	338_2-3	6559.3	9_1-3	7033.0
2	185_2-3	6195.2	141_2-1	6390.5	16_2-3**	6499.4	219_2-10	6681.4	338_2-5	6587.1	9_1-7	6947.3
3	185_2-6	6384.0	141_2-9	6129.4			219_2-11	6596.7	121_1-1*	6656.7	9_2-2	6804.4
4							221_1-4	6036.2	121_2-3*	7027.5	16_1-6**	6025.8
5							221_1-13	6642.8			16_2-3**	6499.4

Standard error (stderr) values for all ILs that were tested in both 2007 and 2008 were 823.8 (except for lines 158_2-12, 219_2-11, and 9_2-10 with 824.6 standard error, and line 185_2-3 with stderr = 831.6) and the degrees-of-freedom was 65. Lsm = least squares mean of yield (kg ha^{-1}) under flooded field conditions. Jefferson, the recurrent parent, yielded 6216.7 with standard error of 809.7. The two ILs that are shaded in gray were entered into the URRN trials. * These 8 lines were omitted from the 2008 field trials due to red pericarp or segregation of red pericarp evident in the previous year. Lsmeans standard error was 861.5 for families 121 and 13 except for 121_1-1 with a standard error of 865.4. ** Note that values for control lines 16_1-6 and 16_2-3 are duplicated to represent control lines for both *yld3.2* and *yld9.1*.

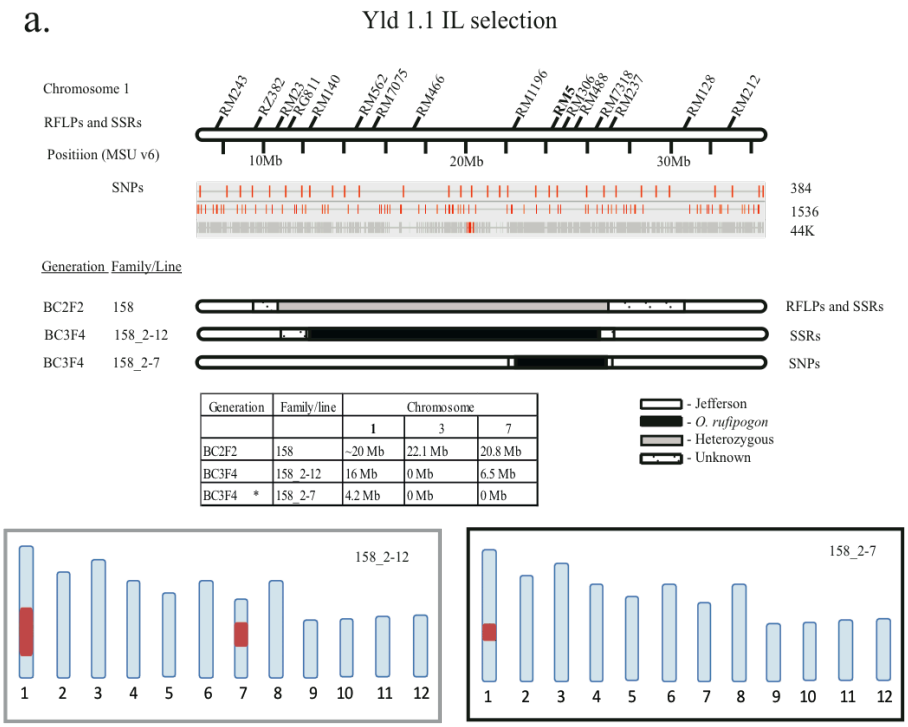
Table 2-6. Yield performance across four locations under flooded conditions during 2007-8 of lines submitted to the Genetic Stocks *Oryza* (GS*OR*) collection.

Line.id	Line name	PA#	Target QTL	% donor genome	Target introg. size (kb)	# Bkg_ introg.	2007-08 lsmean yield	lsmean_std_err	2007-08 %diff to Jeff
yld1_A	158_2-7	PA23899	1.1	3.1%	4158	0	7428.0	823.8	11.3
yld2_A	43_1-2	PA23931	2.1	6.7%	11714	4	7937.7	823.8	27.7
yld2_B	43_2-12	PA23954	2.1	5.6%	11714	3	7534.9	823.8	21.2
yld2_C	43_1-2_7-1	PA29887	2.1	7.8%*	11714	3	NA	NA	NA**
yld2_D	85_2-8_16-8	PA34445	2.1	9.7%	9920	2	NA	NA	NA***
yld3_A	16_2-1	PA23978	3.2	11.94%	6373	5	7306.2	823.8	17.5
yld3_B	16_2-1_17-3	PA29884	3.2	4.9%	7001	2	NA	NA	NA**
yld6_A	219_2-9	PA23984	6.1	5.6%	14160	2	7838.6	823.8	26.1
yld6_B	219_1-5	PA23981	6.1	4.9%	14160	2	7762.6	823.8	24.9
yld6_C	219_1-5_29-7	PA34451	6.1	5.6%*	14160	0	NA	NA	NA
yld8_A	121_2-2	PA23992	8.1	16.38%	8280	6	7242.9	861.5	16.5
yld9_A	13_1-1	PA24032	9.1	2.9%	5295	2	7176.3	823.8	15.4
	Jefferson						6216.8	809.7	

Abbreviations: GSOR_ID = Accession number used to order seeds from GSOR; PA# = Pedigree identification number to track seeds in McCouch lab; % donor genome = Percentage of donor genome calculated by the number of polymorphic markers out of total markers genotyped; Target introg. size (kb) = Size of the introgression at the target QTL; #_Bkg_introg. = Number of *O. rufipogon* introgressions in genetic background; 2007-08 lsmean yield = least squares mean yield for 2007-2008 across four flooded locations; lsmeans_std_err = standard error of the mean; 2007-08 %diff to Jeff = Yield performance in 2007-08 of lines compared to the recurrent parent, Jefferson, calculated as the percent improvement. * BC_4 generation % genome was calculated based on 384 SNP OPA genotypes and appears to have higher genome content than BC_3 due to the uneven coverage of polymorphic SNPs between the parents when using the Illumina 1,536 array ** No data available for yield on lines derived from high performing families and selected genotypically to reduce background introgressions. *** Line yld2_D has lower yield and is

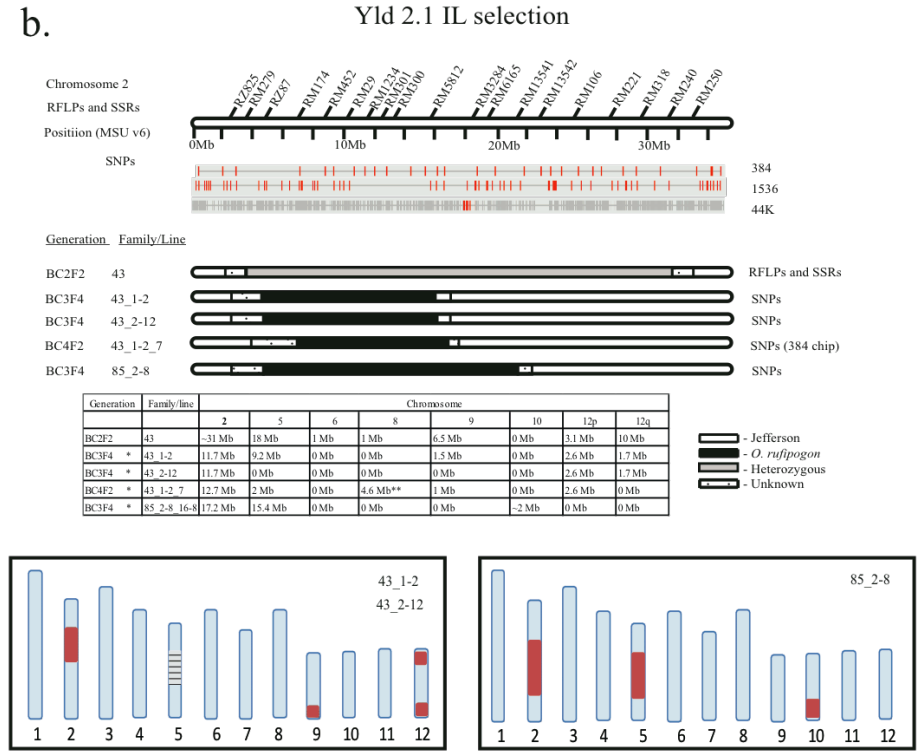
susceptible to straighthead but is interesting genotypically because it has fewer *O. rufipogon* background introgressions and contrasts with other ILs in family 43 lines (yld2_A, B and C). Grey boxes indicate accessions with the greatest yield improvement as compared to Jefferson.

a.



Generation	Family/line	Chromosome		
		1	3	7
BC2F2	158	~20 Mb	22.1 Mb	20.8 Mb
BC3F4	158_2-12	16 Mb	0 Mb	6.5 Mb
BC3F4 *	158_2-7	4.2 Mb	0 Mb	0 Mb

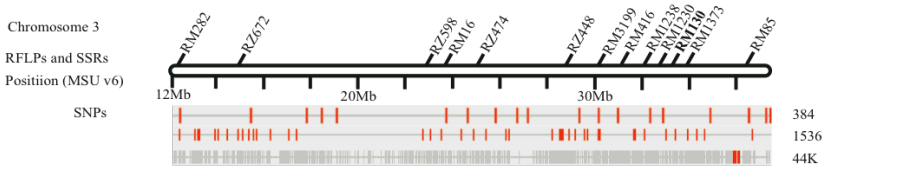
b.



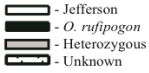
Generation	Family/line	Chromosome							
		2	5	6	8	9	10	12p	12q
BC2F2	43	~31 Mb	18 Mb	1 Mb	1 Mb	6.5 Mb	0 Mb	3.1 Mb	10 Mb
BC3F4 *	43_1-2	11.7 Mb	9.2 Mb	0 Mb	0 Mb	1.5 Mb	0 Mb	2.6 Mb	1.7 Mb
BC3F4 *	43_2-12	11.7 Mb	0 Mb	0 Mb	0 Mb	0 Mb	0 Mb	2.6 Mb	1.7 Mb
BC4F2 *	43_1-2_7	12.7 Mb	2 Mb	0 Mb	4.6 Mb**	1 Mb	0 Mb	2.6 Mb	0 Mb
BC3F4 *	85_2-8_16-8	17.2 Mb	15.4 Mb	0 Mb	0 Mb	0 Mb	~2 Mb	0 Mb	0 Mb

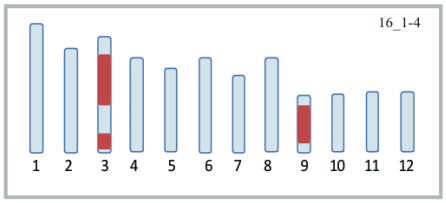
c.

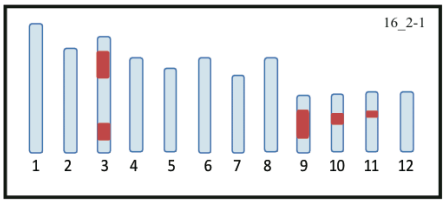
Yld 3.2 IL selection



Generation	Family/line	Chromosome				
		3	7	9	10	11
BC2F2	16	~ 24 Mb	~ 3 Mb	~ 20 Mb	1.5 Mb	14 Mb
BC3F4	16_1-4	17.6 Mb	3Mb/ 0.1 Mb	14.2 Mb	1.2 Mb	0 Mb
BC3F4 *	16_2-1	6 Mb	0.1 Mb	14.2 Mb	1.2 Mb	1.1 Mb

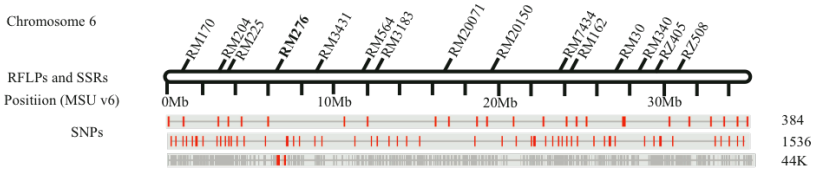




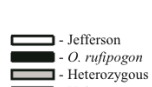


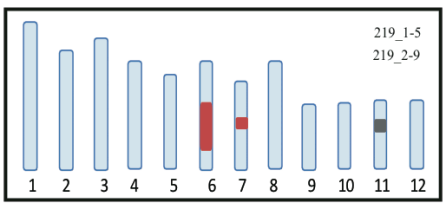
d.

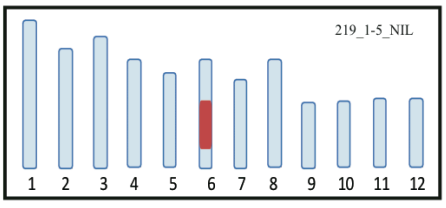
Yld 6.1 IL selection



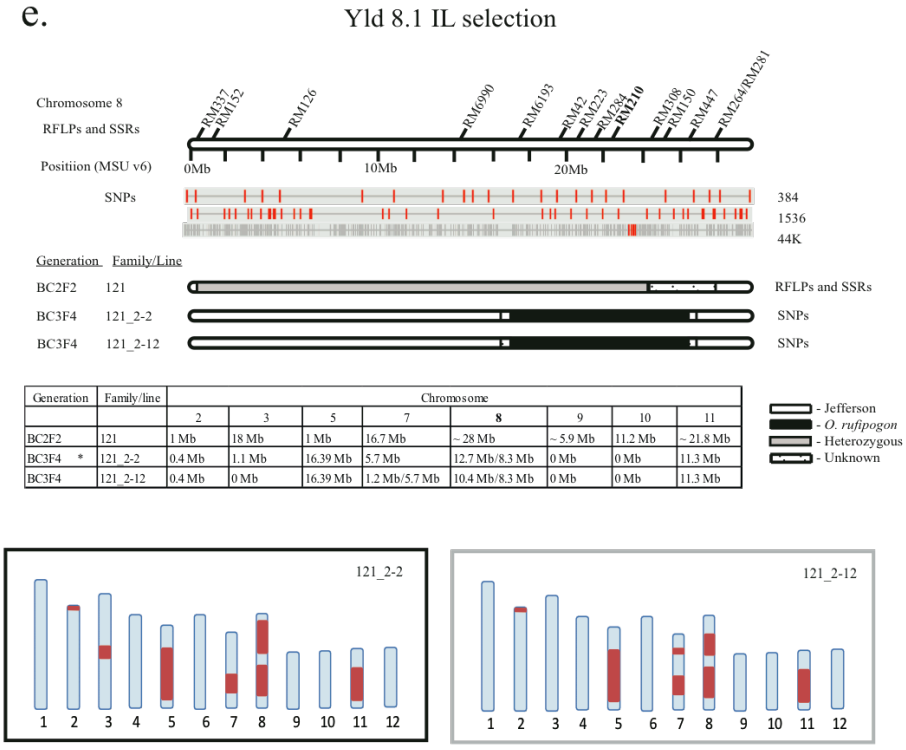
Generation	Family/line	Chromosome					
		3	5	**6**	7	9	11
BC2F2	219	11.4 Mb	~2 Mb	~24 Mb	3.6 Mb	~2 Mb	4.0 Mb
BC3F4 *	219_1-5	0 Mb	0 Mb	14.1 Mb	1.3 Mb	0 Mb	0.3 Mb
BC3F4 *	219_2-9	0 Mb	0 Mb	14.1 Mb	1.3 Mb	0 Mb	3.8 Mb





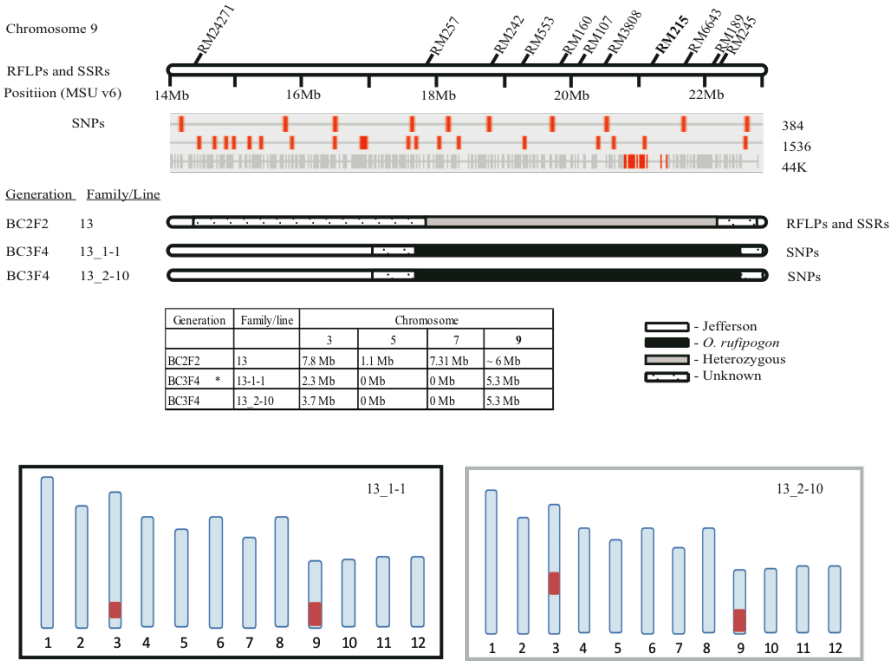


e.



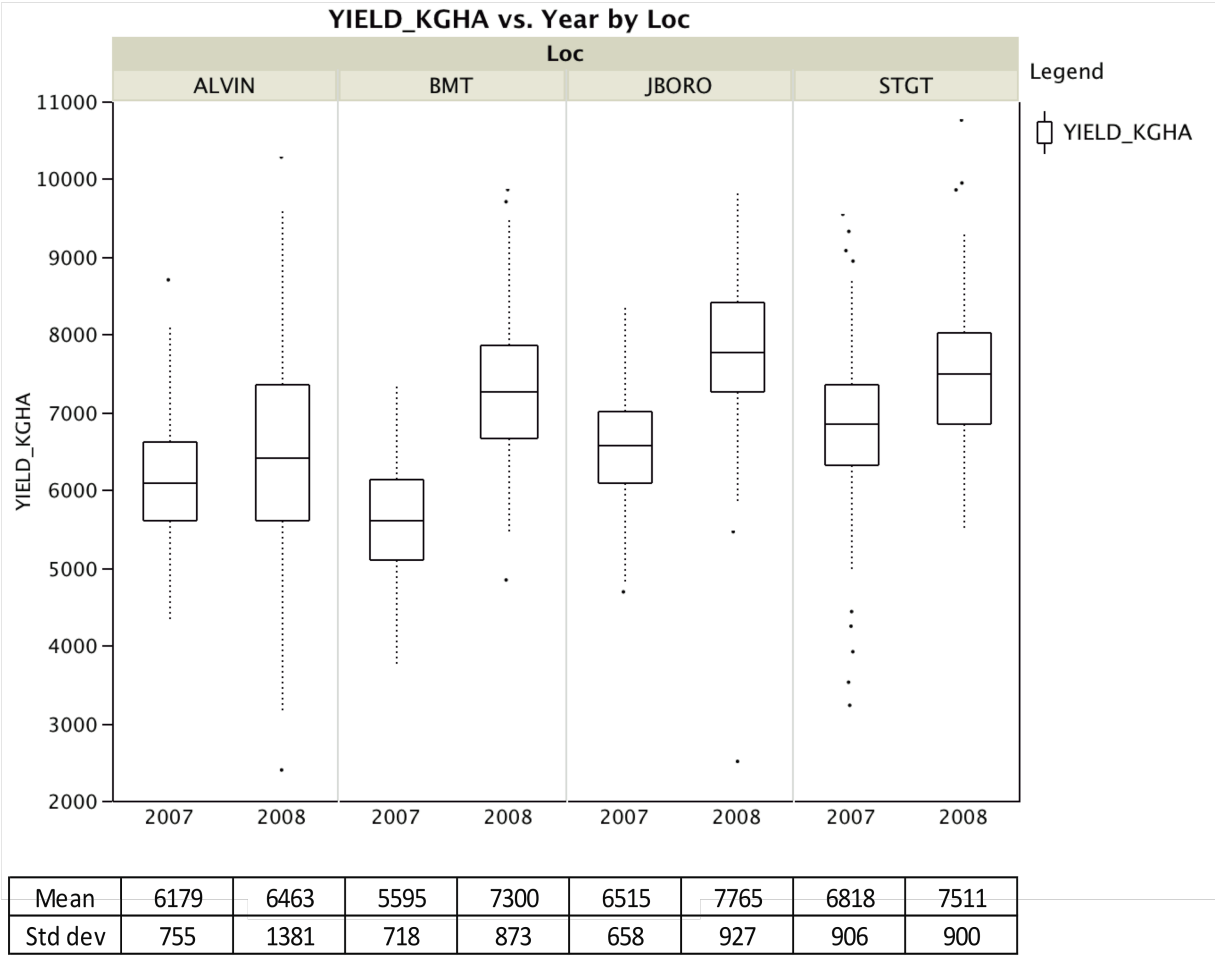
Generation	Family/line	Chromosome							
		2	3	5	7	**8**	9	10	11
BC2F2	121	1 Mb	18 Mb	1 Mb	16.7 Mb	~ 28 Mb	~ 5.9 Mb	11.2 Mb	~ 21.8 Mb
BC3F4 *	121_2-2	0.4 Mb	1.1 Mb	16.39 Mb	5.7 Mb	12.7 Mb/8.3 Mb	0 Mb	0 Mb	11.3 Mb
BC3F4	121_2-12	0.4 Mb	0 Mb	16.39 Mb	1.2 Mb/5.7 Mb	10.4 Mb/8.3 Mb	0 Mb	0 Mb	11.3 Mb

f.



Generation	Family/line	Chromosome			
		3	5	7	**9**
BC2F2	13	7.8 Mb	1.1 Mb	7.31 Mb	~ 6 Mb
BC3F4 *	13-1-1	2.3 Mb	0 Mb	0 Mb	5.3 Mb
BC3F4	13_2-10	3.7 Mb	0 Mb	0 Mb	5.3 Mb

Figure 2-2. Graphical genotypes of IL families showing 6 QTL target regions; a) *yld1.1*; b) *yld2.1*; c) *yld3.2*; d) yld6.1; e) *yld8.1*; f) *yld9.1*; and background introgressions identified by SNPs, SSRs and RFLPs. The top graph illustrates the target QTL region marked by SSR and RFLP markers as originally mapped by Thomson *et al.* (2003); the rows in light gray indicate 384 SNP, 1536 SNP and 44K SNP assays; horizontal chromosome bars indicate regions of heterozygosity (diagonal stripe), homozygous *O. rufipogon* introgressions (black) and Jefferson DNA (white); Table summarizes size and number of *O. rufipogon* introgression in each line and the asterisks indicates the 12 lines selected for submission to the GSOR; bottom graphs contain vertical chromosomes with regions of introgression indicated in dark gray; graphs framed in black indicate lines selected for submission to GSOR. The double asterisks on chromosome 8 introgression are non-parental segment.



Mean	6179	6463	5595	7300	6515	7765	6818	7511
Std dev	755	1381	718	873	658	927	906	900

Figure 2-3. Mean yield performance of 50 ILs in four different environments (Alvin, Beaumont, Jonesboro, Stuttgart) in two years (2007-2008). The number of total entries in each location in 2007 was 216. In 2008, 180 entries were evaluated in Alvin, 189 at Jonesboro, and 192 in both Beaumont and Stuttgart. Std dev = standard deviation.

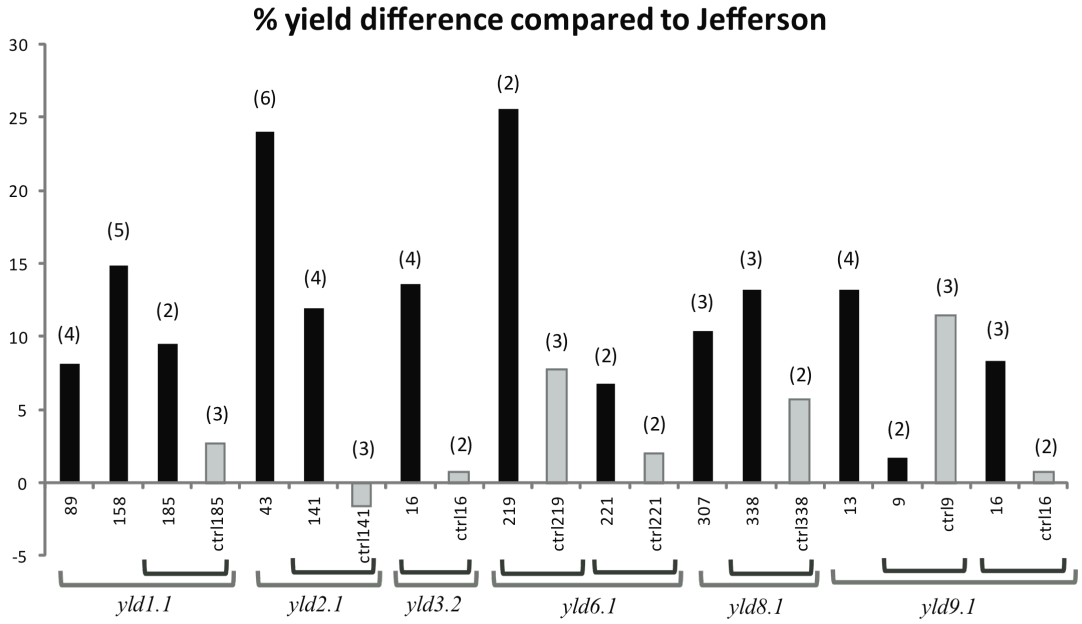


Figure 2-4. Percent difference in yield (kg ha^{-1}) relative to Jefferson of BC_3 families containing (black bars) or lacking (gray bars) the *O. rufipogon* introgression at the target QTL evaluated at four flooded locations during 2007-8. Multiple families containing each QTL are represented by the lower bracket, and pairs of families with (black) or without (gray) the target QTLs are indicated by the upper bracket. The number in parentheses above the bars shows the number of lines in each of the families.

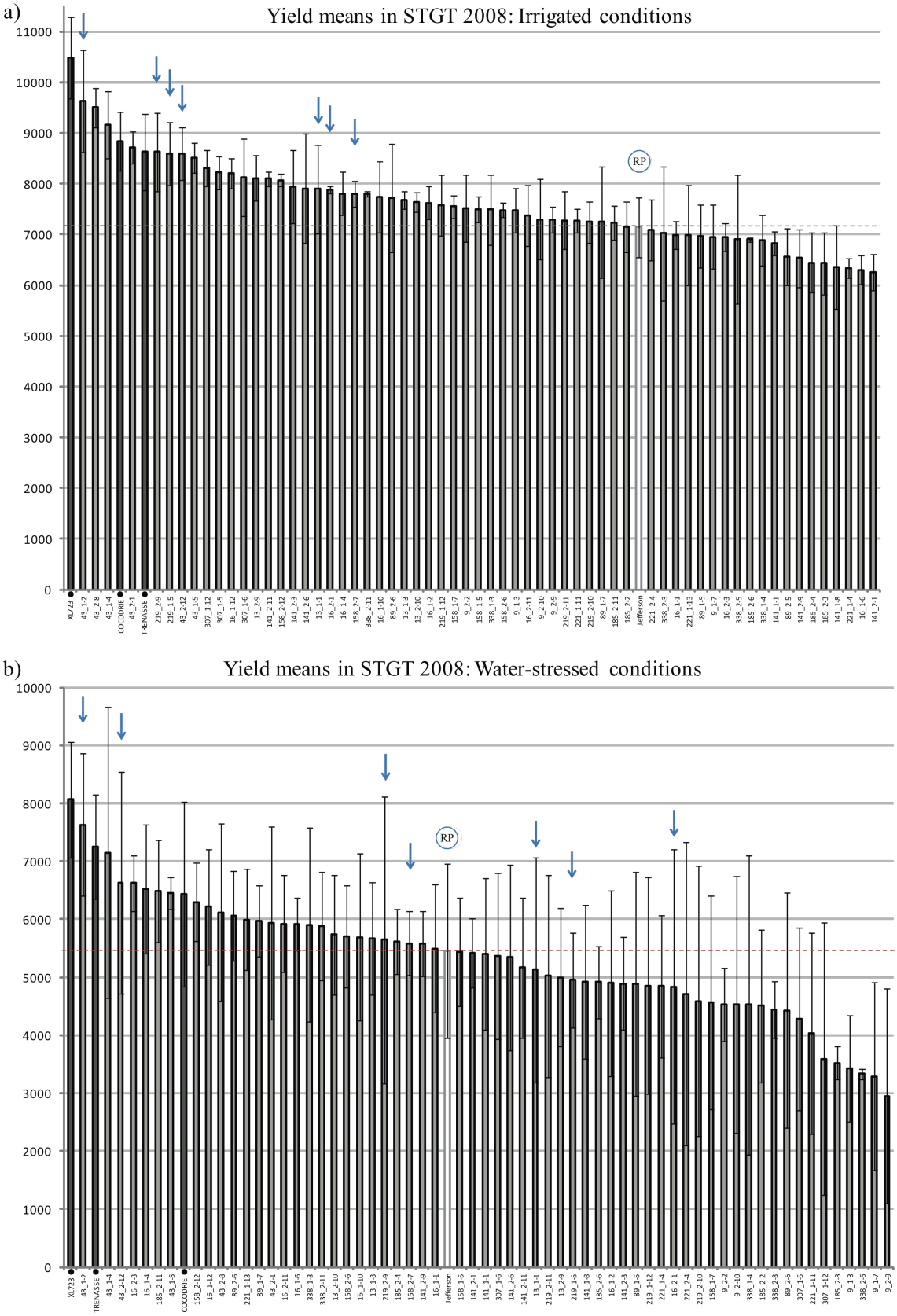

a)
Yield means in STGT 2008: Irrigated conditions
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
RP
XL723
43_1-2
43_2-8
43_1-4
COCODRIE
43_2-1
TRENASSE
219_2-9
219_1-5
43_2-12
43_1-5
307_1-12
307_1-5
16_1-12
307_1-6
13_2-9
141_2-11
158_2-12
141_2-3
141_2-6
13_1-1
16_2-1
16_1-4
158_2-7
338_2-11
16_1-10
89_2-6
13_1-3
13_2-10
16_1-2
219_1-12
158_1-7
9_2-2
158_1-5
338_1-3
158_2-6
9_1-3
16_2-11
9_2-10
9_2-9
219_2-11
221_1-11
219_2-10
89_1-7
185_2-11
185_2-2
Jefferson
221_2-4
338_2-3
16_1-1
221_1-13
89_1-5
9_1-7
16_2-3
338_2-5
185_2-6
338_1-4
141_1-1
89_2-5
141_2-9
185_2-4
185_2-3
141_1-8
221_1-4
16_1-6
141_2-1
b)
Yield means in STGT 2008: Water-stressed conditions
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
RP
XL723
43_1-2
TRENASSE
43_1-4
43_2-12
16_2-3
16_1-4
185_2-11
43_1-5
COCODRIE
158_2-12
16_1-12
43_2-8
89_2-6
221_1-13
89_1-7
43_2-1
16_2-11
16_1-6
338_1-3
338_2-11
13_2-10
158_2-6
16_1-10
13_1-3
219_2-9
185_2-4
158_2-7
141_2-9
16_1-1
Jefferson
158_1-5
141_2-1
141_1-1
307_1-6
141_2-6
141_2-11
13_1-1
219_2-11
13_2-9
219_1-5
141_1-8
185_2-6
16_1-2
141_2-3
89_1-5
219_1-12
221_1-4
16_2-1
221_2-4
219_2-10
158_1-7
9_2-2
9_2-10
338_1-4
185_2-2
338_2-3
89_2-5
307_1-5
221_1-11
307_1-12
185_2-3
9_1-3
338_2-5
9_1-7
9_2-9

Figure 2-5. Yield performance ((kg ha^{-1}) of ILs and commercial checks in Stuttgart A) Irrigated plots during 2008; b) Water-stressed plots in 2008. Values show one-way ANOVA means with error bars representing one standard deviation from the mean. Arrows indicate ILs submitted to GSOR (Table 2); dark gray bars indicated with a filled black circle along the x-axis are checks; dotted line shows average yield performance of recurrent parent, Jefferson (white bar) indicated by RP symbol.

Table 2-7. Regression analysis using IL and control sib-lines identifying significant variables explaining yield under A) irrigated field conditions and B) water-stressed conditions. Variables highlighted in bold font indicate variables explaining most of the model.

A) Summary of Forward Selection: Flooded (IL entries only)								
Step	**Variable Entered**	**Label**	**Number Vars In**	**Partial R-Square**	**Model R-Square**	**C(p)**	**F Value**	**Pr > F**
1	**AVPANL**	**AVPANL**	1	**0.1771**	0.1771	250.192	166.82	<.0001
2	**AVPLTWT**	**AVPLTWT**	2	**0.0755**	0.2527	158.252	78.25	<.0001
3	**AMYLOSE**	**AMYLOSE**	3	**0.0393**	0.292	111.375	42.92	<.0001
4	**KSDWT**	**KSDWT**	4	**0.0254**	0.3174	81.8221	28.7	<.0001
5	**GL**	**GL**	5	**0.034**	0.3514	41.5511	40.41	<.0001
6	WHOLEMY	WHOLEMY	6	0.0164	0.3678	23.1447	19.99	<.0001
7	TOTALMY	TOTALMY	7	0.0089	0.3766	14.1329	10.92	0.001
8	GW	GW	8	0.0057	0.3824	9	7.13	0.0077

B) Summary of Forward Selection: Water-stressed (IL entries only)								
Step	**Variable Entered**	**Label**	**Number Vars In**	**Partial R-Square**	**Model R-Square**	**C(p)**	**F Value**	**Pr > F**
1	**KSDWT**	**KSDWT**	1	**0.3432**	0.3432	137.926	125.93	<.0001
2	**AVPANWT**	**AVPANWT**	2	**0.1025**	0.4457	81.109	44.37	<.0001
3	**PLTHT**	**PLTHT**	3	**0.0654**	0.5111	45.592	31.96	<.0001
4	**AVTILL**	**AVTILL**	4	**0.0548**	0.5659	16.132	30.05	<.0001
5	AVPANL	AVPANL	5	0.0159	0.5818	8.9905	9.03	0.0029
6	STDPCT	STDPCT	6	0.007	0.5888	7	3.99	0.0469

Abbreviations: AVPANL = average panicle length; AVPLTWT = average plant weight; AMYLOSE = amylose content; KSDWT = 1,000-seed weight; GL = grain length; WholeMY = Whole milling yield (filled seeds only); TotalMY = Total milling yield (filled and unfilled seed); GW = grain width; AVPANWT = average panicle weight; PLTHT = plant height; AVTILL = average number of tillers; STDPCT = stand percent.

Table 2-8. Correlation analysis of IL and control sib-line means of traits measured under A) irrigated field conditions in 2007-2008 and B) water-stressed field conditions in 2008. Missing correlations are indicated by periods.
Genetic correlation is significantly different from zero at p<0.05 (*), p<0.01 (**), p<0.001 (***), and p<0.0001 (****and gray coloration).

A	Yield	TotalMY	WholeMY	AVPANL	AVSDPAN	AVPANWT	AVTILL	AVPLTWT	KSDWT	ASV	AMYLOSE	GL	GW	CHKPCT	PLSQM	STDPCT	D2HD	PLTHT	Lodge
Yield	1	0.432****	0.084	0.392****	0.260****	0.324****	0.247****	0.462****	-0.160****	0.156*	-0.169**	-0.010	0.043	0.342****	-0.285****	0.136	0.554****	0.062	0.256****
TotalMY		1	0.263****	0.391****	0.309****	0.312****	0.288****	0.545****	-0.287****	0.178**	-0.086	0.218****	-0.080	-0.089	-0.444****	-0.512****	0.292****	0.102*	0.331****
WholeMY			1	0.391****	0.450****	0.570****	-0.185***	0.266****	-0.021	0.070	-0.017	0.199****	0.155**	-0.035	-0.384****	-0.591****	0.078	-0.022	-0.060
AVPANL				1	0.636****	0.618****	-0.045	0.424****	-0.407****	0.229***	0.187**	0.113*	-0.228****	-0.018	-0.266****	-0.437****	0.410****	0.330****	0.257****
AVSDPAN					1	0.900****	-0.040	0.533****	-0.513****	0.260****	0.179**	0.153**	-0.180***	-0.292****	-0.367****	-0.520****	0.317****	0.236****	0.255****
AVPANWT						1	-0.148**	0.519****	-0.186***	0.178**	0.172**	0.169***	0.047	-0.135**	-0.458****	-0.549****	0.329****	0.113*	0.076
AVTILL							1	0.700****	-0.283****	0.322****	-0.030	-0.006	-0.093	-0.042	-0.250****	0.271**	0.218****	0.140**	0.508****
AVPLTWT								1	-0.396****	0.220***	-0.044	0.163**	-0.067	-0.088	-0.558****	0.115	0.344****	0.313****	0.562****
KSDWT									1	-0.419****	-0.055	0.059	0.487****	0.306****	0.061	0.172	-0.202****	-0.340****	-0.529****
ASV										1	0.353****	-0.186**	-0.309****	-0.223***	-0.150*	.	0.529****	0.218***	0.243****
AMYLOSE											1	-0.341****	0.068	-0.138*	-0.066	.	0.301****	0.006	-0.096
GL												1	-0.247****	-0.428****	-0.175**	-0.509****	-0.277****	0.067	0.185***
GW													1	0.379****	-0.091	0.058	-0.045	-0.191***	-0.182***
CHKPCT														1	-0.048	0.425****	0.331****	-0.078	-0.102*
PLSQM															1	.	-0.488****	-0.164**	-0.219***
STDPCT																1	0.451****	-0.185*	.
D2HD																	1	0.129**	0.197****
PLTHT																		1	0.406****
Lodge																			1

B	Yield	TotalMY	WholeMY	AVPANL	AVSDPAN	AVPANWT	AVTILL	AVPLTWT	KSDWT	GL	GW	CHKPCT	STDPCT	D2HD	PLTHT
Yield	1	-0.374****	-0.161	0.028	0.295***	0.519****	0.581****	0.477****	0.640****	0.329***	0.225*	-0.095	0.224*	0.171	0.390****
TotalMY		1	0.645****	0.163	0.187*	0.094	-0.301***	0.369**	-0.218*	0.107	-0.126	-0.115	0.046	-0.050	-0.389****
WholeMY			1	0.147	0.175	0.128	-0.193*	0.245	-0.095	0.282**	-0.180*	-0.448****	-0.051	0.201*	-0.129
AVPANL				1	0.543****	0.474****	-0.223*	0.413***	-0.134	-0.004	-0.221*	-0.068	0.125	0.053	-0.031
AVSDPAN					1	0.926****	-0.103	0.780****	-0.081	0.002	-0.038	-0.029	0.200*	0.047	-0.107
AVPANWT						1	0.069	0.798****	0.279**	0.211*	0.095	-0.070	0.210*	0.122	0.030
AVTILL							1	0.402**	0.464****	0.180*	0.213*	-0.061	0.094	0.089	0.216*
AVPLTWT								1	0.096	0.103	0.106	-0.053	0.302*	0.001	-0.255*
KSDWT									1	0.555****	0.366****	-0.142	0.039	0.192*	0.381****
GL										1	-0.127	-0.237**	-0.112	0.316***	0.180*
GW											1	0.302***	0.069	-0.226*	-0.053
CHKPCT												1	0.034	-0.400****	-0.278**
STDPCT													1	-0.412****	-0.163
D2HD														1	0.436****
PLTHT															1

Abbreviations: Yield = grain yield in kg ha^{-1}; TotalMY = Total milling yield (filled and unfilled seed); WholeMY = Whole milling yield (filled seeds only); AVPANL = average panicle length; AVSDPAN = average seeds per panicle; AVPANWT = average panicle weight; AVTILL = average number of tillers; AVPLTWT = average plant weight; KSDWT = thousand-seed weight; ASV = alkali spreading value; AMYLOSE = amylose content; GL = grain length; GW = grain width; CHKPCT = percent chalk; PLSQM = plants per square meter; D2HD = days to heading; PLTHT = plant height; Lodge = percent lodge.

Table 2-9. Yield and milling yield of yld2_A and yld2_B measured in six locations at the Uniform Rice Regional Nursery (URRN) trials during 2009, compared with three commercial varieties, Presidio, Cocodrie and Trenasse.

Line	Yield	StdErr	Grouping	TotalMY	StdErr	Grouping	WholeMY	StdErr	Grouping
IL yld2_A	10108.2	282.1	A	74.6	0.4	A	49.0	1.9	B
COCODRIE	9523.4	264.6	A B	72.8	0.2	B	61.2	1.5	A
TRENASSE	9235.3	279.0	A B	69.8	0.3	D	66.3	1.8	A
IL yld6_A	8999.1	271.1	B	71.9	0.2	C	60.2	1.0	A
PRESIDIO	8562.3	221.5	B	71.6	0.2	C	60.4	1.0	A

Yield = least squares mean of grain yield; StdErr = standard error; Grouping = Letter groupings that demarcate the statistical differences between the line as determined by Tukey-Kramer test; TotalMY = total milling yield; WholeMY = whole milling yield.

Table 2-10. URRN 2009 disease resistance ratings of yld2_A (IL 43_1-2) and yld6_A (IL 219_2-9) compared with commercial checks.

	Bacterial Panicle Blight (LA) [1]	Sheath Blight (3 Loc) [1]	Panicle Blast (LA) [2]	Leaf Blast (TX) [1]	Leaf Smut (LA) [2]	Narrow Brown Leaf Spot (LA) [2]	Brown Spot (LA) [2]	Straighthead (AR) [1]
IL yld2_A	4.8	4.9	3.0	2.0	0.5	0.0	2.5	6.0
IL yld6_A	7.0	6.9	5.0	3.0	1.0	2.0	2.0	3.5
Presidio	4.2	6.0	5.0	1.0	3.3	0.8	0.8	2.3
Cocodrie	4.5	7.7	5.0	1.0	4.0	3.5	1.0	8.0
Trenasse	4.8	7.6	5.0	2.0	1.0	1.5	3.0	7.0

Abbreviations: LA = Louisiana; 3 Loc = Arkansas, Louisiana, and Texas; TX = Texas; AR = Arkansas

Dark gray highlights indicate desirable ratings, and light gray indicates susceptibility. Standard deviation unavailable due to nature of data collection in URRN trials.

[1] Replicated trial inoculated with pathogen. Straighthead induced with MSMA.

[2] Replicated trial based on natural incidence of disease

Table 2-11. Subset of LSmeans for sheath blight resistance evaluations in greenhouse conditions during 2009. Asterisks indicate lines with significantly higher resistance to sheath blight than Jefferson.

Line	PCTSB_lsmean	Std Err	Adj. p-value
Jasmine	38.5	6.49	0.0003*
16_2-3	48.8	6.47	0.0304*
16_1-6	49.0	6.75	0.0321*
Spring	57.6	4.82	ns
Cocodrie	58.3	4.72	ns
IL yld2_A (43_1-2)	64.5	6.57	ns
Wells	69.0	4.70	ns
IL yld6_A (219_2-9)	75.3	6.55	ns
Jefferson	76.0	4.97	ns
Lemont	92.1	6.59	ns

Abbreviations: PCTSB_lsmean = Least squares means of percent sheath blight; Std Err = standard error; Adj. p-value = adjusted p-value; ns = not significant.

Table 2-12. Lsmeans of yield, yield components and grain quality measurements in flooded trials conducted in 2007-2008 for lines submitted to GS*OR*.

Lines	Line name	**Yield**	StdErr	Diff	Adjp	**TotalMY**	StdErr	Diff	Adjp	**WholeMY**	StdErr	Diff	Adjp	**KSDWT**	StdErr	Diff	Adjp
yld1_A	158_2-7	7428.02	823.78	1211.26	0.02	70.27	1.25	0.26	1.00	56.37	2.72	0.38	1.00	26.95	0.94	0.10	1.00
yld2_A	43_1-2	7937.68	823.78	1720.92	0.00	71.64	1.25	1.63	0.08	53.82	2.72	-2.18	1.00	24.86	0.94	-1.99	0.00
yld2_B	43_2-12	7534.92	823.78	1318.16	0.01	71.01	1.25	1.00	0.87	53.65	2.72	-2.34	0.99	24.14	0.94	-2.71	0.00
yld3_A	16_2-1	7306.19	823.74	1089.43	0.05	69.18	1.25	-0.83	0.98	48.81	2.72	-7.18	0.00	27.22	0.94	0.37	1.00
yld6_A	219_2-9	7838.58	823.78	1621.82	0.00	69.75	1.25	-0.26	1.00	56.54	2.72	0.55	1.00	22.43	0.94	-4.42	0.00
yld6_B	219_1-5	7762.6	823.78	1545.84	0.00	69.8	1.25	-0.21	1.00	55.55	2.72	-0.44	1.00	23.31	0.94	-3.54	0.00
yld8_A	121_2-2	7242.87	861.47	1026.11	0.40	71.17	1.32	1.16	0.97	58.47	2.98	2.48	1.00	NA	NA	NA	NA
yld9_A	13_1-1	7176.34	823.78	959.58	0.15	70.81	1.25	0.80	0.99	53.80	2.72	-2.19	1.00	27.36	0.94	0.51	1.00
Jefferson	Jefferson	6216.76	809.66			70.01	1.20			55.99	2.58			26.85	0.88		

Lines	Line name	**Amylose**	StdErr	Diff	Adjp	**ASV**	StdErr	Diff	Adjp	**CHKPCT**	StdErr	Diff	Adjp
yld1_A	158_2-7	20.18	0.27	-0.02	1.00	3.94	0.19	-0.23	0.99	6.29	0.72	2.85	0.00
yld2_A	43_1-2	20.94	0.27	0.73	0.13	4.14	0.19	-0.03	1.00	5.37	0.72	1.93	0.14
yld2_B	43_2-12	20.27	0.27	0.06	1.00	4.02	0.19	-0.14	1.00	9.12	0.72	5.68	0.00
yld3_A	16_2-1	20.28	0.27	0.07	1.00	4.21	0.19	0.04	1.00	6.65	0.72	3.21	0.00
yld6_A	219_2-9	19.61	0.27	-0.60	0.42	3.70	0.19	-0.47	0.49	8.07	0.72	4.63	0.00
yld6_B	219_1-5	19.33	0.27	-0.87	0.02	3.52	0.19	-0.65	0.05	6.70	0.72	3.26	0.00
yld8_A	121_2-2	21.02	0.31	0.81	0.19	4.05	0.22	-0.12	1.00	NA	NA	NA	NA
yld9_A	13_1-1	20.79	0.27	0.59	0.46	4.00	0.19	-0.17	1.00	8.11	0.72	4.66	0.00
Jefferson	Jefferson	20.21	0.23			4.17	0.13			3.44	0.53		

Lines	Line name	**GL**	StdErr	Diff	Adjp	**GW**	StdErr	Diff	Adjp	**AVSDPAN**	StdErr	Diff	Adjp
yld1_A	158_2-7	6.69	0.07	-0.11	0.92	2.37	0.01	0.03	0.81	123.01	12.68	-24.34	0.24
yld2_A	43_1-2	6.75	0.07	-0.05	1.00	2.31	0.01	-0.02	0.99	142.80	12.68	-4.56	1.00
yld2_B	43_2-12	6.77	0.07	-0.03	1.00	2.23	0.01	-0.10	0.00	155.32	12.68	7.96	1.00
yld3_A	16_2-1	6.88	0.07	0.08	1.00	2.28	0.01	-0.06	0.05	105.33	12.68	-42.03	0.00
yld6_A	219_2-9	6.33	0.07	-0.47	0.00	2.29	0.01	-0.05	0.20	173.20	12.68	25.85	0.17
yld6_B	219_1-5	6.48	0.07	-0.32	0.00	2.31	0.01	-0.02	1.00	168.36	12.68	21.01	0.49
yld8_A	121_2-2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
yld9_A	13_1-1	6.92	0.07	0.12	0.84	2.36	0.01	0.02	0.99	127.16	12.68	-20.20	0.56
Jefferson	Jefferson	6.80	0.06			2.33	0.01			147.35	11.78		

Lines	Line name	**AVPANL**	StdErr	Diff	Adjp	**AVPANWT**	StdErr	Diff	Adjp	**AVTILL**	StdErr	**PLTHT**	StdDev	**D2HD**	StdDev
yld1_A	158_2-7	18.12	0.55	-0.54	1.00	3.18	0.28	-0.56	0.47	7.25	1.47	89.25	5.67	74.25	6.10
yld2_A	43_1-2	19.86	0.55	1.20	0.09	3.35	0.28	-0.39	0.96	7.51	1.47	90.81	4.36	79.58	6.63
yld2_B	43_2-12	20.48	0.55	1.82	0.00	3.54	0.28	-0.20	1.00	7.12	1.47	87.77	7.12	77.04	6.55
yld3_A	16_2-1	19.22	0.55	0.56	0.99	2.73	0.28	-1.01	0.00	7.63	1.47	90.21	6.45	75.42	6.25
yld6_A	219_2-9	18.03	0.55	-0.63	0.97	3.52	0.28	-0.22	1.00	7.62	1.47	92.35	7.77	74.13	6.35
yld6_B	219_1-5	18.07	0.55	-0.59	0.99	3.62	0.28	-0.12	1.00	7.44	1.47	92.17	6.63	73.42	6.11
yld8_A	121_2-2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	99.75	4.79	70.42	5.62
yld9_A	13_1-1	17.83	0.55	-0.83	0.65	3.24	0.28	-0.50	0.70	7.10	1.47	85.98	5.39	78.00	6.65
Jefferson	Jefferson	18.66	0.50			3.74	0.25			7.42	1.37	91.65	8.83	74.60	6.77

Abbreviations: Yield = grain yield in kg ha^{-1}; Yield_stdErr = Standard error of yield lsmeans; Diff = difference between Jefferson and the line values; TotalMY = Total Milling Yield (filled and unfilled seed); WholeMY = Whole Milling Yield (filled seeds only); KSDWT = 1,000-seed weight; Amylose = amylose content; ASV = alkali spreading value; CHKPCT = percent chalk; GL = grain length (mm); GW = grain width (mm); AVPANL = average panicle length (cm);

AVPANWT = average panicle weight (g); AVSDPAN = average seeds per panicle; AVTILL = average number of tillers. Lsmeans could not be calculated for plant height (PLTHT) and days to heading (D2HD) due to large location effects; instead the averages and standard deviation (StdDev) of 24 replicates for each line (except for yld8_A which was only tested in one year) are shown in the last four columns. Gray colored cells indicate lsmeans with significantly different values from the recurrent parent, Jefferson.

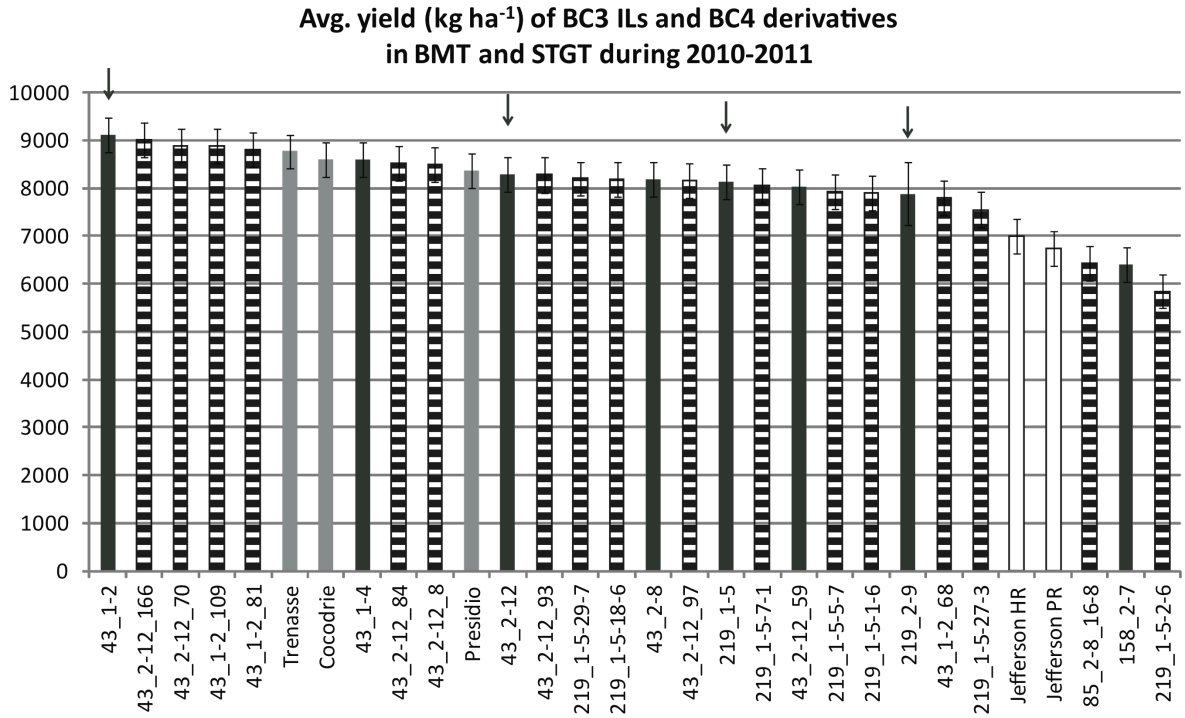


Figure 2-6. ANOVA mean yield performance of *yld2.1* and *yld6.1* BC_3 introgression lines and BC_4 derivatives in Beaumont during 2010 and both Beaumont and Stuttgart during 2011 field trials. Comparison of the yield performance of ILs (black bars) and derived ILs with reduced background introgressions (patterned bars) against commercial checks (Trenasse, Cocodrie and Presidio; highlighted in light gray) and Jefferson, the recurrent parent (white bars). Jefferson HR is the recurrent parent amplified in Texas in summer 2009 and Jefferson PR is the recurrent parent amplified in the winter nursery in Puerto Rico. BC_3 lines submitted to the GS*OR* are highlighted with arrows.

Yld 2.1 IL selection

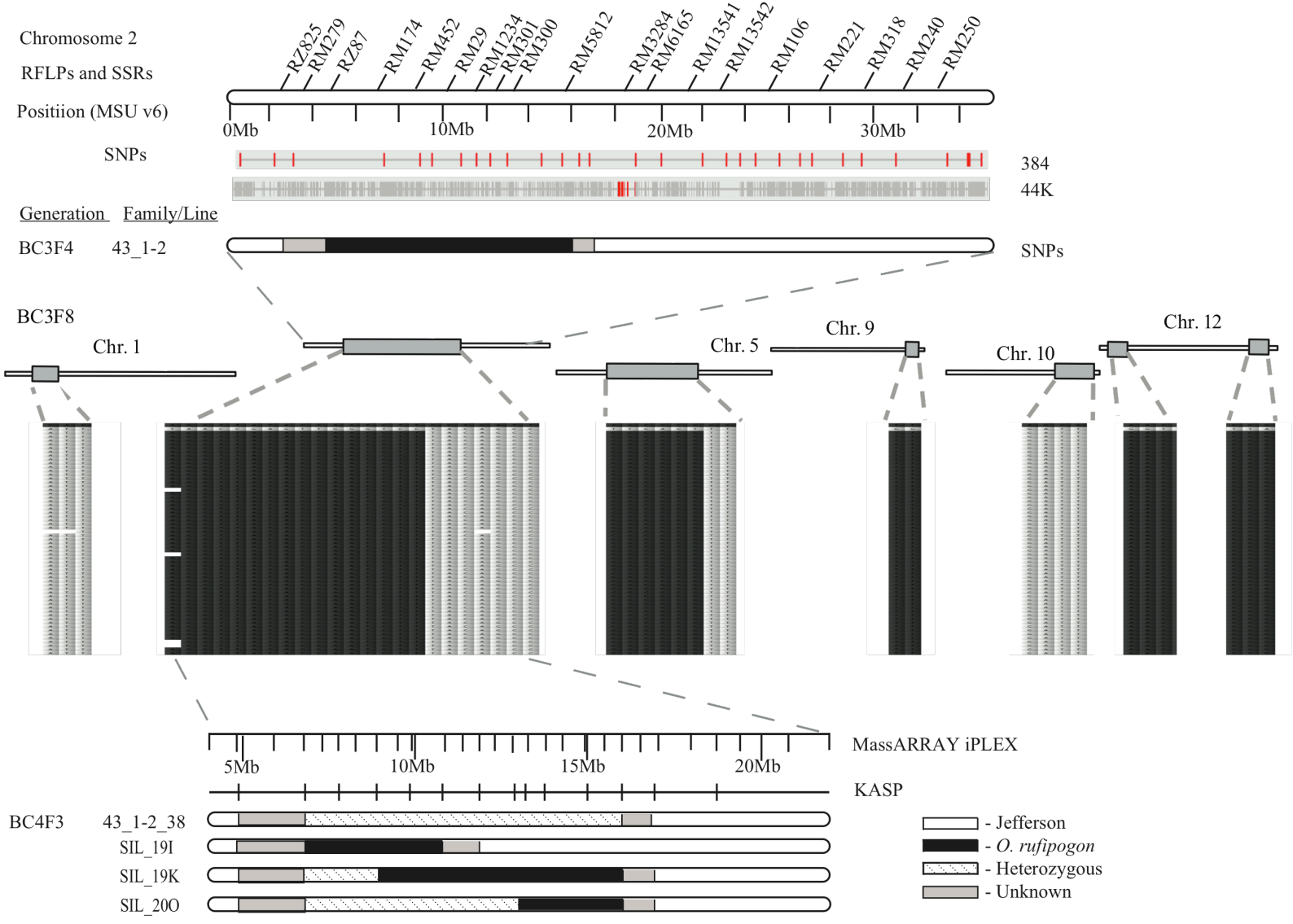


Figure 2-7. Graphical genotype of IL yld2_A at BC_3F_8 generation and derived BC_4 lines. The top graph illustrates the target QTL region marked by SSR and RFLP markers as originally mapped by Thomson *et al.* (2003); rows in light gray indicate 384 SNP and 44K SNP assays; horizontal bar depicts regions of homozygous *O. rufipogon* DNA (black), homozygous Jefferson DNA (white) and regions of recombination (gray) in BC_3F_4 generation. Multiple chromosomes are shown to represent target and background introgressions genotyped using iPLEX MassARRAY. A total of 61 samples representing 12 different headrows of yld2_A grown in 2010 were genotyped using 58 markers. Genotypes of selfed F_8 lines are represented with homozygous *O. rufipogon* introgressions (dark gray squares), homozygous Jefferson (light gray) and missing data (white) with the first two rows illustrating the *O. rufipogon* and Jefferson parental genotypes. Graph shows approximate locations of SNPs in iPLEX MassARRAY and KASP assays above sub-introgression lines (SILs); segregation within the target introgressions shown in bars indicating heterozygosity (diagonal stripe), homozygous *O. rufipogon* DNA (black), homozygous Jefferson DNA (white), and regions of recombination (gray).

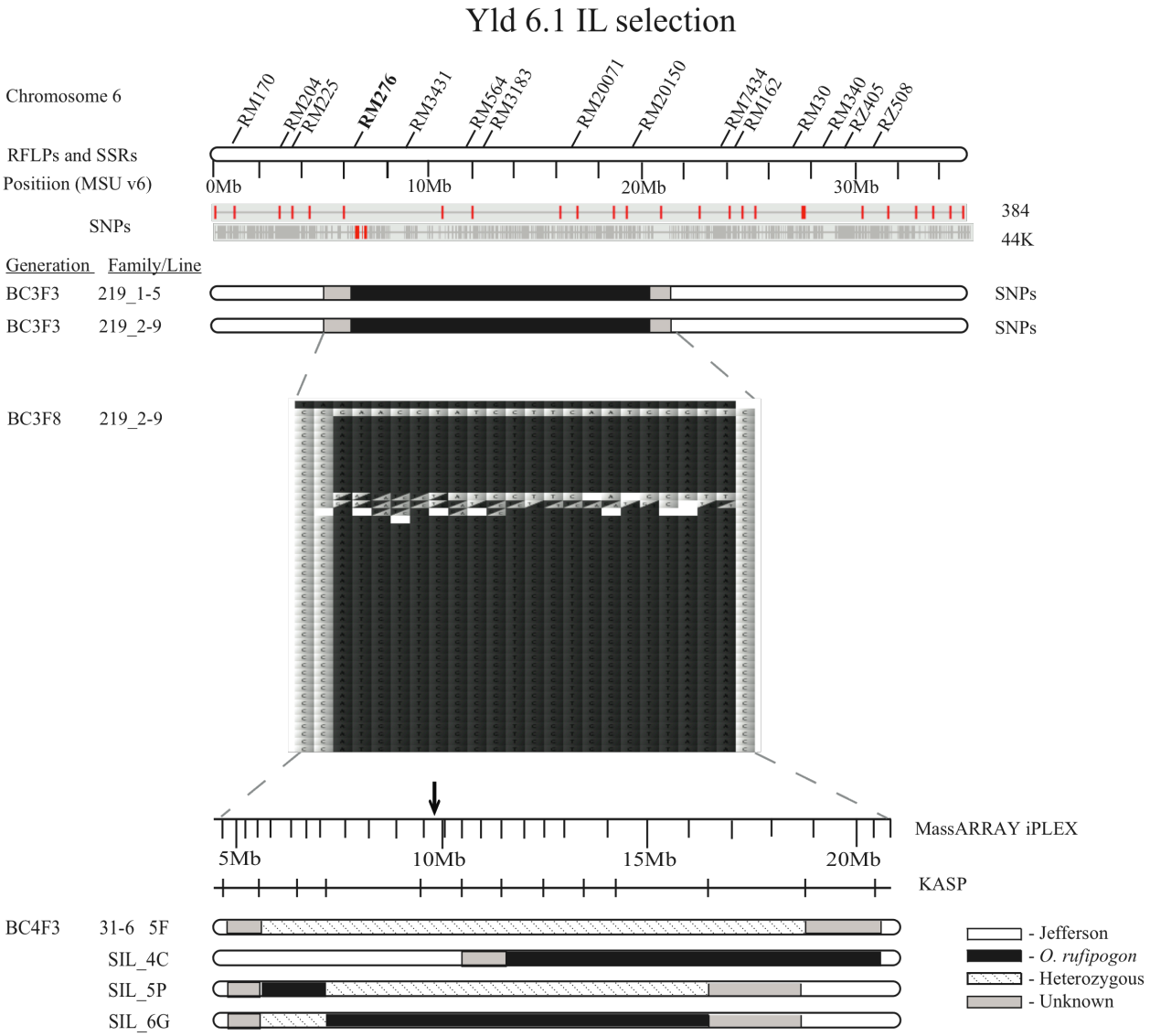

Yld 6.1 IL selection
Chromosome 6
RM170
RM204
RM225
RM276
RM3431
RM564
RM3183
RM20071
RM20150
RM7434
RM162
RM30
RM340
RZ405
RZ508
RFLPs and SSRs
Positiion (MSU v6)
0Mb
10Mb
20Mb
30Mb
SNPs
384
44K
Generation
Family/Line
BC3F3 219_1-5
SNPs
BC3F3 219_2-9
SNPs
BC3F8 219_2-9
MassARRAY iPLEX
5Mb
10Mb
15Mb
20Mb
KASP
BC4F3 31-6 5F
SIL_4C
SIL_5P
SIL_6G
- Jefferson
- O. rufipogon
- Heterozygous
- Unknown

Figure 2-8. Graphical genotype of IL yld6_A at BC_3F_8 generation and derived lines. The top graph illustrates the target QTL region marked by SSR and RFLP markers as originally mapped by Thomson *et al.* (2003); rows in light gray indicate 384 SNP and 44K SNP assays; horizontal bar indicate regions of homozygous *O. rufipogon* introgressions (black), Jefferson DNA (white) and unknown regions (gray) in BC_3F_4 generation. 46 samples representing eight different headrows of yld6_A grown in 2010 were genotyped across the target introgression using 26 markers. Genotypes of these further selfed lines (F_8) are represented with homozygous *O. rufipogon* DNA (dark gray squares), homozygous Jefferson (light gray) and missing data (white) with the first two rows illustrating the *O. rufipogon* and Jefferson parental genotypes. Graph shows approximate locations of SNPs in iPLEX MassARRAY and KASP assays above sub-introgression lines (SILs) with segregation within the target introgression shown in bars indicating heterozygosity (diagonal stripe), homozygous *O. rufipogon* DNA (black), Jefferson DNA (white), and regions of recombination (gray). MRG5836 (a marker associated with *Pi-z* located at 9.3Mb) is annotated as a black arrow between SNP position id6005761 (8.98Mb) and id6006235 (10.01Mb) on the iPLEX MassARRAY track.

Table 2-13: Number of polymorphic SNPs between Jefferson and *O. rufipogon* on the 384 (Thomson *et al.,* 2011), 1,536 and 44K SNP assays (Zhao, *et al.,* 2010; 2011). 384 SNP OPA used in this study is OPA6.0 (VC0011530-OPA).

SNP platform	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 11	Chr 12	Total
384	33	29	27	25	21	29	23	21	22	19	21	20	290
1536	96	66	69	52	37	59	40	42	47	41	53	51	653
44K	3088	2014	2401	1477	1652	1758	1046	1150	1111	834	1198	1031	18760

Table 2-14. Polymorphic SNPs between Jefferson and *O. rufipogon* on 384 SNP chip: Rice OPA 6.0 designed by Thomson et al. (2011) and Zhao *et al* (2011). These 384 SNP positions are shown as tracks in Figure 5.

SNP.id	Chr	MSU V.6 Position (Mb)	Jeff allele	Rufi allele		SNP.id	Chr	MSU V.6 Position (Mb)	Jeff allele	Rufi allele
id1001073	1	1.17	A	G		id7000276	7	1.63	A	C
id1002308	1	2.90	T	C		ud7000253	7	3.64	A	C
id1003138	1	3.78	G	C		id7000840	7	5.73	G	A
id1004759	1	6.10	C	G		ud7000659	7	8.56	C	T
id1006175	1	7.85	A	G		id7001678	7	9.18	G	C
id1006604	1	8.76	A	G		id7001824	7	10.32	T	C
ud1000463	1	9.57	A	G		id7002000	7	11.59	A	G
id1007745	1	10.74	G	T		ud7000886	7	13.09	C	G
id1008184	1	11.77	T	C		id7002255	7	14.41	T	C
id1008679	1	12.85	G	A		id7002580	7	16.77	T	C
id1008872	1	13.44	T	C		wd7002408	7	17.40	G	A
id1009796	1	14.90	C	T		id7002784	7	18.48	A	G
id1010121	1	15.68	T	C		id7002801	7	18.79	C	T
id1010403	1	16.72	T	C		id7002939	7	19.43	A	G
id1010811	1	19.60	C	T		id7003082	7	20.40	A	C
id1012048	1	22.68	T	C		id7003296	7	21.58	G	A
id1012330	1	23.45	A	G		ud7001529	7	22.13	C	G
id1012784	1	24.20	A	G		id7003813	7	23.08	T	C
id1013568	1	25.24	T	A		id7004922	7	26.60	C	T
id1014841	1	26.55	C	T		id7005370	7	27.74	A	G
id1015544	1	28.46	G	A		id7005658	7	28.88	T	C
id1015931	1	29.31	A	C		id8000171	8	0.53	G	A
id1016403	1	29.85	C	T		id8000876	8	2.94	C	T
id1017934	1	31.78	A	G		id8001277	8	3.86	T	C
id1018796	1	32.97	T	A		id8001543	8	4.70	A	G
id1019539	1	33.79	T	A		id8002786	8	8.88	C	G
id1020326	1	34.57	C	T		id8003329	8	10.51	T	C
id1020938	1	35.50	T	C		id8003738	8	12.95	T	C
id1021697	1	36.46	A	G		id8003808	8	14.00	T	G
id1022408	1	37.30	G	C		id8003913	8	14.51	C	A
id1024503	1	40.39	A	T		ud8000972	8	15.24	T	C
id1025292	1	41.56	C	A		ud8001065	8	16.54	A	G
id1026824	1	43.53	A	T		id8004756	8	17.94	C	G
id2001252	2	1.76	A	G		id8004948	8	18.76	C	G
id2001501	2	2.64	A	G		id8005235	8	19.70	G	C
id2003646	2	7.15	T	C		id8005581	8	20.55	A	G
wd2000443	2	8.90	G	C		id8005810	8	21.18	T	A
id2004552	2	9.52	A	T		id8006308	8	22.10	A	G

id2005152	2	10.90	T	G		id8006792	8	24.22	T	C
id2005462	2	11.74	A	G		id8006997	8	25.68	T	A
id2005569	2	12.39	T	A		id8007210	8	26.40	A	G
id2005784	2	13.21	A	G		id8007352	8	27.00	T	A
ud2000759	2	14.90	T	G		id9000233	9	0.88	A	G
id2006175	2	15.91	G	C		id9000515	9	1.76	T	C
id2006394	2	16.74	A	G		wd9000348	9	2.43	T	C
id2006621	2	17.28	T	C		id9000713	9	3.61	T	G
wd2001906	2	19.57	T	C		wd9000677	9	4.41	T	C
id2007797	2	20.84	A	G		id9001633	9	6.64	A	G
id2008716	2	22.81	G	C		id9002027	9	7.57	A	G
id2009246	2	23.99	C	T		id9002453	9	8.49	A	C
id2009772	2	24.71	A	G		id9002693	9	9.97	T	C
id2010498	2	25.50	T	A		ud9000580	9	10.75	C	T
id2011296	2	26.66	T	C		id9003100	9	11.77	A	G
id2011813	2	27.71	A	G		id9003276	9	12.75	C	A
id2012011	2	28.24	A	C		id9003720	9	14.35	G	A
fd12	2	29.78	C	T		id9004347	9	15.89	T	C
id2013174	2	30.74	T	C		id9004843	9	16.65	T	C
id2013975	2	32.36	A	G		id9005501	9	17.84	T	C
id2015227	2	34.93	C	T		id9006058	9	18.41	T	G
id2015934	2	35.92	G	A		id9006361	9	19.02	G	A
id2016013	2	36.01	A	C		id9006968	9	20.02	C	G
id2016439	2	36.58	G	A		wd9002310	9	20.85	G	A
id3000589	3	0.97	T	G		id9007287	9	22.05	A	G
id3001087	3	1.93	A	C		id9007743	9	23.24	A	G
id3001415	3	2.55	A	G		id10000153	10	0.77	G	A
id3002476	3	4.32	A	T		wd10000371	10	1.56	A	G
id3003181	3	5.36	T	C		id10000575	10	2.13	A	G
id3003846	3	6.95	A	G		id10001250	10	4.00	T	C
id3004123	3	7.68	T	C		id10001501	10	4.52	G	A
id3004522	3	8.65	A	G		id10001608	10	5.06	A	C
id3004807	3	9.25	A	G		id10002295	10	6.80	A	G
id3006042	3	11.77	G	A		ud10000451	10	7.33	C	G
id3007120	3	14.74	A	C		wd10002607	10	11.91	C	T
id3008187	3	17.06	T	A		wd10002654	10	12.56	C	G
id3008418	3	17.66	T	C		wd10003015	10	13.27	G	C
id3008619	3	18.28	A	G		id10004110	10	15.80	A	C
id3010248	3	22.84	A	G		id10004515	10	16.61	C	T
id3010545	3	23.76	T	G		id10005370	10	18.66	A	C
id3010802	3	24.95	T	C		id10006161	10	19.98	G	T
id3010951	3	25.83	T	C		id10006389	10	20.64	A	G
id3011064	3	26.33	C	G		id10006610	10	21.40	T	C
id3012168	3	28.45	C	A		id10006910	10	22.34	A	G
id3013325	3	29.30	T	G		id10007079	10	22.94	T	C

id3013806	3	30.07	G	C		id11000268	11	1.19	G	A
id3014586	3	31.45	T	C		id11000778	11	2.50	A	C
id3014850	3	31.99	T	G		ud11000156	11	3.79	G	A
id3015986	3	33.93	T	C		id11001683	11	4.23	A	G
id3017089	3	35.58	A	C		id11001927	11	4.88	C	T
id3017539	3	36.28	G	A		id11002963	11	7.52	T	C
id4000574	4	0.96	G	T		id11003973	11	10.96	T	A
id4001090	4	2.44	A	G		id11004160	11	11.74	A	T
id4002032	4	4.74	T	C		wd11001777	11	15.81	C	T
ud4000481	4	6.66	C	T		id11004890	11	16.71	C	G
id4003430	4	10.81	G	A		id11005855	11	18.69	C	T
wd4001519	4	11.38	G	C		id11006398	11	20.12	A	C
id4003922	4	13.41	C	G		id11007523	11	21.89	G	A
id4004185	4	14.19	G	A		id11007802	11	22.52	T	G
wd4002148	4	14.77	G	A		id11008135	11	23.58	G	A
id4004605	4	16.52	G	A		dd11000336	11	24.12	G	A
id4005120	4	17.69	T	A		id11008888	11	24.93	C	A
id4005404	4	18.76	C	G		id11009575	11	26.65	T	C
id4007105	4	21.84	A	G		id11010097	11	27.46	G	T
id4007444	4	22.83	G	A		id11010475	11	28.35	T	C
id4007720	4	23.72	G	A		id11010846	11	29.23	A	G
id4007907	4	24.36	T	C		id12000405	12	0.94	C	G
ud4001849	4	25.76	T	C		id12000763	12	1.60	A	G
id4008536	4	26.44	A	C		id12001102	12	2.43	A	G
id4008684	4	26.98	A	G		ud12000175	12	3.68	T	C
id4009149	4	28.25	T	C		id12002092	12	4.61	C	G
id4009672	4	29.59	A	T		wd12000490	12	5.82	A	G
id4010220	4	30.73	G	A		id12002740	12	6.82	G	A
id4010433	4	31.32	A	G		id12004293	12	11.22	G	A
id4010985	4	32.44	A	G		id12004456	12	12.29	T	C
id4011398	4	33.22	A	G		id12004787	12	13.58	T	C
id5000013	5	0.02	C	T		id12005245	12	14.70	G	C
id5000678	5	1.00	G	A		id12005567	12	15.99	A	G
id5000986	5	1.52	G	A		id12005912	12	17.72	G	A
id5001700	5	2.97	A	G		id12006657	12	19.93	C	T
id5002212	5	3.82	T	C		id12006773	12	20.94	A	C
id5002987	5	5.89	T	C		id12007161	12	21.87	T	C
id5004086	5	8.01	T	C		id12007504	12	22.38	A	C
id5005143	5	12.08	A	T		id12008796	12	24.85	T	G
ud5000604	5	13.31	A	T		id12009820	12	26.87	T	C
id5005867	5	14.28	A	G		id12009959	12	27.48	T	C
id5006128	5	15.02	T	C						
id5006916	5	17.26	C	T						
id5007172	5	17.86	G	C						
id5008590	5	20.56	T	C						

id5008876	5	21.11	A	G						
id5009967	5	22.70	A	G						
id5010661	5	23.71	C	A						
id5012179	5	25.79	G	A						
id5013326	5	27.50	A	G						
id5014500	5	28.93	A	G						
id5015016	5	29.67	T	G						
id6000606	6	0.83	A	G						
id6002100	6	2.75	A	G						
id6002687	6	3.29	T	C						
id6003050	6	4.02	C	G						
id6003555	6	5.53	G	A						
id6006146	6	9.68	A	T						
id6006754	6	10.99	T	C						
id6008752	6	14.67	C	G						
id6008983	6	15.43	A	T						
wd6002213	6	16.97	C	T						
id6009452	6	17.49	A	G						
id6010102	6	18.97	T	C						
id6010489	6	20.66	G	A						
wd6002805	6	21.91	C	A						
id6011324	6	22.47	A	T						
id6011524	6	23.03	C	T						
wd6003061	6	24.96	C	A						
id6012658	6	25.09	T	A						
id6014975	6	27.59	G	C						
id6015867	6	28.61	T	A						
id6016440	6	29.85	C	T						
id6016625	6	30.54	A	T						
id6016966	6	31.79	T	C						

References

Abedin, M. J., & Meharg, A. A. (2002). Relative toxicity of arsenite and arsenate on germination and early seedling growth of rice (*Oryza sativa* L.). *Plant and Soil, 243*(1), 57-66.

Ali, M., Sanchez, P., Yu, S., Lorieux, M., & Eizenga, G. (2010). Chromosome segment substitution lines: A powerful tool for the introgression of valuable genes from *Oryza* wild species into cultivated rice (*O. sativa*). *Rice, 3*(4), 218-234.

Ando, T., Yamamoto, T., Shimizu, T., Ma, X., Shomura, A., Takeuchi, Y., et al. (2008). Genetic dissection and pyramiding of quantitative traits for panicle architecture by using chromosomal segment substitution lines in rice. *Theoretical and Applied Genetics, 116*(6), 881-890.

Ashikari, M., & Matsuoka, M. (2006). Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends in Plant Science, 11*(7), 344-350.

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005). Cytokinin oxidase regulates rice grain production. *Science, 309*(5735), 741-745.

Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., et al. (2008). Two adjacent nucleotide-binding Site–Leucine-rich repeat class genes are required to confer *pikm*-specific rice blast resistance. *Genetics, 180*(4), 2267-2276.

Bao, J.S., Corke, H., & Sun, M. (2006). Nucleotide diversity in *starch synthase IIa* and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physiochemical properties in rice (*Oryza sativa L.) Theor Appl Genet* 113, 1171-1183.

Bernier, J., Kumar, A., Ramaiah, V., Spaner, D., & Atlin, G. (2007). A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Science, 47*(2), 507-516.

Bernier, J., Atlin, G. N., Serraj, R., Kumar, A., & Spaner, D. (2008). Breeding upland rice for drought resistance. *Journal of the Science of Food and Agriculture, 88*(6), 927-939.

Bernier, J., Kumar, A., Venuprasad, R., Spaner, D., Verulkar, S., Mandal, N. P., et al. (2009a). Characterization of the effect of a QTL for drought resistance in rice, *qtl12.1*, over a range of environments in the Philippines and Eastern India. *Euphytica, 166*(2), 207-217.

Bernier, J., Serraj, R., Kumar, A., Venuprasad, R., Impa, S., Gowda, R. P. V., et al. (2009b). The large-effect drought-resistance QTL *qtl12.1* increases water uptake in upland rice. *Field Crops Research, 110*(2), 139-146.

Briggle, L. W. (1968). Near-isogenic lines of wheat with genes for resistance to *Erysiphe graminis* f. sp. *tritici1*. *Crop Sci., 9*(1), 70-72.

Bryan, G. T., Wu, K., Farrall, L., Jia, Y., Hershey, H. P., McAdams, S. A., et al. (2000). A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *The Plant Cell Online, 12*(11), 2033-2046.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The genetic architecture of maize flowering time. *Science, 325*(5941), 714-718.

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. *J. Cereal Sci*. 47, 536-545.

Chen, M., Fjellstrom, R.G., Christensen, E.F., Bergman, C.J. (2010). Development of three allele-specific codominant rice *Waxy* gene PCR markers suitable for marker assisted selection of amylose content and paste viscosity. *Molecular Breeding*. 26, 513-523

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 363*(1491), 557-572.

Cox, T. S., & Frey, K. J. (1984). Genetic variation for grain yield and related traits in sorghum introgression populations. *Theoretical and Applied Genetics, 68*(1), 145-153.

Eichten, S. R., Foerster, J. M., de Leon, N., Kai, Y., Yeh, C., Liu, S., et al. (2011). B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. *Plant Physiology, 156*(4), 1679-1690.

Eshed, Y., & Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics, 141*(3), 1147-1162.

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). *GS3,* a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theoretical and Applied Genetics, 112*(6), 1164-1171.

Feuerstein, U., Brown, A. H. D., & Burdon, J. J. (1990). Linkage of rust resistance genes from wild barley (*Hordeum spontaneum*) with isozyme markers. *Plant Breeding, 104*(4), 318-324.

Fjellstrom, R.G., Conaway-Bormans, C.A., McClung, A.M., Marchetti, M.A., Shank, A.R., & Park, W.D. (2004). Development of DNA markers suitable for marker assisted selection of three *Pi* genes conferring resistance to multiple *Pyricularia grisea* pathotypes. *Crop Sci*. 44, 1790–1798.

Fjellstrom, R., McClung, A., & Shank, A. (2006). SSR markers closely linked to the *pi-z* locus are useful for selection of blast resistance in a broad array of rice germplasm. *Molecular Breeding, 17*(2), 149-157.

Flint-Garcia, S. A., Thuillet, A., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. *The Plant Journal, 44*(6), 1054-1064.

Fukuoka, S., Nonoue, Y., & Yano, M. (2010). Germplasm enhancement by developing advanced plant materials from diverse rice accessions. *Breeding Science, 60*(5), 509-517.

Gur, A., Semel, Y., Osorio, S., Friedmann, M., Seekh, S., Ghareeb, B., et al. (2011). Yield quantitative trait loci from wild tomato are predominately expressed by the shoot. *Theoretical and Applied Genetics, 122*(2), 405-420.

Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., et al. (2009). Natural variation at the *DEP1* locus enhances grain yield in rice. *Nature Genetics, 41*(4), 494-497.

Heuer, S., Lu, X., Chin, J. H., Tanaka, J. P., Kanamori, H., Matsumoto, T., et al. (2009). Comparative sequence analyses of the major quantitative trait locus *phosphorus uptake 1* (*Pup1*) reveal a complex genetic structure. *Plant Biotechnology Journal, 7*(5), 456-471.

Iyer, A. S., & McCouch, S. R. (2004). The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Molecular Plant-Microbe Interactions, 17*(12), 1348; 7-1354.

Jia, Y., Correa-Victoria, F., McClung, A., Zhu, L., Liu, G., Wamishe, Y., et al. (2007). Rapid determination of rice cultivar responses to the sheath blight pathogen *Rhizoctonia solani* using a micro-chamber screening method. *Plant Disease, 91*(5), 485-489.

Jiang, G., Xia, Z., Zhou, Y., Wan, J., Li, D., Chen, R., et al. (2006). Testifying the rice bacterial blight resistance gene *xa5* by genetic complementation and further analyzing *xa5* (*Xa5*) in comparison with its homolog *TFIIA* gamma 1. *Molecular Genetics and Genomics, 275*(4), 354-366.

Jiang, H., Wang, S., Dang, L., Wang, S., Chen, H., Wu, Y., et al. (2005). A novel short-root gene encodes a glucosamine-6-phosphate acetyltransferase required for maintaining normal root cell shape in rice. *Plant Physiology*, *138*(1), 232-242.

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of *OsSPL14* by *OsmiR156* defines ideal plant architecture in rice. *Nature Genetics, 42*(6), 541-544.

Kandemir, N., Jones, B. L., Wesenberg, D. M., Ullrich, S. E., & Kleinhofs, A. (2000). Marker-assisted analysis of three grain yield QTL in barley (*Hordeum vulgare* L.) using near isogenic lines. *Molecular Breeding, 6*(2), 157-167.

Kovach M.J., Calingacion M.N., Fitzgerald M.A., & McCouch S.R. (2009). The origin and evolution of fragrance in rice (*Oryza sativa L.*). *PNAS* 106, 14444-14449.

Krystkowiak, K., Adamski, T., Surma, M., & Kaczmarek, Z. (2009). Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis effects in wheat (*Triticum aestivum* L.). *Euphytica, 165*(3), 419-434.

Kuczynska, A., Surma, M., Kaczmarek, Z., & Adamski, T. (2007). Relationship between phenotypic and genetic diversity of parental genotypes and the frequency of transgression effects in barley (*Hordeum vulgare* L.). *Plant Breeding, 126*(4), 361-368.

Kumar, A., Verulkar, S., Dixit, S., Chauhan, B., Bernier, J., Venuprasad, R., et al. (2009). Yield and yield-attributing traits of rice (*Oryza sativa* L.) under lowland drought and suitability of early vigor as a selection criterion. *Field Crops Research, 114*(1), 99-107.

Kumar, J., Jaiswal, V., Kumar, A., Kumar, N., Mir, R. R., Kumar, S., et al. (2011). Introgression of a major gene for high grain protein content in some Indian bread wheat cultivars. *Field Crops Research, 123*(3), 226-233.

Kumar, J., Mir, R. R., Kumar, N., Kumar, A., Mohan, A., Prabhu, K. V., et al. (2010). Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. *Plant Breeding, 129*(6), 617-621.

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nature Genetics, 43*(12), 1266-1269.

Little, R. R., Hilder, G. B., & Dawson, E. H. (1958). Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chemistry*, *35*, 111-126

Liu, G., Jia, Y., Correa-Victoria, F., Prado, G. A., Yeater, K. M., McClung, A., et al. (2009). Mapping quantitative trait loci responsible for resistance to sheath blight in rice. *Phytopathology, 99*(9), 1078-1084.

Liu, T., Shao, D., Kovi, M., & Xing, Y. (2010). Mapping and validation of quantitative trait loci for spikelets per panicle and 1,000-grain weight in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics, 120*(5), 933-942.

Liu, X., Lin, F., Wang, L., & Pan, Q. (2007). The in silico map-based cloning of *Pi36*, a rice coiled-Coil–Nucleotide-binding Site–Leucine-rich repeat gene that confers race-specific resistance to the blast fungus. *Genetics, 176*(4), 2541-2549.

Lu, H. Y. (1995). PC-SAS program for estimating Huhn's nonparametric stability statistics. *Agronomy Journal, 87*(5), 888-891.

Mansur, L. M., Orf, J. H., Chase, K., Jarvik, T., Cregan, P. B., & Lark, K. G. (1996). Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Science, 36*(5), 1327-1336.

McClung, A. M., Marchetti, M. A., Webb, B. D., & Bollich, C. N. (1997). Registration of 'Jefferson' rice. *Crop Science, 37*(2), 629-630.

McCouch, S., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septiningsih, E., et al. (2007). Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. *Euphytica, 154*(3), 317-339.

McCouch, S., Zhao, K., Wright, M., Tung, C., Ebana, K., Thomson, M., et al. (2010). Development of genome-wide SNP assays for rice. *Breeding Science, 60*(5), 524-535.

Miedaner, T., Wilde, F., Korzun, V., Ebmeyer, E., Schmolke, M., Hartl, L., et al. (2009). Marker selection for Fusarium head blight resistance based on quantitative trait loci (QTL) from two European sources compared to phenotypic selection in winter wheat. *Euphytica, 166*(2), 219-227.

Miura, K., Ikeda, M., Matsubara, A., Song, X., Ito, M., Asano, K., et al. (2010). *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nature Genetics, 42*(6), 545-549.

Moncada, P., Martínez, C. P., Borrero, J., Chatel, M., Gauch Jr, H., Guimaraes, E., et al. (2001). Quantitative trait loci for yield and yield components in an *Oryza sativa*×*Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. *Theoretical and Applied Genetics, 102*(1), 41-52.

Monforte, A. J., & Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a L. *esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome, 43*(5), 803-813.

Nelson, J., McClung, A., Fjellstrom, R., Moldenhauer, K., Boza, E., Jodari, F., et al. (2011). Mapping QTL main and interaction influences on milling quality in elite US rice germplasm. *Theoretical and Applied Genetics, 122*(2), 291-309.

Oeth, P., Beaulieu, M., Park, C., Kosman, D., del Mistro, G., van den Boom, D., et al. (2005). iPLEX™ assay: Increased plexing efficiency and flexibility for MassARRAY system through single base primer extension with mass-modified terminators. *Sequenom Application Note*.

Ohsumi, A., Takai, T., Ida, M., Yamamoto, T., Arai-Sanoh, Y., Yano, M., et al. (2011). Evaluation of yield performance in rice near-isogenic lines with increased spikelet number. *Field Crops Research, 120*(1), 68-75.

Paterson, A. H., Damon, S., Hewitt, J. D., Zamir, D., Rabinowitch, H. D., Lincoln, S. E., et al. (1991). Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. *Genetics, 127*(1), 181-197.

Pedersen, J. F., & Toy, J. J. (2001). Germination, emergence, and yield of 20 plant–color, seed–color near-isogenic lines of grain sorghum. *Crop Science, 41*, 107.

Perez, C. M., & Juliano, B. O. (1978). Modification of the simplified amylose test for milled rice. *Starch - Stärke, 30*(12), 424-426.

Pinto, Y. M., Kok, R. A., & Baulcombe, D. C. (1999). Resistance to rice yellow mottle virus (RYMV) in cultivated African rice varieties containing RYMV transgenes. *Nature, 17*, 702.

Poland, J. A., Bradbury, P. J., Buckler, E. S., & Nelson, R. J. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences, 108*(17), 6893-6898.

RoyChowdhury, M., Jia, Y., Jackson, A., Jia, M., Fjellstrom, R., & Cartwright, R. Analysis of rice blast resistance gene *Pi-z* in rice germplasm using pathogenicity assays and DNA markers. *Euphytica*, 1-12.

Sarla, N., & Mallikarjuna Swamy B.P. (2005). Oryza glaberrima: A source for the improvement of *Oryza sativa*. *Current Science, 89*(6), 955.

Septiningsih, E. M., Prasetiyono, J., Lubis, E., Tai, T. H., Tjubaryat, T., Moeljopawiro, S., et al. (2003). Identification of quantitative trait loci for yield and yield components in an advanced backcross population

derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theoretical and Applied Genetics, 107*(8), 1419-1432.

Septiningsih, E. M., Pamplona, A. M., Sanchez, D. L., Neeraja, C. N., Vergara, G. V., Heuer, S., et al. (2009). Development of submergence-tolerant rice cultivars: The *Sub1* locus and beyond. *Annals of Botany, 103*(2), 151-160.

Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., et al. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. *Nature Genetics, 40*(8), 1023-1028.

Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Holsten, T., et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science, 270*(5243), 1804-1806.

Song, X., Huang, W., Shi, M., Zhu, M., & Lin, H. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature Genetics, 39*(5), 623-630.

Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., Kanamori, H., et al. (2009). Evolutionary history of *GS3*, a gene conferring grain length in rice. *Genetics, 182*(4), 1323-1334.

Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science, 277*(5329), 1063-1066.

Thomson, M. J., Tai, T. H., McClung, A. M., Lai, X., Hinga, M. E., Lobos, K. B., et al. (2003). Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and Applied Genetics, 107*(3), 479-493.

Thomson, M. J., Edwards, J. D., Septiningsih, E. M., Harrington, S. E., & McCouch, S. R. (2006). Substitution mapping of *dth1.1*, a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. *Genetics, 172*(4), 2501-2514.

Thomson, M. J., Ismail, A. M., McCouch, S. R., & Mackill, D. J. (2011). Marker assisted breeding. In A. Pareek, S. K. Sopory, H. J. Bohnert & Govindjee (Eds.), *Abiotic stress adaptation in plants: Physiological, molecular and genomic foundation* (pp. 451-469). Dordrecht, the Netherlands: Springer. doi:10.1093/aob/mcr053

Thomson, M., Zhao, K., Wright, M., McNally, K., Rey, J., Tung, C., et al. (2011). High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding*,12.

Tian, F., Li, D., Fu, Q., Zhu, Z., Fu, Y., Wang, X., et al. (2006). Construction of introgression lines carrying wild rice (*Oryza rufipogon* griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. *Theoretical and Applied Genetics, 112*(3), 570-580.

Venuprasad, R., Dalid, C., Del Valle, M., Zhao, D., Espiritu, M., Sta Cruz, M., et al. (2009). Identification and characterization of large-effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. *Theoretical and Applied Genetics, 120*(1), 177-190.

Venuprasad, R., Impa, S. M., Gowda, R. P. V., Atlin, G. N., & Serraj, R. (2011). Rice near-isogenic-lines (NILs) contrasting for grain yield under lowland drought stress. *Field Crops Research, 123*(1), 38-46.

Venuprasad, R., Bool, M., Quiatchon, L., & Atlin, G. (2012). A QTL for rice grain yield in aerobic environments with large effects in three genetic backgrounds. *Theoretical and Applied Genetics, 124*(2), 323-332.

Wan, X., Weng, J., Zhai, H., Wang, J., Lei, C., Liu, X., et al. (2008). Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5. *Genetics, 179*(4), 2239-2252.

Wang, P., Zhou, G., Yu, H., & Yu, S. (2011). Fine mapping a major QTL for flag leaf size and yield-related traits in rice. *Theoretical and Applied Genetics,* (8), 1319-1330.

Wang, X., Fjellstrom, R. G., Jia,Y., Yan, W., Jia, M. H., Scheffler, B. E., Wu, D., Shu, O., & McClung, A. M. (2010). Characterization of the *Pi-ta* blast resistance gene in an international rice core collection. Plant Breeding 129, 491-505.

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008). Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Research, 18*(12), 1199-1209.

Xiao, J., Li, J., Grandillo, S., Ahn, S. N., Yuan, L., Tanksley, S. D., et al. (1998). Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics, 150*(2), 899-909.

Xie, X., Jin, F., Song, M., Suh, J., Hwang, H., Kim, Y., et al. (2008). Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa* × *O. rufipogon* cross. *TAG Theoretical and Applied Genetics, 116*(5), 613-622.

Xu, J., Zhao, Q., Du, P., Xu, C., Wang, B., Feng, Q., et al. (2010). Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome re-sequencing in rice (oryza sativa L.). *BMC Genomics, 11*(1), 656.

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., et al. (2006). *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature, 442*(7103), 705-708.

Yan, W., Dilday, R. H., Tai, T. H., Gibbons, J. W., McNew, R. W., & Rutger, J. N. (2005). Differential response of rice germplasm to straighthead induced by arsenic. *Crop Science, 45*(4), 1223-1228.

Yano, M., & Sasaki, T. (1997). Genetic and molecular dissection of quantitative traits in rice. *Plant Molecular Biology, 35*(1), 145-153.

Young, N. D., & Tanksley, S. D. (1989). Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theoretical and Applied Genetics, 77*(1), 95-101.

Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology, 17*(2), 155-160.

Yuan, M., Chu, Z., Li, X., Xu, C., & Wang, S. (2010). The bacterial pathogen *Xanthomonas oryzae* overcomes rice defenses by regulating host copper redistribution. *Plant Cell, 22*(9), 3164-3176.

Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., et al. (2010). Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS ONE, 5*(5), e10780.

Zhao, K., Tung, C., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun, 2*, 467.

Zhou, B., Qu, S., Liu, G., Dolan, M., Sakai, H., Lu, G., et al. (2006). The eight amino-acid differences within three leucine-rich repeats between *Pi2* and *Piz-t* resistance proteins determine the resistance specificity to *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions, 19*(11), 1216-1228.

Zou, J. H., Pan, X. B., Chen, Z. X., Xu, J. Y., Lu, J. F., Zhai, W. X., et al. (2000). Mapping quantitative trait loci controlling sheath blight resistance in two rice cultivars (*Oryza sativa* L.). *Theoretical and Applied Genetics, 101*(4), 569-573.

CHAPTER 3

Jefferson x *O. rufipogon* IL project: Addendum analysis

Progenitor species of rice are repositories of valuable traits that can be efficiently utilized for crop improvement through the transfer of small introgressions into elite genetic backgrounds. In the previous chapter, we began steps toward validating several yield QTLs detected in the Jefferson x *O. rufipogon* BC_2F_2 population (Thomson *et al*., 2003) through the development of ILs. The segregation of phenotypes and QTLs detected in this population has also given rise to fine-mapping efforts for flowering time (Thomson *et al*., 2006; Maas *et al*., 2010), grain weight (Li *et al*., 2004; Takano-Kai *et al.,* 2009) and red pericarp (Sweeney *et al*., 2006; 2007). With future fine-mapping we aim to identify non-target 'background' introgressions from *O. rufipogon* that confer useful GxG interactions when transferred along with target yield-QTL into adapted cultivars to produce superior germplasm.

The *Oryza rufipogon* allele(s) at the *yld2.1* and *yld6.1* loci increased yield in the Jefferson background despite the fact that *O. rufipogon* was a low yielding parent. To understand the underlying mechanism behind this transgressive variation and to harness this knowledge to enhance the efficiency of future crop improvement efforts, we aim to further analyze the best-performing ILs to determine which combinations of *O. rufipogon* introgressions confer optimal performance in the Jefferson background. Although none of the QTLs in our study are predicted to explain >17% of the phenotypic variation, and thus all six are considered relatively minor QTLs, the multi-location field data from 2007 and 2008 indicates that several of the BC_3F_4 lines exceeded the performance of the elite parent, Jefferson**.** Due to the large introgression sizes across the target QTL regions and the presence of several background introgressions in each of

the ILs, our goal is to identify the genes underlying both the target QTLs and the background introgressions that interact favorably with the QTLs. To address this goal, the two top-performing ILs were selected for backcrossing as a prerequisite for fine-mapping purposes and eventual cloning of the genes from *O. rufipogon* that contribute to the yield advantage of the ILs.

In this chapter, I aim to elaborate on the discussion from Chapter 2 based on personal observations of data that were collected by myself and others members of the project.

Comparisons between original QTL study and current study

As mentioned in Chapter 2, the yield QTLs as described by Thomson *et al.,* (2003) was not grain yield per unit area, but grain weight per plant. The fact that we targeted these QTLs but evaluated for grain yield (kg ha^{-1}) is one of many crucial distinctions that need to be made to understand the importance of these yield QTLs and how we interpret the resulting phenotyping data for ILs.

The six QTLs targeted in our study were not all identified as yield QTLs *per se.* The traits corresponding to these QTLs included several yield components, including grains per panicle (*gpp*), spikelets per panicle (*spp),* panicles per plant (*ppl),* and grain weight (*gw),* and in several cases, clusters of yield component traits were co-located in a single QTL region (Fig. 3-1, Table 3-1). In the case of the QTL referred to as *yld1.1* in the present study, Thomson *et al.,* (2003) reported that alleles from the *O. rufipogon* parent were expected to reduce the number of grains per panicle and the number of spikelets per panicle at this locus (*gpp1.2* and *spp1.2,* respectively), suggesting that the *O. rufipogon* allele(s) are undesirable at this QTL, and would be expected to reduce yield compared to the Jefferson recurrent parent. *Yld2.1* represents a cluster of QTLs (*gpp2.1, spp2.1,* and *yld2.1*), with the desirable allele coming from *O. rufipogon* for all three

traits. *Yld3.2* encompasses a cluster of six QTLs where *O. rufipogon* contributed the desirable allele for four traits (days to heading, *dth3.4, gpp3.2, spp3.1* and *yld3.2*) and the undesirable allele for two traits (percent seed set, *pss3.1,* and grain weight, *gw3.2*). *Yld6.1* is a cluster of two QTLs where the *O. rufipogon* allele is predicted to increase yield (*yld6.1*) but to lower the percent seed set (*pss6.1*). *Yld8.1* is also a cluster of two QTLs where the *O. rufipogon* allele increased the number of grains per panicle (*gpp8.1*) but increases plant height (*ph8.1*) (an undesirable trait). *Yld9.1* is a cluster of five QTLs, of which the *O. rufipogon* allele is beneficial for four traits (panicle length, *pl9.1, gpp9.1, spp9.1* and *yld9.1,*) and undesirable for tiller type (*tt9.1*). Of the six target QTLs identified by Thomson *et al*., (2003), only *yld2.1* carried a favorable combination of *O. rufipogon* alleles for all traits. This may be the reason why the ILs carrying *O. rufipogon* alleles at *yld2.1* performed better than the ILs carrying wild alleles at other target QTLs.

Further, each QTL was not detected in all environments tested in the original QTL study (Thomson *et al.,* 2003). There were a total of three different field locations: Alvin (Texas), Beaumont (Texas), and Newport (Arkansas). Of the 19 QTLs (marker-trait associations) originally identified by Thomson *et al*. (2003) in the six QTL target regions that were the focus of the current study, only one, *spp3.1,* was consistently detected in all three locations. Five QTLs, *gpp1.2*, *gpp3.1*, *yld3.2, gpp9.1* and *spp9.1,* were detected in two of the three locations. The rest of the QTLs were only detected in one location. In the current study, the field locations in Alvin and Beaumont, TX, were the same as those used by Thomson *et al.* (2003), but locations in Jonesboro and Stuttgart, AR, represent new locations in the current study, while Newport, AR, was used by Thomson *et al.* (2003). These sites in AR are not comparable with each other, given

the significant weather, temperature and elevation differences between the field locations (Appendix II: Field location weather, temperature and elevation differences).

Field management schemes were also different between the original study by Thomson *et al*. (2003) and our large-scale replicated field trials in 2007-2008 and the URRN trials in 2009-2011. The Thomson *et al.* (2003) study included plots that were either drill seeded and/or transplanted; both Beaumont and Newport environments were drill seeded, while Alvin had both transplanted and drill seeded plots. In all plots in that study, plants were thinned to 42 plants per square meter to attain uniform spacing. Our field trials were all drill seeded in 2007-2008, but a very low seeding rate was used (45 kg ha^{-1}). This is the seeding rate designed for growing hybrids and introduced a bias that favored the yield of the hybrids compared to the inbred ILs. In the URRN plots evaluated in 2009-2011, the seeding rate was increased to 125 kg ha^{-1}, approximately 2.8 times the spacing of the 2007-2008 field trials. This drastic difference in field management may be a factor explaining why the yield measurements in 2007-2008 could not be accurately compared with 2009-2011 trials, and why some lines that looked promising under less dense seeding rates (such as IL yld1_A) did not perform as well in the more recent trials.

Despite the fact that the ILs in the current study may not have attained optimal yield per area due to the lower than optimal seeding rate for inbred lines, the dynamic nature of rice development makes it possible for plants to at least partially compensate for this aspect of environmental variation. Rice plants typically compensate for low stand within a plot by higher rates of tillering and enhanced biomass production. We addressed this component of variation in our statistical analysis. However, when stand percent (STDPCT) and plants per square meter (PLSQM) were considered as possible covariates in our model, these variables were found to be not significant (K. Yeater, USDA-ARS, pers. comm., 2011). This may be due to the fact that we

had only approximate measurements of stand in 10% increments, rather than records of the exact density of plants within each plot, and the fact that stand was not measured in all environments. In future experiments, this variable will be targeted for more in-depth analysis.

There is also substantial work to be done to determine exactly what makes IL yld2_A and IL yld6_A perform so well. We have confirmed that the yield potential of these lines is genetically due to a small number of introgressions from a low yielding wild ancestor, and that the yield of these lines is comparable to the best elite commercial rice cultivars available today, 15 years after the release of Jefferson. We detect no significant differences in grain or panicle morphology compared to Jefferson, and hypothesize that the yield advantage may be due to a larger number of panicles per plant. Unfortunately, this yield component was not measured in these field trials. Though we measured tiller number per plant, this did not provide an estimate of the number of productive tillers per plant and therefore, could not be used to approximate the number of panicles per plant.

Check performance across years and locations

ILs were phenotyped alongside Jefferson and three other commercial checks (Cocodrie, Trenasse and a hybrid, XL723). Cocodrie and Trenasse are U.S. long grain cultivars that were released more recently than Jefferson and are comparable to currently used varieties. Cocodrie (CCDR, PI 606331) is a very early maturing long-grain rice with high milling yield and high amylose content (Linscombe *et al.,* 2000). XL723 is a hybrid variety released by RiceTec and marketed for its superior yield. We expected relatively stable performance of the released varieties, Cocodrie, Trenasse, XL723 and Jefferson across environments, and used these as controls to gauge the stability of our segregating ILs. However, the number of replications for each check was different in each environment, making comparisons difficult. What is perceived as stability (small standard deviations) may actually be a product of having more replication (Table 3-2).

Jefferson, the recurrent parent performed better in Beaumont, TX than in other locations in the 2007-8 field trials, reflecting a potential bias for the location where it was developed. Our experiments also provided insight into the considerable morphological plasticity of Jefferson. In Jonesboro, AR, we observed unusually high tiller number per plant in 30 plots of Jefferson during 2008, with observations ranging from 14 to 29 tillers per plant. Because of a problem in seed germination that year, these same 30 plots had a low number of plants per square meter, ranging from 3 to 19 plants. These data document the ability of Jefferson to compensate for low stand by producing more tillers per plant. This suggests that Jefferson may be a good candidate for studies in alternative field management where greater spacing of plants may lead to greater water and nutrient use efficiency, as suggested by the System of Rice Intensification (SRI).

The hybrid, XL723, yielded better than any of the ILs or inbred varieties, including the checks, Cocodrie and Trenasse in all the trials. As mentioned earlier, the 2007-2008 field trials were conducted with wide plant spacing such as is commonly used for hybrid production. Knowing that the management was biased towards optimizing the yield advantage of the hybrids, the performance of XL723 is less surprising. It will be interesting to compare XL723 with our best IL, yld2_A, in plots with seeding rates that are commonly used for inbreds. The mean yield for XL723 was highest in Jonesboro, AR, where RiceTec is based, with 12500 kg ha^{-1} in 2007 and 11000 kg ha^{-1} in 2008.

IL performance under water-stressed conditions

In comparison to other cereals such as wheat, sorghum or maize, the irrigated rice system is highly wasteful of freshwater resources, therefore developing rice varieties and management systems that utilize less water is a major goal of rice research today. We investigated the performance of the ILs developed on this project under water-stressed conditions (WS) during 2008. WS trials were conducted using 44 ILs and 18 control lines in both Beaumont (TX) and Stuttgart (AR) during 2008. WS were imposed by flush irrigation only after the ground had dried to the point of cracking (refer to Chapter 2), so areas with more rainfall would have less cracking of the soil, and will not be flush irrigated as frequently. Combining data for all lines across the two locations showed that the average reduction in rice yield under water-stressed conditions was 23.3%. When comparing these locations separately, we found that the water-stress conditions were more severe in Beaumont than in Stuttgart (Fig. 3-2 and 3-3), with an average drop of 65.0% (WS 2608; Irrigated 7436) in Beaumont, and only 28.3% (WS 5604; Irrigated 7815) in Stuttgart.

Plants grown in WS conditions were stunted compared to their counterparts in the irrigated system (Table 3-3). In Beaumont, the mean height of the ILs under WS was reduced 13.5% compared to flooded conditions. The difference in height was not significantly different in the two irrigation schemes in Stuttgart, because of the relatively mild water-stressed conditions of Stuttgart compared to Beaumont (see Appendix II for rainfall data). In addition, the average panicle length was shorter and the average panicle weight was lighter, with fewer seeds per panicle, under WS compared to flooded conditions. Interestingly, we also observed that the total milling yield was higher under WS than flooded conditions (Table 3-3).

We saw an interesting trend in grain morphology during 2008, where entries grown under WS conditions in Beaumont, TX had shorter grain length (GL) than those grown in flooded conditions, while the reverse was observed in Stuttgart, AR, where plants grown under flooded conditions had shorter grain length than those grown under WS. Grain width, on the other hand, was always smaller (i.e., grains were narrower) when plants were grown under WS compared to irrigated conditions.

All commercial checks also experienced reductions in yield under WS conditions but to different degrees. Cocodrie was evaluated in both Stuttgart and Beaumont under WS conditions (N=17) and the combined average yield performance was 6020 kg ha^{-1}, which was 25.3% less than the yield of Cocodrie grown under irrigated conditions (8070 kg ha^{-1}). Jefferson, the recurrent parent, did not do as well under WS conditions, with a 31.4% drop in yield (5150 kg ha^{-1} under WS conditions compared to 7500 kg ha^{-1} under irrigated conditions). Both Trenasse and XL723 were only tested in the WS environment in Stuttgart, so the drop in yield can only be calculated for that location. Trenasse experienced a 15.8% yield reduction (7260 kg ha^{-1} in WS

conditions compared to 8630 kg ha^{-1} in irrigated conditions), while yield of XL723 was reduced by 23.1% (8060 kg ha^{-1} in WS conditions compared to 10500 kg ha^{-1} under irrigation).

There were several outliers among the ILs in terms of the percentage reduction under WS compared to irrigated conditions. Several ILs carrying the QTL *yld3.2* in family 16 were very stable across environments, showing the least difference in yield regardless of irrigation type. Line 16_1-6 (Flooded: 6025 kg ha^{-1}; WS: 5912 kg ha^{-1}) and line 16_2-3 (Flooded: 6499 kg ha^{-1}; WS: 6327 kg ha^{-1}) had less than 5% reduction in yield under WS conditions. In addition, 16_1-6 is the only line among all the ILs evaluated in this experiment that had less chalk than the recurrent parent (Jefferson). We hypothesize that line 16_1-6 has superior qualities of both low chalk and water-stress tolerance due to an ability to stabilize the process of laying down starch in the grain, regardless of water stress. In contrast, ILs in family 9 carrying an *O. rufipogon* introgression across *yld9.1* had the most severe reduction in yield under WS conditions. Line 9_1-7 (Flooded: 6947 kg ha^{-1}; WS: 3746 kg ha^{-1}) and line 9_1-3 (Flooded: 7033 kg ha^{-1}; WS: 3604 kg ha^{-1}) experienced > 45% yield reduction under WS conditions. None of these outliers were high yielding under flooded conditions to begin with, but they do highlight some features of phenotypic plasticity observed among the ILs and controls in response to WS conditions.

Differences in rainfall between Beaumont and Stuttgart are likely to have affected the yield measurements in the two locations (Appendix II: Weather). The drastic difference is evident in April where Beaumont had only 0.89 inches and Stuttgart had 9.77 inches of rainfall. This likely affected the establishment (tillering and root structure) of the plants, where those in Beaumont would have experienced greater WS. However, Stuttgart was substantially drier than Beaumont during the field season, and for that reason, we might expect that plants grown in Beaumont would show less difference in yield performance between WS and irrigated

environments. Yet, at the same time, WS plants in Stuttgart were frequently flash flooded by field managers when ground-cracking was observed, so the difference between flooding and WS conditions may not be as pronounced as we might expect. With data from two locations in only one year, and at the current level of replication, it is difficult to make any firm conclusions about how the ILs in our study respond to WS, but it will be interesting to follow up on these preliminary results with additional experiments in the future.

Correlation and regression analysis for 2007-2008 field trials

Regression analyses (Table 2-7 and 3-4) show only forward regression and omitted some yield components, such as stand and lodging, due to missing values or because of non-significant p-values ($p > 0.1$). It is important to note that out of a total of 1500 observations, only 700 were useful due to missing data. When only the IL entries were considered, average panicle length (AVPANL) explained the largest portion of the phenotypic variation associated with yield ($R^2 = 0.1771$), followed by average plant weight (AVPLTWT, $R^2=0.0755$), amylose content (AMYLOSE, $R^2=0.0393$), average thousand seed weight (KSDWT, $R^2=0.0254$) and grain length (GL, $R^2=0.0340$) (Table 2-7A). When all checks as well as ILs were included, AVPANL still explained the largest portion of the phenotypic variation, but the R^2 value increased to 0.4208, most likely due to the extremely long panicles of the hybrid, XL723 (Table 3-4A).

Under water-stressed conditions, the only variable that was significant in both groups was average panicle weight (AVPANWT), which explained 10.25% of the phenotypic variation when IL-entries only were included, and 20.8% when both IL-entries and checks were included in the model (Table 2-7B and 3-4B). Four variables were significant for IL entries only under water-stressed conditions (Table 2-7B): thousand grain weight (KSDWT, $R^2=0.3432$), average

panicle weight (AVPANWT, R^2=0.1025), plant height (HT, R^2=0.0654), and average tillers (AVTILL, R^2=0.0548). When considering all entries and checks (Table 3-4B), only two variables were significant, average panicle weight (AVPANWT, R^2=0.208) and chalk percent (CHAKPCT, R^2=0.0271).

Several yield components were consistently correlated with yield performance (kg ha^{-1}) under irrigated conditions, but there were profound differences when considering the IL entries separately compared to combining both ILs and checks together (Table 2-8A and 3-5A). The most notable change was that plant height was not correlated with yield (kg ha^{-1}) for all ILs and checks, but there was a strong correlation when considering the ILs only. We also observed that days to heading (determined by the field site location) was highly correlated with all yield components except for grain width and whole milling yield.

Under WS conditions, there were fewer significant correlations among yield components overall (Table 2-8B and 3-5B), suggesting that water stress disrupted many of the trait associations observed under irrigated conditions. Only 15 components were considered in the WS trials (due to missing data for AMYLOSE (amylose content), PLSQM (plants per square meter), STDPCT (stand percent), LODGE (percent lodging)). Of those that were significantly correlated with yield performance (kg ha^{-1}), the only components that were significant under WS conditions but not under irrigated conditions were grain length (R^2=0.329 for ILs only and R^2=0.363 for ILs plus checks) and grain width (R^2=0.225 for ILs only and R^2=0.181 for ILs plus checks).

Physiological observations of ILs in field and greenhouse

Yield, measured as kilograms of seed produced per unit of harvested area, describes the plant's ability to convert radiation energy into fixed carbon and to mobilize photosynthate from the leaves and stem into the developing grains along the panicle. Effective canopy structure allows the plant to intercept light efficiently, and this can be accomplished through a variety of plant architectural designs, involving differences in tiller number, tiller angle, leaf angle, plant height and plant biomass (Loomis and Amthor, 1999). Observations of the plant morphology can be a hint to why these lines were most productive under field conditions.

Prolonged photosynthetic period (or delayed senescence) is one tactic to increase the total amount of photosynthate available to the plant. 'Staygreen' is a characteristic of the flag leaf, whereby it retains its ability to photosynthesize when other plants senesce. Staygreen is observed as a prolonged presence of chlorophyll or green color in the flag leaf, and is thought to extend the plant's photosynthetic capabilities, aiding in grain filling as the plant matures. Staygreen was observed in IL 43_1-2 and its derivatives (carrying *yld2.1)* but not in IL 219_1-5 (carrying *yld6.1)* (Fig. 3-4).

Another tactic for plants to effectively enhance levels of photosynthate available for grain filling is to store more carbohydrates in the stem (or culm) and subsequently to remobilize them to the panicle during grain filling. To evaluate and compare the capacity for carbohydrate storage among ILs, tiller number and stem diameter were measured on greenhouse-grown plants simultaneously with harvesting in Oct 2010 (approx. 3.5 months after planting). Stem diameter was measured at 20 cm height and appeared to differ between ILs carrying *yld2.1* and *yld6.1*. In this experiment, stem diameter was measured with a caliper, but the method needs to be standardized because too much pressure would cause the culm to collapse, but too little would

give an inaccurately large stem diameter reading. One way to get around the problem was to peel back the leaves surrounding the culm.

When peeling back the sheaths, I observed three specimens carrying the *yld2.1* target introgression that had adventitious roots growing upwards against gravity (Fig. 3-5). This was not observed in any of the ILs carrying the *yld6.1* QTL that were grown at the same time. These three specimens were also the most productive (high seed weight) plants in the greenhouse trial. Adventitious roots growing inside the leaf sheath along the culm may be due to a combination of environment (e.g. high moisture within the culm) and genetic background (adventitious roots are related to growth of stolons, a plant characteristic seen in wild *Oryza* relatives). Adventitious root growth that goes against gravity is suggestive of genetic control through hormonal pathways. This morphological feature associated with some of the highest yielding families in our study was a striking discovery, but to date it has only been observed under greenhouse conditions. These observations should be followed up to determine whether this characteristic is observed in the field.

Previously documented QTLs

Another way to search for possible mechanisms underlying the yield advantage of our best ILs is to search the literature to identify QTLs or candidate genes that co-localize with the regions of introgression in our lines. As outlined below, several of the introgressed regions in our ILs coincide with previously reported QTLs, but each of the introgressions is several megabases in length, and therefore we are proceeding to fine map the regions of interest in an attempt to narrow down the search space before investing time in the search for candidate genes.

Yld1.1

There was no *yld1.1* in the original QTL study by Thomson *et al*. (2003). What we call *yld1.1* in this project actually corresponds to the QTL *gpp1.2* (grains per panicle), where the *O. rufipogon* introgression was expected to decrease the number of grains per panicle. We included this target as a negative control, hypothesizing that the *O. rufipogon* introgression across the *yld1.1* QTL region would confer lower yield. Surprisingly, some of the ILs carrying *O. rufipogon* alleles at *yld1.1* were high yielding during the 2007-2008 multi-location trials, with line 158_2-7 significantly out-yielding the recurrent parent and ranking ninth among all 50 *IL*s. During the 2009-2011 trials, the yield performance of 158_2-7 decreased to a level below the performance of the recurrent parent, Jefferson. This may have been related to the fact that the plant spacing was decreased in the 2009-2011 URRN trials, compared to the wide spacing in 2007-2008 multi-location field trials.

Yld2.1

Of the 50 lines tested in the field, two of the best yielding lines targeted *yld2.1,* and both lines have continued to show remarkable yield performance in the URRN trials. The consistently outstanding performance of these lines prompted further study. We have developed several sub-ILs derived from the original lines, and have prioritized this QTL for further backcrossing and fine mapping to genetically dissect the region of interest.

Some of the *yld2.1* material, including the best performing IL 43_1-2, has a potentially important background introgression close to *yld9.1*. Recently published work shows that there is a 1,000-grain weight (TGW) QTL on chromosome 9 between the rice microsatellite markers, RM215 and RM1013, identified in a single environment and explaining 7.8% of the phenotypic

variation (Liu *et al.,* 2010). This QTL coincides with our *yld9.1* QTL peak marker RM215. Other QTLs in the same region on chromosome 2 are associated with markers RM279 (panicle length, seeds per panicle and secondary branch number) (Luo *et al.,* 2009) and RG256b – RM207 (*Gp1*, grains per plant) (Moncada *et al.,* 2001).

Yld3.2

Although *yld3.2* was originally reported with the highest LOD score of the six yield QTLs in our study, the ILs developed to target *yld3.2* did not perform as well as other ILs in the field. Numerous background introgressions remain in IL 121_1-1, but the most prominent one is a 14Mb background introgression that includes *yld9.1*. With such extensive background noise, linkage drag may have lowered the yield potential of the lines carrying *yld3.2*. It can be hypothesized that *O. rufipogon* alleles at *yld3.2* may act in the opposite manner as *yld9.1,* or that when *O. rufipogon* alleles are introgressed in both target regions, they interact to negate the yield advantage that each confers if introgressed individually into the Jefferson background. We aim to test this hypothesis by separating the *yld3.2* and *yld9.1* target introgressions via further backcrossing and marker assisted selection of progeny derived from IL 121_1-1.

Yld6.1

Yld6.1 colocalized with another yield component QTL, percent seed set (*pss6.1*) (Thomson *et al*., 2003). Percent seed set is calculated by dividing the number of grains per panicle by the number of spikelets per panicle. The original QTL study showed that the Jefferson allele at the *pss6.1* locus conferred increased percent seed set, therefore the *O. rufipogon* introgression at

yld6.1 can be predicted to lower the percent seed set. In field trials, we did not measure the number of spikelets per panicle, therefore we could not verify this yield component QTL.

The ILs carrying *yld6.1* that are being submitted to the GS*OR* represent families 219_2-9 and 219_1-5. These lines have lower amylose content, lower alkali spreading value, shorter grain length and more seeds per panicle than Jefferson. Several candidate genes associated with amylose content and ASV are located in the *yld6.1* region; *O. rufipogon* alleles at *Waxy* and *Alk* may contribute to the grain quality traits observed in ILs yld6_A and yld6_B.

Yld8.1

Yld8.1 co-localized with two other yield component QTLs: grains per panicle (*gpp8.1*) and plant height (*ph8.1*) (Thomson *et al.,* 2003). *O. rufipogon* alleles at these loci were expected to increase the number of grains per panicle and to increase plant height. In the greenhouse, we did not observe any obvious height differences in the ILs carrying *yld8.1*, but in the irrigated field conditions, ILs 121_1-1 and 121_2-2 were significantly taller than Jefferson (100 cm, p-value = 0.0093; 99.7 cm, p-value = 0.036, respectively). Several lines carrying *yld8.1* were eliminated from the 2008 trials due to red pericarp, so it is difficult to evaluate the range of phenotypes we might have observed in lines with an introgression across *yld8.1*.

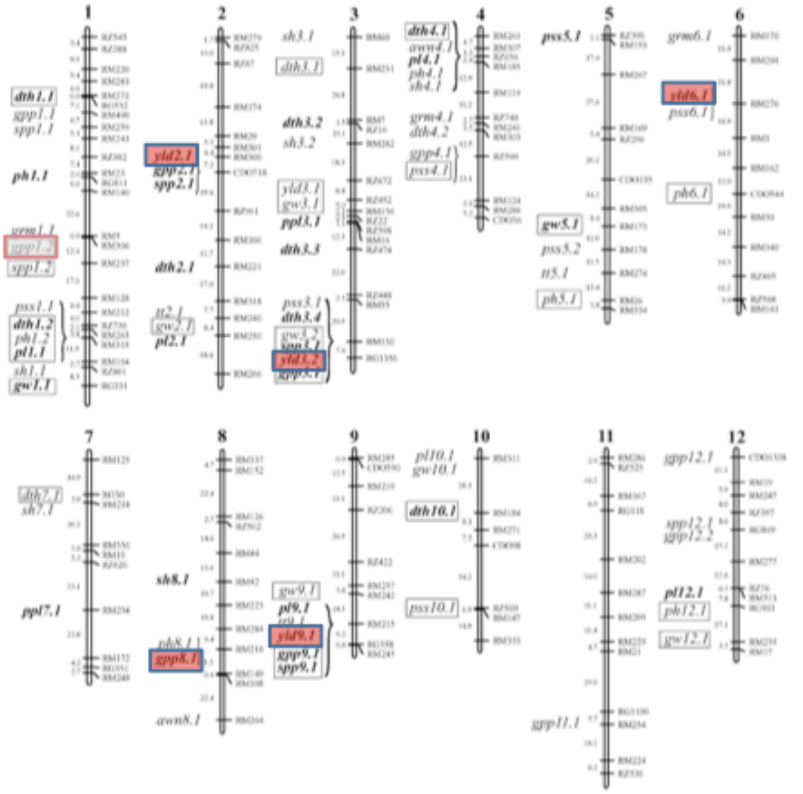
One of the most striking characteristics of this family 121 carrying *yld8.1* was observed in greenhouse trials, where highly vigorous and extensive root growth could be seen extending underneath the pots. We were not able to accurately measure root length in these experiments, but it would be interesting to document root growth patterns of ILs carrying *yld8.1* using the gellan gum system for time-lapse photography (Clark *et al.,* 2011). We also noted that plants carrying *yld8.1* segregated for awns.

The *yld8.1* QTL was previously fine-mapped in a Korean elite cultivar, Hwaseongbyeo (Xie *et al.,* 2006). An introgression from the same *O. rufipogon* donor enhanced the yield of this *temperate japonica* cultivar as seen in ILs 121_2-2 selected for submission to GS*OR* collection. Yet IL 121_2-2 containing *O. rufipogon* alleles at *yld8.1* in the Jefferson background did not yield as well as other ILs in this study. This is likely due to the large number of non-target *O. rufipogon* introgressions in the genetic background of IL 121_2-2.

Yld9.1

Yld9.1 coincided with four other yield component QTLs in the study by Thomson *et al*. (2003); panicle length (*pl9.1*), tiller type (*tt9.1*), grains per panicle (*gpp9.1*), and spikelets per panicle *(spp9.1).* The *O. rufipogon* alleles across this region were predicted to confer longer panicle length, open tillering (lazy plant type), increased number of grains per panicle and spikelets per panicle. The most significant QTL was *tt9.1,* with a LOD score of 45.24 and an R^2 value of 0.463. Morphological traits such as tiller type (*tt*) were originally measured under controlled conditions in the greenhouse (Thomson *et al.,* 2003). At no point in our study did we observe the lazy tiller type in any of the material targeting *yld9.1* grown in the greenhouse or in the field. This may be due to the phenotypic selection that was subsequently imposed during backcrossing to create the ILs used in our study.

Yld9.1 coincides with other known QTLs such as thousand grain weight 9 (*TGW9)* detected in a population of RILs developed from a cross between Minghui 63 and Teqing (Liu *et al.,* 2010), and a QTL cluster including *gw9.1*, *hd9.1* and *ph9.1* reported in a BC_2 population derived from a cross between Hwaseongbyeo and *O. rufipogon* (Xie *et al.,* 2008).



	LOD	R_2
gpp1.2	6.35	9.7
yld2.1	7.35	12.3
yld3.2	11.56	16.6
yld6.1	3.74	6.5
gpp8.1	4.44	7.2
yld9.1	4.28	7.4

Fig. 3-1. Location of six yield QTLs detected by Thomson *et al.* (2003). Highlighted QTLs in boxes were targeted for NIL development. The table to the right shows the respective LOD score and R^2 value.

Table 3-1. Summary of the QTL targets based on Thomson *et al*. (2003) split by QTL analysis method (IM and CIM) and location.

Abbreviations: QTL = yield QTL that is the target of our study; Thomson QTL = QTL of yield and QTLs for yield components that were detected in the 2003 study; Beneficial allele = Parental allele that confers the more desirable trait; Peak marker = SSR or RFLP marker closest to the highest LOD score; Flanking markers = SSR or RFLP marker flanking the loci with the highest LOD score; Alvin, transplanted = Plots that were transplanted in Alvin, Texas; Alvin, drilled = Drill seeded plots in Alvin, Texas; Beaumont, drilled = Drill seeded plots in Beaumont, Texas; Newport, drilled = Drill seeded plots in Newport, Arkansas (only 1 replication); Greenhouse =

				QTL analysis	Interval mapping										Composite interval mapping									
				Environment	Alvin, transplant		Alvin, drilled		Beaumont, drilled		Newport, drilled		Greenhouse		Alvin, transplant		Alvin, drilled		Beaumont, drilled		Newport, drilled		Greenhouse	
QTL	Thomson QTL	Beneficial allele	Peak marker	Flanking markers	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %	LOD	R^2 %
yld1.1	*gpp1.2*	Jefferson	RM5	RM5-RM237	4.51	7.7	3.58	6.2	3.09	5.4	-	-	NA	NA	6.35	9.7	4.34	6.9	3.34	6.5	-	-	NA	NA
yld2.1	*gpp2.1*	*O. rufipogon*	CDO718	CDO718	4.57	7.8	-	-	-	-	-	-	NA	NA	-	-	-	-	-	-	-	-	NA	NA
yld2.1	*spp2.1*	*O. rufipogon*	CDO718	CDO718	5.91	10.0	-	-	-	-	-	-	NA	NA	3.69	5.7	-	-	-	-	-	-	NA	NA
yld2.1	*yld2.1*	*O. rufipogon*	CDO718	CDO718	7.35	12.3	-	-	-	-	-	-	NA	NA	4.77	9.8	-	-	-	-	-	-	NA	NA
yld3.2	*gpp3.1*	*O. rufipogon*	RG1356	RM130-RG1356	4.53	7.8	2.96	5.2	-	-	-	-	NA	NA	5.51	9.5	7.02	12.4	-	-	3.02	5.3	NA	NA
yld3.2	*spp3.1*	*O. rufipogon*	RM130	RM130-RG1356	4.73	8.1	5.58	9.5	-	-	2.82	5.9	NA	NA	5.82	9.5	10.46	15.1	3.30	4.8	3.30	6.5	NA	NA
yld3.2	*pss3.1*	Jefferson	RM130	RM130	-	-	4.00	6.9	-	-	-	-	NA	NA	-	-	4.31	6.0	-	-	-	-	NA	NA
yld3.2	*dth3.4*	*O. rufipogon*	RM130	RM130	3.63	6.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
yld3.2	*gw3.2*	Jefferson	RM130	RM130-RG1356	3.75	6.5	6.69	11.3	NA	NA	-	-	NA	NA	3.72	6.0	6.42	8.8	NA	NA	-	-	NA	NA
yld3.2	*yld3.2*	*O. rufipogon*	RG1356	RM130-RG1356	-	-	3.80	6.6	10.1	16.5	NA	NA	NA	NA	-	-	5.60	8.4	11.56	16.6	NA	NA	NA	NA
yld6.1	*pss6.1*	Jefferson	RM276	RM276-RM3	-	-	-	-	-	-	-	-	NA	NA	-	-	5.69	18.6	-	-	-	-	NA	NA
yld6.1	*yld6.1*	*O. rufipogon*	RM274	RM276	-	-	-	-	3.74	6.5	NA	NA	NA	NA	-	-	-	-	-	-	NA	NA	NA	NA
yld8.1	*gpp8.1*	*O. rufipogon*	RM210	RM210	-	-	-	-	-	-	-	-	NA	NA	-	-	-	-	4.44	7.2	-	-	NA	NA
yld8.1	*ph8.1*	Jefferson	RM210	RM210-RM308	NA	NA	-	-	-	-	NA	NA	-	-	NA	NA	-	-	-	-	NA	NA	5.52	5.3
yld9.1	*pl9.1*	*O. rufipogon*	RM215	RM215	-	-	4.87	8.3	NA	NA	-	-	-	-	-	-	5.15	7.5	NA	NA	-	-	-	-
yld9.1	*tt9.1*	Jefferson	RM215	RZ422-RM245	NA	NA	NA	NA	NA	NA	NA	NA	32.19	34.3	NA	NA	NA	NA	NA	NA	NA	NA	45.24	46.3
yld9.1	*yld9.1*	*O. rufipogon*	RM215	RM215	4.28	7.4	-	-	-	-	NA	NA	NA	NA	-	-	-	-	-	-	NA	NA	NA	NA
yld9.1	*gpp9.1*	*O. rufipogon*	RM215	RM215	4.03	6.9	-	-	3.61	6.2	-	-	NA	NA	-	-	-	-	-	-	-	-	NA	NA
yld9.1	*spp9.1*	*O. rufipogon*	RM215	RM215	4.68	8.0	-	-	4.59	7.9	-	-	NA	NA	-	-	-	-	4.84	8.2	-	-	NA	NA

Table 3-2. The number of replications of each IL and check in each field trial location.

Year	2007				2008					
Location	Alvin	BMT	JBORO	STGT	Alvin	BMT	JBORO	STGT	BMT-WS	STGT-WS
# Reps	3	3	3	3	3	3	3	3	2	3
Cocodrie	18	6	18	12	18	12	18	15	2	15
Jefferson	18	12	18	15	44	21	45	15	2	15
Trenasse	18	6	18	6	18	9	18	6	0	6
XL723	18	6	18	18	18	12	18	18	0	18

Abbreviations: Alvin = Alvin, Texas; BMT= Beaumont, Texas; JBORO = Jonesboro, Arkansas; STGT = Stuttgart, Arkansas; BMT-WS = water-stressed field conditions at BMT; STGT-WS = water-stressed field conditions at STGT; # Reps = number of replications at each location.

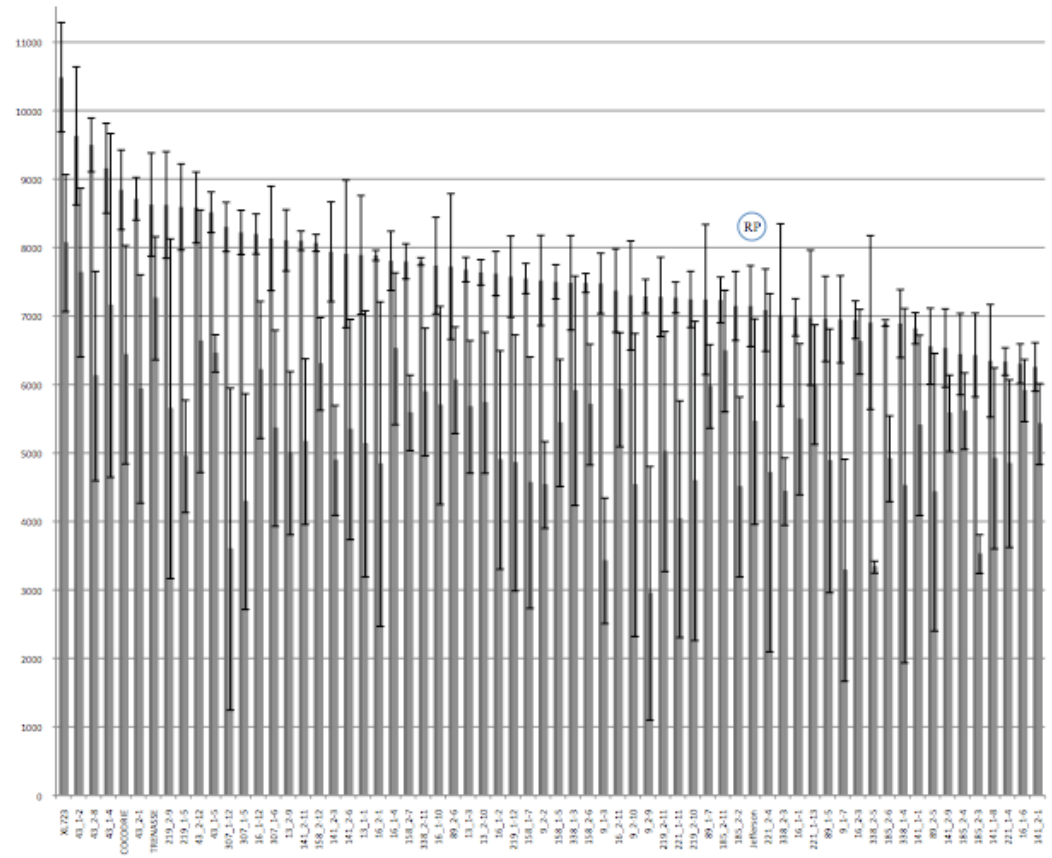


Fig. 3-2. Stuttgart flooded and water-stressed yield performance distribution in 2008. In both flooded (light gray bars) and water-stressed conditions (dark gray bars), there were three replications per entry. For controls, there were 15 replications each of Jefferson and Cocodrie, six replications of Trenasse and 18 of XL723.

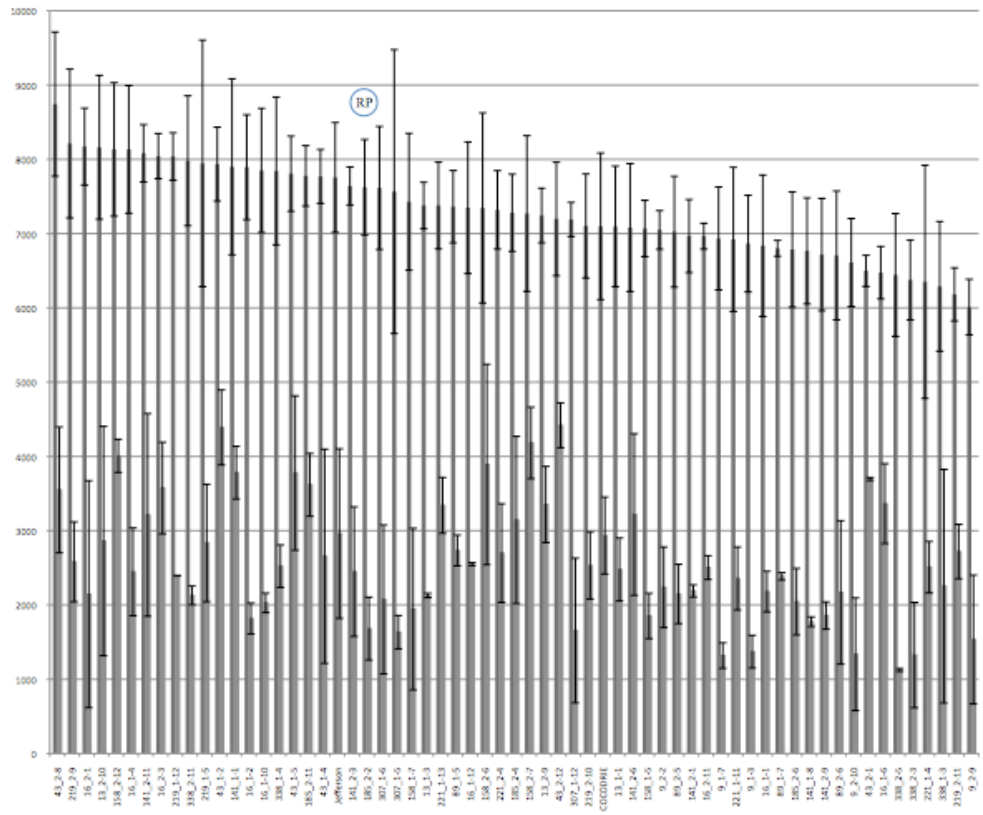


Fig. 3-3. Beaumont flooded and water-stressed yield performance distribution in 2008. In flooded conditions (light gray bars), there were three replications per entry, 12 replications for Cocodrie, and 21 replications for Jefferson. In water-stressed conditions (dark gray bars), there were two replications for each entry, Jefferson and Cocodrie.

Table 3-3. Average of yield and yield component traits for all entries tested in 2008 under water-stressed and irrigated conditions in Beaumont (BMT) and Stuttgart (STGT). Values are rounded to show only three significant figures.

	Irrigated		Water-stressed	
Trait mean (Std dev)	BMT	STGT	BMT	STGT
Yield	6630 (861)	6970 (107)	2330 (847)	5000 (1490)
Yield (kg ha^{-1})	7440 (965)	7810 (1200)	2610 (950)	5600 (1670)
Total milling yield	71.5 (1.94)	68.7 (1.62)	74.2 (2.41)	72.0 (1.50)
Whole milling yield	61.0 (4.06)	45.6 (7.65)	55.0 (16.6)	50.0 (9.68)
Thousand seed weight	24.8 (3.09)	27.6 (3.33)	22.7 (2.37)	27.0 (2.52)
Percent stand	69.6 (9.42)	79.4 (10.6)	80.6 (2.99)	81.1 (7.73)
Days to heading	78.4 (2.33)	82.2 (2.07)	79.2 (3.12)	80.5 (4.82)
Plant height	98.6 (12.9)	96.1 (14.0)	65.1 (8.76)	83.0 (14.0)
Average panicle length	21.0 (2.04)	18.6 (2.36)	17.3 (1.74)	17.5 (3.03)
Average seeds per panicle	150 (37.6)	89.8 (21.9)	71.6 (28.4)	73.4 (25.3)
Average panicle weight	3.71 (0.97)	2.46 (0.59)	1.62 (0.64)	1.94 (0.67)
Average number of tillers	6.33 (2.15)	7.81 (3.02)	4.74 (1.32)	6.61 (2.12)
Average plant weight	NA	16.0 (7.38)	NA	9.71 (4.57)
Grain length	6.63 (0.27)	6.21 (0.22)	6.29 (0.29)	6.48 (0.25)
Grain width	2.31 (0.09)	2.35 (0.08)	2.21 (0.09)	2.23 (0.08)
Percent chalk	5.76 (3.55)	12.28 (8.28)	5.12 (3.53)	3.55 (3.84)

Abbreviations: Trait mean = average of all entries (ILs, sib-lines, and checks) in corresponding irrigation scheme; Std dev = standard deviation; BMT = Beaumont, Texas; STGT = Stuttgart, AR; NA = not available because trait was not measured at the location.

Table 3-4. Regression analysis using all ILs, controls and checks identifying significant variables explaining yield under A) irrigated field conditions in 2007-2008 and B) water stressed conditions in 2008. Variables highlighted in bold font indicate variables with significant partial R-squared values.

A) Summary of Forward Selection: Flooded (All entries and checks)								
Step	**Variable Entered**	**Label**	**Number Vars In**	**Partial R-Square**	**Model R-Square**	**C(p)**	**F Value**	**Pr > F**
1	**AVPANL**	**AVPANL**	1	**0.4208**	0.4208	399.314	809.25	<.0001
2	**AVPANWT**	**AVPANWT**	2	**0.0951**	0.5158	153.306	218.5	<.0001
3	**WHOLEMY**	**WHOLEMY**	3	**0.0266**	0.5424	85.9819	64.56	<.0001
4	PLTHT	PLTHT	4	0.0099	0.5523	62.1361	24.58	<.0001
5	ASV	ASV	5	0.0079	0.5602	43.4052	20.06	<.0001
6	AVPLTWT	AVPLTWT	6	0.0053	0.5656	31.5246	13.58	0.0002
7	AVSDPAN	AVSDPAN	7	0.0044	0.57	22.0373	11.34	0.0008
8	CHAKPCT	CHAKPCT	8	0.0037	0.5737	14.3093	9.68	0.0019
9	TOTALMY	TOTALMY	9	0.0024	0.5761	10	6.31	0.0122

B) Summary of Forward Selection: Water stressed (All entries and checks)								
Step	**Variable Entered**	**Label**	**Number Vars In**	**Partial R-Square**	**Model R-Square**	**C(p)**	**F Value**	**Pr > F**
1	**AVPANWT**	**AVPANWT**	1	**0.208**	0.208	23.9464	73	<.0001
2	D2HD	D2HD	2	0.016	0.2239	19.9001	5.7	0.0176
3	**CHAKPCT**	**CHAKPCT**	3	**0.0271**	0.251	11.6477	9.98	0.0018
4	WHOLEMY	WHOLEMY	4	0.0174	0.2684	7.0662	6.53	0.0111
5	AVPANL	AVPANL	5	0.0081	0.2765	6	3.07	0.0811

Abbreviations: AVPANL = average panicle length; AVPANWT = average panicle weight; WHOLEMY = whole milling yield; PLTHT = plant height; ASV = alkali spreading value; AVPLTWT = average plant weight; AVSDPAN = average seeds per panicle; CHAKPCT = percent chalk; TOTALMY = total milling yield; D2HD = days to heading.

Table 3-5. Correlation analysis of combined IL entries and check means of traits measured under A) irrigated field conditions in 2007-2008 and B) water-stressed field conditions in 2008. Missing correlations are indicated by periods.
Genetic correlation is significantly different from zero at p<0.05 (*), p<0.01 (**), p<0.001 (***), and p<0.0001 (****and gray coloration).

	Yield	TotalMY	WholeMY	AVPANL	AVSDPAN	AVPANWT	AVTILL	AVPLTWT	KSDWT	ASV	AMYLOSE	GL	GW	CHKPCT	PLSQM	STDPCT	D2HD	PLTHT	Lodge
Yield	1	0.377****	0.133**	0.516****	0.259****	0.250****	0.247****	0.382****	-0.195****	0.010	-0.056	0.032	-0.043	0.258****	-0.137*	0.066	0.482****	0.145***	0.222****
TotalMY		1	0.270***	0.352****	0.315****	0.319****	0.285****	0.500****	-0.290****	0.181**	-0.074	0.218****	-0.082	-0.085	-0.443****	-0.454****	0.306****	0.112**	0.333****
WholeMY			1	0.423****	0.472****	0.565****	-0.128**	0.302****	-0.063	0.034	0.053	0.224****	0.105*	-0.052	-0.380****	-0.581****	0.078	0.007	-0.061
AVPANL				1	0.610****	0.548****	0.059	0.455****	-0.418****	0.100	0.154**	0.157**	-0.276****	-0.049	-0.209***	-0.454****	0.348****	0.372****	0.222****
AVSDPAN					1	0.887****	0.018	0.536****	-0.517****	0.232****	0.199***	0.185***	-0.207****	-0.304****	-0.386****	-0.521****	0.304****	0.246****	0.240****
AVPANWT						1	-0.096	0.537****	-0.174***	0.177**	0.128*	0.178***	0.041	-0.138**	-0.493****	-0.535***	0.322****	0.107*	0.068
AVTILL							1	0.732****	-0.300****	0.304****	-0.038	0.026	-0.112*	-0.058	-0.303****	0.167	0.225****	0.171***	0.475****
AVPLTWT								1	-0.365****	0.183**	-0.063	0.188***	-0.083	-0.101	-0.570****	-0.178	0.320****	0.309****	0.460****
KSDWT									1	-0.380****	-0.115	0.025	0.507****	0.311****	0.058	0.208*	-0.199****	-0.356****	-0.515****
ASV										1	0.265****	-0.204***	-0.270***	-0.207***	-0.178**	.	0.525****	0.182**	0.243****
AMYLOSE											1	-0.254****	-0.006	-0.141*	-0.032	.	0.261****	-0.003	-0.105
GL												1	-0.263****	-0.439****	-0.197**	-0.515****	-0.287****	0.083	0.179***
GW													1	0.386****	-0.094	0.132	-0.043	-0.209****	-0.172
CHKPCT														1	-0.022	0.453****	0.330****	-0.085	-0.097*
PLSQM															1	.	-0.471****	-0.136*	-0.193**
STDPCT																1	0.438****	-0.189*	.
D2HD																	1	0.125**	0.196****
PLTHT																		1	0.400****
Lodge																			1

	Yield	TotalMY	WholeMY	AVPANL	AVSDPAN	AVPANWT	AVTILL	AVPLTWT	KSDWT	GL	GW	CHKPCT	STDPCT	D2HD	PLTHT
Yield	1	-0.377****	-0.128	0.157	0.299***	0.526****	0.574****	0.469****	0.644****	0.363****	0.181*	-0.126	0.216**	0.116	0.447****
TotalMY		1	0.616****	0.092	0.161	0.073	-0.291***	0.353**	-0.208*	0.082	-0.078	-0.094	0.046	-0.028	-0.392****
WholeMY			1	0.164	0.188*	0.141	-0.182*	0.268*	-0.092	0.295***	-0.198*	-0.455****	-0.052	0.176*	-0.104
AVPANL				1	0.493****	0.450****	-0.153	0.379*	-0.060	0.083	-0.246	-0.100	0.131	-0.028	0.111
AVSDPAN					1	0.924****	-0.090	0.773****	-0.082	0.023	-0.068	-0.047	0.190*	0.031	-0.085
AVPANWT						1	0.086	0.795****	0.283***	0.233**	0.068	-0.091	0.199*	0.098	0.061
AVTILL							1	0.425***	0.470****	0.186*	0.209*	-0.071	0.085	0.095	0.239**
AVPLTWT								1	0.094	0.142	0.071	-0.074	0.280*	-0.024	-0.208
KSDWT									1	0.551****	0.364****	-0.148	0.041	0.171	0.403****
GL										1	-0.148	-0.256**	-0.112	0.255**	0.216*
GW											1	0.313***	0.068	-0.181*	-0.072
CHKPCT												1	0.038	-0.365****	-0.291***
STDPCT													1	-0.398****	-0.143
D2HD														1	0.372****
PLTHT															1

Abbreviations: Yield = grain yield in kg ha^{-1}; TotalMY = Total milling yield (filled and unfilled seed); WholeMY = Whole milling yield (filled seeds only); AVPANL = average panicle length; AVSDPAN = average seeds per panicle; AVPANWT = average panicle weight; AVTILL = average number of tillers; AVPLTWT = average plant weight; KSDWT = 1,000-seed weight; ASV = alkali spreading value; AMYLOSE = amylose content; GL = grain length; GW = grain width; CHKPCT = percent chalk; PLSQM = plants per square meter; D2HD = days to heading; PLTHT = plant height; Lodge = percent lodge.

Fig. 3-4. Segregation of staygreen phenotype in 2009 field evaluations: 43_1-2_68 (short erect flag leaf with staygreen); 43_1-2_81 (long erect flag leaf with staygreen); 219_1-5_29-7 (long erect senescent flag leaf).

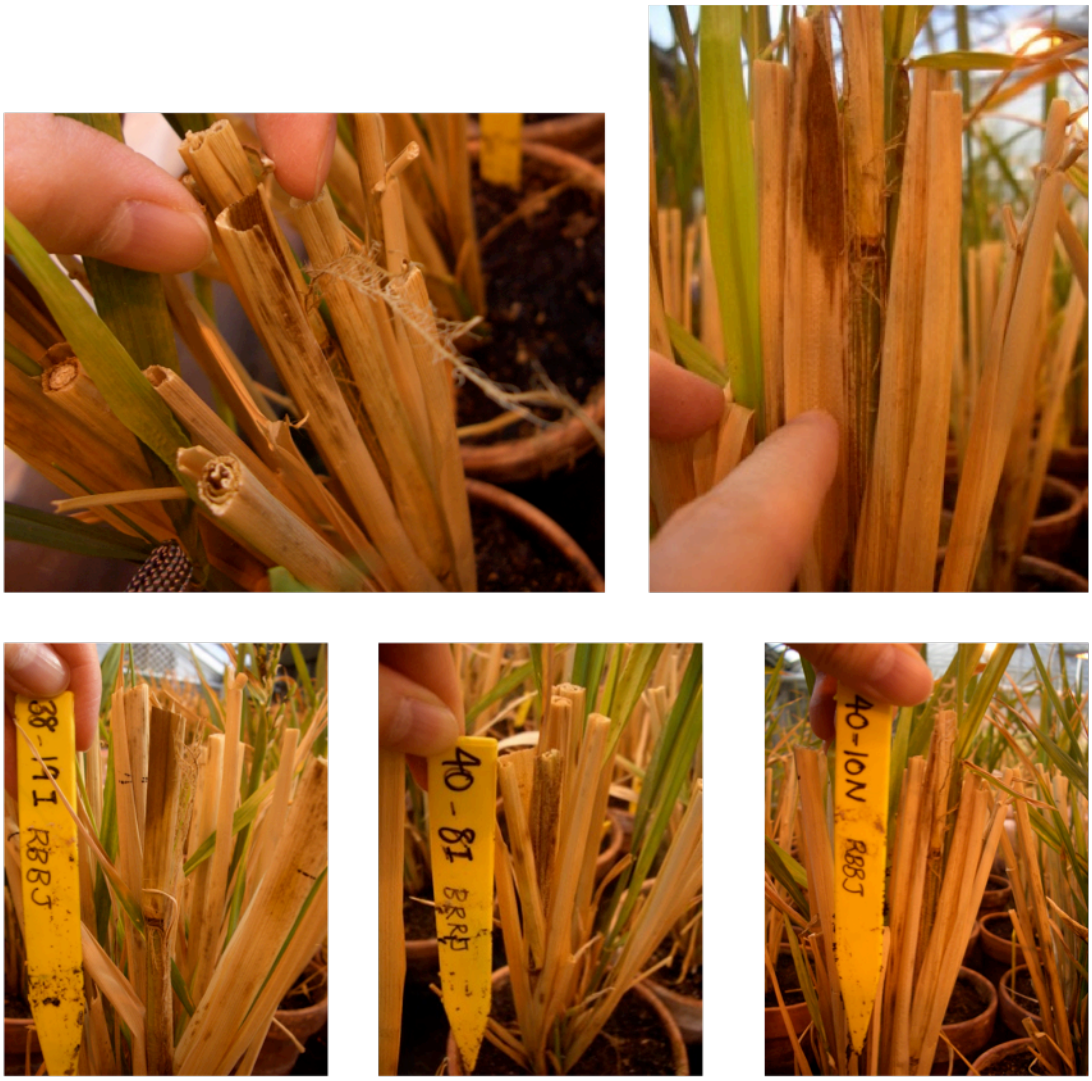


Fig. 3-5. Adventitious roots growing upward within the sheath of the culm of most productive plants in greenhouse during 2011 summer. Black marker lines on surrounding leaf sheath marks 20cm above soil level.

References

Clark, R. T., MacCurdy, R. B., Jung, J. K., Shaff, J. E., McCouch, S. R., Aneshansley, D. J., et al. (2011). Three-dimensional root phenotyping with a novel imaging and software platform. *Plant Physiology, 156*(2), 455-465.

Li, J., Thomson, M., & McCouch, S. R. (2004). Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. *Genetics, 168*(4), 2187-2195.

Linscombe, S. D., Bollich, P. K., Groth, D. E., White, M., & Dunand, R. T. (2000). Registration of 'Cocodrie' rice. . *Crop Science, 40*(294).

Liu, T., Shao, D., Kovi, M., & Xing, Y. (2010). Mapping and validation of quantitative trait loci for spikelets per panicle and 1,000-grain weight in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics, 120*(5), 933-942.

Loomis, R. S., & Amthor, J. S. (1999). Yield potential, plant assimilatory capacity, and metabolic efficiencies. *Crop Science, 39*(6), 1584-1596.

Luo, X., Tian, F., Fu, Y., Yang, J., & Sun, C. (2009). Mapping quantitative trait loci influencing panicle-related traits from Chinese common wild rice (*Oryza rufipogon*) using introgression lines. *Plant Breeding, 128*(6), 559-567.

Maas, L., McClung, A., & McCouch, S. (2010). Dissection of a QTL reveals an adaptive, interacting gene complex associated with transgressive variation for flowering time in rice. *Theoretical and Applied Genetics, 120*(5), 895-908.

Moncada, P., Martínez, C. P., Borrero, J., Chatel, M., Gauch Jr, H., Guimaraes, E., et al. (2001). Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. *Theoretical and Applied Genetics, 102*(1), 41-52.

Sweeney, M. T., Thomson, M. J., Pfeil, B. E., & McCouch, S. (2006). Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell, 18*(2), 283-294.

Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., et al. (2007). Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genetics, 3*(8), e133.

Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., Kanamori, H., et al. (August 2009). Evolutionary history of GS3, a gene conferring grain length in rice. *Genetics, 182*(4), 1323-1334.

Thomson, M. J., Tai, T. H., McClung, A. M., Lai, X., Hinga, M. E., Lobos, K. B., et al. (2003). Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and Applied Genetics, 107*(3), 479-493.

Thomson, M. J., Edwards, J. D., Septiningsih, E. M., Harrington, S. E., & McCouch, S. R. (2006). Substitution mapping of dth1.1, a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. *Genetics, 172*(4), 2501-2514.

Xie, X., Song, M., Jin, F., Ahn, S., Suh, J., Hwang, H., et al. (2006). Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*. *Theoretical and Applied Genetics, 113*(5), 885-894.

Xie, X., Jin, F., Song, M., Suh, J., Hwang, H., Kim, Y., et al. (2008). Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa* × *O. rufipogon* cross. *Theoretical and Applied Genetics, 116*(5), 613-622.

CHAPTER 4

Advances in SNP genotyping

Advantages of SNP genotyping platforms

We were able to draw upon genotyping information generated using an Affymetrix array to select SNP markers for use in the *Jefferson x O. rufipogon* populations (Table 2-9). Using data from the 44K SNP chip (Zhao *et al.* 2011), we identified 18,760 SNPs that were polymorphic in the Jefferson and *O. rufipogon* parents, and from these, we selected subsets that were distributed uniformly across the 12 chromosomes, or that targeted particular introgressions of interest, and developed marker assays for use in characterizing the introgression lines (ILs) (Fig. 4-1; see Appendix for using plink to extract SNPs from 44K). We were able to extract useful SNPs from the Affymetrix genotypic dataset, and found it could be used reliably and interchangeability for developing smaller assays.

Approximately 90% of SNPs selected from the 44K SNP chip converted well to the KASP and iPLEX MassARRAY platforms. Thus, using prior knowledge from the 44K (Zhao *et al.* 2011) was a very cost-effective strategy for developing new, tailor-made SNP assays for this project. Using the KASP and iPLEX MassARRAY assays allowed us to clearly define the size and positions of *O. rufipogon* introgressions in the Jefferson ILs.

This underscores the versatility of the SNP diversity data available for rice, and the value of having a large diversity database as the foundation from which to design more targeted assays for specific projects. The SNP markers we selected will also be useful for transferring favorable *O. rufipogon* introgressions from Jefferson into other *tropical japonica* backgrounds.

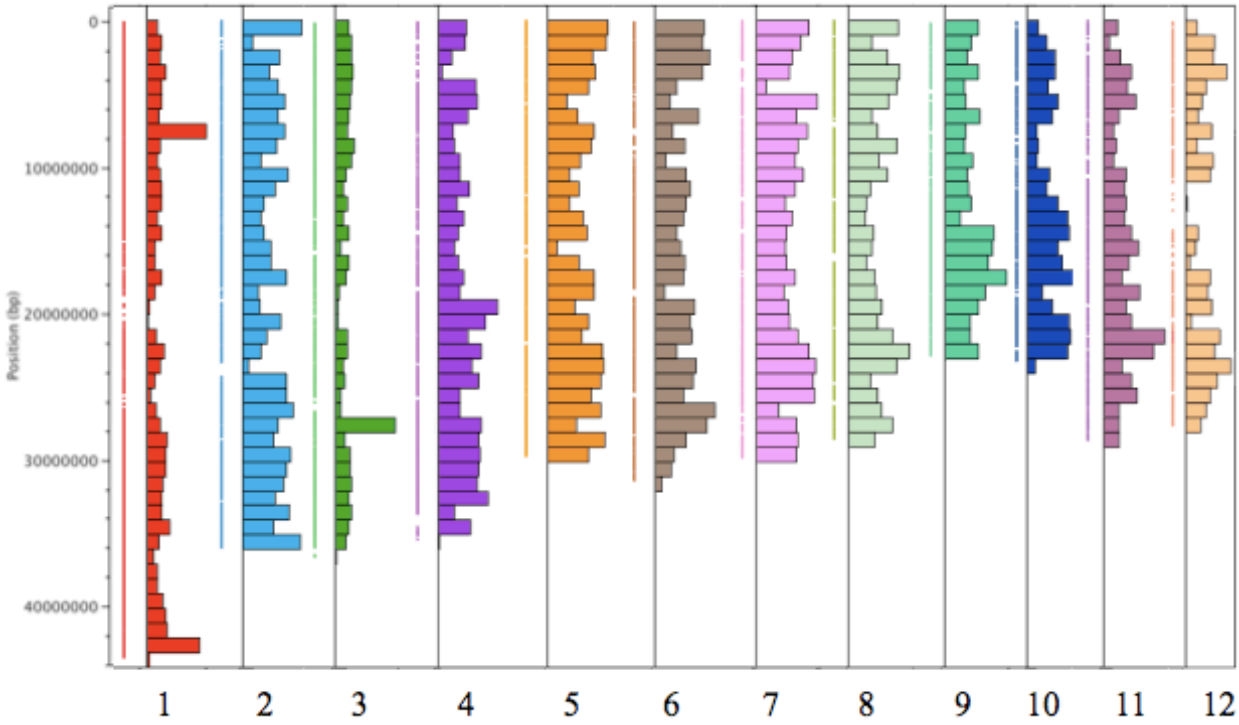
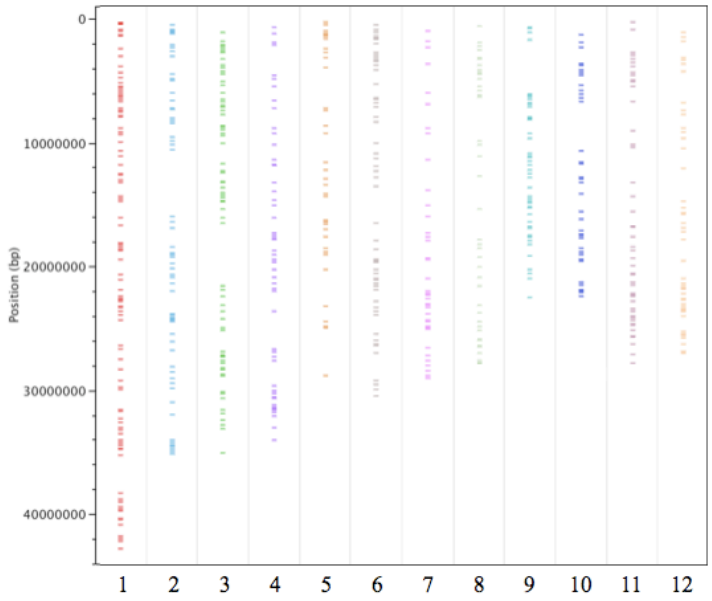


Fig. 4-1. Density of polymorphic SNPs between Jefferson and *O. rufipogon* on the 44K SNP array. Numbers along the bottom of the graph indicate the chromosome number; Y-axis shows the genetic position of SNPs along each chromosome; histograms indicate the density of SNPs in each 1 megabase block.

a)



b)

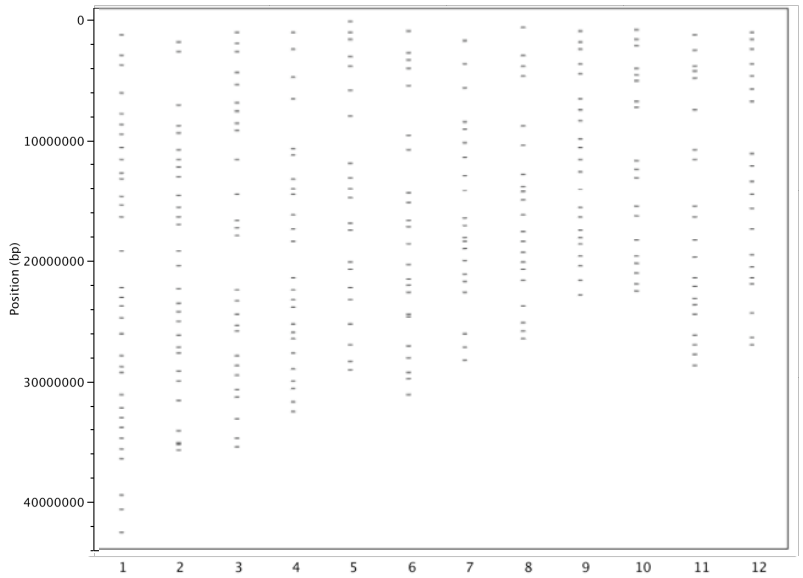


Fig. 4-2. Location of SNPs that are polymorphic between Jefferson and *O. rufipogon* on the a) 1536 SNP array and b) 384 SNP OPA 6.0. Numbers along the bottom of the graph indicate the chromosome number; Y-axis shows the genetic position of polymorphic SNPs along each chromosome.

Validating SNP selections using the 384 SNP chip:

Selected *IL*s were analyzed using a 384-SNP Illumina (BeadXpress) assay, RiceOPA6.0 (VC0011530-OPA) (Thomson *et al.* 2011). This SNP assay was designed to be informative for *Japonica*/*O. rufipogon* populations. For the Jefferson x *O. rufipogon* population used in this project, 282 SNPs out of the 384 (73.7%) were polymorphic between the recurrent parent (Jefferson) and the wild donor (*O. rufipogon*). The 384 SNP 'breeder's chip' had the resolution to reveal previously undetected background introgressions in lines previously genotyped with SSR and InDel markers. SNP-based genotyping increased the likelihood of finding small (previously undetected) background introgressions that may affect the phenotype.

The 384 SNP chip was developed as described by Thomson *et al*. (2011) to provide genome-wide coverage at low resolution using the Illumina BeadXpress platform, mainly for breeding purposes. Each 384-SNP assay was tailored for use with specific populations or subgroups. The chips have proven useful in developing NILs by reducing the number of generations of backcrossing (Boualaphanh *et al.,* 2011). If one were to generate the same amount of data using conventional SSR genotyping, the amount of time and labor required to collect 384 different markers for each sample would be prohibitive in terms of both time and money. The SNP chip lowers the cost of generating marker data, from 40 cents to 10 cents per data point (or 40 dollars per sample), and researchers are obliged to spend a greater portion of their time analyzing the data.

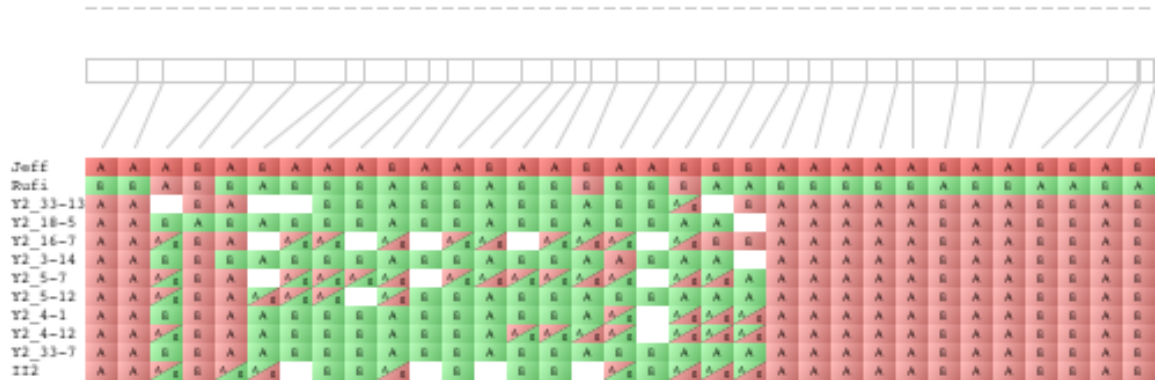
In our work, use of the custom 384 SNP OPA was more informative than using the larger 1536 SNP assay, because there were fewer gaps on a genome-wide scale (Fig. 4-2). This is due to the fact that the 384 SNPs were designed specifically to detect polymorphism between *tropical japonica* and *O. rufipogon* samples. The high quality of the OPA ensured that there was

little missing data and allowed for direct comparison across samples, and easy visualization using Flapjack software (Milne *et al.,* 2010).

Bulking DNA samples for genotyping using the 384 SNP chip

To minimize the number of chips used to genotype a population, we considered bulking several DNA samples in equal proportions to form a single bulk DNA sample. Nine individuals from family 85 and eight individuals from family 43 were each bulked into one mixture after normalizing the DNA concentration (Fig. 4-3). Within a bulk, we were able to detect heterozygous allele calls when there were two or more segregating individuals out of 9 total (22%), but not when there was only one in an eight-plant bulk (12.5%). We have yet to identify the exact number of plants that is best suited to genotyping using bulked DNA samples, but this study confirmed that it is possible to detect recombinants, heterozygotes and/or off-types using bulked DNA samples with the 384 Illumina BeadXpress SNP assay.

a)



b)

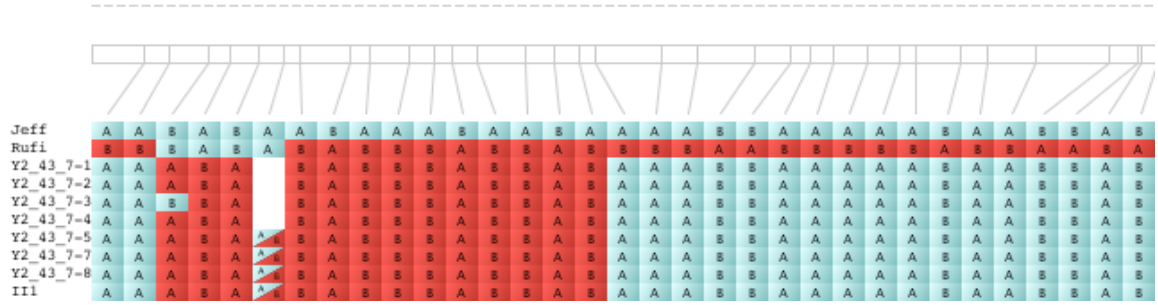


Figure 4-3. Bulked DNA sample using 384 SNP chip of a) nine individuals from family 85 and b) eight individuals from family 43. SNPs are indicated as columns and each sample is a row. First two rows show the parental genotypes, Jefferson and *O. rufipogon*, and the last row is the bulked sample genotype. II2 is a bulk of nine individuals, and II1 is a bulk of eight individuals.

Biosciences KASP markers

Genotyping of introgression lines (ILs) for *yld2.1* and *yld6.1* was conducted by competitive allele-specific PCR KASP chemistry (KBiosciences Ltd, Hertfordshire, UK). KASP is a fluorescence-based SNP detection system that uses two SNP-specific primers tagged with VIC and FAM tags for each SNP allele, and a common reverse primer. For purposes of IL

development, a total of 42 primer sets were designed (see Chapter 2 for list). Based on previous knowledge of polymorphic SNP positions from the 44K and 1536 SNP arrays, we designed one marker per megabase within the target regions of interest for *yld2.1* and *yld6.1*, and designated the VIC dye (appears red on the y-axis in analysis software) for *O. rufipogon* and the FAM dye (blue, x-axis) for Jefferson. Due to the presence of florescent dyes, the PCR product is photosensitive and temperature-sensitive so the plates were stored in cardboard boxes in the refrigerator. Plates were read using the RT-PCR machine located in the Cornell BioResource Center.

The single most time-conserving quality of KASP is that the template DNA quality does not have to be high. We used low quality Extract-N-Amp DNA, which takes less than 20 minutes to prepare, from tissue to concentrated DNA. To reduce the amount of reagents used, we first extracted DNA into 96 well plates; diluted DNA (2 ul of >5 ng/μl) was then manually or machine pipetted into a 384 well plate and placed under the fume hood to dry overnight (>8hrs). Only one PCR reaction was required (approximately 2 hours), after which the florescence could be scanned immediately. The system required that we assay at least 24 samples per SNP to ensure that there were enough samples to cluster manually using the lasso tool (Fig. 4-4). The KASP platform is ideal for genotyping hundreds or thousands of samples for a few SNPs, and would be perfect for fine mapping purposes.

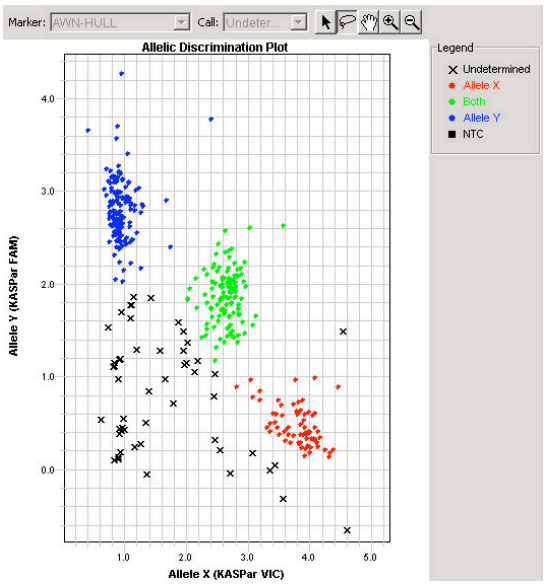


Fig. 4-4. SNP call cluster plots for KASP markers. Dots indicate the emission intensity of each fluorescent dye, FAM (y-axis) and VIC (x-axis). Alleles are called manually by clustering the samples into three different groups: Homozygous for FAM is the blue cluster (upper left), homozygous for VIC is the red cluster (lower right) and heterozygous (equal quantity of FAM and VIC dyes between the two clusters and highlighted in green). No-calls are indicated by the black X-marks.

iPLEX MassARRAY assay optimization

As an alternative to KASP, we developed markers for use with the iPLEX MassARRAY Gold SNP detection platform. Unlike KASP that is fluorescence-based, MassARRAY detects the alleles by molecular weight difference through a mass spectrometer (Buetow *et al.,* 2001) and has the ability for multiplexing SNP targets within a single well. There are two PCR steps to generate the amplicons for the mass spectrometer (Fig. 4-5). First, the region of interest containing the SNP is amplified through PCR using a forward and reverse SNP specific primer. After a cleaning step, the PCR products are mixed with the extension primer that aligns next to the SNP and by single based extension, the complementary nucleotide to the SNP is attached to the extension primer. The extension primers with the additional nucleotide are flown through the mass spectrometer (lighter, short sequence travel faster than heavier, longer sequences), and alleles are differentiated by the weight differences between the complementary nucleotide to the SNP. Based on molecular weight, all four nucleotides for each SNP can be detected; tri-allelic and quad-allelic SNPs can be tested as well. The protocol is available through Sequenom application notes online (Oeth *et al.,* 2005). iPLEX MassARRAY is known for high precision. It has been used quantitatively for maize (Liu *et al.,* 2010), and has the potential to quantify alleles in bulked DNA samples when genotyping many lines.

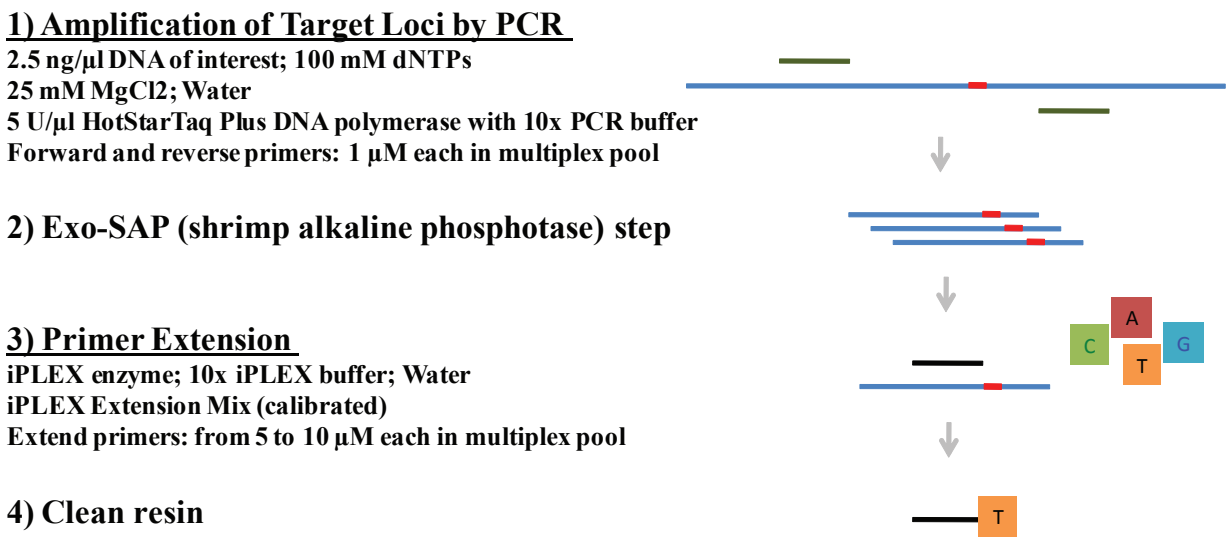


Fig. 4-5. Overview of PCR reaction steps before placing into iPLEX MassARRAY system.

Genotyping 384 samples using the iPLEX MassARRAY system at the BioResource Center at Cornell cost $1000 as of March 2011. This cost included reagents (the chip, PCR reagents, single base extension PCR reagents), the use of the MALDI-TOF machine to read the samples, and labor ($299 to cover the cost of processing by core staff and cleaning of resin). It is advisable to use all 384 wells at once due to the humidity-sensitive nature of the physical chip. In order to split the chip within a week, the facility must be notified ahead of time so that the staff can preserve the chip in a vacuum to prevent the chip from getting stale (3-HPA strength will diminish with time and increase the noise level).

Primers were purchased through IDT (Integrated Device Technologies, Inc., San Jose, California) with specifications for the SNP-specific primers to be at the 25 nm (nanomole) synthesis scale for 100 uM (micromole) final concentration, and extension primers at the 100 nmole synthesis scale for 400 uM concentration ($0.18/Base and $0.28/base, respectively, at Nov 16th 2010 purchase). The SNP-specific primers were mixed (forward and reverse) at a pre-set

volume and normalized concentration in water in a V-bottom 96-well format (there are three plate formats for different amount of liquid: deep well 800 ul, V-bottom 300 ul and PCR plate 150 ul). The average price of a three primer set for each SNP target was $21.21 in a 26-plex mix. With larger plexes, the longer extension primer will cost more than this average. Ordering the primers by plate significantly reduced the price compared to ordering in tubes. Initially we proposed to design a 384 SNP assay using Sequenom but the upfront primer cost of $4,000 did not justify the design of such an assay.

Pilot study design

For designing the trial experiment, the Genomics core facility suggested that we replicate the parents twice, with two blanks, and try different DNA extraction methods. Sequenom's assay designing specialist, Hema Liyange, also advised us to include two replicates of each of the parents, as well as at least four lines that were known to be segregating across the target and background introgressions (including some that were missing introgressions) for each of the families. Since we were on a tight deadline, one tactic to ensure enough data for selection was to include 'doublets' or two SNPs per target locus, so that even if one fails there would be another that could succeed close by. For the purposes of this project, chromosome 2, 6 and 9 were the most important regions, encompassing three different yield QTLs. We employed the 'doublet' strategy and increased the density to one SNP per half megabase. Originally, a single multiplex was designed to include both target and background introgressions but this was impossible with the additional SNPs, therefore we settled for three different 26 to 28 multiplexes and separating the background introgressions from the target introgression containing *yld2.1*. The Sequenom assays built for *yld2.1* and *yld6.1* materials will prove helpful for fine mapping because of the

increased density across the sequence of interest. Also with the pending release of *yld2.1 IL* 43_1-2, the genotypic data and information about the Sequenom assay can be released for use by breeders.

Effect of DNA quality on signal strength

Various DNA extraction methods were compared using DNA from the two parents, Jefferson and *O. rufipogon.* We used two replications for each DNA extraction method for the pilot assay. Three different extraction methods were compared: (1) Qiagen DNAeasy Plant tissue kit following the protocols provided by (Qiagen Inc., Valencia, CA), (2) CTAB method as described by Dietrich *et al*. (2002) with 1μl of 10mg/ml RNAse step, and (3) Extract-N-Amp Plant kit (Sigma-Aldrich, Saint Louis, MO) with the dry-down method. In addition to the parents, I also included a 'pseudo-F1' sample (mixing equal portions of Jefferson and *O. rufipogon* DNA), as well as previously genotyped recombinants using KASP to test the quality of detecting heterozygous material.

In terms of call rate, the Qiagen kit gave the most reliable results, with CTAB method a close second; Extract-N-Amp proved useless, regardless of concentration differences (Table 4-1). All three SNP detection assays were considered 'successful' by Sequenom standards because over 85% of the assays gave a call rate of over 90% using the Qiagen and CTAB-extracted DNA. Interestingly, the pseudo-F1 samples worked using CTAB-extracted DNA but not with Qiagen. Most of the errors were caused by the program calling one allele over the other, because of the uneven SNP peaks seen in pseudo-F1 samples. It is likely that the Jefferson and *O. rufipogon* DNA concentrations were not exactly equal, and the differences were further exaggerated after

the first round of PCR where the target loci were amplified. This highlighted the sensitivity of the platform for detecting heterozygous alleles.

Table 4-1. Call rate and concordance of the recurrent parent Jefferson and *O. rufipogon* DNA samples in iPLEX MassARRAY pilot study. The pilot study involved a 28-plex assay targeting the *yld2.1* target introgression and compared three different DNA extraction methods. Five of the assays were problematic.

Extraction Type	Sample.name	# Good Call	% success	Assay 2 (id)		Assay 12 (ud)		Assay 15 (id)		Assay 16 (wd)		Assay 21 (id)	
Qiagen	Jeff_rep1	28	100.0										
Qiagen	Jeff_rep2	27	96.4			1	no call						
Qiagen	Rufi_rep1	23	82.1	1	hetz call	1	Jeff call			1	Hetz call	1	no call
Qiagen	Rufi_rep2	24	85.7	1	hetz call	1	Jeff call			1	Hetz call		
CTAB	Jeff_rep1	27	96.4			1	no call						
CTAB	Jeff_rep2	27	96.4			1	no call						
CTAB	Rufi_rep1	25	89.3	1	hetz call	1	Jeff call			1	hetz call		
CTAB	Rufi_rep2	23	82.1	1	hetz call	1	Jeff call	1	no call	1	hetz call	1	no call
ExNAmp	Jeff_rep1	21	75.0	1	no call	1	no call					1	no call
ExNAmp	Jeff_rep2	23	82.1			1	no call						
ExNAmp	Rufi_rep1	11	39.3	1	no call	1	no call	1	hetz call	1	hetz call		
ExNAmp	Rufi_rep2	1	3.6	1	no call	1	no call	1	no call	1	no call	1	no call

One of the limitations to Sequenom technology for genotyping purposes was the sensitivity of the assay. Although this sensitivity is what makes the platform ideal for CNV (copy number variation) and methylation studies, there are less costly options that are easier to implement for genotypic selection. In order to get clear peaks indicating genotype calls, extra precautions were undertaken to keep the assays as contaminant free as possible. This included avoiding salts that are commonly found in the TE and TAE buffers used to suspend DNA in

common extraction methods, adding an RNAse step to standard chloroform extractions, as well as using deionized water stored in plastic (not glass) containers. Any of these contaminants would cause the peaks to shift and/or multiple peaks that would lower the quality of the genotyping calls.

Comparisons between KASP, iPLEX MassARRAY and 384 SNP chip

The ideal SNP marker assay is one that is tailored to the type of population under consideration, the number of SNPs to be genotyped per sample, and the type of platform available. The cost per data point using KASP is $0.25 cents regardless of how many SNPS are genotyped, while iPLEX MassARRAY prices fluctuate depending on the number of SNPs that are multiplexed in a single well. Each well in a 384 plate costs $2.60 (because each chip costs $1000 to purchase and run), so the more plexes that can be fit into a single assay, the cheaper is each data point to generate (albeit if too high, there are more SNP calling failures). For example, in a 25-plex and a fully occupied 384-sample set, there are a total of 9,600 data points, each costing roughly $0.10. If fewer than 10 SNPs are being assayed, then KASP is cheaper than MassARRAY. Therefore, if one needs to genotype 11 or more SNPs per sample and there are 384 samples or more, then the use of iPLEX MassARRAY is favorable over KASP. If genotyping only a small number of samples, such that one cannot use up all 384 wells on the chip, the price per data point price skyrockets. For both KASP and MassARRAY iPLEX, the main cost is born up-front, when ordering the primer sets (Table 4-2); if you order too many, you cannot recover the price of genotyping. The advantage of using fixed arrays, such as the 384 SNP chips, is that one can genotype one or two samples at a time at a fixed cost per data point

(but there is a minimum number of chips that must be purchased initially and one saves money by ordering a larger number of chips up-front).

Conventional SSR marker genotyping using polyacrylamide gel electrophoresis (PAGE) has comparable 'per-data point' cost as these SNP markers. In the McCouch lab, we estimate $0.04 per data point, if genotyping 9,600 data points using 44 PAGE gels to do both positive and negative selection (Diane Wang, McCouch lab, Cornell University, personal communication), but this is without taking into account the extensive labor time and cost. Using SNP instead of SSR markers, the same data can be generated faster and be analyzed in a much more efficient manner. Further, using the Sequenom assay there is less room for error because high quality DNA and robotic procedures practically ensure high quality data, while silver staining of PAGE gels requires both expert technique and a long training period to ensure correct calling of alleles. The opportunity to use SNP assays also avoids the health risks of exposure to polyacrylamide gels and silver staining reagents. Acrylamide (in its unpolymerized form) and silver nitrate are toxic and corrosive materials that require care and caution when handling PAGE gels. The use of Extract-N-Amp also eliminates exposure to chloroform when extracting DNA.

Allele calling procedures are another important point of comparison. The iPLEX MassARRAY uses an automatic genotype calling system based on molecular weight (Fig. 4-6), while KASP requires the use of the manual lasso drawing tool to designate clusters within the analysis software. It is up to the KASP user to designate how large each cluster should be, and where the cutoff point will be between calling an allele (colored dots) and no-call (black X symbols); this subjectivity is a cause of error when deciding whether to call an allele or designate a data point as a no-call. This manual step will generate different calls depending on the user.

Also, while KASP is used only for bi-allelic SNPs (due to the use of only two dyes), MassARRAY iPLEX can detect up to four alleles for each SNP. There are other notable differences in the system requirements for both methods: the RT-PCR machines required to detect KASP florescence readers are substantially cheaper and more readily available in many institutions compared to Sequenom's iPLEX MassARRAY (which is essentially an entire mass spectrometer tailored for analyzing DNA sequences). For our purposes, both systems were available on the Cornell campus and this feature was not an impediment in implementing either system.

Table 4-2. Comparison of primers for iPLEX MassARRAY and KASP marker assays. Number of reactions possible using SNP marker platforms with different primer purchasing formats. 3000 ul is an approximation based on the yield of primer in a single tube. Price was calculated in Spring 2011.

Primers in plate format (iPLEX MassARRAY):	Total volume (ul)	ul use per 384 samples	# of 384 plates that can be genotyped	# of samples that can be genotyped
Extension primer	87.5	1	87	33408
SNP specific primer	150	5	30	11520
Primers in tube format (KASP):	Volume diluted (ul)	ul used per 384 samples	# of 384 plates that can be genotyped	# of samples that can be genotyped
Allele-specific	3000	3.5	846	324864
Common reverse	3000	8.9	338	129792

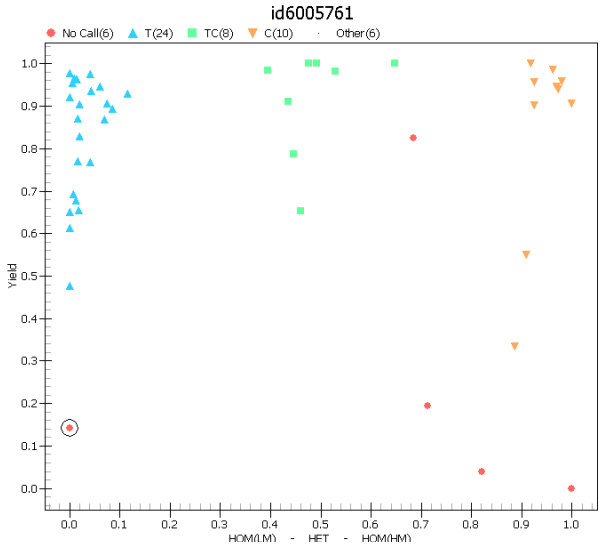


Fig. 4-6. SNP call cluster plots for MassARRAY iPLEX. X-axis indicates the molecular weight of extension primers and Y-axis shows the quantity of corresponding compound detected. The triangles (cyan) in the left of the graph correspond to extension primers with additional adenine indicating that the SNP allele was T (thymine), while upside-down triangles (orange) on the right correspond to the extension primer with additional guanine, which means the SNP allele was C (cytosine). The green squares between the two weights indicate a mix of the two extension primer weights meaning the sample was heterozygous at the SNP. Round (red) points are no-call due to low quantity of extension primers or ambiguous position on the x-axis.

SNP calling methodology and its effect on SNP marker quality

The quality of the dataset and subsequent SNP calling and selection strongly affect the success rate when designing SNP markers and assays. Of the 81 SNP markers tested in our trial of the Sequenom marker system, SNPs from the "intersection set" of the OryzaSNP dataset had a success rate of 90%, compared to a success rate of only ~50-56% for SNPs that were selected from the machine-learning only OryzaSNP dataset. Thus, stringent SNP calling during SNP

discovery is strongly recommended before setting out to make targeted SNP-assays for high throughput genotype detection.

Table 4-3. List of total SNP markers tested divided by different SNP pools. SNP names in bold succeeded when used for targeted-SNP assays. First three columns have names starting with id, which is the prefix for the "intersection set" consisting of 160,00 SNPs from the OryzaSNP dataset. SNPs in the fourth column (prefix ud) were also from the OryzaSNP dataset but were called using only one of two algorithms (~400,000 SNPs). The last column (wd) is a set of SNPs selected from BAC end sequencing of *O. rufipogon* generated by the OMAP project.

id2002162	id6003397	id1003932	**ud2000643**	**wd2001042**
id2002637	id6003502	id1004294	ud12001503	wd6001025
id2003133	id6003555	id1004698		**wd6002805**
id2003553	id6003812	id5003627	50% failure	wd5001098
id2003785	id6004051	id5004393		wd5001400
id2004031	id6004356	id5005425		**wd5002158**
id2004232	id6004650	id5006506		wd5002587
id2004457	id6004968	id5007105		wd10003936
id2004548	id6005428	**id5008654**		**wd12000212**
id2004718	id6005761	id9007407		
id2005118	**id6006095**	id9007821		44% failure
id2005498	id6006235	id10005037		
id2005558	id6006541	**id10005801**		
id2005731	id6006838	id10006386		
id2005915	id6007245	id10006761		
id2005978	id6007445	id12000380		
id2006124	id6007872	id12000824		
id2006241	id6008426	id12001409		
id2006450	id6008871	id12009772		
id2006621	**id6009213**	id12009955		
id2006798	id6009639			
id2007218	id6010081			
id2007468	id6010178			
id2007602	id6010766			
id2008351				
id2008815		10% failure		

Using SNPs to differentiate multiple stock sources of Jefferson

One of the discoveries that exemplifies the usefulness and robustness of SNP genotyping is the detection of 'foreign alleles' on chromosome 8 in the yield *IL*s. The 384 SNP chip indicated a 7.6Mb introgression (8.9 – 16.5 Mb), where the SNPs did not correspond to either of the parental controls that were simultaneously placed on the chip. The source of the discrepancy was traced back to different Jefferson stock being used in the backcrossing efforts in 2007, and to different introductions of Jefferson seed into the McCouch lab (indicated by different Rice Accession (RA) numbers; RA2747, RA163, and the foundation seed RA4829). Using the available SNP markers, we were able to track which stock of Jefferson seed should accompany a given project in order to avoid contamination where there are multiple seed sources (Table 4-4). As of August 2011, there were 11 different accessions of Jefferson available in the lab (Table 4-5), and the use of these SNP markers was useful to figure out which Jefferson seed stock(s) were used in this study. Maintaining the integrity of the lines under study proved to be an essential part of developing NILs, particularly where genotyping was being used as the basis for selection every generation. The use of genome-wide SNP markers was effective in diagnosing non-parental introgressions that could potentially obscure results and lead to false interpretations of the experiment.

Table 4-4. Diagnostic SNPs used to identify the original parental Jefferson line. The 384 SNP chip revealed large non-parental regions on chromosome 4 (~10Mb) and 8 (7.8Mb) and small introgressions on 6 and 7 that differentiate the original line from other sources of Jefferson in the breeding community.

						Non-parental Jefferson	**True Jefferson**	**O. rufipogon (105491)**
sample id\snp id	Chr	TIGR.MSU 6 (bp)	TIGR (Mb)	A allele	B allele	4303623023_R007_C007	4543053077_R008_C011	4313215088_R006_C005
id4003791	4	12215278	12.215	T	C	BB	AA	BB
id4006172	4	20209086	20.209	T	C	BB	AA	BB
id4007645	4	22941779	22.942	A	C	BB	AA	BB
id6015530	6	27381493	27.381	A	G	AA	BB	BB
id7004205	7	23509537	23.510	T	C	BB	AA	BB
id8000575	8	2019555	2.020	T	C	AA	BB	BB
id8001641	8	5310591	5.311	T	C	AA	BB	AA
id8002632	8	8507411	8.507	A	G	BB	AA	BB
id8003014	8	9370078	9.370	T	C	BB	AA	BB
id8003626	8	11882169	11.882	A	G	AA	BB	AA
id8003808	8	13778779	13.779	T	G	BB	AA	BB
id8003881	8	14188295	14.188	T	C	AA	BB	AA
id8004111	8	15222443	15.222	A	G	AA	BB	AA

Table 4-5. Eleven different accessions of Jefferson available as of August 2011. Screenshot from <http://rice.generationcp.org/germplasm/>

germplasm search for - JEFFERSON - (11 records found)

#	Name	Method Of Creation	Germplasm Location	Studies
1	IRTP 21321, JEFFERSON	Entry in international nursery	International Rice Testing Program, IRRI,Philippines	View: 3 Studies
2	IRIS 1-77223, JEFFERSON	Cultivar release	Texas,United States	
3	IRIS 146-25067, JEFFERSON	Import	ITQB-INSTITUTE OF TECHNOLOGY, BIOLOGY AND CHEM	
4	IRIS 149-26270, JEFFERSON	Import	IRRI-International Rice Research Institute, Los Baños,Philippi	
5	IRIS 251-27214, JEFFERSON	Unknown selection method	BANGLADESH RICE RESEARCH INSTITUTE;GAZIPUR,Bar	
6	IRIS 251-27241, JEFFERSON	Import	IRRI-International Rice Research Institute, Los Baños,Philippi	
7	IRIS 251-27357, JEFFERSON	Import	IRRI-International Rice Research Institute, Los Baños,Philippi	
8	IRIS 251-27465, JEFFERSON	Import	IRRI-International Rice Research Institute, Los Baños,Philippi	
9	WAB 16408, JEFFERSON	Accession into genebank	AfricaRice-Africa Rice Centre, Bouaké,Côte d'Ivoire	
10	PI 593892, JEFFERSON	Import	National Small Grain Collection USDA,ARS,United States	
11	IRIS 307-45199	Unknown	United States	

Homogeneity of IL families

We used targeted SNP markers and whole genome SNP assays to measure the level of homozygosity of ILs that had undergone field-based seed propagation for four years (2007-2010) and to confirm the genetic identity and composition of the lines. Using SSR data from 2006 (159 markers), we identified the size and location of both target and background introgressions in the BC_2F_3 lines and confirmed this information using a 1536-SNP genome-wide assay in 2008 on the BC_3F_4 generation. In 2009, we developed a 384 SNP genome-wide assay tailored for *tropical japonica* x *O. rufipogon* populations (Thomson *et al*., 2011) and used it to derive ILs with fewer background introgressions. MassARRAY assays were designed to target only regions of introgression in the two top-performing ILs with 24 markers positioned along the *yld2.1* target QTL (~17 Mb) and 23 markers in background introgressions in IL 43_2-1. Similarly, 26 markers distributed along the *yld6.1* QTL (~14 Mb) were used in the development of the second QTL target. Both of these lines are being prepared for submission to the Genetic Stock Center for *Oryza* (GS*OR*).

Results confirmed that by the BC_3F_8 generation, yld2_A (IL 43_1-2) was highly homogeneous and had been fixed for *O. rufipogon* alleles in the *yld2.1* region as well as in background introgressions on chromosomes 5, 9 and 12 (see Chapter 2 for introgression information). On the other hand, in the BC_3F_8 generation, three yld6_A (IL 219_2-9) out of 46 samples representing eight different headrows were still segregating across the target region on chromosome 6. As a result, only lines that carried a homozygous *O. rufipogon* introgression were selected and bulked for submission to the GS*OR*. Two of the headrows 2169 and 2173 contained samples with segregating individuals (one sample from 2169 and two from 2173), and will be backcrossed for use in fine-mapping.

References

Boualaphanh, C., Daygon, V. D., Calingacion, M. N., Sanitchon, J., Jothityangkoon, D., Mumm, R., et al. (2011). Use of new generation single nucleotide polymorphism genotyping for rapid development of near-isogenic lines in rice. *Crop Science, 51*

Buetow, K., Edmonson, M., MacDonald, R., Clifford, R., Yip, P., Kelley, J., et al. (2001). High-throughput development and characterization of a genome-wide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 581-584.

Liu, S., Chen, H. D., Makarevitch, I., Shirmer, R., Emrich, S. J., Dietrich, C. R., et al. (2010). High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. *Genetics, 184*(1), 19-26.

Milne, I., Shaw, P., Stephen, G., Bayer, M., Cardle, L., Thomas, W. T. B., et al. (2010). Flapjack-graphical genotype visualization. *Bioinformatics, 26*(24), 3133-3134.

Oeth, P., Beaulieu, M., Park, C., Kosman, D., del Mistro, G., van den Boom, D., et al. (2005). iPLEX™ assay: Increased plexing efficiency and flexibility for MassARRAY system through single base primer extension with mass-modified terminators. *Sequenom Application Note,*

Thomson, M., Zhao, K., Wright, M., McNally, K., Rey, J., Tung, C., et al. (2011). High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding,* , 12.

Zhao, K., Tung, C., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun, 2*, 467.

Future opportunities

Fine-mapping

Our field data allowed us to identify the highest yielding ILs across all environments and genotypic data were used to select for lines retaining *O. rufipogon* alleles across the six target regions while retaining the fewest background introgressions. We are currently fine mapping the most prominent yield-enhancing QTLs, *yld2.1* and *yld6.1*, to identify underlying genes and to resolve the genetic interactions between the target introgression(s) and the various background introgressions. SILs (sub-introgression lines) will be generated to dissect the megabase long target introgressions while keeping track of the numerous background introgressions. Preliminary field data from the 2009 to 2011 seasons show that further backcrossing of our best BC4 lines to obtain derivatives with fewer background introgressions performed equally or slightly less well compared to the BC3 ILs. We aim to identify the trait-enhancing alleles coming from the low-yielding wild rice relative that confer transgressive variation for yield and yield components and use those alleles for the improvement of U.S. rice cultivars.

Although 50 introgression lines and 20 control lines were assessed, only nine introgression lines were significantly higher yielding than Jefferson across locations and years between 2007-2008. Eight of these nine lines targeted only two yield QTLs, *yld2.1* and *yld6.1*. To expedite fine-mapping of these two, most productive QTL targets, as well as to facilitate marker assisted backcrossing of the *O. rufipogon* introgressions into other elite materials, we have developed high resolution SNP markers and SNP multiplex assays that are freely available.

Comparing ILs and control sib-lines

Each of the six QTL targets had accompanying 'control' lines that were lacking the *O. rufipogon* allele at the target marker. The control lines were derived from the same families as the 50 original lines chosen for field trials. The purpose of these control lines was to test whether the target introgression was responsible for altering the yield of the ILs, on the assumption that the control lines had the same background introgression(s) as the 50 chosen ILs. New 384 SNP (OPA 6.1) results are pending for a series of control lines, and we expect these data to help explain the different yield performance observed among families targeting the same yield QTL.

Other yield component traits

When selecting lines to submit to the For GS*OR* as genetic stocks (discussed in Chapter 2), we submitted the most well-rounded lines, selected based on field performance. Each of the six QTL targets were represented, and in each case, we selected the line that recorded the highest yield across all four environments over two years. Of ths\e 50 introgression lines and 20 control lines derived from the *O. rufipogon* x Jefferson backcross population, we identified several lines that were significantly different from the recurrent parent and may have interesting breeding applications (Appendix III: ILs showing transgressive variation in yield component traits).

Line 16_1-6 was the only line with significantly less chalk than Jefferson in multi-location trials in 2007 and 2008, and this line also has the best sheath blight resistance rating according to the 2009 disease trials. This was surprising because 16_1-6 was a control line for *yld3.2* and *yld9.1* (i.e., it lacks the *O. rufipogon* introgression(s) at the target QTL) and was originally to be used as a comparison to other lines that contained the target introgression(s). As soon as it is available, the new 384 SNP OPA data for this line will be used to determine the number and size

of introgressions across the genome. This will provide the foundation for identifying the region(s) associated with low chalk in this material.