DISSERTATION

UNDERSTANDING BROAD-SPECTRUM DISEASE RESISTANCE IN RICE: PROMPTING A GENOME-WIDE UPRISING

Submitted by

Bradley William Tonnessen

Department of Bioagricultural Sciences and Pest Management

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2017

Doctoral Committee:

Advisor: Jan E. Leach

Asa Ben-Hur
Courtney E. Jahn
Cristiana T. Argueso
Daniel R. Bush

ABSTRACT

UNDERSTANDING BROAD-SPECTRUM DISEASE RESISTANCE IN RICE:

PROMPTING A GENOME-WIDE UPRISING

Rice is the main staple food crop of the world, and thus, the detriments caused by rice diseases are a threat to international food security. The emergence of new virulent strains of pathogens can significantly reduce yields, and there are continual efforts to develop more resistant rice cultivars. Utilization of single R-genes is effective, but has proven inadequate due to rapid pathogen evolution. Thus, there is a need for breeding multigenic, broad-spectrum disease resistance in new varieties.

This study aims to understand the aspects of basal resistance and its contribution to tolerance to multiple, diverse pathogens. Phenylalanine ammonia-lyase is a key enzyme in phenylpropanoid metabolism, which contributes to the basal defense response (DR). In this project, the DR gene, *OsPAL4*, which colocalizes with a disease resistance Quantitative Trait Loci (QTL), was shown to contribute to resistance to three important rice diseases, rice blast, bacterial blight, and sheath blight, in experiments using an *ospal4* mutant. The functional element of resistance QTL haplotypes of DR genes such as *OsPAL4* are largely unknown, and this work searched for sequence patterns in the promoters of DR genes to discern a regulatory mechanism specific to DR. Multiple *cis*-regulatory Modules (CRMs), or groups of DR-related sequence motifs were identified in promoters of DR genes. These CRMs harbor structural organizations of *cis*-elements known to be involved in the DR, and also motifs involved in a putative epigenetic regulatory mechanism. Polymorhpisms in CRMs are found in resistant

relative to susceptible QTL haplotypes in DR gene promoters. These CRMs are sequence patterns found across DR gene promoters. Thus, we hypothesize that DR-associated CRM can be used as breeding markers to select loci on a genome scale that encode traits supporting broad spectrum basal resistance to important rice diseases.

ACKNOWLEDGEMENTS

First, I would like to thank the founders and supporters of the Program in Molecular Plant Biology (PMPB) for awarding me a fellowship, which allowed me to start my PhD program here at CSU. That opportunity was everything I needed to get started on a path that was so new to me. I'd also like to warmly thank my committee members who were there in those first years helping get my feet on the ground. Dan Bush, for allowing me to do a rotation in your lab, and always greeting me with a warm smile. Courtney Jahn, for showing me what it means to do great science and stay grounded while doing it. Cris Argueso, for being a great lab bench mentor that first year, and giving me that healthy dose of skepticism. I'd also like to thank Anireddy Reddy, for allowing me the opportunity to learn from him and his lab for a rotation. Lastly, I'd like to thank Asa Ben-Hur for humoring my bioinformatic woes for weeks on end, but also for providing much needed help and wisdom.

I also cannot forget all the staff and faculty at the International Rice Research Institute. Spending many months in the Philippines was an enlightening experience, and I hope to always keep that close to my heart (karaoke, too). Especially, Ramil Mauleon and Nickolai Alexandrov, who were both great mentors to me during my computational rookie years.

As for the people in BSPM, there are too many to thank. Janet Dill has saved my anxiety from blowing out of proportion many times. All those graduate students with me in the basement office. Of our lab, Paul Tanger was a senior student I could look up to. Jillian Lang helped me get my feet wet in the Leach lab. Paul Langlois, Rene Coral, and others were great support. Finally, Jan Leach for taking me on, and allowing me to learn what I had to learn, but making sure I was looking in the right direction.

TABLE OF CONTENTS

Chapter 1: Introduction

Agriculture today and the threat of diseases in rice

The farming and distribution of staple foods allows us to sustain the world's population. These crops mainly consist of maize, wheat, and rice, occupying 40 percent of global farmland (Oerke & Dehne, 2004). As globalization continues, the types of agricultural crops grown across the world, including the staple grains, are heading towards homogenization (Khoury et al., 2014). The increasing similarity of crops worldwide brings a great risk to our global food security, especially due to the spread of plant diseases. On the field level, close proximity of genetically similar hosts facilitates plant-to-plant pathogen transmission (Plantegenest et al., 2007). On the regional and global scale, crop pathogens, notably fungal, are spreading rapidly into new host ranges, mostly due to anthropogenic introductions (Anderson et al., 2004; Bebber et al., 2014). This ensures that all crops will face a continuous offensive of newly emergent, virulent pathogens in the future.

Rice feeds 50 to 60 percent of all people, and stands as the most important staple food for humanity (Sharma et al., 2012; White, 1994). There exist three major diseases of rice: the bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* (bacterial blight), and two fungal pathogens *Magnaporthe oryzae* (rice blast) and *Rhizoctonia solani* (sheath blight). The potential yield of rice can be decimated by 70 percent due to bacterial blight caused by *X. oryzae* pv. *oryzae* (Goto, 1992; K. Reddy, 1979; Ou, 1985). Rice blast can reduce yields from 20 to 30 percent, and sheath blight similarly can reduce yield by 25 percent (Banniza & Holderness, 2001; Jia et al., 2013; Skamnioti & Gurr, 2009). The severity of disease conditions can be attributed to the farming methods used. In many instances, given the right climatic conditions, rice farmers

implement as many as three croppings per year of only rice, which can increase disease prevalence due to the continuous abundance of hosts (Tilman et al., 2002). One solution for staving off fungal infection and insects is through application of fungicides and pesticides, but the detriments often outweigh the benefits. These applications can negatively affect the agricultural ecosystem, by eliminating beneficial predators and competitors of pathogens and pests (Wilson & Tisdell, 2001). Secondly, the effectiveness of these chemical treatments can be overcome by rapid evolution of fungal pathogens and pests due to a strong selection on the target molecular mechanism (Ma & Michailides, 2005; Su et al., 2014). One example involves the resistance of *M. oryzae* to a rice fungicide, carpropamid, which occurred from a single, non-synonymous point mutation in the scytalone dehydrase gene that disabled the inhibitory action of carpropamid on the encoded enzyme (Sawada et al., 2004). Given the continuous selection pressure on pathogens to overcome disease resistance, the threats to crop production are daunting for farmers and stakeholders, and there has been a continuous call for new solutions, either through genetic resistance in the seed or changes to farming and cultural methods (Gilligan, 2008).

One method for the reduction of disease pressures on rice involves planting multiple, genetically dissimilar varieties to diffuse selective pressures and mitigate pathogen spread. Intercropping to exploit the benefits of genetic diversity across the rice germplasm has been studied in multiple scenarios. Due to the cool and wet climate, farms in Yunnan Province, China experience frequent outbreaks of rice blast disease. Many farmers here grow a glutinous rice for the high market value, but it is highly susceptible to this fungal pathogen. When these glutinous varieties were grown in rows interspersed with resistant, non-glutinous hybrid rice, the susceptible varieties showed as much as 89% increase in yield and a 94% reduction in blast

disease incidence compared to growth in monoculture (Zhu et al., 2000). A similar result was found in Japan, with the interplanting of multilines to reduce rice blast disease (Koizumi & Ashizawa, 2004). Multilines are defined as a group of Near-Isogenic Lines (NILs) which are phenotypically similar, yet contain different compositions of resistance genes among them. Planting these multilines greatly reduced the severity of blast disease in each component pure line (Koizumi & Ashizawa, 2004). The use of varietal mixtures is advantageous against fungal pathogens due to the low gradient of spore dispersal by wind, but in the bacterial blight pathosystem, splashing rain droplets aid dissemination. This results in increased autoinfection, or the re-inoculation of a host by bacteria already present on the plant, resulting in a high dispersal gradient (Mundt & Leonard, 1986). Intercropping two rice varieties with distinct *X. oryzae* pv. *oryzae* resistance genotypes was effective in certain allelic combinations, though the resulting resistance phenotype was similar to planting single varieties containing the two resistance loci (Ahmed et al., 1997). Since dispersal of *X. oryzae* pv. *oryzae* is concentrated to the plants directly surrounding the infected host, a more complex mixture of varieties could increase the effectiveness of this method (Ahmed et al., 1997). There are many reasons why including genetic diversity in the field reduces disease. Planting rice in genetic mixtures can prevent a pathogen from evolving to specialize against a particular variety, reduce host-to-host dispersal, and result in overall reduction of pathogen populations on the canopy and within the soil (Mundt, 2002). This method of rice farming has promising potential, but there are problems with adoption, since machine harvests would be impaired by large phenotypic variation in plant structure and differences in maturation time. The use of multilines as described by Koizumi and Ashizawa, 2004, circumvents this issue due to the selection of phenotypic homogeneity of NIL populations. It is important to also note that issues such as plant height difference are only

difficult if the deviation is extreme, and the timing of harvest among anthesis variants can be lenient enough to accommodate for any "late bloomers" (Mundt, 2002). There lies great promise in exploiting genetic diversity to improve the adaptability of our agricultural system to new diseases. There is also a great need to understand how rice varieties harbor resistance, specifically at the molecular level of plant-microbe interactions. Discerning the genetic factors that improve rice resistance to a broad spectrum of pathogens will inform breeders and farmers of what genotypic varieties to use as they creatively find new ways to sustain crop health.

The molecular plant-microbe interaction

The mode of infection for the hemibiotrophic, vascular pathogenic bacteria, *X. oryzae* pv. *oryzae*, begins with entrance into the rice leaf through hydathodes at the leaf tip and margin or through wounds (Ou, 1985). Moving into the hydathode can be passive as water exudes and retracts over the day/night cycle, or occurs actively by swimming into the intercellular space (Nino-liu et al., 2006). The bacteria multiply in the epitheme and eventually infect the xylem vessels. Conversely, the rice blast fungus, *M. oryzae*, can directly penetrate through rice tissue surfaces of the leaf, panicle, nodes, and stems (Wilson & Talbot, 2009). Spores (conidia) land on the rice plant and initiate germ tube growth. The end of the germ tube differentiates into an appressorium through cues such as contacting a hydrophobic surface or cutin monomers (Talbot, 2003). The appressorium generates turgor and penetrates the cell surface and through this site the hyphae grow throughout the intercellular space. Ultimately, the hyphae modify behavior into more invasive and necrotrophic intracellular growth. Symptoms of disease can occur during either stage (Wilson & Talbot, 2009). Sheath blight disease is caused by *R. solani* and this fungus begins the infection progress within the soil after initial germination (Keijer, 1996). The

mycelia grow towards the rice plant and up along the surface of epidermal cells, eventually forming appressorial structures. Upon penetration into the intracellular space, the hyphae secrete enzymes including proteinases that result in host tissue necrosis (Gvozdeva et al., 2006).

Plant recognition of virulent pathogens starts with cell-to-cell contact, and the molecular reactions that ensue are both complex and comprehensive (Azizi et al., 2016; Corwin & Kliebenstein, 2017; Cui et al., 2015; Jones & Dangl, 2006; Katagiri & Tsuda, 2010; Nejat & Mantri, 2017; Nishimura & Dangl, 2010). During infection, the plant cell recognizes molecular compounds that originate from the pathogen, known as Pathogen Associated Molecular Patterns (PAMPs); these PAMPs are highly conserved across fungal and bacterial genera, respectively (Corwin & Kliebenstein, 2017; Jones & Dangl, 2006). PAMPs are recognized by plant cell membrane-localized receptor-like kinases (RLKs) (Boller & Felix, 2009; Monaghan & Zipfel, 2012). These proteins are known as Pattern Recognition Receptors (PRRs) and contain an extracellular Leucine-Rich Repeat (LRR) domain for reception of the PAMP, a transmembrane domain, and an intracellular kinase domain (Couto & Zipfel, 2016). The recognition of a PAMP is the first step in the process known as PAMP-Triggered Immunity (PTI), and the subsequent course of events that prevent pathogen infection is known as the plant Defense Response (DR) (Corwin & Kliebenstein, 2017; Jones & Dangl, 2006). The role of PRRs is to initiate a signaling cascade of mitogen-activated protein kinases (MAPKs) (Couto & Zipfel, 2016; Pitzschke et al., 2009). After this initial signaling step, the plant cell begins an elaborate modulation of cellular processes contributing resistance.

The plant DR resulting from PTI is complex, but can be measured by changes in pH and/or increased oxidative compounds. For example, the bacterial pathogen *Pseudomonas syringae* contains a conserved 22-amino acid sequence, flg22, within its flagellin protein. One of

the PRRs, known as FLS2, binds to flg22, and functional homologs reside in Arabidopsis, rice and tomato (Boller & Felix, 2009; Gómez-Gómez & Boller, 2000; Takai et al., 2008). The FLS2-flg22 interaction elicits a strong alkalinization and an oxidative burst in suspension-cultured cells of multiple crop species (Felix et al., 1999). Another PAMP, the bacterial elongation factor, EF-Tu, binds and activates a different PRR, the transmembrane protein EFR (Zipfel et al., 2006). This causes a response in Arabidopsis with oxidative bursts and ethylene production (Kunze, 2004). The oligosaccharide chitin, a component of fungal cell walls, also functions as a PAMP (Felix et al., 1993). Chitin is recognized by the rice Chitin Elicitor Binding Protein, CEBiP, which resides in the plant cell membrane, but lacks an intracellular kinase domain (Kaku et al., 2006). The receptor CEBiP requires the rice protein kinase activity of CERK1 to activate the chitin-induced MAPK reactions, and they form a heterodimer when cells are treated with chitin oligosaccharides (Miya et al., 2007; Shimizu et al., 2010). This interaction elicits a DR witnessed as oxidative bursts in many crops including rice (Felix et al., 1999; Felix et al., 1993; Kaku et al., 2006). Each of these described PTI-inducing interactions are due to non-species-specific PAMP-PRR recognition. Thus, the downstream mechanism of PTI is a broad-spectrum DR reaction, meaning that pathogens of many species, either fungal or bacterial, can trigger the response (Dodds & Rathjen, 2010). This is known as basal resistance, and has been realized by breeders in the past as incomplete, horizontal, or field resistance, due to the tolerance rather than total immunity that emerges in the field relative to fully susceptible crop varieties (Kiyosawa, 1982).

The PTI reaction can be circumvented by pathogen-derived secreted proteins known as effectors that either suppress the DR or enhance susceptibility (Jones & Dangl, 2006; Koeck et al., 2011; Liu et al., 2013). One case of an effector inhibiting PTI is during the biotrophic stage

of *M. oryzae*. As the hyphae grow through the host intercellular space, the fungus secretes an effector, Secreted LysM protein 1 (Slp1), that binds to chitin oligosaccharides. This activity is in direct competition with the chitin-binding PRR, CEBiP. The result is chitin sequestration by Slp1, thereby lessening the severity of DR caused by CEBiP/CERK1-mediated PTI (Mentlak et al., 2012). Many pathogen effectors are recognized directly by the plant cell, or indirectly by monitoring effector activity, which frequently, but not always, initiates a Hypersensitive Response (HR) or rapid localized cell death, to inhibit infection spread (Nejat & Mantri, 2017). The effector triggered responses are known as Effector-Triggered Immunity (ETI). Host genes that encode effector-sensing proteins are known as Resistance (R-) genes; many R genes encode nucleotide-binding leucine-rich repeat (NBS-LRR) proteins (Cui et al., 2015). One example of an *M. oryzae* effector triggering ETI involves the secreted protein, AvrPiz-t, which inhibits DR and prevents induced cell death in tobacco (Li et al., 2009). The rice R-gene, *Piz-t*, produces a NBS-LRR that recognizes AvrPiz-t, with the aid of the rice E3 ligase APIP10, producing an HR phenotype (Park et al., 2016). A number of effectors secreted from the bacterial pathogen *X. oryzae* pv. *oryzae* encode Transcription Activator-Like Effectors (TALEs) that are injected into host cells via the type III secretion system, and modulate transcription of various host "susceptibility genes" to promote infection (Hutin et al., 2015). An interesting form of ETI plays out between the *X. oryzae* pv. *oryzae* TALE, AvrXa27, and the rice R-gene *Xa27*. Transcription of *Xa27* is activated by the binding of AvrXa27 with the R-gene's native promoter, which induces cell death, resulting in resistance (Gu et al., 2005). These examples of a single R-gene interacting (directly or indirectly) with a single effector protein are known as gene-for-gene resistance (Brown & Tellier, 2011; Flor, 1942). Contrary to disease tolerance, the strong response from ETI is more qualitative than quantitative, classically named as vertical or

complete resistance due to the strong inhibition of pathogen proliferation in the field. Since the nature of ETI is a gene-for-gene interaction, it is also usually species- and strain-specific.

The downstream responses of ETI or PTI both involve modulation of the expression of functionally diverse "DR genes," which are genes that enhance resistance or mediate other cellular processes to avoid fitness costs during disease (Kou & Wang, 2010). The processes to which DR genes contribute is expansive. For instance, DR genes modulate cellular secondary metabolism while also subsequently reacting to those modulations. The production of Reactive Oxygen Species (ROS), specifically superoxide and hydrogen peroxide, are increased in response to pathogen infection in what is called an oxidative burst (Lamb & Dixon, 1997; Torres et al., 2006). This can aid in strengthening the cell wall and activating DR genes (Bradley et al., 1992; Levine et al., 1994). Other secondary metabolites include the phenylalanine-derived phytoalexins produced from the phenylpropanoid pathway, many of which have antimicrobial properties (Piasecka et al., 2015). Hormonal flux also plays a major role in DR (Shigenaga & Argueso, 2016). Auxin, in the form of Indole-3-Acetic Acid (IAA), secreted by *X. oryzae* pv. o*ryzae*, and *M. oryzae*, induces the production of expansins in the host, weakening the cell wall (Ding et al., 2008; Fu et al., 2011). Rice DR genes, *GH3-1*, *GH3-2*, *GH3-8* and *GH3-13* all are able to prevent this by inactivating IAA (Fu et al., 2011). Salicylic Acid (SA) functions positively on rice defense. Treatment with SA activates the DR gene transcription factor, *OsWRKY45*. When *OsWRKY45* is overexpressed, the DR is enhanced against *M. oryzae* (Shimono et al., 2007). Jasmonic Acid (JA) activates the expression of DR genes, and plays a primary role in defense against necrotrophic pathogens (Browse, 2009; Yang et al., 2012). Another plant hormone, ethylene, activates DR genes in response to fungal pathogens, such as

M. oryzae, but has a negative effect on resistance to the bacteria, *X. oryzae* pv. *oryzae* (Iwai et al., 2006; Shen et al., 2011; Singh et al., 2004).

Many DR genes are transcription factors themselves that mediate plant resistance. The WRKY family of transcription factors have a large contribution. The suppression of *OsWRKY13* results in susceptibility to both *X. oryzae* pv. *oryzae* and *M. oryzae* (Cheng et al., 2015; Qiu et al., 2009). Along with WRKY45 (*OsWRKY45-1*), *OsWRKY45-2* is also a DR gene involved in hormonal flux, but when overexpressed, the gene conversely increases resistance to *X. oryzae* pv. *oryzae* and enhances susceptibility to *M. oryzae* (Tao et al., 2009). Another WRKY factor, *OsWRKY30*, is activated by both JA and SA, and involved in resistance to *R. solani* and *M. oryzae* (Peng et al., 2012). A CCCH-type zinc finger transcription factor, *OsC3H12*, enhances resistance to *X. oryzae* pv. *oryzae* when expressed, as well as increases JA and expression of JA-related DR genes (Deng et al., 2012). A member of the NAC transcription factor family, *OsNAC111*, is upregulated during *M. oryzae* infection, and when overexpressed, increases resistance to the rice blast fungus (Yokotani et al., 2014). In addition, OsNAC111 binds to and activates the promoters of DR genes chitinase and β-1,3-glucanase (Yokotani et al., 2014). DR genes are also involved in many other fitness-related processes such as carbon metabolism for mediating energy usage to maintain homeostasis (Hulsmans et al., 2016). The diversity of DR gene function illustrates the adaptability and complexity of plant defense, and understanding this is critical for future breeding efforts.

A multitude of DR genes are regulated by both ETI and PTI (Bozsó et al., 2016; Navarro et al., 2004; Tao et al., 2003). The cellular processes involved in inter- and intra-cellular signaling, along with producing anti-fungal and anti-microbial compounds are pertinent to any immune reaction. Thus, DR genes are important when breeding for broad spectrum and durable

basal resistance from both fronts. The conventional approach to developing resistant rice varieties mainly consists of selecting plants with strong resistance owing to single R-genes. The fault in this method lies in the gene-for-gene mechanism, which a pathogen easily overcomes over a few generations of rapid effector protein evolution (Ballini et al., 2008). Contrary to the functional restrictions of ETI, the PAMPs that bind to PRRs are highly conserved, non-species-specific, and usually critical for pathogen survival. Thus, selective pressures will not favor the mutation of PAMPs (Dodds & Rathjen, 2010; Kou & Wang, 2010). Additionally, breeding for R-gene-mediated resistance overlooks allele variants of strong basal resistance DR genes, thus these genotypes can be lost during the breeding process. This generational loss of quantitative resistance due to a focus on R-gene resistance is known as the Vertifolia effect (Van der Plank, 1963). Instead, the focus of breeding should expand to encompass strong quantitative resistance, which includes the downstream genes. A more sustainable and holistic approach to breeding for broad-spectrum and durable resistance focuses not only on both PTI and ETI, but also importantly, on the underlying DR genes and their regulation.

DR genes within Quantitative Trait Loci for durable resistance

The short-lived resistance from single R-genes is historically well-known and the search for more durable forms of disease resistance is a long-lived and venerable proposal (Johnson, 1984). Durable resistance depends on the environmental context, but generally signifies that a cultivar harbors resistance to multiple strains of a pathogen species, and lasts for an extended period. Disease resistance Quantitative Trait Loci (QTL) are genomic regions associated with a resistant phenotype variant. These QTL are often partial contributors to basal resistance, and DR genes reside within these loci (Kou & Wang, 2012; Ramalingam et al., 2003). A QTL for partial

resistance to stripe rust in wheat, *Lr34*, encodes an ABC transporter that was effective for fifty years (Dyck, 1966; Krattinger et al., 2009). A durable partial resistance locus, *Rpg1*, has persisted in barley against the stem rust pathogen since 1942 (Sun & Steffenson, 2005). The barley *MLO* locus is a negative regulator of DR, and the mutant allele has exhibited long-lasting resistance to powdery mildew (Piffanelli, 2002). Selecting for these partial resistance loci can produce a longer lasting phenotype than an R-gene mediated hypersensitive response, possibly due to the lower selective pressure that QTL DR genes have on the pathogen. Though, these loci are all single genes, and thus the event of resistance breakdown in the future is possible. To strengthen QTL-mediated basal resistance, many studies have attempted stacking QTL containing DR loci into a single variety, known as QTL pyramiding (Boyd et al., 2013; Kou & Wang, 2012). Wheat varieties were crossed that harbored QTL for "slow rusting", a resistance phenotype attributed to various races of leaf and stripe rusts, and the current results of the breeding program show durable resistance. This is due to the pyramiding of four to five different DR gene loci, and there exists a positive relationship between the level of resistance and the number of DR genes present in different lines (Singh et al., 2011). Strong basal resistance in winter wheat to *Septoria tritici* blotch was achieved by incorporating DR gene QTL from four separate chromosomes. The winter wheat lines that contained all four QTL displayed the lowest mean disease scores (Vagndorf et al., 2017). In rice, basal resistance to *M. oryzae* was attributed to five QTL, with a candidate DR gene within each locus (Liu et al., 2004). Incorporating each of the effective QTL alleles incrementally decreased the diseased leaf area across multiple environments and locations, and has retained durability (Liu et al., 2004; Manosalva et al., 2009). The success of QTL pyramiding reinforces our understanding of the polygenic nature of a robust DR, and the importance of considering many loci when breeding for durable resistance. In these

studies, it is nearly impossible to consider all possible biotic stresses that can harm a crop, so single disease threats are the focus. The results of many different experiments in identifying resistance QTL exist, and subsequent analysis of overlapping loci can shed light on DR genes of broad spectrum resistance.

Can broad spectrum resistance QTL be identified?

There are several resistance-associated QTL analyses for the rice diseases caused by *M. oryzae*, *X. oryzae* pv. *oryzae* and *R. solani*. It is probable that overlapping QTL for resistance to these three pathogens may contain DR genes that impart broad-spectrum resistance. For this review, the publicly available data for QTL against these three diseases was obtained from curated databases and recently published results (Table S1-1: Supplemental). Each QTL was then mapped to the twelve chromosomes of the rice reference genome Nipponbare (Figure 1-1) (rice.plantbiology.msu.edu). An enrichment analysis to find QTL-dense regions within each chromosome was conducted by comparing QTL density in ~3 Mb sections to the entire respective chromosome (Figure 1-2, Table S1-2: Supplemental). The results were insignificant, with no P-value less than the Benjamini-Hochberg critical value. Even though no genomic regions were identified as "QTL-enriched," some segments such as the long arm of chromosome 1 and chromosomal segments of 3, 4, 5, and 11 show a relatively high density in QTL (Figure 1). There may be functional DR genes within these regions. Indeed, previous studies have mapped large effect R-genes and DR genes to regions where multiple resistance QTL overlap (Kou & Wang, 2012; Ramalingam et al., 2003). The enrichment analysis is likely impacted by the variability of QTL sizes, some spanning the majority of a chromosome. This is due to the technological restraints of QTL analyses over the years, with differences in marker density and

sequencing availability. This meta-analysis illustrates the difficulty of understanding the contribution of polygenic DR through QTL mapping. On one hand, marker density can be very low, thus encompassing many DR genes along with non-DR loci. Oppositely, a high marker density, such as the QTL peaks seen in Figure 1-1, are found using sophisticated methods such as Genome-Wide Association Studies (GWAS). The improvement in resolution will only find the most significant players in DR and overlook the small-effect DR genes involved in quantitative resistance. Another issue in identifying broad-spectrum DR genes involves the inadequate assumptions made about what constitutes a DR gene, because we lack a comprehensive understanding of all genes involved in basal resistance. In a GWAS study for *M. oryzae* resistance, 16 strains of the pathogen were tested on a total of 366 varieties from the indica subpopulation (Wang et al., 2014). Using ~800,000 SNP variants, a total of thirty loci were identified as significant. Four of these loci were mapped to known R-genes, *Pia, Pik,* and *Pif* (x2) (Wang et al., 2014). The candidate genes within these loci were attributed to basal immunity as well as R-gene-mediated resistance, but the significant loci were few. The precision of GWAS is helpful in finding loci with a large-effect, but the phenotypic contributions of these genes mask any small-effect DR genes, even though they collectively play a large role in basal resistance.

Transcriptome analysis of differential gene expression during infection has contributed to understanding the wide scope of genes involved in the DR. In a study measuring the transcriptome of the response to *M. oryzae* infection where the varieties tested were resistant or susceptible with a similar genetic background, a total of 755 differentially expressed genes were identified relative to the mock inoculation (Wei et al., 2013). The resistant variety significantly up- and down-regulated ~500 more genes than the susceptible rice after infection. These

differentially expressed genes may not yet be functionally validated in the DR, but their expression is DR-specific, thus each one could be small contributors to the mediation of quantitative resistance. When identifying broad spectrum QTL, there stands a dichotomy between very small-effect DR genes hidden at a low resolution, and largely-influential DR and R-genes found at high resolution. It is not feasible to identify and functionally validate each broad-spectrum DR gene in the genome, thus it is best to understand what characteristics a DR gene harbors to better contribute to DR. These characteristics may be overarching, allowing easier, blanketed search for "good" DR gene traits.

Searching for "good" DR genes

The polymorphisms that distinguish a DR gene between resistant and susceptible varieties often do not involve the coding sequence, but rather the upstream region behind the DR gene. Promoter sequence differences can affect the expression of a downstream gene due to factors such as transcription factor binding, polymerase recruitment, or structural epigenetic modifications. In a QTL for rice blast resistance on rice chromosome 8, the candidate functional genes are a family of twelve Germin-Like Proteins (GLPs), *OsGLP8-1* to *-12*. The functionality of GLPs has to do with cell wall localization and superoxide dismutase enzyme activity to produce hydrogen peroxide, but they lack an oxalate oxidase domain as seen in normal germins (Christensen et al., 2004; Davidson et al., 2009). A subset of *OsGLP8* DR genes has been shown to contribute to resistance, and the most influential, *OsGLP8-6*, confers resistance to *M. oryzae, X. oryzae* pv. *oryzae* and *R. solani* (Davidson et al., 2010, 2009; Manosalva et al., 2009). In the resistant QTL donor variety, SHZ-2, there exists an 858 bp promoter insertion and upon pathogen inoculation, expression of *OsGLP8-6* is higher and faster relative to the susceptible

variety, LTH (Davidson et al., 2010). The promoter insertion contains known *cis*-elements, i.e. transcription factor binding sites, for DR-related transcription factors such as the WRKYs (Davidson et al., 2010). Within a rice blast resistance QTL on rice chromosome 3, a family of four functional Oxalate Oxidases (*OsOXOs*) reside, and one of them, *OsOXO4*, is involved in resistance to both *M. oryzae* and *R. solani* (Carrillo et al., 2009; Karmakar et al., 2016). In the resistant variety, Moroberekan, *OsOXO4* is expressed to higher levels than that of the susceptible variety, Vandana (Carrillo et al., 2009). The resistant haplotype contains a 26 bp promoter insertion in *OsOXO4*, containing DR-associated *cis*-elements (Carrillo et al., 2009). The IAA-suppressing DR gene mentioned before, *OsGH3-2*, colocalizes with resistance QTL against both *X. oryzae* pv. *oryzae* and *M. oryzae* (Fu et al., 2011). Induced expression of *OsGH3-2* enhances resistance to *X. oryzae* pv. *oryzae* as well (Fu et al., 2011). The *OsGH3-2* promoter also contains distinct polymorphisms between resistant Minghui 63 and susceptible Zhenshan 97 that introduce auxin-responsive and DR-related *cis*-elements in the resistant haplotype (Fu et al., 2011). A rice CCCH-type zinc finger nucleic acid-binding protein, *OsC3H12*, colocalizes with a *X. oryzae* pv. *oryzae* resistance QTL on chromosome 1 (Deng et al., 2012). The *OsC3H12* resistant QTL donor, Mudanjiang 8, uniquely accommodates 12 SNPs and an 8 bp insertion in the promoter. Each of these examples illustrates the possibility of DR gene regulation being the key to enhance basal resistance. It is also significant to note that most of these genes are found in multiple disease resistance QTL and function in broad-spectrum resistance. The adaptability of DR genes to transcriptionally modulate during different biotic stresses is key to broad-spectrum functionality, thus where to look when searching for "good" DR genotypes hides in the promoters.

Towards improving broad spectrum resistance

Developing rice cultivars with broad spectrum resistance can be approached with many different strategies. Introgression of single DR genes into an elite variety by crossing with a DR gene donor and subsequent backcrossing is one solution. Another proposal is to incorporate a "good" DR gene promoter into the upstream space of known DR genes using transgenics. Both methods are focused on one or a few genes with a fixed genotype. Reliance on a few loci brings the same risks as reliance on single R gene resistance, i.e., the possibility that the gene will be overcome by the pathogen in future generations. Continued focus on single genes neglects the potential that lies in resistance alleles of small effect DR genes throughout the genome. Sequencing technologies continue to improve and become more accessible. Rice genomic resources also expand, such as the 3000 rice genomes project (www.irri.org). With these advances, a more holistic view of broad spectrum resistance is coming into focus, helping us to consider the whole genome as we breed rice. However, for that to happen, new ways to analyze the rice DR need to be explored.

Thesis summary

Future directions in rice breeding will need to consider how to sustain crop productivity in a changing environment, while also considering the constant threat of new biotic stresses. We now fully understand that single genes will not achieve this goal, but the genotypes for durable resistance do exist, and the responsible genes lie throughout the genome in different rice varieties. Identifying these alleles is key to future breeding efforts, but they are largely unknown. Broad spectrum resistance has been attributed to known DR genes, but understanding what regulates diverse DR genes is yet to be understood. The following study asks the

questions: (1) What is a broad-spectrum DR gene? (2) How is diverse DR gene regulation orchestrated? (3) Are there similar resistance alleles attributed to many DR genes?

In the first chapter, I address the role of candidate DR genes underlying QTL in broad spectrum resistance using as an example one DR gene family, Phenylalanine Ammonia Lyase (*OsPAL*). To approach this, an *OsPAL* mutant (*ospal4*) was identified from a rice mutant population, and after genetic characterization, the response of the mutant to three diverse rice pathogens (*M. oryzae, X. oryzae* pv. *oryzae,* and *R. solani*) was assessed. In the second chapter, I used a bioinformatics analysis to address how DR genes coregulate to control different diseases, investigate if promoter signatures are shared among DR genes, and if these signatures are associated with their function in broad-spectrum disease resistance. In the final chapter, I provide a synopsis of what new findings and important breeding resources came from this work, and the future directions in understanding rice disease resistance.

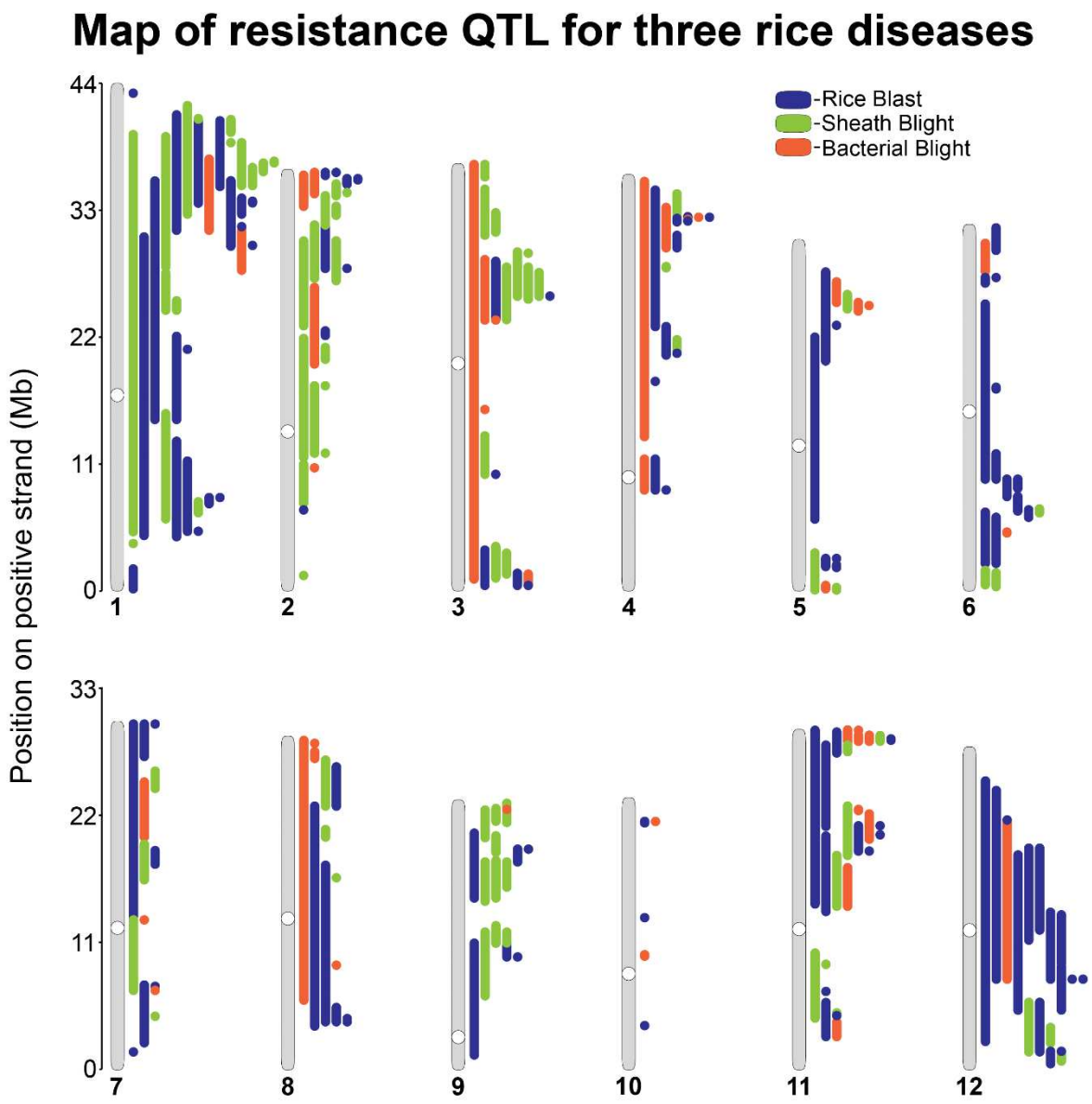
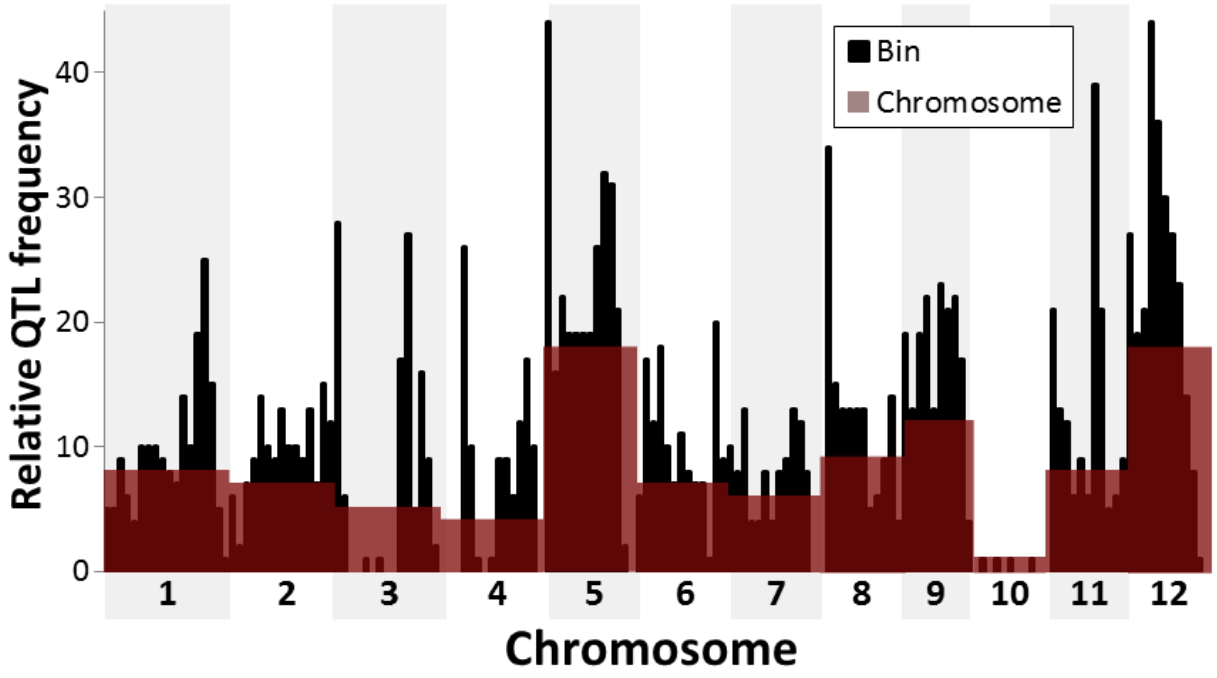


Figure 1-1: Rice genomic locations of disease resistance QTL for three major rice diseases. The genomic positions for each QTL against *M. oryzae* (blue), *X. oryzae* pv. *oryzae* (orange), and *R. solani* (green) are shown as vertical bars to the right of each respective chromosome. Centromeres are shown as white circles on each grey chromosome.



$$ChrFreq = \frac{ChrQTL(n.t.)}{ChrSize(n.t.) \times TotalChrQTL(\#)} \times 100$$

$$BinFreq = \frac{BinQTL(n.t.)}{BinSize(n.t) \times TotalChrQTL(\#)} \times 100$$

Figure 1-2: Disease resistance QTL enrichment analysis using chromosomal bins.
Each chromosome was separated into bins of size 3.1Mb, and the relative frequency of QTL was calculated using the given formula (black bars). The total chromosomal QTL frequency is shown as the height of the red shaded area, and was calculated given the equation below. Both *BinQTL* and *ChrQTL* were calculated as the total number of nucleotide coverage for every QTL within the bin. Some bins show high percentage frequency, yet none show statistical enrichment.

References

Ahmed, H. U., Finckh, M. R., Alfonso, R. F., & Mundt, C. C. (1997). Epidemiological effect of gene deployment strategies on bacterial blight of rice. *Phytopathology*, *87*(1), 66–70. https://doi.org/10.1094/PHYTO.1997.87.1.66

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution*, *19*(10), 535–544. https://doi.org/10.1016/j.tree.2004.07.021

Azizi, P., Rafii, M. Y., Abdullah, S. N. A., Nejat, N., Maziah, M., Hanafi, M. M., … Sahebi, M. (2016). Toward understanding of rice innate immunity against *Magnaporthe oryzae*. *Critical Reviews in Biotechnology*, *36*(1), 165–174. https://doi.org/10.3109/07388551.2014.946883

Ballini, E., Morel, J.-B., Droc, G., Price, A., Courtois, B., Notteghem, J.-L., & Tharreau, D. (2008). A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. *Molecular Plant-Microbe Interactions : MPMI*, *21*(7), 859–68. https://doi.org/10.1094/MPMI-21-7-0859

Banniza, S., & Holderness, M. (2001). *Pathogen biology and diversity in major fungal diseases of rice: recent advances*. The Netherlands: Kluwer Academic Publishers.

Bebber, D. P., Holmes, T., & Gurr, S. J. (2014). The global spread of crop pests and pathogens. *Global Ecology and Biogeography*, *23*(12), 1398–1407. https://doi.org/10.1111/geb.12214

Boller, T., & Felix, G. (2009). A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annual Review of Plant Biology*, *60*(1), 379–406. https://doi.org/10.1146/annurev.arplant.57.032905.105346

Boyd, L. a, Ridout, C., O'Sullivan, D. M., Leach, J. E., & Leung, H. (2013). Plant-pathogen interactions: disease resistance in modern agriculture. *Trends in Genetics : TIG*, *29*(4), 233–40. https://doi.org/10.1016/j.tig.2012.10.011

Bozsó, Z., Ott, P. G., Kámán-Tóth, E., Bognár, G. F., Pogány, M., & Szatmári, Á. (2016). Overlapping Yet Response-Specific Transcriptome Alterations Characterize the Nature of Tobacco–Pseudomonas syringae Interactions. *Frontiers in Plant Science*, *7*(March), 1–22. https://doi.org/10.3389/fpls.2016.00251

Bradley, D. J., Kjellbom, P., & Lamb, C. J. (1992). Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: A novel, rapid defense response. *Cell*, *70*(1), 21–30. https://doi.org/10.1016/0092-8674(92)90530-P

Brown, J. K. M., & Tellier, A. (2011). Plant-Parasite Coevolution: Bridging the Gap between Genetics and Ecology. *Annual Review of Phytopathology*, *49*(1), 345–367. https://doi.org/10.1146/annurev-phyto-072910-095301

Browse, J. (2009). Jasmonate Passes Muster: A Receptor and Targets for the Defense Hormone. *Annual Review of Plant Biology*, *60*(1), 183–205. https://doi.org/10.1146/annurev.arplant.043008.092007

Carrillo, M. G. C., Goodwin, P. H., Leach, J. E., Leung, H., & Vera Cruz, C. M. (2009). Phylogenomic Relationships of Rice Oxalate Oxidases to the Cupin Superfamily and Their Association with Disease Resistance QTL. *Rice*, *2*(1), 67–79. https://doi.org/10.1007/s12284-009-9024-0

Cheng, H., Liu, H., Deng, Y., Xiao, J., Li, X., & Wang, S. (2015). The WRKY45-2 WRKY13 WRKY42 Transcriptional Regulatory Cascade Is Required for Rice Resistance to Fungal Pathogen. *Plant Physiology*, *167*(3), 1087–1099. https://doi.org/10.1104/pp.114.256016

Christensen, A. B., Thordal-Christensen, H., Zimmermann, G., Gjetting, T., Lyngkjaer, M. F., Dudler, R., & Schweizer, P. (2004). The germinlike protein GLP4 exhibits superoxide dismutase activity and is an important component of quantitative resistance in wheat and barley. *Mol Plant Microbe Interact*, *17*(1), 109–117. https://doi.org/10.1094/MPMI.2004.17.1.109

Corwin, J. A., & Kliebenstein, D. J. (2017). Quantitative Resistance: More Than Just Perception of a Pathogen. *The Plant Cell*, *29*(4), 655–665. https://doi.org/10.1105/tpc.16.00915

Couto, D., & Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. *Nature Reviews Immunology*, *16*(9), 537–552. https://doi.org/10.1038/nri.2016.77

Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. *Annual Review of Plant Biology*, *66*(1), 487–511. https://doi.org/10.1146/annurev-arplant-050213-040012

Davidson, R. M., Manosalva, P. M., Snelling, J., Bruce, M., Leung, H., & Leach, J. E. (2010). Rice Germin-Like Proteins: Allelic Diversity and Relationships to Early Stress Responses. *Rice*, *3*(1), 43–55. https://doi.org/10.1007/s12284-010-9038-7

Davidson, R. M., Reeves, P. a., Manosalva, P. M., & Leach, J. E. (2009). Germins: A diverse protein family important for crop improvement. *Plant Science*, *177*(6), 499–510. https://doi.org/10.1016/j.plantsci.2009.08.012

Deng, H., Liu, H., Li, X., Xiao, J., & Wang, S. (2012). A CCCH-Type Zinc Finger Nucleic Acid-Binding Protein Quantitatively Confers Resistance against Rice Bacterial Blight Disease. *Plant Physiology*, *158*(2), 876–889. https://doi.org/10.1104/pp.111.191379

Ding, X., Cao, Y., Huang, L., Zhao, J., Xu, C., Li, X., & Wang, S. (2008). Activation of the Indole-3-Acetic Acid-Amido Synthetase GH3-8 Suppresses Expansin Expression and Promotes Salicylate- and Jasmonate-Independent Basal Immunity in Rice. *The Plant Cell Online*, *20*(1), 228–240. https://doi.org/10.1105/tpc.107.055657

Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews. Genetics*, *11*(8), 539–48. https://doi.org/10.1038/nrg2812

Dyck, P. (1966). Inheritance of adult-plant leaf rust resistance derived from the common wheat varieties Exchange and Frontana. *Canadian Journal of …*, 665–671. Retrieved from http://www.nrcresearchpress.com/doi/pdf/10.1139/g66-082

Felix, G., Duran, J. D., Volko, S., & Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant Journal*, *18*(3), 265–276. https://doi.org/10.1046/j.1365-313X.1999.00265.x

Felix, G., Regenass, M., & Boller, T. (1993). Specific perception of subnanomolar concentrations of chitin fragments by tomato cells: induction of extracellular alkalinization, changes in protein phosphorylation, and establishment of a refractory state. *The Plant Journal*. https://doi.org/10.1046/j.1365-313X.1993.04020307.x

Flor, H. H. (1942). Inheritance of pathogenicity in Melampsora lini. *Phytopathology*, *28*, 365–391.

Fu, J., Liu, H., Li, Y., Yu, H., Li, X., Xiao, J., & Wang, S. (2011). Manipulating Broad-Spectrum Disease Resistance by Suppressing Pathogen-Induced Auxin Accumulation in Rice. *Plant Physiology*, *155*(1), 589–602. https://doi.org/10.1104/pp.110.163774

Fu, J., Yu, H., Li, X., Xiao, J., & Wang, S. (2011). Rice GH3 gene family: Regulators of growth and development. *Plant Signaling & Behavior*, *6*(4), 570–574. https://doi.org/10.4161/psb.6.4.14947

Gilligan, C. A. (2008). Sustainable agriculture and plant diseases: an epidemiological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1492), 741–759. https://doi.org/10.1098/rstb.2007.2181

Gómez-Gómez, L., & Boller, T. (2000). FLS2: An LRR Receptor-like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. *Molecular Cell*, *5*(6), 1003–1011. https://doi.org/10.1016/S1097-2765(00)80265-8

Goto, M. (1992). *Fundamentals of Bacterial Plant Pathology*. San Diego: Academic Press.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., … Yin, Z. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*, *435*(7045), 1122–1125. https://doi.org/10.1038/nature03630

Gvozdeva, E. L., Volotskaya, A. V, Sof'in, A. V, Kudryavtseva, N. N., Revina, T. A., & Valueva, T. A. (2006). Interaction of proteinases secreted by the fungal plant pathogen Rhizoctonia solani with natural proteinase inhibitors produced by plants. *Applied Biochemistry and Microbiology*, *42*(5), 502–507. https://doi.org/Doi 10.1134/S0003683806050103

Hulsmans, S., Rodriguez, M., De Coninck, B., & Rolland, F. (2016). The SnRK1 Energy Sensor in Plant Biotic Interactions. *Trends in Plant Science*, *21*(8), 648–661. https://doi.org/10.1016/j.tplants.2016.04.008

Hutin, M., Pérez-Quintero, A. L., Lopez, C., & Szurek, B. (2015). MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. *Frontiers in Plant Science*, *6*(July). https://doi.org/10.3389/fpls.2015.00535

Iwai, T., Miyasaka, A., Seo, S., & Ohashi, Y. (2006). Contribution of Ethylene Biosynthesis for Resistance to Blast Fungus Infection in Young Rice Plants. *Plant Physiology*, *142*(3), 1202–1215. https://doi.org/10.1104/pp.106.085258

Jia, Y., Liu, G., Park, D., & Yang, Y. (2013). Inoculation and scoring methods for rice sheath blight disease. In Y. Yang (Ed.), *Rice Protocols* (Vol. 956, pp. 257–268). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-194-3

Johnson, R. (1984). A Critical Analysis of Durable Resistance. *Annual Review of Phytopathology*, *22*(1), 309–330. https://doi.org/10.1146/annurev.py.22.090184.001521

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*, 323–329. https://doi.org/10.1038/nature05286

K. Reddy, a. P. (1979). Relationship Between Nitrogen Fertilization, Bacterial Leaf Blight Severity, and Yield of Rice. *Phytopathology*, *69*(9), 970. https://doi.org/10.1094/Phyto-69-970

Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., … Shibuya, N. (2006). Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proceedings of the National Academy of Sciences*, *103*(29), 11086–11091. https://doi.org/10.1073/pnas.0508882103

Karmakar, S., Molla, K. A., Chanda, P. K., Sarkar, S. N., Datta, S. K., & Datta, K. (2016). Green tissue-specific co-expression of chitinase and oxalate oxidase 4 genes in rice for enhanced resistance against sheath blight. *Planta*, *243*(1), 115–130. https://doi.org/10.1007/s00425-015-2398-x

Katagiri, F., & Tsuda, K. (2010). Understanding the plant immune system. *Molecular Plant-Microbe Interactions: MPMI*, *23*(12), 1531–1536. https://doi.org/10.1094/MPMI-04-10-0099

Keijer, J. (1996). The initial steps of the infection process in Rhizoctonia solani. *Rhizoctonia Species: Taxonomy, Molecular Biology, Ecology, Pathology and Disease Control*, 149–162.

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., … Struik, P. C. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences*, *111*(11), 4001–4006. https://doi.org/10.1073/pnas.1313490111

Kiyosawa, S. (1982). Genetics and Epidemiological Modeling of Breakdown of Plant Disease Resistance. *Annual Review of Phytopathology*, *20*, 93–117.

Koeck, M., Hardham, A. R., & Dodds, P. N. (2011). The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cell Microbiology*, *13*(12), 1849–1857. https://doi.org/10.1111/j.1462-5822.2011.01665.x.The

Koizumi, S., & Ashizawa, T. (2004). Durable Control of Rice Blast Disease with multilines. *Rice Blast: Interaction with Rice and Control*, 191–199.

Kou, Y., & Wang, S. (2010). Broad-spectrum and durability: understanding of quantitative disease resistance. *Current Opinion in Plant Biology*, *13*(2), 181–5. https://doi.org/10.1016/j.pbi.2009.12.010

Kou, Y., & Wang, S. (2012). Toward an understanding of the molecular basis of quantitative disease resistance in rice. *Journal of Biotechnology*, *159*(4), 283–90. https://doi.org/10.1016/j.jbiotec.2011.07.002

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., McFadden, H., … Keller, B. (2009). A Putative ABC Transporter Confers Durable Resistance to Multiple Fungal Pathogens in Wheat. *Science*, *323*(5919), 1360–1363. https://doi.org/10.1126/science.1166453

Kunze, G. (2004). The N Terminus of Bacterial Elongation Factor Tu Elicits Innate Immunity in Arabidopsis Plants. *The Plant Cell Online*, *16*(12), 3496–3507. https://doi.org/10.1105/tpc.104.026765

Lamb, C., & Dixon, R. A. (1997). the Oxidative Burst in Plant Disease Resistance. *Annual Review of Plant Physiology and Plant Molecular Biology*, *48*(1), 251–275. https://doi.org/doi:10.1146/annurev.arplant.48.1.251

Levine, A., Tenhaken, R., Dixon, R., & Lamb, C. (1994). H2O2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell*, *79*(4), 583–593. https://doi.org/10.1016/0092-8674(94)90544-4

Li, W., Wang, B., Wu, J., Lu, G., Hu, Y., Zhang, X., … Zhou, B. (2009). The Magnaporthe oryzae avirulence gene AvrPiz-t encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene Piz-t. *Molecular Plant-Microbe Interactions : MPMI*, *22*(4), 411–420. https://doi.org/10.1094/MPMI-22-4-0411

Liu, B., Zhang, S., Zhu, X., Yang, Q., Wu, S., Mei, M., … Leung, H. (2004). Candidate defense genes as predictors of quantitative blast resistance in rice. *Molecular Plant-Microbe Interactions : MPMI*, *17*(10), 1146–52. https://doi.org/10.1094/MPMI.2004.17.10.1146

Liu, W., Liu, J., Ning, Y., Ding, B., Wang, X., Wang, Z., & Wang, G.-L. (2013). Recent Progress in Understanding PAMP- and Effector-Triggered Immunity against the Rice Blast Fungus Magnaporthe oryzae. *Molecular Plant*, *6*(3), 605–620. https://doi.org/10.1093/mp/sst015

Ma, Z., & Michailides, T. J. (2005). Advances in understanding molecular mechanisms of fungicide resistance and molecular detection of resistant genotypes in phytopathogenic fungi. *Crop Protection*, *24*(10), 853–863. https://doi.org/10.1016/j.cropro.2005.01.011

Manosalva, P. M., Davidson, R. M., Liu, B., Zhu, X., Hulbert, S. H., Leung, H., & Leach, J. E. (2009). A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiology*, *149*, 286–96. https://doi.org/10.1104/pp.108.128348

Mentlak, T. A., Kombrink, A., Shinya, T., Ryder, L. S., Otomo, I., Saitoh, H., … Talbot, N. J. (2012). Effector-Mediated Suppression of Chitin-Triggered Immunity by Magnaporthe oryzae Is Necessary for Rice Blast Disease. *The Plant Cell*, *24*(1), 322–335. https://doi.org/10.1105/tpc.111.092957

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., … Shibuya, N. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19613–8. https://doi.org/10.1073/pnas.0705147104

Monaghan, J., & Zipfel, C. (2012). Plant pattern recognition receptor complexes at the plasma membrane. *Current Opinion in Plant Biology*, *15*(4), 349–357. https://doi.org/10.1016/j.pbi.2012.05.006

Mundt, C. C. (2002). U <scp>SE OF</scp> M <scp>ULTILINE</scp> C <scp>ULTIVARS AND</scp> C <scp>ULTIVAR</scp> M <scp>IXTURES FOR</scp> D <scp>ISEASE</scp> M <scp>ANAGEMENT</scp>. *Annual Review of Phytopathology*, *40*(1), 381–410. https://doi.org/10.1146/annurev.phyto.40.011402.113723

Mundt, C. C., & Leonard, K. J. (1986). Analysis of Factors Affecting Disease Increase and Spread in Mixtures of Immune and Susceptible Plants in Computer-Simulated Epidemics. *Phytopathology*. https://doi.org/10.1094/Phyto-76-832

Navarro, L., Navarro, L., Zipfel, C., Zipfel, C., Rowland, O., Rowland, O., … Jones, J. D. G. (2004). The Transcriptional Innate Immune Response to g22. Interplay and Overlap with Avr Gene-Dependent Defense Responses and Bacterial Pathogenesis. *Plant Physiology*, *135*(June), 1113–1128. https://doi.org/10.1104/pp.103.036749.1

Nejat, N., & Mantri, N. (2017). Plant immune system: Crosstalk between responses to biotic and abiotic stresses the missing link in understanding plant defence. *Current Issues in Molecular Biology*, *23*(March). https://doi.org/10.21775/cimb.023.001

Nino-liu, D. O., Ronald, P. C., & Bogdanove, A. J. (2006). Xanthomonas oryzae pathovars: model pathogens of a model crop. *Moledular Plant Pathology*, *7*(5), 303–324. https://doi.org/10.1007/s00216-009-2643-x

Nishimura, M. T., & Dangl, J. L. (2010). Arabidopsis and the plant immune system. *Plant Journal*, *61*(6), 1053–1066. https://doi.org/10.1111/j.1365-313X.2010.04131.x

Oerke, E. C., & Dehne, H. W. (2004). Safeguarding production - Losses in major crops and the role of crop protection. *Crop Protection*, *23*(4), 275–285. https://doi.org/10.1016/j.cropro.2003.10.001

Ou, S. H. (1985). *Rice Diseases* (2nd Editio). Kew, Surrey, England: Commonwealth Mycological Institute.

Park, C. H., Shirsekar, G., Bellizzi, M., Chen, S., Songkumarn, P., Xie, X., … Wang, G. L. (2016). The E3 Ligase APIP10 Connects the Effector AvrPiz-t to the NLR Receptor Piz-t in Rice. *PLoS Pathogens*, *12*(3), 1–23. https://doi.org/10.1371/journal.ppat.1005529

Peng, X., Hu, Y., Tang, X., Zhou, P., Deng, X., Wang, H., & Guo, Z. (2012). Constitutive expression of rice WRKY30 gene increases the endogenous jasmonic acid accumulation, PR gene expression and resistance to fungal pathogens in rice. *Planta*, *236*(5), 1485–98. https://doi.org/10.1007/s00425-012-1698-7

Piasecka, A., Jedrzejczak-Rey, N., & Bednarek, P. (2015). Secondary metabolites in plant innate immunity: Conserved function of divergent chemicals. *New Phytologist*, *206*(3), 948–964. https://doi.org/10.1111/nph.13325

Piffanelli, P. (2002). The Barley MLO Modulator of Defense and Cell Death Is Responsive to Biotic and Abiotic Stress Stimuli. *Plant Physiology*, *129*(3), 1076–1085. https://doi.org/10.1104/pp.010954

Pitzschke, A., Schikora, A., & Hirt, H. (2009). MAPK cascade signalling networks in plant defence. *Current Opinion in Plant Biology*, *12*(4), 421–426. https://doi.org/10.1016/j.pbi.2009.06.008

Plantegenest, M., Le May, C., & Fabre, F. (2007). Landscape epidemiology of plant diseases. *Journal of The Royal Society Interface*, *4*(16), 963–972. https://doi.org/10.1098/rsif.2007.1114

Qiu, D., Xiao, J., Xie, W., Cheng, H., Li, X., & Wang, S. (2009). Exploring transcriptional signalling mediated by OsWRKY13, a potential regulator of multiple physiological processes in rice. *BMC Plant Biology*, *9*(1), 74. https://doi.org/10.1186/1471-2229-9-74

Ramalingam, J., Vera Cruz, C. M., Kukreja, K., Chittoor, J. M., Wu, J. L., Lee, S. W., … Leung, H. (2003). Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. *Molecular Plant-Microbe Interactions : MPMI*, *16*(1), 14–24. https://doi.org/10.1094/MPMI.2003.16.1.14

Sawada, H., Sugihara, M., Takagaki, M., & Nagayama, K. (2004). Monitoring and characterization of Magnaporthe grisea isolates with decreased sensitivity to scytalone dehydratase inhibitors. *Pest Management Science*, *60*(8), 777–785. https://doi.org/10.1002/ps.858

Sharma, T. R., Rai, A. K., Gupta, S. K., Vijayan, J., Devanna, B. N., & Ray, S. (2012). Rice Blast Management Through Host-Plant Resistance: Retrospect and Prospects. *Agricultural Research*, *1*(1), 37–52. https://doi.org/10.1007/s40003-011-0003-5

Shen, X., Liu, H., Yuan, B., Li, X., Xu, C., & Wang, S. (2011). OsEDR1 negatively regulates rice bacterial resistance via activation of ethylene biosynthesis. *Plant, Cell and Environment*, *34*(2), 179–191. https://doi.org/10.1111/j.1365-3040.2010.02219.x

Shigenaga, A. M., & Argueso, C. T. (2016). No hormone to rule them all: Interactions of plant hormones during the responses of plants to pathogens. *Seminars in Cell and Developmental Biology*, *56*, 174–189. https://doi.org/10.1016/j.semcdb.2016.06.005

Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii-Minami, N., Nishizawa, Y., … Shibuya, N. (2010). Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *Plant Journal*, *64*(2), 204–214. https://doi.org/10.1111/j.1365-313X.2010.04324.x

Shimono, M., Sugano, S., Nakayama, A., Jiang, C.-J., Ono, K., Toki, S., & Takatsuji, H. (2007). Rice WRKY45 Plays a Crucial Role in Benzothiadiazole-Inducible Blast Resistance. *The Plant Cell*, *19*(6), 2064–2076. https://doi.org/10.1105/tpc.106.046250

Singh, M. P., Lee, F. N., Counce, P. a, & Gibbons, J. H. (2004). Mediation of partial resistance to rice blast through anaerobic induction of ethylene. *Phytopathology*, *94*(8), 819–825. https://doi.org/10.1094/PHYTO.2004.94.8.819

Singh, R. P., Huerta-Espino, J., Bhavani, S., Herrera-Foessel, S. A., Singh, D., Singh, P. K., … Crossa, J. (2011). Race non-specific resistance to rust diseases in CIMMYT spring wheats. *Euphytica*, *179*(1), 175–186. https://doi.org/10.1007/s10681-010-0322-9

Skamnioti, P., & Gurr, S. J. (2009). Against the grain: safeguarding rice from rice blast disease. *Trends in Biotechnology*, *27*(3), 141–150. https://doi.org/10.1016/j.tibtech.2008.12.002

Su, J., Zhang, Z., Wu, M., & Gao, C. (2014). Changes in Insecticide Resistance of the Rice Striped Stem Borer (Lepidoptera: Crambidae). *Journal of Economic Entomology*, *107*(1), 333–341. https://doi.org/10.1603/EC13285

Sun, Y., & Steffenson, B. J. (2005). Reaction of barley seedlings with different stem rust resistance genes to Puccinia graminis f. sp. tritici and Puccinia graminis f. sp. secalis. *Can. J. Plant Pathol.*, *27*, 80–89. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/07060660509507198

Takai, R., Isogai, A., Takayama, S., & Che, F.-S. (2008). Analysis of Flagellin Perception Mediated by flg22 Receptor OsFLS2 in Rice. *Molecular Plant-Microbe Interactions*, *21*(12), 1635–1642. https://doi.org/10.1094/MPMI-21-12-1635

Talbot, N. J. (2003). On the Trail of a Cereal Killer: Exploring the Biology of *Magnaporthe grisea*. *Annual Review of Microbiology*, *57*(1), 177–202. https://doi.org/10.1146/annurev.micro.57.030502.090957

Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H.-S., Han, B., … Katagiri, F. (2003). Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen Pseudomonas syringae. *The Plant Cell*, *15*(2), 317–30. https://doi.org/10.1105/tpc.007591.can

Tao, Z., Liu, H., Qiu, D., Zhou, Y., Li, X., Xu, C., & Wang, S. (2009). A Pair of Allelic WRKY Genes Play Opposite Roles in Rice-Bacteria Interactions. *Plant Physiology*, *151*(2), 936–948. https://doi.org/10.1104/pp.109.145623

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature*, *418*(6898), 671–677. https://doi.org/10.1038/nature01014

Torres, M. A., Jones, J. D. G., & Dangl, J. L. (2006). Reactive oxygen species signaling in response to pathogens. *Plant Physiology*, *141*(2), 373–378. https://doi.org/10.1104/pp.106.079467.conjunction

Vagndorf, N., Nielsen, N. H., Edriss, V., Andersen, J. R., Orabi, J., Jørgensen, L. N., & Jahoor, A. (2017). Genomewide association study reveals novel quantitative trait loci associated with resistance towards Septoria tritici blotch in North European winter wheat. *Plant Breeding*, (July 2016), 474–482. https://doi.org/10.1111/pbr.12490

Van der Plank, J. E. (1963). *Plant diseases: epidemics and control*. New York & London: Academic Press.

Wang, C., Yang, Y., Yuan, X., Xu, Q., Feng, Y., Yu, H., … Wei, X. (2014). Genome-wide association study of blast resistance in indica rice. *BMC Plant Biology*, *14*(1), 311. https://doi.org/10.1186/s12870-014-0311-6

Wei, T., Ou, B., Li, J., Zhao, Y., Guo, D., Zhu, Y., … Qu, L.-J. (2013). Transcriptional profiling of rice early response to Magnaporthe oryzae identified OsWRKYs as important regulators in rice blast resistance. *PloS One*, *8*(3), e59720. https://doi.org/10.1371/journal.pone.0059720

White, P. T. (1994). Rice: the essential harvest. *Natl. Geogr.*, *185*, 48–79.

Wilson, C., & Tisdell, C. (2001). Why farmers continue to use pesticides despite environmental, health and sustainability costs. *Ecological Economics*, *39*(3), 449–462. https://doi.org/10.1016/S0921-8009(01)00238-5

Wilson, R. A., & Talbot, N. J. (2009). Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. *Nature Reviews Microbiology*, *7*(3), 185–195. https://doi.org/10.1038/nrmicro2032

Yang, C.-Q., Fang, X., Wu, X.-M., Mao, Y.-B., Wang, L.-J., & Chen, X.-Y. (2012). Transcriptional regulation of plant secondary metabolism. *Journal of Integrative Plant Biology*, *54*(10), 703–12. https://doi.org/10.1111/j.1744-7909.2012.01161.x

Yokotani, N., Tsuchida-Mayama, T., Ichikawa, H., Mitsuda, N., Ohme-Takagi, M., Kaku, H., … Nishizawa, Y. (2014). OsNAC111, a Blast Disease–Responsive Transcription Factor in Rice, Positively Regulates the Expression of Defense-Related Genes. *Molecular Plant-Microbe Interactions*, *27*(10), 1027–1034. https://doi.org/10.1094/MPMI-03-14-0065-R

Zhu, Y., Chen, H., Fan, J. J. J., Wang, Y., Li, Y., Chen, J., … Mundt, C. C. C. C. (2000). Genetic diversity and disease control in rice. *Nature*, *406*(August), 718–722. https://doi.org/10.1038/35021046

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D. G., Boller, T., & Felix, G. (2006). Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. *Cell*, *125*(4), 749–760. https://doi.org/10.1016/j.cell.2006.03.037

Chapter 2: Rice phenylalanine ammonia-lyase gene *OsPAL4* is associated with broad spectrum disease resistance

Introduction

Much of the world depends on rice (*Oryza sativa*) as a primary food source, but sustaining this high demand can be challenging for growers because rice yields can be reduced on an average of 37% due to the effects of pathogens and pests (Savary et al., 2011). Thus, understanding the molecular mechanisms underlying rice disease resistance is critical to maintain high production levels.

Higher plants defend themselves against pathogens through diverse responses. One type of defense, Effector Triggered Immunity (ETI), is facilitated by disease resistance (R) proteins that are involved in detection of specific pathogen effector proteins (Gassmann & Bhattacharjee, 2012). This type of resistance is often overcome by rapidly evolving virulent pathogen populations due to high selective pressures placed on the pathogen from a single host gene (Ballini et al., 2008). Conversely, a second type of resistance, called Pattern Triggered Immunity (PTI), is activated after the recognition of a conserved feature of the invading pathogen, called a Pathogen Associated Molecular Pattern or PAMP (Jones and Dangl, 2006; Chisholm et al., 2006). PTI is a form of basal resistance, and is thought to be longer lasting across plant generations because it recognizes highly conserved pathogen features and involves a multigenic response, thus diffusing the selective pressure on pathogens (Boyd et al. 2013). Basal resistance

is mediated by enzymes and other proteins encoded by defense response (DR) genes that can co-localize within Quantitative Trait Loci (QTL), which are genomic regions associated with a resistant interaction (Boyd et al., 2013; Kim et al., 2014). Understanding the genetic mechanisms of QTL-based resistance will help to improve molecular marker-assisted breeding approaches.

Increasing evidence supports the role of DR genes in long-lasting and broad-spectrum resistance associated with QTL. Several mapping studies in rice using molecular markers have shown that a variety of DR genes co-localize with QTL that confer resistance against fungal diseases (*Magnoporthe oryzae*: rice blast and *Rhizoctonia solani*: sheath blight), bacterial disease (*Xanthomonas oryzae* pv. *oryzae*: bacterial blight), and brown plant hopper (Wang et al., 1994; Ramalingam et al., 2003; Wu et al., 2004; Liu et al., 2004; Kou and Wang, 2010). Furthermore, accumulation of genomic regions with QTL containing candidate DR genes has been used as a strategy to improve disease resistance. Marker-assisted selection for five QTL that contain candidate DR genes was used to demonstrate that rice lines containing more effective DR genes were more resistant to rice blast than lines with fewer effective genes (Liu et al., 2004; Manosalva et al., 2009).

DR genes have a direct effect on host resistance. An RNAi silencing approach was used to evaluate the contributions of one group of DR gene candidates, a family of 12 genes encoding germin-like proteins (*OsGLP*) associated with a QTL on chromosome 8 (Manosalva et al., 2009). Increased silencing of *OsGLP* members correlated with enhanced susceptibility to two important and distinct rice pathogens, *M. oryzae* and *R. solani*. These experiments provide strong evidence of the direct contribution that DR genes such as the *OsGLPs* have on QTL-governed resistance.

Other candidate DR genes include the phenylalanine ammonia-lyase (PAL) gene family. PAL is a conserved homotetrameric protein that is a key enzyme in the phenylpropanoid pathway of higher plants (Bowles, 1990; Bate et al., 1994; Reichert et al., 2009; Rawal et al., 2013). Products of this biochemical pathway, which include soluble phenolics, flavonoids and lignin, are key contributors to disease resistance (La Camera et al., 2004; Vogt, 2010). Enhanced deposition of lignin can reinforce the plant cell wall providing a structural barrier to pathogen spread, and the toxic phenolic precursors produced during lignin biosynthesis or polymerization can directly inhibit pathogen multiplication and movement (Venere, 1980; Ride, 1983; Reimers and Leach, 1991; Naoumkina et al., 2010). PAL is also a key enzyme for biosynthesis of salicylic acid (SA), a plant hormone required to initiate systemic acquired resistance (SAR) in plants (Lee et al., 1995; Mauch-Mani and Slusarenko, 1996; Coquoz et al., 1998; Achnine et al., 2004; Beckers and Spoel, 2006; Gruner et al., 2013; Malamy et al., 2014). In rice, *PAL* genes are activated during both PTI and ETI, and *PAL* mRNA accumulation and enzyme activity is induced by diverse types of rice pathogens (Cramer et al., 1985; Becker-Andre et al., 1991; Joos and Hahlbrock, 1992; Logemann et al., 1995; Shiraishi et al., 1995; Cui et al., 1996; Sana et al., 2010; Giberti et al., 2012; Gupta et al., 2012; Li et al., 2013). Although multiple rice PALs (*OsPALs*) are known to exist, little is known about which, if any, contribute to resistance and the relative contributions of each.

Our interest is to determine the roles of the *OsPAL* gene family in QTL-based resistance in rice. An *in silico* analysis revealed nine closely related *OsPAL* loci within the rice genome, their colocalization with resistance QTL from multiple studies, and the changes in *OsPAL* gene expression levels across disease pressures and tissue types. We screened an IR64 mutant collection for deletions in members of the *OsPAL* gene family using a PCR-based screening

strategy with pooled mutant DNA. We identified one rice mutant line (*ospal4*) that harbors a deletion in the *OsPAL4* gene associated with increased susceptibility to three distinct pathogens, *X. oryzae* pv. *oryzae*, *R. solani* and *M. oryzae*. The mutation also affects basal expression levels of other *OsPAL* gene family members in rice seedlings. Our data suggest that *OsPAL4* contributes to broad spectrum resistance controlled by disease resistance QTL on chromosome 2 in rice.

Results

Analysis of the OsPAL gene family

Nine *OsPAL* gene family members were predicted using the MSU version 7 Nipponbare reference genome database (Table 2-1). Predicted gene sizes ranged from 2,400-4,005 nucleotides, and predicted proteins ranged from 690-718 amino acids. Four genes (*OsPAL1*, *2*, *3*, and *4*) are clustered on rice chromosome 2. Two of these genes (*OsPAL1*, *2*) have several alternatively-spliced forms as documented in the MSU rice annotation database (Ouyang et al., 2007). Two *OsPAL* genes are arranged in tandem on chromosome 4 (*OsPAL5*, *6*), and single genes are found on chromosomes 5 (*OsPAL7*), 11 (*OsPAL8*), and 12 (*OsPAL9*). Most of the gene models have two exons and one intron, with the exception of genes *OsPAL3* and *OsPAL7*, which each contain one exon (Figure 2-1).

Phylogenetic reconstruction using the unrooted neighbor-joining method revealed sequence identity and evolutionary relationships among all the *OsPAL* family of proteins (Figure 2-2). Genomic DNA sequence identities among the nine members ranged from 36% to 92%, and amino acid sequence identity ranged from 68% to 95%. *OsPAL* gene members group into three major clusters. One cluster contains the most closely related among the nine members, *OsPAL8*

and *OsPAL9*, which share 95% amino acid identity. The second cluster held *OsPAL1* and *OsPAL5*, which share 87% identity in predicted amino acid sequences. Genes that showed divergence from the other two groups included five gene family members, *OsPAL2*, *3*, *4*, *6*, and *7*. Three closely related members of this group *OsPAL3*, *4* and *7*, share from 89% to 94% amino acid sequence identity. The predicted *OsPAL* genes are orthologous between the japonica cv. Nipponbare and the recently assembled IR64 genome (Schatz et al., 2014) except for *OsPAL8* and *OsPAL9*, which are not present in the IR64 genome.

Expression of the *OsPAL* genes in various tissues and treatments show similarities and differences among the paralogs. *OsPAL4*, *OsPAL1*, *OsPAL*5, and *OsPAL*2 all show similar expression patterns across different tissues (Figure 2-2). Most *OsPAL* genes are up-regulated during resistant interactions with *X. oryzae*, *M. oryzae* and *R. solani* (Figure 2-2).

The phylogenetic relationships among *PAL* genes across species were analyzed using amino acid sequence homologs in rice (www.rice.plantbiology.msu.edu), barley (mips.helmholtz-muenchen.de/plant/barley), and *Arabidopsis thaliana* (www.arabidopsis.org) (Figure 2-3). PAL within the dicot *Arabidopsis* are diverged from monocots, with all four PAL proteins forming their own clade. OsPAL8 and OsPAL9 are highly dissimilar from any barley or *Arabidopsis* PAL. Only the OsPAL1 and OsPAL5 proteins formed a clade with a single barley PAL; the remaining rice and barley PAL formed a large group. Interestingly, the relationships of the genes encoding OsPAL were not predictable from their chromosomal positions since *OsPAL1* (chromosome 2) and *OsPAL5* (chromosome 4) were more similar to each other than they were to the other closely linked genes on the same chromosomes.

OsPAL gene family members co-localize with disease resistance QTL

QTL data were mined from public databases and were mapped to the Nipponbare reference genome. Each of the nine *OsPAL* gene family members resides within multiple resistance-associated QTL, aside from *OsPAL8* and *OsPAL9*, which were not in any of the queried regions (Figure 2-4). Physical positions of each QTL identified are given in Table S1. *OsPAL1, OsPAL2, OsPAL3* and *OsPAL4* are within three resistance-associated QTL for sheath blight, qSB-2 (Pinson et al., 2005), qSBR2-2 (Fu et al., 2011), QSbr2a (Wang et al., 2001) and one for bacterial blight (*Xoo*) qBbr2a (Zhou et al., 2012). *OsPAL5* and *OsPAL6* are within one QTL for bacterial blight, qBbr4b (Zhou et al., 2012), and one QTL for resistance to rice blast, CQAC2 (Fukuoka & Okuno, 2001). *OsPAL7* is located inside two QTL for rice blast resistance, AQEN004 (Wang et al., 1994) and qNBL-5 (Bagali et al., 1998) as well as one QTL for resistance to bacterial blight, qBbr5 (Chen et al., 2012). *OsPAL9* is within two QTL; one for bacterial blight, AQBT030 (Li et al., 1999) and one for rice blast, qDLA-12-3 (Bagali et al., 1998). The QTL loci localized on the chromosome 12 region associated with *OsPAL9* were identified in studies using *indica* cultivars as the resistant parent, yet, *indica* cultivars do not contain *OsPAL9* (Figure 2-4). Therefore, *OsPAL9* is not associated with any resistance QTL.

Identification and characterization of the ospal4 mutant

An *ospal* mutant was identified using a PCR-based DNA pooling screen of rice mutants generated by DEB mutagenesis using *OsPAL* gene family primers (Table 2-3). Amplification revealed polymorphisms in an *OsPAL* gene in super-pool #17 (Figure 2-5a). Amplification of the 10-line pools that correspond to super-pool 17 identified a single mutant line, 1982, that contained the variant 200 bp fragment (Figure 2-5a). Both wild type and mutant fragments were

amplified from line 1982, suggesting either that the mutant line is heterozygous for the deleted allele or that the wild type fragment corresponds to other *OsPAL* genes amplified with the gene-family primers. A second fragment larger than the 200 bp fragment that resolved in the gels was likely an artifact of secondary structure formation of the amplicon, because sequence analysis showed it was identical to the 200 bp band. The IR64 background of the *ospal* mutant line 1982 was confirmed by genotyping with several simple sequence repeat (SSR) markers (data not shown).

Characterization of the mutation at the OsPAL4 locus

Sequence of 200 bp mutant amplicons revealed a gene fragment similar to the second exon of *OsPAL4*, but with a deletion between nucleotides 889 and 1,639 of *OsPAL4* (Figure 2-5b). To confirm that only *OsPAL4* was mutated in line 1982, we used primers specific to each member of the *OsPAL* gene family (Table 2-3) to amplify wild type and variant fragments after amplification with the *OsPAL* gene family primers (Figure 2-6). Only the *OsPAL4* gene-specific primers amplified a 200 bp mutant fragment (Figure 2-6). These primers also amplified a wild type-sized fragment (Figure 2-6). Relative to the other family members, both the mutant and wild type amplicons showed highest similarity to *OsPAL4* from Nipponbare (LOC_Os02g41680) (Figure 2-7).

No gross morphological differences were observed between the *ospal4* mutant lines and the wild type *OsPAL4* segregant lines. However, the *ospal4* mutant lines showed not only low seed germination but also low seed production as indicated by poor grain filling and seed abortion in the inflorescences of the rice mutant lines, compared with the wild type lines or mutant segregants containing the wild type *OsPAL4* allele.

Genetic segregation of the ospal4 mutation

To understand the inheritance of the *ospal4* mutation, we determined the segregation of the mutation in 91, 122, and 52 individual progeny of the M3, M4, and M5 generations, respectively, using gene specific primers corresponding to the *OsPAL4* gene member. In the segregation analysis, we only saw progeny with the smaller fragment of the mutant allele and the entire *OsPAL4* gene. If the mutant gene is derived from the *OsPAL4* gene, one quarter of the progeny would have been expected to not carry the wild type *OsPAL4* gene, unless homozygotes for the mutant allele are lethal. In either case, the transmission, or homozygosity of the mutant gene in the progeny appears to be selected against. Three quarters of the progeny would be expected to carry at least one copy of the mutant allele in progeny of these families and a large deficiency was observed in each generation (Table 2-4).

To confirm that no individuals homozygous for the *OsPAL4* mutation could be obtained, we followed segregation of the mutants in M4 and M5 generations. Ten individuals from the M3 progeny were examined from each of 20 M3 plants that were presumed to be heterozygous for the variant allele. We also selected 10 lines from the progeny of 20 M3 plants that did not have the variant allele. DNA was extracted from these progeny and amplified fragment sizes were examined. Each of the 20 M3 lines presumed to be heterozygous segregated the variant allele among their progeny, but the variant band was not observed in all of the individual progeny. On the other hand, progeny from the 20 M3 lines without the variant allele never showed the 200 bp band. Similar results were obtained in the progeny of the following M5 generation. The inability to identify individuals homozygous for the variant allele and the deviation from a 3:1 segregation ratio indicate that the variant allele either has poor transmission through the male or female gametes or that homozygous progeny are inviable.

Reciprocal crosses between an M4 heterozygous *ospal4* mutant line and wild type IR64 resulted in 23 progeny with IR64 as the female parent and 10 progeny with *ospal4* mutant as the female parent. All of these germinated, however, two of the progeny from the *ospal4* female parent lines failed to grow (20%) along with one of the IR64 female parent lines (4.4%). Of the remaining lines, seven of the eight *ospal4* female parent progeny contained one copy of the mutant allele as did 11 of 20 lines tested from the IR64 female parent progeny. The results demonstrate transmission of the mutant allele through both types of gametes. The inflorescence of the rice *ospal4* mutant and the F_1 progeny between the mutant line and IR64 all exhibited seed abortion (less grains filled in the inflorescence), supporting the hypothesis of lethality associated with a homozygous *ospal4* mutation. Further investigation is needed to determine why this phenomenon occurs.

Differential expression of OsPAL genes in ospal4 mutant seedlings

To understand the effect of *OsPAL4* mutation on the activity of other *OsPAL* gene homologs, expression analysis was performed on two-week-old rice plants containing the heterozygous *ospal4* mutation and compared to a WT segregant of the F1 IR64 x *ospal4* cross not containing the deletion in *OsPAL4*. Differential expression in the mutant relative to WT was measured by quantitative real-time PCR (qPCR) on the genes, *OsPAL4* (LOC_Os02g41680), *OsPAL2* (LOC_Os02g41650), *OsPAL6* (LOC_Os04g43800), and *OsPAL7* (LOC_Os05g35290) (Figure 2-8). A greater than two-fold drop in expression was exhibited in the *ospal4* mutant, as expected due to the loss-of-function mutation in one allele for the gene. Interestingly, *OsPAL2* showed increased expression in *ospal4* plants, whereas *OsPAL6* displayed lower relative transcript abundance in the mutant compared to WT seedlings. *OsPAL7* showed no change in

expression. These results infer an association between a loss of *OsPAL4* transcript abundance to the altered regulation of other members of the *OsPAL* gene family.

Response of ospal4 to bacterial blight and sheath blight disease

Considering that *PAL* genes are involved in plant defense in several pathosystems, and that *OsPAL4* co-localized with bacterial blight and sheath blight disease resistance QTL (Figure 2-4), we tested whether the heterozygous *ospal4* rice mutant lines were more susceptible to a virulent strain of *X. oryzae* pv. *oryzae* (PXO339). In three screens, individual plants (M3, M4, and M5) were genotyped for the presence of the mutation in *OsPAL4*, and grouped into two groups, those with (*ospal4* mutants) and without the altered gene fragment (*OsPAL4* segregant). In all three generations, the group with the mutant *ospal4* showed a higher level of susceptibility to *X. oryzae* pv. *oryzae* PXO339 than those with only wild type fragments (Figure 2-9). Because *OsPAL4* was also associated with chromosome 2 QTL for sheath blight resistance, we inoculated the M5 progeny with *R. solani*. The *ospal4* mutants exhibited more disease than the wild type segregants ($P<0.0001$) (Figure 2-9).

Response of the F1 to bacterial blight and rice blast disease

Our *in silico* analysis of previous transcriptome studies indicated that *OsPAL6* is up-regulated during a resistant interaction with rice blast along with bacterial blight and sheath blight (Figure 2-2), and that it resides within rice blast QTL (Figure 2-4). Thus, we asked if the *ospal4* mutant was more or less susceptible to rice blast. The F1 generation from a cross between wild type IR64 and a heterozygous M5 *ospal4* mutant were inoculated with the bacterial blight pathogen, *X. oryzae* pv. *oryzae* (PXO339) and the rice blast pathogen, *M. oryzae* (P06-6).

The *ospal4* mutants were more susceptible to both pathogens relative to the wild type segregants, confirming that *ospal4* is more susceptible to bacterial blight, and suggesting that the *ospal4* mutant negatively affects disease resistance through changes in expression of other *OsPAL* genes, including *OsPAL6* (Figure 2-10).

Discussion

We present evidence that phenylalanine ammonia-lyase genes, and in particular, *OsPAL4,* are involved in rice defense responses. Meta-analysis of previous transcriptome studies shows increased expression of *OsPAL4* and several other *OsPAL* family members during resistance to diverse pathogens (Figure 2-2). In addition, seven of the *OsPAL* family members, including *OsPAL4*, co-localize with multiple disease resistance QTL in rice (Figure 2-4). Furthermore, a rice mutant containing a deletion in the *OsPAL4* gene was more susceptible to rice bacterial blight, blast and sheath blight diseases. Our findings are consistent with previous results from multiple pathosystems that implicate PAL in induced plant defense responses (Hahlbrock and Scheel, 1989; Zhu et al., 1995; Tanaka et al., 2003; Zhu et al., 2004; Naoumkina et al., 2010; Sana et al., 2010; Giberti et al., 2012; Gupta et al., 2012; Li et al., 2013; Zhang et al., 2013; Kim and Hwang, 2014; Riaz et al., 2014; Bhat et al., 2014).

OsPAL4 and multiple members of the *OsPAL* gene family could contribute to resistance differently, depending on the type of disease. Each *OsPAL* locus (with the exception of *OsPAL8* and *OsPAL9*) is found within disease resistance QTL for at least two different diseases (Figure 2-4). Six *OsPAL* genes contained in clusters on chromosomes 2 (*OsPAL1-OsPAL4*) and 4 (*OsPAL5-OsPAL6*) are located within QTL for resistance to bacterial blight (Figure 2-4). The only *OsPAL* genes associated with sheath blight QTL include *OsPAL1-OsPAL4*, which are

clustered on chromosome 2. Our demonstration that a mutation of *OsPAL4* enhances susceptibility to sheath blight and bacterial blight provides evidence that at least this family member contributes to QTL-governed resistance on chromosome 2. Because enhanced basal expression of *OsPAL2* (Figure 2-8), a physical neighbor of *OsPAL4*, did not compensate for the loss of disease resistance in the *ospal4* mutant, we conclude that *OsPAL2* is not a direct contributor to disease resistance governed by the chromosome 2 QTL. Alternatively, the increased susceptibility in the *ospal4* line may have been due to down-regulation of *OsPAL6* in the *ospal4* mutant lines (Figure 2-8). Since *OsPAL6,* not *OsPAL4,* co-localizes with a blast resistance QTL, the decrease in *OsPAL6* expression may have caused susceptibility to this particular disease. Conclusions made on the roles of *OsPAL4*, *OsPAL6*, and *OsPAL2* in resistance are summarized in Table 2-5. These examples suggest that a specialization of *OsPAL* genes on different chromosomes could yield genes more effective towards certain pathogens.

Evolutionary divergence in the *PAL* gene family was observed across plant species. The organization of rice *OsPAL* gene clusters on chromosomes 2 and 4 is consistent with the organization of similar *PAL* gene members on chromosomes of potato (Trognitz et al., 2002), wheat (Liao et al., 1996) and cucumber (Shang et al., 2012). Phylogenetic analysis shows that barley PAL proteins are organized into multiple groups similar to rice OsPALs, suggesting that some of the gene duplication events at these loci predate the divergence of clades containing these two monocots, but after monocots diverged from dicots (Figure 2-3). There may also be a mutational bias in the genomic regions containing these gene clusters, meaning that these duplication events were the result of parallel evolution after divergence of rice and barley (Stern, 2013). The patterns of similarities between the rice genes suggest that an ancestral *OsPAL* gene was duplicated and the two members diverged before a second duplication event gave rise to the

clusters of adjacent genes on chromosomes 2 and 4. The *OsPAL4* gene (chromosome 2) was more similar to the *OsPAL5* gene (chromosome 4) than it was to the other members in the chromosome 2 gene cluster. Such gene duplication can be a major driving force for recruitment of genes for secondary metabolism (Rensing, 2014), where gene copies are gradually modified to create genes with new specificities and expression patterns adapted to the needs of the pathway in which they are involved. The modifications among these highly conserved duplications may reflect differences in regulation (promoter), subcellular localization, enzyme properties, and even feedback mechanisms. These modifications could explain the changes in expression patterns of other *OsPAL* gene family members that occur in *ospal4* mutant (Figure 2-8). Regulation of *OsPAL* genes could be due to feedback from transcriptional patterns of other gene family members across the genome.

Increased disease susceptibility in the *ospal4* mutant may be due to pleiotropic effects involving genes outside of the phenylpropanoid pathway. Using transcript-profiling approaches, Rohde and coworkers (Rohde et al., 2004) demonstrated that a single *PAL* mutation profoundly affected the transcription of a plethora of genes outside of the phenylpropanoid pathway. Arabidopsis *pal1*, *pal2*, and *pal1pal2* double mutants exhibited significant changes in transcription of genes that encode other enzymes of the phenylpropanoid pathway as well as enzymes in carbohydrate and amino acid metabolism, establishing a link between primary and secondary metabolism. Components of most of the pathways altered in these Arabidopsis *pal* mutants were previously described for their responsiveness to pathogen attack or elicitor activation (Sommssich & Hahlbrock, 1998). Tobacco lines epigenetically suppressed for *PAL* exhibited reduced PAL activity and increased susceptibility to virulent *Cercospora nicotinae* as well as to avirulent Tobacco Mosaic Virus (TMV), and those plants were also impaired in

systemic acquired resistance induction due to reduced salicylic acid (Maher et al., 1994; Pallas et al., 1996). Plant responses to pathogens are complex and multifaceted, and a mutation in *OsPAL4* could disrupt defense responses both within and outside the phenylpropanoid pathway.

The enzyme PAL functions as a homotetramer with four subunits (Ritter and Schulz, 2004; Calabrese et al., 2004; Moffitt et al., 2007). In wheat, the enzyme is made up of two pairs of different subunits, two subunits with a molecular weight of 75 kDa and two subunits of 80 kDa, suggesting that wheat PAL proteins are derived from at least two structural genes (Havir & Hanson, 1973). Logeman et al., (1995) suggested the possibility that PAL functions as part of a multi-enzyme complex formed by three enzymes from the phenylpropanoid pathway (Logemann et al., 1995). Supporting this hypothesis, Harvis and Hanson (1973) described the formation of high molecular weight aggregates of PAL enzyme with other proteins in mustard seedlings *in vitro*. Sarma and Sharma (1999) demonstrated that only one isoform of OsPAL was found both during and after UV-B induction of *OsPAL* in rice seedlings, although it is not clear which gene(s) gave rise to the monomer (Sarma & Sharma, 1999). The recent solving of the crystal structure of PAL in parsley, yeast and cyanobacteria revealed that the enzyme is highly conserved among these species (Ritter and Schulz, 2004; Calabrese et al., 2004; Moffitt et al., 2007). The *OsPAL4* gene deletion spans ~750 bp, which results in deletion of most of the second exon, and likely abolishes the function of the encoded OsPAL4 enzyme. Subunits transcribed from normal copies of the *OsPAL4* gene in the heterozygous mutant lines could associate to form an active homotetramer. Subunits encoded from mutant alleles, however, might form truncated, non-functional units, if they are able to be transcribed. If these subunits associate with each other they would form non-functional tetramers. Heterozygous progeny would have both normal and mutant subunits. Tetramers formed between these normal and

mutant subunits may or may not have activity. However, because disease susceptibility increased in the *ospal4* mutant line, it is likely that enzyme activity was significantly impaired.

The importance of understanding *PAL* and the phenylpropanoid pathway applies to the field of biotechnology as well as pathology. Lignin biosynthesis is catalyzed by PAL and therefore, the *PAL* gene is a desirable bioenergy target for optimizing downstream processing of biomass (Boerjan et al., 2003; Vanholme et al., 2012; Craven-Bartle et al., 2013). High lignin content is preferred for thermochemical conversion, while low lignin content is more advantageous for enzymatic approaches targeting cellulose embedded in the cell wall matrix (Tanger et al., 2013). Reduced lignin content, perhaps due to altered *PAL* expression, could minimize the need for acid pretreatment of feedstocks (Chen et al., 2006; Chen and Dixon, 2007).

The *ospal4* mutant identified in this study will provide important insights into rice primary and secondary metabolism and stress responses. Further characterization of the rice mutation, including inheritance analysis, histochemical studies, complementation studies, and responses against other rice pathogens such as brown leaf spot (*Bipolaris oryzae*) or bacterial leaf streak (*X. oryzae* pv. *oryzicola)* could help to further elucidate the broad effects of *OsPAL4* in the plant defense response.

Materials and Methods

OsPAL sequence analysis and primer design

Sequences corresponding to *OsPAL* genes were obtained from the MSU Rice Genome Annotation Project release 7 (Ouyang et al., 2007) by searching the gene annotation field using the keywords “phenylalanine ammonia lyase”. Primers were designed manually, and primer

properties were analyzed using IDT's Oligo Analyzer 3.0 using default settings. Primer specificity was verified by performing blastn (www.ncbi.nlm.nih.gov) analyses of the primer sequences against the MSU release 7.0 rice genome database. Phylogenetic analysis was performed using the program MEGA6 on alignments of amino acid sequences from MSU 7.0 polypeptide predictions using ClustalW with a Neighbor-Joining method (Saitou and Nei, 1987; Tamura et al., 2013). Bootstrapping values (1,000 replicates) were computed (Felsenstein, 1985) and evolutionary distances were found using the Poisson correction method (Zuckerlandl & Pauling, 1965). Expression data to assess tissue and disease-specific expression of the *OsPAL* gene family was mined from the NCBI Sequence Read Archive (www.ncbi.nlm.nih.gov/sra), the MSU digital gene expression library (www.rice.plantbiology.msu.edu), the NCBI Gene Expression Omnibus (www.ncbi.nlm.nih.gov), and from a rice transcriptome analysis after *R. solani* infection (data provided by Yulin Jia, USDA-ARS) (Venu et al., 2007). *OsPAL* gene expression during susceptible and resistant interactions with *X. oryzae* pv. *oryzae* strain PXO99A and PXO99A (pavrXa7), respectively, at 24 h post infiltration, or untreated rice cultivar IRBB61 (contains resistance gene *Xa7*) was extracted from RNASeq data from duplicate samples (Leach, unpublished). Raw transcript reads were normalized and expressed as reads per kilobase per million reads using Cufflinks v2.0.2 (Trapnell et al., 2010) with default settings. This normalization allowed for comparison between the inoculated and untreated samples with $\log_2$ fold change values.

To identify associations of *OsPAL* gene family members with disease resistance QTL for sheath blight, rice blast and bacterial blight, marker data for QTL regions were obtained from GRAMENE and Q-TARO databases (www.gramene.org, release #39, qtaro.abr.affrc.go.jp), as well as recent literature (Table 2-2). QTL positions in the rice Nipponbare reference genome

were located by mapping the marker sequences to the genome pseudomolecules using an electronic PCR approach (www.ncbi.nlm.nih.gov/tools/epcr/; max indels=4, max mismatches=5, word size=10) and a custom python script (Schuler, 1997).

Plant material and OsPAL4 mutant discovery

Rice deletion mutants were generated by the International Rice Research Institute using the IR64 cultivar and diepoxybutane (DEB) as the mutagen as described (Wu et al., 2005). The same IR64 seed stock was used as a wild type control in all experiments. The *OsPAL* mutant was discovered using a PCR-based gDNA pooling strategy (Figure 2-11). Genomic DNA for pooling was extracted from rice leaves using a modified hexadecetyltrimethylammonium bromide (CTAB) procedure (Saghai-Maroof et al., 1984), and was quantified by UV absorbance using a NanoDrop ND-1000 spectrophotometer (Rockland, DE, USA).

Prior to PCR screening, to optimize the DNA pool size for detection of mutant amplicons among wild-type amplicons, we performed a reconstruction experiment using a natural deletion in the *Xa21* gene in rice line IR24 (gift from Guo-Liang Wang, Ohio State University). In those experiments, the 135 bp *Xa21* deletion was clearly detected up to a template DNA dilution of 1:200 (Figure 2-12). Based on this data, we organized two levels of gDNA pools for each of 3,000 DEB-mutagenized lines, a pool of 10 lines (10-line-pool) and a pool of 100 lines (super-pools). In this strategy, gDNA from each line was included in the 10 line pool and the 100 line pool.

Primers that correspond to highly conserved regions among *OsPAL* gene family members (PAL-F and PAL-R, Table 2-3) were used to screen all DEB-mutants. PCR reactions used HotStar Taq DNA polymerase (Qiagen, Valencia, CA, USA), and resulting amplicons were

separated in 4% polyacrylamide-7.5M urea gels and visualized by silver staining. When putative mutant amplicons, i.e., those that migrated further than the wild type amplicon, were detected in the products of a super-pool, the *OsPAL* gene family primers (PAL-F and PAL-R) were used to amplify the corresponding 10-line-pools. Finally, once the deletion line was identified in the 10-line-pool, PCR was performed on DNA from the 10 corresponding individual lines to identify the specific mutant line.

To characterize the mutation, wild type and mutant bands detected with the *OsPAL* gene family primers (PAL-F and PAL-R) were excised from the polyacrylamide gel, and DNA was eluted by diffusion into 100 ul distilled water at 4^{o}C overnight. Using 10 ul of this solution as a template, the fragments were reamplified first with the *OsPAL* gene family primers. The PCR products were separated in agarose and the fragments were excised from the gel and purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA). These fragments were inserted in pCR 2.1-TOPO vector using the TOPO TA Cloning kit (Invitrogen, Carlsbad, CA, USA) and introduced in the TOP10 One Shot Chemically Competent Cells (Invitrogen, Carlsbad, CA, USA). Plasmids from positive clones were sequenced. Sequences corresponding to the wild type and mutant fragment were aligned to determine the deletion junctions (Figure 2-7). In addition, gene specific forward primers for each of the nine *OsPAL* gene members predicted from the rice genome were designed in the most divergent regions (Table 2-3) and used with a common reverse primer (PAL-R) to amplify genomic DNA corresponding to four M4 non-mutant lines, two M4 mutant lines, and the IR64 control (Figure 2-6).

Segregation analysis and genetic crosses

Production of M2 and M3 mutant seeds was as described (Wu et al., 2005); M4 and M5 seed were derived from a single M3 and M4 line, respectively. A total of 91 M3, 122 M4 and 84 M5 progeny were genotyped by PCR amplification using *OsPAL4*-specific primers to determine the presence or absence of the *ospal4* mutation. PCR products were resolved in 4% polyacrylamide gel and visualized by silver staining. Segregation ratio (3:1) was tested by Chi Square (X^2) test (Table 2-4).

Reciprocal genetic crosses were performed under greenhouse conditions between an *ospal4* mutant line and cultivar IR64. F_1 plants were screened for the PAL mutation. SSR markers (RM266, RM334, RM278, RM333, and RM224) (www.gramene.org, release #39) were used to confirm that the F_1 plants were the IR64 background.

Pathogen inoculations

To determine if the *ospal4* deletion affected bacterial blight resistance, 182 M3, 85 M4 and 25 M5 progeny that had been genotyped to distinguish *OsPAL4* segregant and *ospal4* mutant lines, were inoculated with a virulent strain of *X. oryzae* pv. *oryzae* PXO339 in the greenhouse. The second youngest fully expanded leaf was inoculated with a bacterial suspension ($5x10^9$ CFU/ml) using the leaf clipping method (Kauffman et al., 1973). Disease was assessed at 2 wk after inoculation by measuring the lesion length. To evaluate sheath blight resistance, 2-wk-old M5 plants were inoculated with *R. solani* RR0140-1, anastomosis group 2 (AG2) (Wamishe et al., 2007) following published methods (Jia et al., 2013). After 7 days, lesion lengths were measured and disease index was calculated as lesion length divided by plant height multiplied by

100. Differences between the test groups were analyzed using the Student's t-test with an alpha of 0.05.

The effect of the *ospal4* deletion on resistance to bacterial blight and rice blast was tested in the F1 progeny of the IR64 X *ospal4* backcross. Plants were inoculated with *X. oryzae* pv. *oryzae* PXO339 as described above. The F1 seedlings at the 4-leaf stage (2-wk-old) were inoculated with *M. oryzae* (P06-6) using the detached leaf assay (Jia et al., 2003). Briefly, sections from the two youngest, fully expanded rice leaves were treated with either *M. oryzae* or a mock inoculum solution of 0.01% Tween-20. Disease was assessed 10 days post-inoculation, using a visual index (Manosalva et al., 2009) that gave a combined score for relative lesion size and mycelia production on the leaf surface, with 0 being a resistant interaction, and 14 being a susceptible interaction. Differences between the two test groups were analyzed using the Student's t-test with an alpha of 0.05.

Quantitative real time PCR (qPCR) of OsPAL gene family members

Three *ospal4* mutant lines backcrossed to IR64 were genotyped by collecting genomic DNA from 10 plants each and amplifying across the deleted region using primers PAL-F8 (5' TAACGTTTACCTGGTCACTGC 3') and PAL-R (5' TTGACGTCCTGGTTGTGCTGC 3') using Standard Taq polymerase (New England Biolabs, Ipswich, MA, USA). Products were visualized in an Experion by automated electrophoresis in a 1 Kb DNA analysis chip (Figure 2-13) (Bio-Rad, Hercules, CA, USA).

RNA was extracted using the Spectrum Plant Total RNA Kit (Sigma Aldrich, St. Louis, MO, USA) according to manufacturer's instructions. RQ1 RNase-free DNase (Promega, Fitchburg, WI, USA) was added to RNA to remove genomic DNA at a proportion of 1 uL per ug

of RNA and incubated at 37°C for 30 min. RNA quality and purity were confirmed with gel electrophoresis and a NanoDrop ND-1000 spectrophotometer (ThermoScientific, Wilmington, DE, USA). cDNA was synthesized using RNase-inhibiting iScript reverse transcriptase and a blend of oligo(dT) and random hexamer primers (Bio-Rad, Hercules, CA, USA) according to the manufacturer's protocol.

Relative quantification of transcripts in mutant *ospal4* plants was conducted using the BioRad iCycler IQ multicolor Real-Time PCR detection system. The reaction was carried out using the SsoAdvanced SYBR Green Supermix (Bio-Rad, Hercules, CA, USA). Three biological replications of the calibrator (wild type plants) and test samples (plants of the same *ospal4* mutant line) as well as three technical replicates of each sample were performed for each assay. Gene primers were designed specifically for qPCR (Table 2-6). The housekeeping gene, EF-1α, was used to normalize the data qPCR (Jain et al., 2006). Primer concentration (200 nM) and annealing temperatures (*OsPAL4* - 53°C, *OsPAL2* - 53°C, *OsPAL6* - 55°C, *OsPAL7* - 55°C , eEF-1α - 60°C) were optimized by conventional gradient PCR followed by agarose gel electrophoresis of products. Cycle parameters for the reaction were 95°C for 2 min, followed by 40 cycles of 95°C for 15 sec, and 53/55/60°C for 30 sec. Melt curves were generated by increasing the temperature by 0.5°C in 10 sec increments for 80 cycles, with the initial temperature at 53/55/60°C. Standard curves were generated based on amplification of wild type (WT) cDNA dilution series for each primer set (*OsPAL4, OsPAL2, OsPAL6, OsPAL7, EF-1α*). Primers were only used if efficiencies were within the range of 95-105%. Threshold values were calculated using a regression analysis for individual samples. Normalized absolute $\Delta\Delta C_t$ was calculated as an average of all biological replicates as well as individual plants separately.

Relative expression ratios were also calculated using the Pfaffl method, $2^{-\Delta\Delta Ct}$ (Pfaffl, 2001). P-values were calculated using a Student's two-sample t-test, alpha=0.05.

Table 2-1: *OsPAL* Characterization. Nine *OsPAL* genes predicted in the rice genome and their corresponding MSU version 7.0 locus ID and chromosomal positions.

Gene	Locus ID	Chromosomal position
OsPAL1	LOC_Os02g41630	24973386 – 24977336
OsPAL2	LOC_Os02g41650	24984871 – 24989388
OsPAL3	LOC_Os02g41670	24992827 – 24995414
OsPAL4	LOC_Os02g41680	25006467 – 25009121
OsPAL5	LOC_Os04g43760	25906849 – 25910074
OsPAL6	LOC_Os04g43800	25926988 – 25929732
OsPAL7	LOC_Os05g35290	20953410 – 20955923
OsPAL8	LOC_Os11g48110	29012490 – 29014744
OsPAL9	LOC_Os12g33610	20292893 – 20295410

Table 2-2: Rice disease resistance QTL that co-localize with *OsPAL* genes.

QTL name	Chr	Start	Stop	Pathogen	Colocalized *OsPALs*	Published study
qBbr2a	2	19342016	25871437	*X. oryzae*	*OsPAL1-4*	Zhou *et al.*, 2012
qBbr5	5	189782	24587132	*X. oryzae*	*OsPAL7*	Chen *et al.*, 2012
qBbr4b	4	21414515	29054517	*X. oryzae*	*OsPAL5-6*	Zhou *et al.*, 2012
AQEN004	5	6132827	21633475	*M. oryzae*	*OsPAL7*	Wang *et al.*, 1994
CQAC2	4	22534597	34157504	*M. oryzae*	*OsPAL5-6*	Fukuoka and Okuno, 2001
qNBL-5	5	19607926	27183777	*M. oryzae*	*OsPAL7*	Bagali *et al.*, 1998
qSBR2-2	2	1326946	31191246	*R. solani*	*OsPAL1-4*	Fu *et al.*, 2011
QSbr2a	2	24571578	28695037	*R. solani*	*OsPAL1-4*	Wang *et al.*, 2001
qSB-2	2	22599386	29824873	*R. solani*	*OsPAL1-4*	Pinson *et al.*, 2005

Table 2-3: *OsPAL* family and gene-specific primers.

Primer name	*Gene*	*Primer sequence 5' --> '3*	*Position*	*Location*	*Product size (bp) with PAL-R*
PAL-R	*OsPAL*	TTGACGTCCTGGTTGTGCTGC	N/A	Exon	N/A
PAL-F	*OsPAL*	AAGCTGCTCAACGCGAACG	N/A	Exon	931
F10	*OsPAL6*	TGCTGTTCGTCTGGTGAGAGC	556	Intron	1148
F12	*OsPAL1*	GTTTGGGCTGCAACTTGGCAG	1664	Intron	1199
F13	*OsPAL5*	CAACGGCTCCGATGGCAACTC	1285	5' 2nd exon	1052
F19	*OsPAL3*	GAACGGTCAGGTTGCTGCCGAT	154	5' 2nd exon	1460
F22	*OsPAL7*	CCGACGGCCACGTCC	566	Exon	1044
F24	*OsPAL8*	CGTACGTACTCGGCGACCTGTACGTGCCGC	1286	5' UTR	1658
F25	*OsPAL2*	AGTCGGCACGGCGGCAGTGTGTATGTAC	1905	Intron	1231
F8	*OsPAL4*	TAACGTTTACCTGGTCACTGC	558	Intron	1168
F15	*OsPAL4*	TCACACCGTGCCTGCCGCTCC	814	5' 2nd exon	912
F21	*OsPAL4*	TTCTCTCGACGCTTTCTGTGCTAGG	84	5' UTR	1500

Table 2-4: Chi squared analysis of M3, M4, and M5 progeny of rice IR64 DEB mutant *ospal4.*

	M3 Progeny	M4 Progeny	M5 Progeny
Obs. Mutant gene	38	64	27
Obs. No Mutant gene	53	54	25
Total	91	118	52
Exp. Mutant gene	68.25	88.5	39
Exp. No Mutant gene	22.75	29.5	13
X2	53.6	27.1	14.8
P	<.001	<.001	<.001

Table 2-5: Predicted roles of *OsPAL2*, *OsPAL4*, and *OsPAL6* in disease resistance.

	X. oryzae			*M. oryzae*			*R. solani*		
	OsPAL2	***OsPAL4***	***OsPAL6***	***OsPAL2***	***OsPAL4***	***OsPAL6***	***OsPAL2***	***OsPAL4***	***OsPAL6***
Co-localizes with disease resistance QTL	+	+	+	-	-	+	+	+	+
Expression in resistant interactions[a]	+	+++	+++	-	-	+	+	+++	+
Expression in *ospal4* mutant[b]	↑	↓	↓	↑	↓	↓	↑	↓	↓
Putative role in resistance[c]	Low	High	High	Low	Low	Good	Low	High	Good

[a]Gene regulation in rice-pathogen resistant interactions is based on data in Fig. 2 is categorized with levels indicated as none (-), weak (+), and strong (+++).

[b]Expression of the paralog in the *ospal4* mutant is given as up- or down-regulated.

[c]Predictions on the involvement of the three *OsPAL* genes in resistance are based on the above three criteria.

Table 2-6: *OsPAL* gene-specific primers used for quantitative real-time PCR of *OsPAL* genes.

Gene	Sequence (5' – 3')
OsPAL2	F – GCA TCA GCT TCC AAC TCG
	R – GGT TTC GCA CTC CAT TAC AGA
OsPAL4	F – CTT CAC AAC AGC TAA TCG AG
	R – CGC ACT CCA TTT CAG TAC CA
OsPAL6	F – AGA TTG AGG TCA TCC GTG
	R – GAA CAT GAG CTT ACC GAT C
OsPAL7	F – ATC GAC ATC CTC AAG CTC ATG
	R – AGT TGG TGC TCA GCG TCT TCT
EF-1α	F – TTT CAC TCT TGG TGT GAA GCA GAT
	R – GAC TTC CTT CAC GAT TTC ATC GTA A

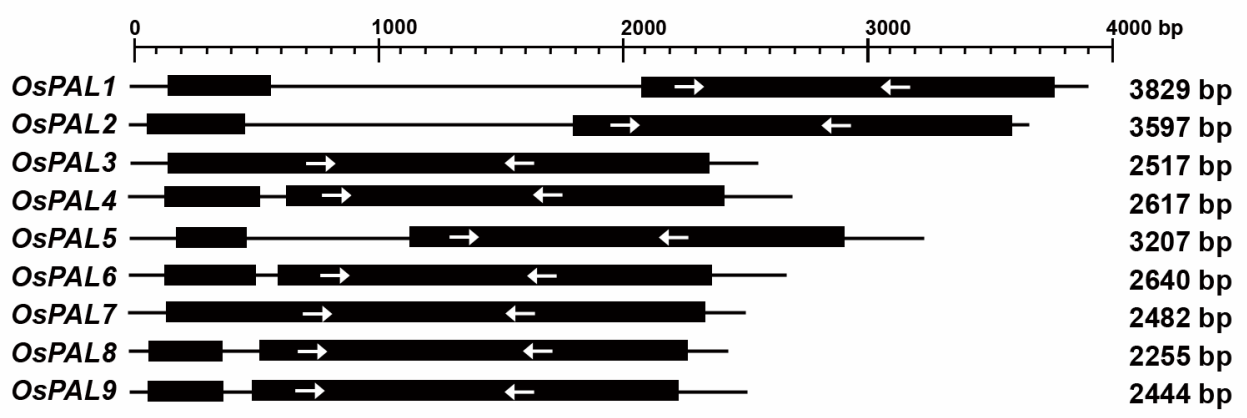


Figure 2-1: Structures of rice phenylalanine ammonia-lyase (*OsPAL*) gene family members. Black boxes indicate exons, and single lines indicate introns. The arrows represent positions of primers PAL-F and PAL-R (gene family specific primers) and their locations in the most highly conserved region among all *OsPAL* genes. The expected amplicon size for most of the genes is 923 bp; the exception is *OsPAL8* with an amplicon size of 938 bp. Lengths in bp of the predicted genes are shown at the right.

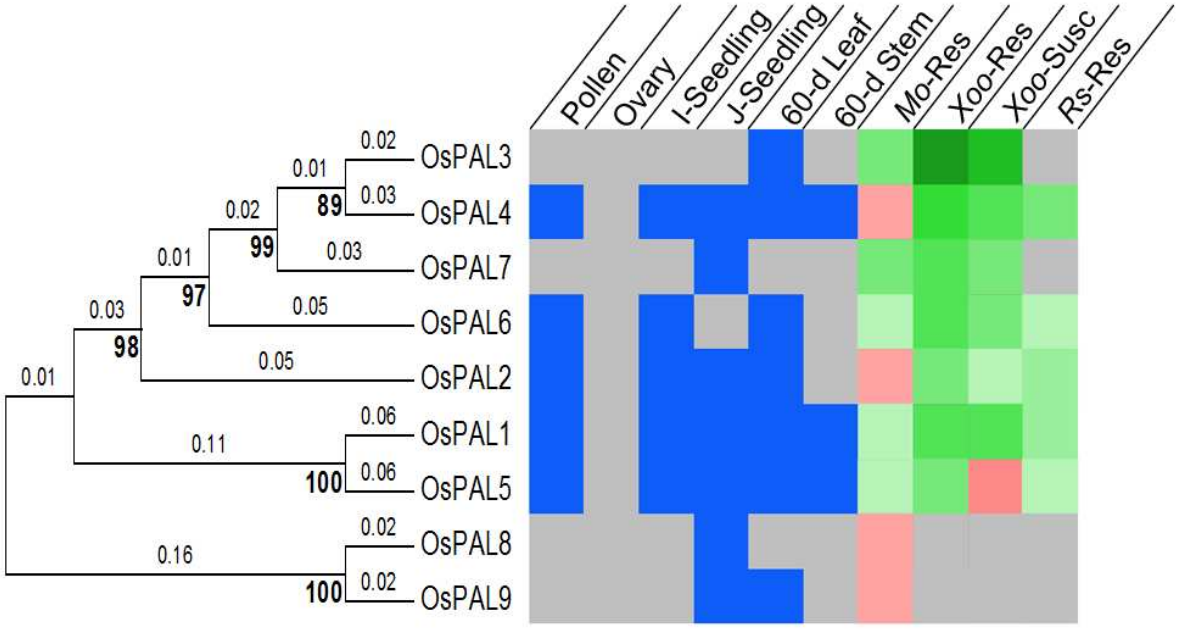


Figure 2-2: Phylogenetic relationships of Nipponbare OsPAL family of enzymes. The tree is based on OsPAL amino acid sequences obtained from the MSU rice genome database vV.7.0 and was constructed using an un-rooted neighbor-joining method. Branch lengths are given on top of each branch and represent amino acid sequence divergence. Bootstrapping values, based on 1,000 iterations, are shown underneath each branch in bold. The first six columns of the gene expression heat map are based on RNA-seq experiments, with transcript presence (blue) and absence (grey) indicated. The different tissues are: mature pollen ('Pollen') (NCBI Acc: SRR074147), mature stigma and ovaries ('Ovary') (NCBI Acc: SRR074170), indica 9311 seedling ('I-Seedling') (NCBI Acc: SRX016111), japonica Nipponbare seedling ('J-Seedling') (NCBI Acc: SRX016110), 60-day mature leaf ('60-d Leaf') (NCBI Acc: SRR074146), and 60-day mature stem ('60-d Stem') (NCBI Acc: SRR074151). The last four columns are based on log base 2 fold-change values of treatment vs. mock leaf tissue, darker shades indicating relatively higher differential expression (green – "up", red – "down"). The treatments are a resistant interaction with *M. oryzae* ('*Mo*-Res') (NCBI GEO Acc: GSE16470), a resistant interaction with *X. oryzae* pv. *oryzae* (PXO99A) ('*Xoo*-Res') (unpublished), a susceptible interaction with *X. oryzae* pv. *oryzae* (PXO99A) ('*Xoo*-Susc') (unpublished), and a resistant interaction with *R. solani* ('*Rs*-Res') (Venu et al., 2007).

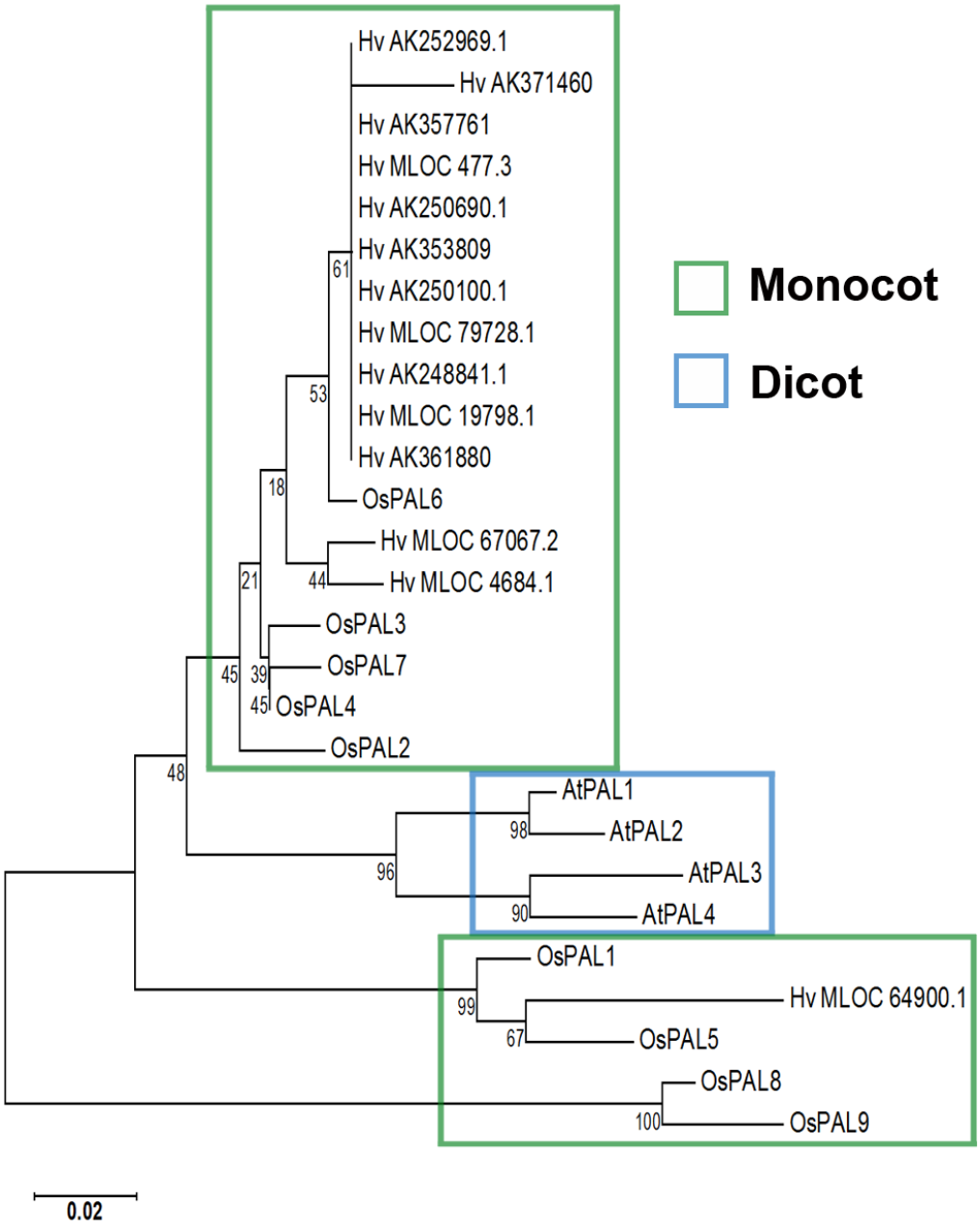


Figure 2-3: Phylogenetic tree of PAL protein families in *Arabidopsis thaliana*, *Hordeum vulgaris*, and *Oryza sativa*. The PAL amino acid sequence homologs from two cereals (rice and barely) and the dicot *Arabidopsis* were inferred using the Neighbor-Joining method. The tree is drawn to scale, with the sum of branch length = 0.64. The bootstrapping values, based on 1,000 iterations, are shown below the branches.

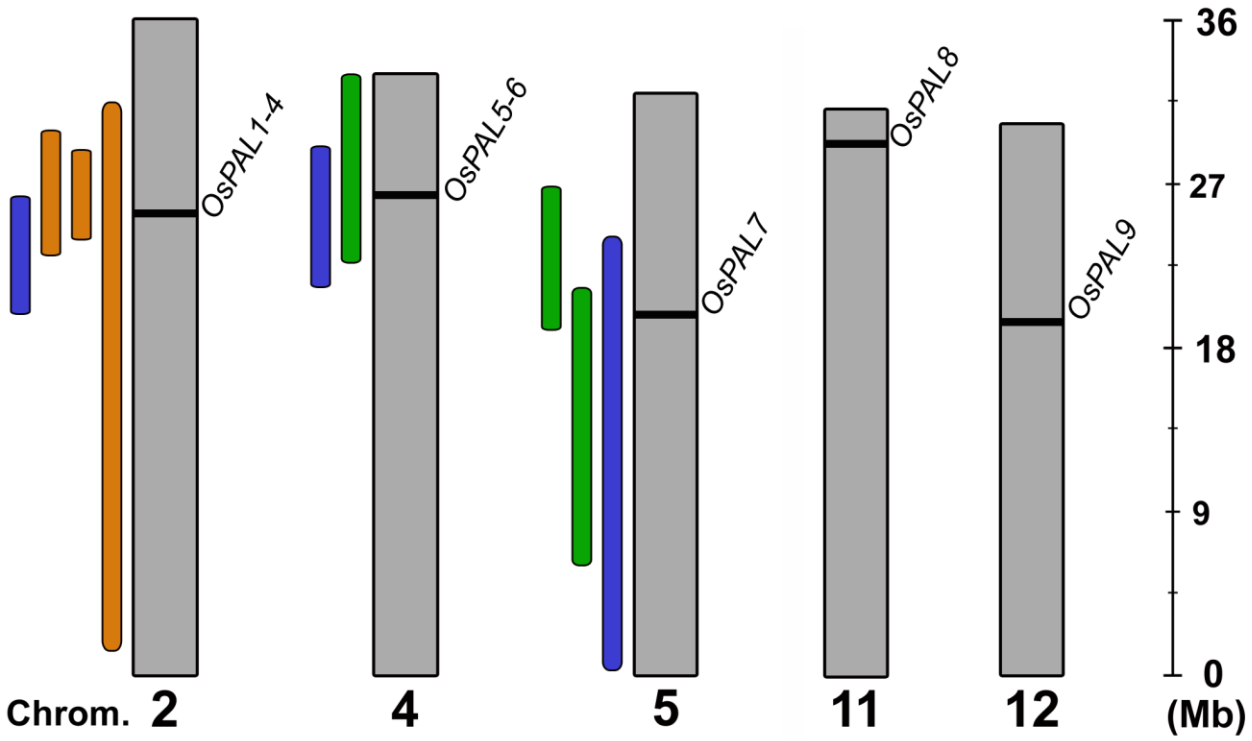


Figure 2-4: Association of disease resistance QTL regions with *OsPAL* gene family loci. Relative chromosomal positioning of the *OsPAL* genes and the positions of disease resistance QTL are shown along each chromosome, scaled in megabases (references in Supplemental Table 3). Blue bars are different *X. oryzae* pv. *oryzae* resistance QTL, green bars are *M. oryzae* resistance QTL, and orange bars are *R. solani* resistance QTL.

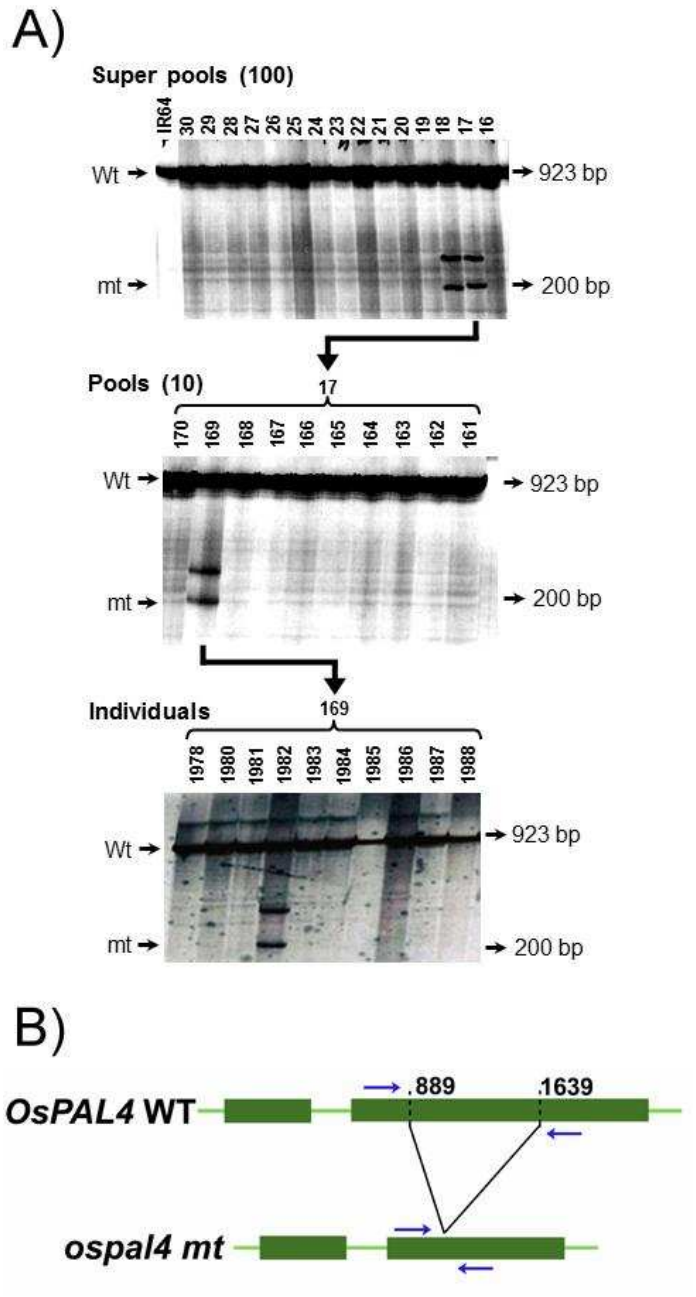


Figure 2-5: Identification of a deletion in an *OsPAL* gene family member detected in rice DEB mutants. (a) Amplification of 30 DEB 1:100 super-pools and IR64 (wild type) using the *OsPAL* gene family specific primers (PAL-F and PAL-R) detected polymorphisms in super-pool 17. The wild type band (WT) *OsPAL* is 923 bp and the mutant bands (mt) are ~200 bp. Subsequent amplification of the 10-line pools corresponding to super-pool 17 is shown. Pool 169 amplified the *OsPAL4* mutant band. Amplifications of DNA from ten individual plants corresponding to the 10-line pool, 169, were analyzed by PCR. Individual mutant 1982 contained the variant bands. (b) Model showing deleted region of *OsPAL4*.

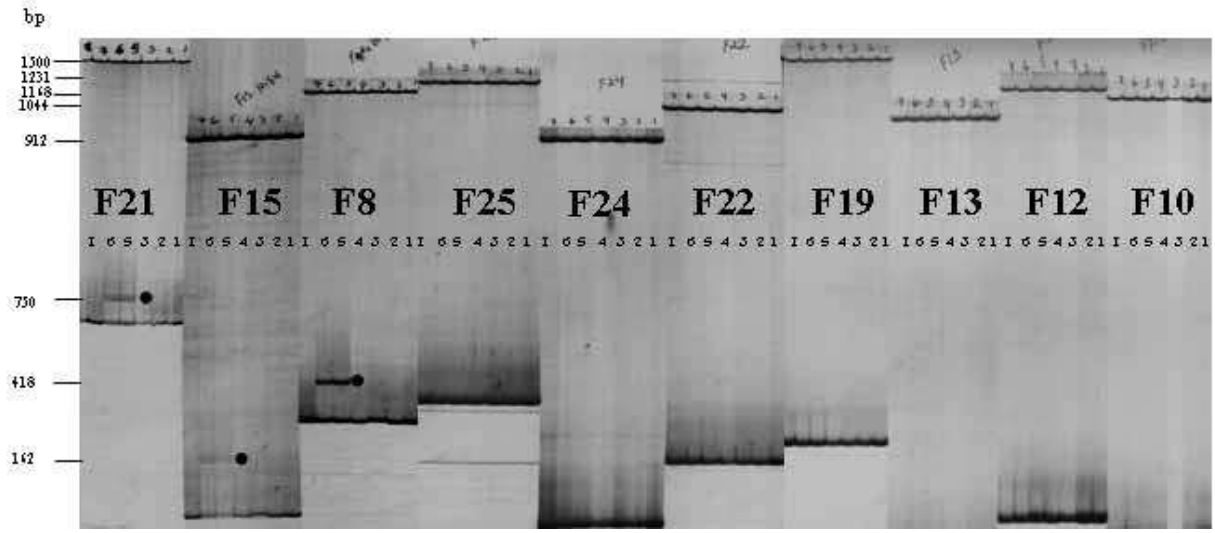


Figure 2-6: PCR confirmation of the mutated *OsPAL4* on chromosome 2. PCR products using the *OsPAL* gene specific primers (Table 2-2) were separated in 4% polyacrylamide gels. Primers were used to amplify DNA from seven plants (for each primer set, lanes1 through 4 are wild type *OsPAL4* segregants, lanes 5 and 6 are *ospal4* mutants, and I is IR64). Only the primers specific for the *OsPAL4* (F21, F15, and F8) amplified the smaller fragment from the mutant plants (● in lanes 5 and 6 for each primer set), but not from the plants with wild type allele *OsPAL4* (lanes 1-4 and IR64). Primers designed to amplify the other *OsPAL* family members (F25, F24, F22, F19, F13, F12, and F10) did not detect deletions in any plant. The PCR product size of the wild type and the mutant bands are indicated in bp at the left.

CLUSTAL O (1.2.1) multiple sequence alignment

```
Nipponbare      CTCAACGCGAACGTCACACCGTGCCTGCCGCTCCGGGGCACGATCACCGCCTCCGGTGAC    60
IR64WT          CTCAACGCGAACGTCACACCGTGCCTGCCGCTCCGGGGCACGATCACCGCCTCCGGTGAC    60
IR64mt          CTCAACGCGAACGTCACACCGTGCCTGCCGCTCCGGGGCACGATCACCGCCTCCGGTGAC    60
                ************************************************************

Nipponbare      CTCGTCCCGCTGTCCTACATTGCCGGCCTTGTCACTGGGCGCGAGAACGCCGTGGCGGTT    120
IR64WT          CTCGTCCCGCTGTCCTACATTGCCGGCCTTGTCACTGGGCGCGAGAACGCCGTGGCGGTT    120
IR64mt          CCCGTCCCGCTGTCCTACACTGCCGGC---------------------------------    87
                * ***************** *******

Nipponbare      GCACCAGATGGCAGCAAGGTGAACGCCGCTGAGGCGTTCAAGATTGCTGGCATCCAGGGC    180
IR64WT          GCACCAGATGGCAGCAAGGTGAACGCCGCTGAGGCGTTCAAGATTGCTGGCATCCAGGGC    180
IR64mt          -----------------------------------------------------------    87

Nipponbare      GGCTTCTTCGAGCTGCAGCCCAAGGAAGGCCTTGCCATGGTCAATGGCACTGCCGTGGGC    240
IR64WT          GGCTTCTTCGAGCTGCAGCCCAAGGAAGGCCTTGCCATGGTCAATGGCACTGCCGTGGGC    240
IR64mt          -----------------------------------------------------------    87

Nipponbare      TCTGGCCTTGCATCGACCGTGCTCTTTGAGGCTAACATTCTTGCCAT-CCTCGCCGAGGT    299
IR64WT          TCTGGCCTTGCATCGACCGTGCTCTTTGAGGCTAACATTCTTGCCATTCCTCGCCGAGGT    300
IR64mt          -----------------------------------------------------------    87

Nipponbare      CCTCTCGGCCGTGTTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGAC    359
IR64WT          CCTCTCGGCCGTGTTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGAC    360
IR64mt          -----------------------------------------------------------    87

Nipponbare      TCACAAGCTCAAGCACCATCCAGGACAGATCGAGGCCGCCGCCATCATGGAGCACATCTT    419
IR64WT          TCACAAGCTCAAGCACCATCCAGGACAGATCGAGGCCGCCGCCATCATGGAGCACATCTT    420
IR64mt          -----------------------------------------------------------    87

Nipponbare      GGAGGGAAGCTCCTACATGAAGCATGCCAAGAAGCTTGGTGAGCTCGACCCACTGATGAA    479
IR64WT          GGAGGGAAGCTCCTACATGAAGCATGCCAAGAAGCTTGGTGAGCTCGACCCGTTGATGAA    480
IR64mt          -----------------------------------------------------------    87

Nipponbare      GCCGAAGCAAGACCGGTACGCGCTCCGGACATCCCCACAGTGGCTCGGCCCTCAAATTGA    539
IR64WT          GCCGAAGCAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCACAGATCGA    540
IR64mt          -----------------------------------------------------------    87

Nipponbare      GGTTATCCGCGCCGCCACCAAGTCCATCGAGCGTGAGATCAACTCCGTGAACGACAACCC    599
IR64WT          GGTCATCCGCTTCGCCACCAAGTCGATCGAGCGCGAGATCAACTCCGTCAACGACAACCC    600
IR64mt          -----------------------------------------------------------    87

Nipponbare      GCTCATCGACGTCTCCCGCGGCAAGGCGCTGCACGGTGGCAACTTCCAGGGCACGCCCAT    659
IR64WT          GCTCATCGACGTCTCCCGCGGCAAGGCGCTGCACGGTGGCAACTTCCAGGGCACGCCCAT    660
IR64mt          -----------------------------------------------------------    87

Nipponbare      CGGCGTGTCCATGGACAACACCCGCCTCGCCATCGCTGCCATCGGCAAGCTCATGTTCGC    719
IR64WT          CGGCGTGTCCATGGACAACACCCGCCTCGCCATCGCTGCCATCGGCAAGCTCATGTTCGC    720
IR64mt          -----------------------------------------------------------    87
```

```
Nipponbare      GCAGTTCTCGGAGCTCGTGAACGACTTCTACAACAACGGCCTGCCATCCAACCTGTCTGG    779
IR64WT          GCAGTTCTCGGAGCTCGTGAACGACTTCTACAACAACGGCCTGCCATCCAACCTGTCTGG    780
IR64mt          ------------------------------------------------------------    87

Nipponbare      CGGTCGCAACCCGAGCTTGGACTACGGGTTCAAGGGCGCCGAGATCGCCATGGCCTCCTA    839
IR64WT          CGGTCGCAACCCGAGCTTGGACTACGGGTTCAAGGGCGCCGAGATCGCCATGGCCTCCTA    840
IR64mt          ------------------------------------------------------------CA    89
                                                                              *

Nipponbare      CTGCTCCGAGCTGCAGTTCTTGGGCAACCCAGTGACCAACCACGTCCAGAGCGCCGAGCA    899
IR64WT          CTGCTCCGAGCTGCAGTTCTTGGGCAACCCAGTGACCAACCACGTCCAGAGCGCCGAGCA    900
IR64mt          CTGCTCCGAGCTGCAGTTCTTGGGCAACCCAGTGACCAACCACGTCCAGAGCGCCGAGCA    149
                ************************************************************

Nipponbare      GCACAACCAGGACG    913
IR64WT          GCACAACCAGGACG    914
IR64mt          GCACAACCAGGACG    163
                **************
```

Figure 2-7: Characterization of the *OsPAL* deletion (Sequence alignment). Multiple sequence alignment of wild type band (1982WT) and mutant band (1982Mt) with the predicted *OsPAL* gene family members from *japonica* and *indica* genomes. The deletion was in gene family member *OsPAL4*.

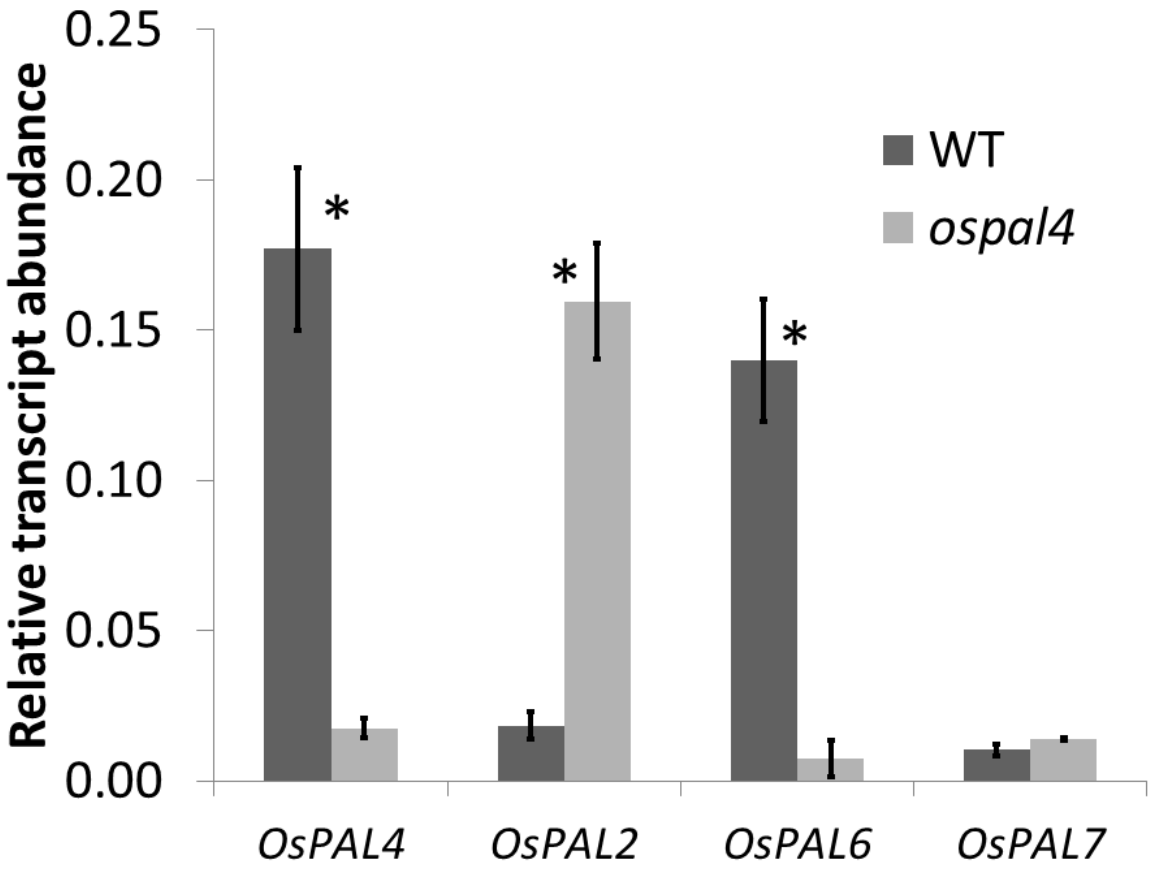


Figure 2-8: Relative expression analysis of different *OsPAL* genes due to *ospal4* mutation. Relative expression of other PAL gene family members in 14-day-old wild type and *ospal4* mutant IR64 rice leaves. Relative transcript abundance was determined using the ΔC_t method ($2^{CT(reference) - CT(target)}$) (Pfaffl, 2001). Asterisks indicate a statistically significant difference (Student's t-test, $p<0.05$).

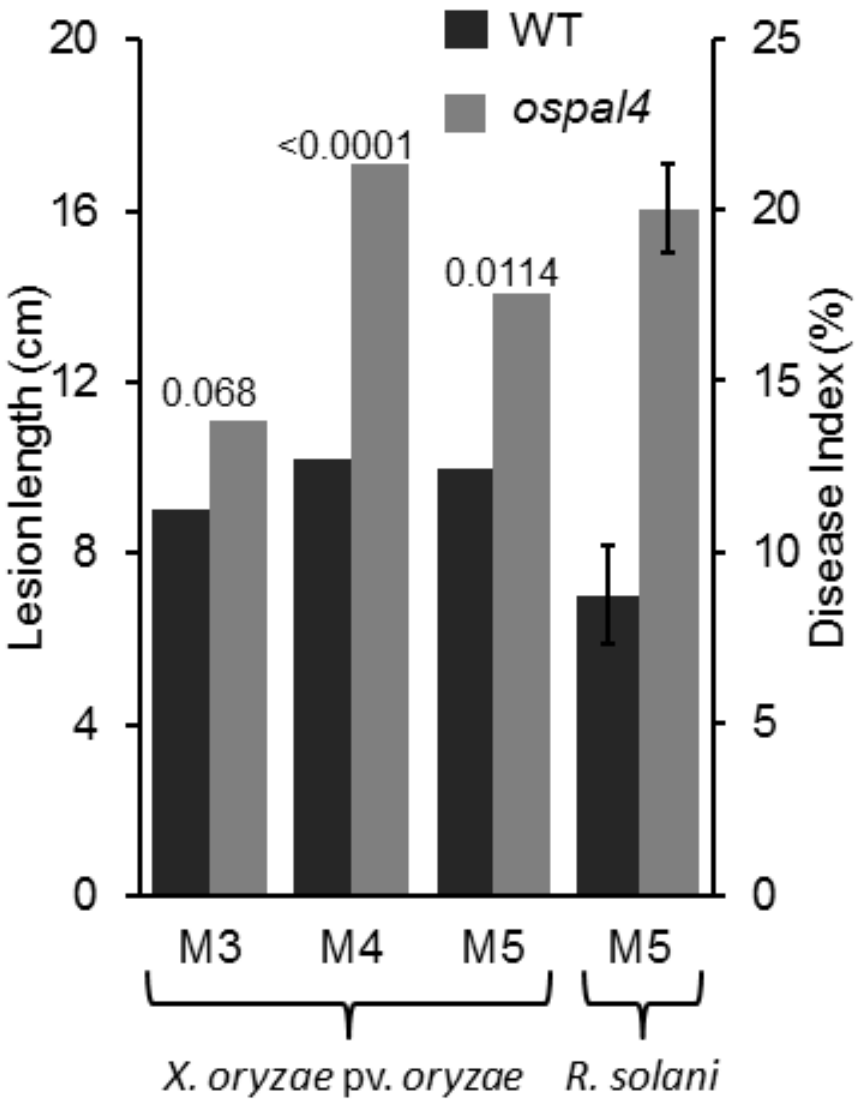


Figure 2-9: The *ospal4* mutant exhibits enhanced susceptibility to the virulent strain of *X. oryzae* pv *oryzae* and *R. solani.* 182 M3 rice progeny of the heterozygous *ospal4* mutant 1982 rice line and 85 M4 progeny, and 25 M5 progeny were genotyped for the presence of the *OsPAL4* deletion and grouped into two groups (*OsPAL4* segregants and *ospal4* mutants). Those plants were inoculated with a virulent *Xoo* isolate PXO339. Plants were scored for disease (lesion length in cm). Analysis was done using GLM procedure in SAS (P-values shown). Responses of M5 plants to *R. solani* are reported as Disease index (lesion length (cm) / plant height x 100) of plants at 14 days after inoculation with *R. solani*. Bars represent SE +/- mean, n = 29; 18 WT segregants, 11 mutants (Student's t-test, P<0.0001).

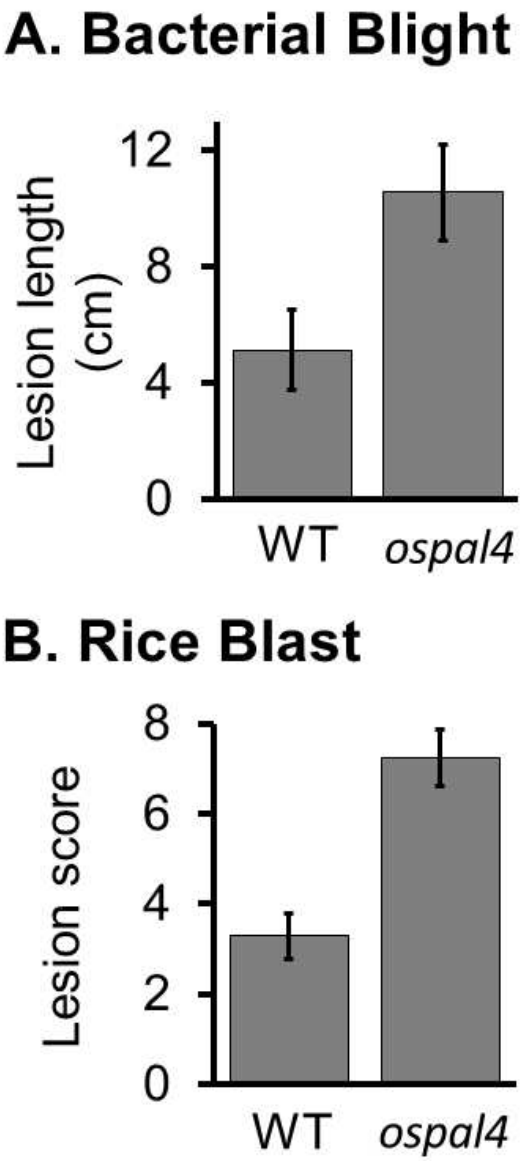


Figure 2-10: The *OsPAL4* deletion is correlated with increased disease susceptibility to *M. oryzae* and *X. oryzae* pv. *oryzae* in the F1 (IR64 x *ospal4*) generation. A) Lesion lengths (cm) after inoculation with *X. oryzae* pv. *oryzae* strain PXO339 (n = 28; 10 WT segregants, 18 mutants). Standard error bars shown (Student's t-test, P<0.0001). B) Lesion score of plants inoculated with *M. oryzae* (P06-6). Disease was assayed 10 days post-inoculation (n = 23; 4 WT segregants, and 19 mutants). Standard error bars shown (Student's t-test, P=0.012).

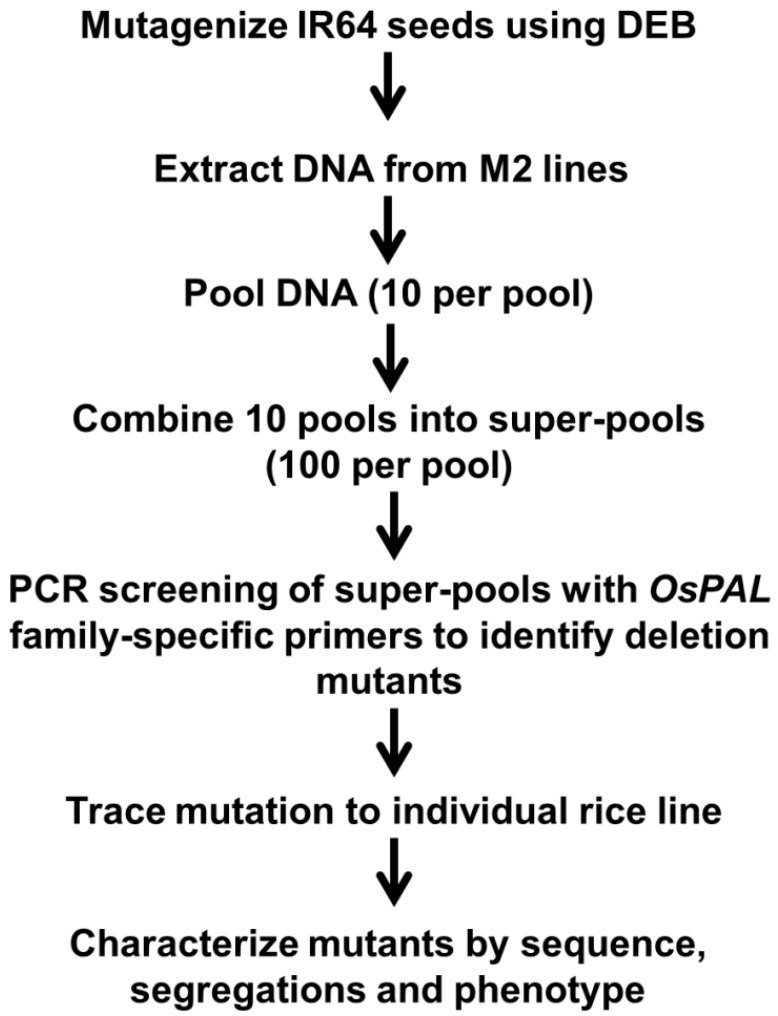


Figure 2-11: Pooling and screening strategy: Flow chart of the pooling and PCR screening processes showing the steps from mutagenesis to the characterization of the gene deletion in the individual mutant line. The PCR-based screening strategy used for deletion mutants involved amplification of both the wild type and mutant gene, because the primer pair is specific to the targeted *OsPAL4* locus. In a population of 3,000 DEB-induced rice deletion lines, we identified one *OsPAL* mutant. In rice where the mean size of genes is ~2.6 Kb, ~460,000 insertions are required for a 95% chance of mutating any gene (Hirochika et al.., 2004). This is approximately 2.6 times more than the number of *Arabidopsis* insertions required to detect a mutated gene with the same probability. The estimation of the population size of rice DEB mutants in this study to detect deletions with a 95% or 99% probability is 2.6 times more than the population size of the *Arabidopsis* Fast Neutron (FN) mutants estimated for the same probabilities (Li et al.., 2001).

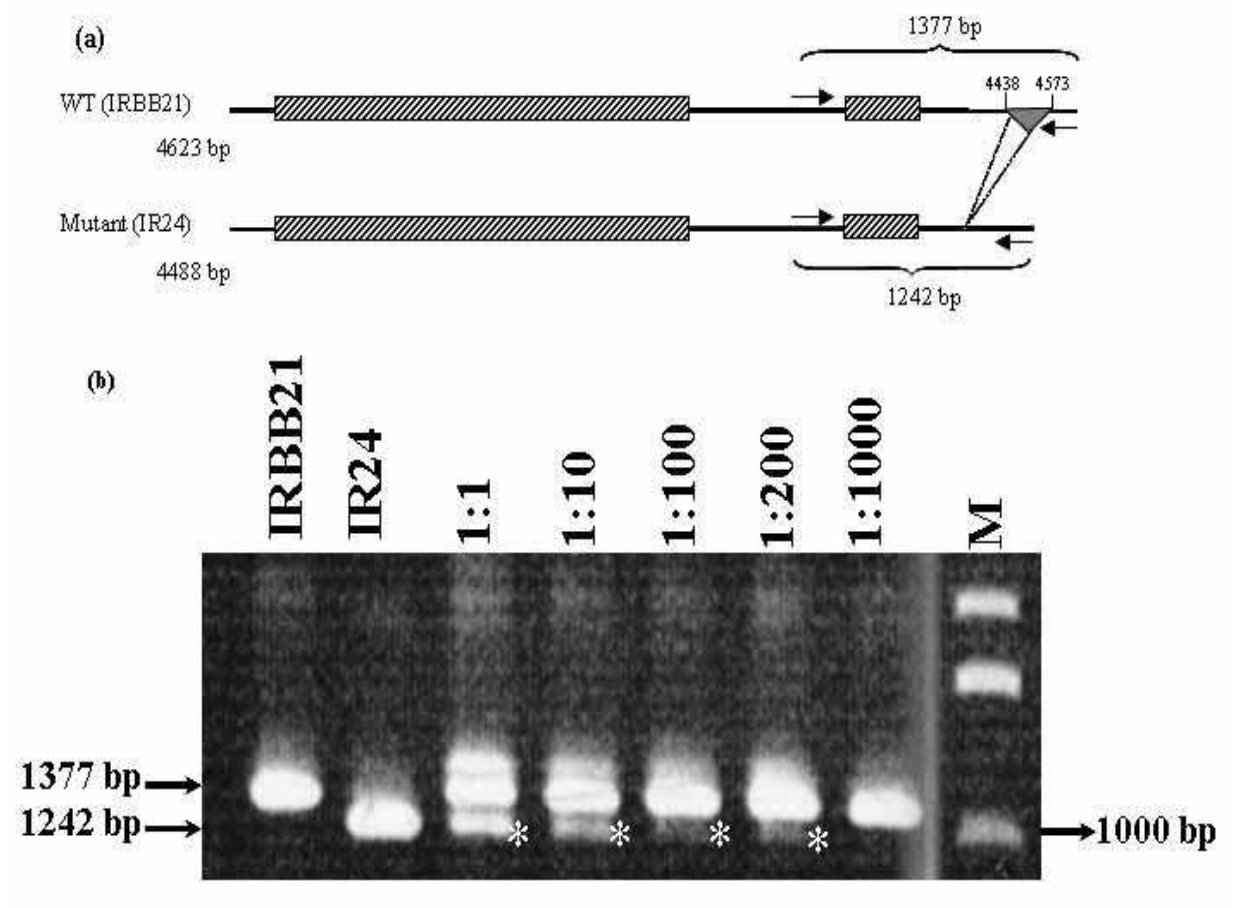


Figure 2-12: Reconstruction experiments using a natural 135 bp deletion in the *Xa21* gene of rice line IR24. (a) Wild type copy of the *Xa21* gene present in IRBB21 rice line. The gene is 4623 bp long and has two exons indicated by the boxes. The arrows show positions of the primers, the gray triangle indicates the position of the natural deletion in the IR24. The distance between the two primers in the wild type is 1377 bp and in the mutant is 1242 bp. (b) PCR was performed to determine the optimal pool size for the rice mutants. Mutant DNA was pooled with wild type DNA solutions to generate pools with ratio of 1:10, 1:100, 1:200, and 1:1000 (mutant DNA:wild type DNA). These DNA pools were used as a template for the PCR reaction (lanes 3-7) together with IRBB21 (lane 1) and IR24 (lane 2) DNA. PCR products were analyzed by agarose gel electrophoresis followed by ethidium bromide staining. 1Kb plus DNA ladder (M) was loaded in lane 8. The IR24 mutant band was detected in all the pools through 1:200.

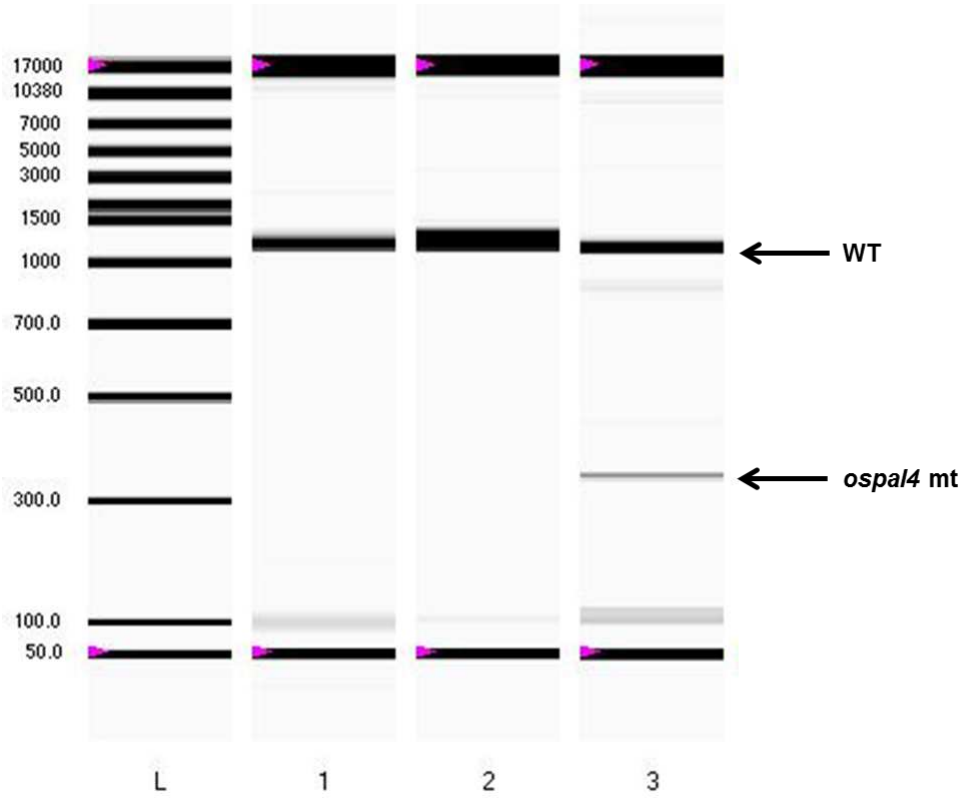


Figure 2-13: *OsPAL4* PCR products resolved using a Bio-Rad Experion. Ladder in lane L, wild type segregants in lanes 1 and 2 and a representative *ospal4* mutant in lane 3 containing the 400 bp diagnostic product. Primers used were F8 and PAL-R (Table 2-3).

References

Achnine L, Blancaflor EB, Rasmussen S, et al (2004) Colocalization of L-Phenylalanine Ammonia-Lyase and Cinnamate 4-Hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell 16:3098–3109. doi: 10.1105/tpc.104.024406

Bagali PG, Hittalmani S, Srinivasachary, et al (1998) Genetic markers associated with field resistance to leaf and neck blast across locations in rice (Oryza sativa L.). Rice Genet Newsl 15:128–131

Ballini E, Morel J-B, Droc G, et al (2008) A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. Mol Plant Microbe Interact 21:859–68. doi: 10.1094/MPMI-21-7-0859

Bate NJ, Orr J, Ni W, et al (1994) Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate-determining step in natural product synthesis. Proc Natl Acad Sci USA 91:7608–7612

Becker-Andre M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate:CoA ligase genes in potato. J Biol Chem 266:8551–8559

Beckers GJM, Spoel SH (2006) Fine-tuning plant defence signalling: salicylate versus jasmonate. Plant Biol (Stuttg) 8:1–10. doi: 10.1055/s-2005-872705

Bhat WW, Razdan S, Rana S, et al (2014) A phenylalanine ammonia-lyase ortholog (PkPAL1) from Picrorhiza kurrooa Royle ex. Benth: molecular cloning, promoter analysis and response to biotic and abiotic elicitors. Gene 547:245–56. doi: 10.1016/j.gene.2014.06.046

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54:519–46. doi: 10.1146/annurev.arplant.54.031902.134938

Bowles DJ (1990) Defense-related proteins in higher plants. Annu Rev Biochem 59:873–907

Boyd L a, Ridout C, O'Sullivan DM, et al (2013) Plant-pathogen interactions: disease resistance in modern agriculture. Trends Genet 29:233–40. doi: 10.1016/j.tig.2012.10.011

Calabrese JC, Jordan DB, Boodhoo A, et al (2004) Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis. Biochemistry 43:11403–16. doi: 10.1021/bi049053

Chen F, Dixon R a (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol 25:759–61. doi: 10.1038/nbt1316

Chen F, Srinivasa Reddy MS, Temple S, et al (2006) Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wall-bound ferulic acid in alfalfa (Medicago sativa L.). Plant J 48:113–24. doi: 10.1111/j.1365-313X.2006.02857.x

Chen L-N, Yang Y, Yan C-Q, et al (2012) Identification of quantitative trait loci for bacterial blight resistance derived from Oryza meyeriana and agronomic traits in recombinant inbred lines of Oryza sativa. J Phytopathol 160:461–468. doi: 10.1111/j.1439-0434.2012.01931.x

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124:803–14. doi: 10.1016/j.cell.2006.02.008

Coquoz J, Buchala A, Metraux J (1998) The biosynthesis of salicylic acid in potato plants. Plant Physiol 117:1095–101

Cramer CL, Ryder TB, Bell JN, et al (1985) Rapid switching of plant gene-expression induced by fungal elicitor. Science (80) 227:1240–3. doi: 10.1126/science.227.4691.1240

Craven-Bartle B, Pascual MB, Cánovas FM, Avila C (2013) A Myb transcription factor regulates genes of the phenylalanine pathway in maritime pine. Plant J 74:755–66. doi: 10.1111/tpj.12158

Cui Y, Magill J, Frederiksen R, Magill C (1996) Chalcone synthase and phenylalanine ammonia-lyase mRNA levels following exposure of sorghum seedlings to three fungal pathogens. Physiol Mol Plant Pathol 49:187–199. doi: 10.1006/pmpp.1996.0048

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution (N Y) 39:783–791.

Fu D, Chen L, Yu G, et al (2011) QTL mapping of sheath blight resistance in a deep-water rice cultivar. Euphytica 180:209–218. doi: 10.1007/s10681-011-0366-5

Fukuoka S, Okuno K (2001) QTL analysis and mapping of pi21 , a recessive gene for field resistance to rice blast in Japanese upland rice. Theor Appl Genet 103:185–190. doi: 10.1007/s001220100611

Gassmann W, Bhattacharjee S (2012) Effector-triggered immunity signaling: from gene-for-gene pathways to protein-protein interaction networks. Mol Plant Microbe Interact 25:862–868

Giberti S, Bertea CM, Narayana R, et al (2012) Two phenylalanine ammonia lyase isoforms are involved in the elicitor-induced response of rice to the fungal pathogen Magnaporthe oryzae. J Plant Physiol 169:249–254. doi: 10.1016/j.jplph.2011.10.008

Gruner K, Griebel T, Návarová H, et al (2013) Reprogramming of plants during systemic acquired resistance. Front Plant Sci 4:1–28. doi: 10.3389/fpls.2013.00252

Gupta SK, Rai AK, Kanwar SS, et al (2012) The single functional blast resistance gene Pi54 activates a complex defence mechanism in rice. J Exp Bot 63:757–72. doi: 10.1093/jxb/err297

Hahlbrock K, Scheel D (1989) Physiology and molecular biology of phenylpropanoid metabolism. Annu Rev Plant Physiol Plant Mol Biol 40:347–69

Havir EA, Hanson KR (1973) L-phenylalanine ammonia-lyase (maize and potato). Evidence that the enzyme is composed of four subunits. Biochemistry 12:1583–91

Jain M, Nijhawan A, Tyagi AK, Khurana JP (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. Biochem Biophys Res Commun 345:646–51. doi: 10.1016/j.bbrc.2006.04.140

Jia Y, Liu G, Park D, Yang Y (2013) Inoculation and scoring methods for rice sheath blight disease. In: Yang Y (ed) Rice Protoc. Humana Press, Totowa, NJ, pp 257–268

Jia Y, Valent B, Lee FN (2003) Determination of host responses to Magnaporthe grisea on detached rice leaves. Plant Dis 87:129–133

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444:323–329. doi: 10.1038/nature05286

Joos HJ, Hahlbrock K (1992) Phenylalanine ammonia-lyase in potato (Solanum tuberosum L) - genomic complexity, structural comparison of 2 selected genes and modes of expression. Eur J Biochem 204:621–629

Kauffman HE, Reddy APK, Hsiek SP V, Marca SD (1973) An improved technique for evaluating resistance of race varieties to Xanthomonas oryzae. Plant Dis Rep 57:537–541

Kim DS, Hwang BK (2014) An important role of the pepper phenylalanine ammonia-lyase gene (PAL1) in salicylic acid-dependent signalling of the defence response to microbial pathogens. J Exp Bot 65:2295–306. doi: 10.1093/jxb/eru109

Kim Y, Tsuda K, Igarashi D, et al (2014) Signaling mechanisms underlying the robustness and tunability of the plant immune network. Cell Host Microbe 15:84–94. doi: 10.1016/j.chom.2013.12.002

Kou Y, Wang S (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. Curr Opin Plant Biol 13:181–5. doi: 10.1016/j.pbi.2009.12.010

La Camera S, Gouzerh G, Dhondt S, et al (2004) Metabolic reprogramming in plant innate immunity : the contributions of phenylpropanoid and oxylipin pathways. Immunol Rev 198:267–284

Lee H, Leon J, Raskin I (1995) Biosynthesis and metabolism of salicylic acid. Proc Natl Acad Sci USA 92:4076–4079

Li B, Liu B, Shan C, et al (2013) Antibacterial activity of two chitosan solutions and their effect on rice bacterial leaf blight and leaf streak. Pest Manag Sci 69:312–320. doi: 10.1002/ps.3399

Li ZK, Luo LJ, Mei HW, et al (1999) A "defeated" rice resistance gene acts as a QTL against a virulent strain of Xanthomonas oryzae pv. oryzae. Mol Gen Genet 261:58–63

Liao Y, Li H, Kreuzaler F, Fischer R (1996) Nucleotide sequence of one of two tandem genes (Accesion No.X99705) encoding phenylalanine ammonia-lyase in Triticum aestivum. Plant Physiol 112:1398

Liu B, Zhang S, Zhu X, et al (2004) Candidate defense genes as predictors of quantitative blast resistance in rice. Mol Plant Microbe Interact 17:1146–52. doi: 10.1094/MPMI.2004.17.10.1146

Logemann E, Parniske M, Hahlbrock K (1995) Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. Proc Natl Acad Sci U S A 92:5905–5909. doi: DOI 10.1073/pnas.92.13.5905

Maher E a, Bate NJ, Ni W, et al (1994) Increased disease susceptibility of transgenic tobacco plants with suppressed levels of preformed phenylpropanoid products. Proc Natl Acad Sci U S A 91:7802–6

Malamy J, Carr JP, Klessig DF, Raskin I (2014) Salicylic acid : a likely endogenous signal in the resistance response of tobacco to viral infection. Science (80-) 250:1002–1004

Manosalva PM, Davidson RM, Liu B, et al (2009) A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. Plant Physiol 149:286–96. doi: 10.1104/pp.108.128348

Mauch-Mani B, Slusarenko a. J (1996) Production of Salicylic Acid Precursors Is a Major Function of Phenylalanine Ammonia-Lyase in the Resistance of Arabidopsis to Peronospora parasitica. Plant Cell 8:203–212. doi: 10.1105/tpc.8.2.203

Moffitt MC, Louie G V, Bowman ME, et al (2007) Discovery of two cyanobacterial phenylalanine ammonia lyases: kinetic and structural characterization. Biochemistry 46:1004–1012

Naoumkina MA, Zhao Q, Gallego-giraldo L, et al (2010) Genome-wide analysis of phenylpropanoid defence pathways. Mol Plant Pathol 11:829–846. doi: 10.1111/J.1364-3703.2010.00648.X

Ouyang S, Zhu W, Hamilton J, et al (2007) The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res 35:D883–7. doi: 10.1093/nar/gkl976

Pallas JA, Paiva NL, Lamb C, Dixon RA (1996) Tobacco plants epigenetically suppressed in phenylalanine ammonia-lyase expression do not develop systemic acquired resistance in response to infection by tobacco mosaic virus. Plant J 10:281–293

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:2002–2007

Pinson SRM, Capdevielle FM, Oard JH (2005) Confirming QTLs and finding additional loci conditioning sheath blight resistance in rice using recombinant inbred lines. Crop Sci 45:503–510

Ramalingam J, Vera Cruz CM, Kukreja K, et al (2003) Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. Mol Plant Microbe Interact 16:14–24. doi: 10.1094/MPMI.2003.16.1.14

Rawal HC, Singh NK, Sharma TR (2013) Conservation, divergence, and genome-wide distribution of PAL and POX A gene families in plants. Int J Genomics 2013:1–10. doi: 10.1155/2013/678969

Reichert AI, He X-Z, Dixon R a (2009) Phenylalanine ammonia-lyase (PAL) from tobacco (Nicotiana tabacum): characterization of the four tobacco PAL genes and active heterotetrameric enzymes. Biochem J 424:233–42. doi: 10.1042/BJ20090620

Reimers P, Leach JE (1991) Race-specific resistance to Xanthomonas oryzae pv . oryzae conferred by bacterial blight resistance gene Xa-10 in rice (Oryza sativa) involves accumulation of a lignin-like substance in host tissues. Physiol Mol Plant Pathol 38:39–55

Rensing S a (2014) Gene duplication as a driver of plant morphogenetic evolution. Curr Opin Plant Biol 17:43–8. doi: 10.1016/j.pbi.2013.11.002

Riaz A, Riaz A, Rattu AUR, et al (2014) Phenylalanine ammonia-lyase (PAL) and peroxidase activity in brown rust infected tissues of Pakistani wheat cultivars. Pakistan J Bot 46:1101–1107

Ride JP (1983) Cell walls and other structural barriers in defence. Biochem Plant Pathol (Callow, JA, ed) 215–236

Ritter H, Schulz GE (2004) Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. Plant Cell 16:3426–3436

Rohde A, Morreel K, Ralph J, et al (2004) Molecular Phenotyping of the pal1 and pal2 Mutants of Arabidopsis thaliana Reveals Far-Reaching Consequences on Phenylpropanoid , Amino Acid , and Carbohydrate Metabolism. 16:2749–2771. doi: 10.1105/tpc.104.023705.phospho-

Saghai-Maroof M a, Soliman KM, Jorgensen R a, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–25

Sana TR, Fischer S, Wohlgemuth G, Fiehn O (2010) Metabolomic and transcriptomic analysis of the rice response to the bacterial blight pathogen Xanthomonas oryzae pv. oryzae. Metabolomics 6:451–465. doi: 10.1007/s11306-010-0218-7

Sarma AD, Sharma R (1999) Purification and characterization of UV-B induced phenylalanine ammonia-lyase from rice seedlings. Phytochemistry 50:729–737

Savary S, Mila a, Willocquet L, et al (2011) Risk factors for crop health under global change and agricultural shifts: a framework of analyses using rice in tropical and subtropical Asia as a model. Phytopathology 101:696–709. doi: 10.1094/PHYTO-07-10-0183

Schatz MC, Maron LG, Stein JC, et al (2014) New whole genome de novo assemblies of three divergent strains of rice (O . sativa) documents novel gene space of aus and indica. BioRxiv http://biorxiv.org/.

Schuler GD (1997) Sequence mapping by electronic PCR. Genome Res 7:541–50

Shang Q-M, Li L, Dong C-J (2012) Multiple tandem duplication of the phenylalanine ammonia-lyase genes in Cucumis sativus L. Planta 236:1093–105. doi: 10.1007/s00425-012-1659-1

Shiraishi T, Yamada T, Nicholson RL, Kunoh H (1995) Phenylalanine ammonia-lyase in barley: activity enhancement in response to Erysiphe graminis f.sp. hordei (race 1) a pathogen, and Erysiphe-pisi, a nonpathogen. Physiol Mol Plant Pathol 46:153–162. doi: 10.1006/pmpp.1995.1012

Sommssich IE, Hahlbrock K (1998) Pathogen defence in plants- a paradigm of biological complexity. Trends Plant Sci 3:86–90.

Stern DL (2013) The genetic causes of convergent evolution. Nat Rev Genet 14:751–64. doi: 10.1038/nrg3483

Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–9. doi: 10.1093/molbev/mst197

Tanaka N, Che F-S, Watanabe N, et al (2003) Flagellin from an incompatible strain of Acidovorax avenae mediates H2O2 generation accompanying hypersensitive cell death and expression of PAL, Cht-1, and PBZ1, but not of Lox in rice. Mol Plant Microbe Interact 16:422–8. doi: 10.1094/MPMI.2003.16.5.422

Chapter 3: Shared *cis*-regulatory architecture across defense response genes predicts broad spectrum quantitative resistance in rice

Introduction

Food security is now a greater issue across the world as the global population increases, arable farmland decreases, and staple food dependency continues to rise (FAO 2010). Rice (*Oryza sativa*) feeds more than 60% of the world's population (Sharma et al., 2012). Thus, assuring stability of this crop requires attention to factors that hinder plant growth and health, such as diseases. Depending upon environmental conditions, the three main diseases of rice, *Magnaporthe oryzae*, *Rhizoctonia solani*, and *Xanthomonas oryzae* pv. *oryzae* can cause up to 30%, 25%, and 70% yield loss, respectively (Banniza & Holderness, 2001; Chen et al., 2006; Goto, 1992; Jia et al., 2013; Reddy, 1979; Liu et al., 2010; Ou, 1985). Identifying and understanding the genetic and molecular mechanisms of rice defense against these pathogens are important steps in the development of durable varieties.

Plant-microbe interactions occur through a two-tiered response system, starting with plant cell recognition of conserved Pathogen-Associated Molecular Patterns (PAMPs) by cell surface receptors known as Pattern Recognition Receptors (PRRs) (Jones & Dangl, 2006). This elicits a downstream defense response (DR) known as Pattern-Triggered Immunity (PTI). The quantitative nature of PTI involves many diverse metabolic functions that work together to stave off infection. This process involves multiple plant genes, and is also known as basal or

quantitative resistance. Pathogens are able to overcome PTI through the secretion of effector proteins, which inhibit aspects of the DR or increase host susceptibility, allowing for stronger infection. If plants contain resistance (*R*) genes, some of which encode nucleotide-binding, leucine-rich repeat (NBS-LRR) domain proteins, the presence or activity of specific pathogen effectors is detected, and a defense response known as Effector Triggered Immunity (ETI) is activated (Kou & Wang, 2010; H. Zhang & Wang, 2013). In crops such as rice, breeding for ETI in elite varieties is the main strategy for developing disease resistant cultivars; however, this single gene resistance is frequently unstable, only lasting a few seasons, because the pathogen evolves to circumvent the R-gene mechanism (Ballini et al., 2008).

Strengthening the effectiveness of quantitative, durable resistance was discovered to be influenced by the diverse functions of downstream DR genes (Liu et al., 2004; Ramalingam et al., 2003). The activities of DR genes are non-species specific, meaning they contribute to Broad Spectrum Resistance (BSR) against multiple pathogens (Ke et al., 2017; Liu et al., 2004; Ramalingam et al., 2003). The functions of DR gene products include pathways in Jasmonic Acid (JA), Salicylic Acid (SA), Indole-3-acetic acid (IAA), Mitogen-Activated Protein Kinase (MAPK) signaling, thiamine biosynthesis, carbon metabolism, oxidative bursts, phenylpropanoid biosynthesis, and other processes (Fu et al., 2011; Hulsmans et al., 2016; Shah, 2009; Torres et al., 2006; Wang et al., 2006; Zipfel & Robatzek, 2010). The genomic regions associated with this type of resistance, Quantitative Trait Loci (QTL), contain DR genes or DR gene families as seen in rice, pepper, common bean, and wheat (Davidson et al., 2009; Faris et al., 1999; Geffroy et al., 2000; Kou & Wang, 2010, 2012; Liu et al., 2004; Ramalingam et al., 2003; Wu et al., 2004). Each QTL imparts a partial contribution to resistance, and pyramiding multiple QTL to include as many functional DR genes as possible predicts durable resistance (Boyd, 2006). Pyramiding

DR gene-containing QTL has been successful in producing durable disease resistance in wheat and rice (Liu et al., 2004; Manosalva et al., 2009; Singh et al., 2011; Vagndorf et al., 2017).

In several cases, the functional difference between the resistant and susceptible DR gene haplotypes of a QTL are not in the coding region, but are likely due to polymorphisms in the promoters. The germin-like protein, *OsGLP8-6*, is one family member of 12 *OsGLP8* genes found on a QTL for rice blast resistance on chromosome 8 (Manosalva et al., 2009). This DR gene is involved in BSR, thus classified as a BS-DR gene, with expression of *OsGLP8-6* enhancing resistance to the pathogens *M. oryzae*, *R. solani*, and *X. oryzae* (Davidson et al., 2010, 2009; Manosalva et al., 2009). An 856 bp promoter insertion that contains known defense-responsive *cis*-elements is present in the resistant haplotype of *OsGLP8-6,* and the gene shows faster and higher expression relative to the susceptible haplotype (Davidson et al., 2010). Another BS-DR gene, *OsOXO4*, is an oxalate oxidase and a member of a gene family of four in a QTL for resistance to *M. oryzae;* this gene also contributes to resistance to *R. solani* (Carrillo, Goodwin, Leach, Leung, & Vera Cruz, 2009; Karmakar et al., 2016). The resistant *OsOXO4* promoter haplotype also contains an insertion (26 bp) with known defense *cis*-elements (Carrillo et al., 2009). A rice IAA-amido synthetase BS-DR gene, *OsGH3-2,* is within a QTL for resistance against both *X. oryzae* and *M. oryzae*, and contains distinct promoter differences between resistant and susceptible alleles (Fu et al., 2011). Promoter differences are also present between resistant and susceptible haplotypes in the CCCH-type zinc finger nucleic acid-binding protein, *OsC3H12,* which contributes to resistance to *X. oryzae* (Deng et al., 2012). In all of these examples, the differences between resistance and susceptibility are polymorphisms that affect transcription, not protein function, suggesting that the timing and intensity of BS-DR gene regulation are key to effective basal resistance. The *cis*-element structure of a promoter is

important for gene responsiveness to an individual stimulus. For instance, in *Arabidopsis*, a specific position and combination of certain *cis*-elements affect the transcription of a subset of genes during environmental stress (Zou et al., 2011). Thus, identifying promoter elements associated with highly responsive BS-DR genes may provide genetic markers to facilitate accumulation of effective QTL.

The BS-DR genes described above are large effect genes, meaning disease phenotype is noticeably altered if the gene is mutated or overexpressed. However, the intrinsic nature of BS-DR genes is quantitative, so many contributing genes will have a small effect that is difficult to detect. Understanding BS-DR genes and their regulation on a genomic scale, rather than restricted by QTL, can lead to a more comprehensive set of markers for broad spectrum, quantitative resistance. Here, we identify promoter motifs and promoter architectures that are shared across co-expressed BS-DR genes, and identify patterns that may predict novel BS-DR genes throughout the genome.

Results

Identifying co-expressed BS-DR genes

- *Co-expression analysis*

Gene expression data compiled from 44 studies (Table S3-1: Supplemental) was used to identify condition-dependent co-expressed gene clusters, with the condition being biotic stress DR elicitors. All included studies measured transcriptomes of rice responses to the diseases bacterial blight (*X. oryzae*), rice blast (*M. oryzae*), and sheath blight (*R. solani*), as well as the chemical defense elicitors benzothiadiazole (BTH), jasmonic acid, cellulose, and chitin (Table S3-1: Supplemental). Each study was converted to a log base 2 differential expression value of

treatment relative to control, normalized, and scaled to then allow comparison across studies (Figure 3-1).

Average linkage hierarchical clustering of a distance matrix from Pearson correlation was used to construct a gene dendrogram of 14,688 loci (Figure 3-2). Identifying the optimal parameters for tree-cutting of the dendrogram included optimizing the scores of Dunn Index, Connectivity, and Silhouette Width, as well as choosing the tree-cutting method that produced the most clusters that harbor BS-DR characteristics (Table 3-1, Figure 3-3). BS-DR gene cluster characteristics were identified based on two criteria, i.e., enrichment of DR-related Gene Ontology (GO) terms (DR-GO terms) in each cluster (Table S3-2: Supplemental), and enrichment of DR genes functionally associated (FA) with plant defense (FA-DR genes) in each cluster (Table S3-3: Supplemental). DR-GO terms were chosen based on the pertinence of the term to plant stress responses. The enrichment tests for FA-DR genes and DR-GO terms, respectively, were done using Benjamini-Hochberg (BH) correction with a False Discovery Rate (FDR) of 0.05 (Table S3-4: Supplemental, Table S3-5: Supplemental). The tree-cutting parameters that produced the most BS-DR gene clusters created 65 co-expressed gene clusters with a mean cluster size of 226 genes (Table 3-1).

- *Identification of the BS-DR gene cluster*

From the 65 co-expressed clusters, 17 were enriched in DR-GO terms (Table 3-2), and, of these 17 clusters, one containing 385 genes, deemed as the "BS-DR cluster" also was enriched in FA-DR genes (Figure 3-2, Table S3-6: Supplemental). We hypothesized that co-expression of genes within this BS-DR cluster suggests common regulation during DR, and therefore, their

promoters might contain common sequence patterns. The rest of this analysis focuses on identifying promoter elements associated with this single BS-DR cluster given in Figure 3-2.

Promoterome annotation for two representative rice varieties

Position in the promoter relative to the downstream gene is important for deducing functionality of putative *cis*-elements (Zou et al., 2011). A first step in identifying similar promoter motif architecture specific to BS-DR genes was to ensure correct alignment to the transcription start site (TSS) across the genome for accurate cross-comparison of promoters. The occurrences of A, T, C, and G for every promoter from the proposed TSS to 2 kb upstream were counted (Figure 3-4). The occurrence of A/T is consistently much higher than C/G from about -400 to -2 kb, but within -35 to -24 of the TSS the A/T counts increase for both Nipponbare and IR64. This is the TATA box site, and it occurs in 24% of total genes in both varieties, a value comparable to the expected value of 19% in rice (Civáň & Švec, 2009). Cytosine was located right before the TSS, at position -1, in about 47% of promoters, and is indicative of transcribed genes (Troukhan et al., 2009). Thus, the promoterome for both IR64 and Nipponbare are aligned to the TSS, and can be used in position-specific motif searches.

Short sequence de novo motifs are enriched in BS-DR cluster genes

To identify shared regulatory signatures of BS-DR co-expression cluster genes, we first detected overrepresented short sequence motifs in their respective promoters relative to the rest of the promoterome. Motifs were mined using Gimmemotifs, an "ensemble method" that takes advantage of the best results from many motif-finding algorithms (van Heeringen & Veenstra, 2011). We identified short sequence motifs ranging from 6 to 15 bases in the promoters of genes

in the co-expressed BS-DR cluster. Using Gimmemotifs, the top motifs found from each component algorithm were clustered by similarity, merged using weighted information scores (van Heeringen & Veenstra, 2011), and then were tested for enrichment in the BS-DR cluster in both IR64 and Nipponbare rice varieties using a Bonferroni P-value correction ($P < 0.00156$) (Figure 3-5). Nine out of sixteen motifs found in this *de novo* method matched known *cis*-elements. Only motifs that were enriched in at least one variety were included. Many of the longer motifs had alignments with segments of miRNA stem loops, or more particularly, a processed miRNA (motif K). Motifs matching *cis*-elements associated with the plant defense response, such as motifs A and B (W-box) and motif C (lectin gene), were identified. However, the short sequence motifs enriched in BS-DR genes were also prevalent in many promoters across the genome. Thus, using these motifs as the foundation, we asked if specific groupings of these motifs were distinct to BS-DR promoters.

BS-DR motifs are organized into cis-regulatory modules

Commonalities in proximity and organization of the BS-DR-associated single motifs (Figure 3-5) across BS-DR cluster genes were discovered by local alignment. Promoter sequences were converted to a list of position-specific locations of each motif ("motif profile") and then this profile was used to align each promoter using a Smith-Watermann alignment as described and applied by the Regulatory region Local Alignment tool (ReLA) (González et al., 2012). Sequence segments containing a group of motifs within promoters, called *cis* regulatory modules (CRMs), in at least ten promoters in the BS-DR cluster were selected (Figure 3-6). Each CRM has a specific window size and may occur anywhere across the 2kb promoter. Each constituent motif found within a respective CRM window occurs anywhere in the CRM (CRM2,

CRM4, CRM5) or in sub-regions (CRM1, CRM3). The refined locations of constituent motifs were identified by calculating the relative frequency of motif occurrence at each nucleotide position for CRMs across the Nipponbare promoterome, the BS-DR cluster, or known DR genes (Figure 3-7). Constituent motif sub-region specificity in CRM1 and CRM3 was consistent across the promoterome, and the ranges are refined in known DR genes that harbored the respective CRM (Figure 3-8). No other CRMs showed position specificity of constituent motifs.

Overrepresentation of CRMs in BS-DR cluster genes and known DR genes in both IR64 and Nipponbare suggests their association with DR (Table 3-3). CRM1 shows enrichment in known DR genes, but not in the BS-DR cluster, whereas CRM2 is enriched in both varieties in both sets of genes. CRM3, which is more prevalent in Nipponbare than IR64 in general, shows enrichment in Nipponbare in the BS-DR cluster only. CRM4 and CRM5 are both highly enriched in BS-DR genes, but not in FA-DR genes. Each CRM is associated with BS-DR genes, FA-DR genes, or both gene sets, and thus, their presence in a promoter may indicate involvement of the downstream gene in basal resistance.

CRMs are found in broad-spectrum DR genes located in resistance QTL

Previous work identified BS-DR genes located within disease resistance QTL and demonstrated their contribution to resistance (Carrillo et al., 2009; Davidson et al., 2010; B. Liu et al., 2004; Manosalva et al., 2009; Tonnessen et al., 2014). Three different DR gene families within disease resistance QTL contain the CRMs, including the Oxalate Oxidases (*OsOXO*s), Phenylalanine Ammonia Lyases (*OsPAL*s), and the Germin-Like Proteins (*OsGLP*s) (Figure 3-9). Some members of these gene families underlie QTL conferring resistance to diverse pathogens, including *X. oryzae* pv. *oryzae*, *M. oryzae*, and *R. solani* (Table S3-7: Supplemental).

Intriguingly, particular members of DR gene families shown to contribute to disease resistance are the members whose promoters contain CRMs (Figure 3-9). The BS-DR-associated motifs found in this analysis are present in the majority of promoters throughout the genome, but the specific organization of CRMs are associated with BS-DR genes, and specifically within resistant alleles.

The *OsGLP8* family includes BS-DR genes functionally validated by gene silencing (*OsGLP8-6*, *OsGLP8-7*, *OsGLP8-9*; (Manosalva et al., 2009)) which all have an instance of at least one CRM in their promoters (Figure 3-9a). Family members that do not contribute to the DR in gene silencing experiments ("Non-DR"), *OsGLP8-2*, *OsGLP8-3*, and *OsGLP8-12*, lack any CRM. Interestingly, occurrence of CRM between resistant and susceptible QTL haplotypes differed in the three DR-related *OsGLP8* genes. The upstream region (~1.9 kb) of *OsGLP8-9* has a 45 bp deletion that removes the two constituent motifs, P and O, from CRM2. The resistant promoter of *OsGLP8-7* contains four instances of Motif B, constituents of CRM4 and CRM5, in a 37 bp region at about position -700. Two SNPs in the susceptible QTL donor alter two of the four motifs, reducing the number of intact instances of Motif B to two within this sequence segment. Finally, *OsGLP8-6* has a large insertion (856bp) in its promoter at position -517 that is associated with higher and earlier expression of *GLP8-6* in response to pathogen inoculation (Davidson et al., 2010). This insertion contains CRM1 at the 3' end, as well as three instances of Motif B, which are W-box motifs. The CRM3 occurrence shown in the susceptible haplotype is unchanged in the resistant haplotype, but it is upstream of the visual window of -2 kb due to the promoter insertion. Other motifs are present across the promoters of these genes, but the occurrence CRMs is specific to DR-related *OsGLP8* genes, and the sites are polymorphic between resistant and susceptible haplotypes.

The family of *OsOXO* genes in resistance QTL contain one BS-DR (*OsOXO4*) and three "non-DR" (*OsOXO1-3*) genes. The genes *OsOXO1-3* are not expressed during rice blast infection (Carrillo et al., 2009). Though, the BS-DR gene, *OsOXO4*, is involved in resistance, and the resistance haplotype of *OsOXO4* is expressed to a greater extent upon induction by disease (Carrillo et al., 2009). The promoter of *OsOXO3* contains an extra occurrence of CRM5 in the susceptible variety, whereas *OsOXO4* gains CRM4 in the resistant haplotype (Figure 3-9b). No significant differences in promoter structure affect CRMs in *OsOXO1* or *OsOXO2* promoters. The resistant haplotype of *OsOXO4* is more highly expressed during rice blast infection, and CRM4 is only present in this haplotype, not in the susceptible variety.

Members of the *OsPAL* gene family co-localize with a resistance QTL on chromosome 2; one of these (*OsPAL4*) functions in BS-DR, while another (*OsPAL2*) does not (Tonnessen et al., 2014). *OsPAL6* was inferred to be involved in BS-DR by expression analysis (Tonnessen et al., 2014). The CRM profile of *OsPAL4* differs between resistant and susceptible QTL donor haplotypes (Figure 3-9c). In susceptible haplotypes (Azucena), there is a 229 bp deletion at position -1141, removing two constituent motifs from CRM2; CRM2 remains intact in the resistant haplotype. Along with the removal of a complete CRM2, the CRM4 in the IR64 promoter located at position -532 has a single nucleotide deletion in the first Motif B in the susceptible haplotype. The promoters of *OsPAL6* and *OsPAL2* showed no difference in motif structure. The CRM motif profiles of other members of gene families mentioned above with unknown function in DR are presented (Figure 3-10). Taken together, these three examples of CRMs being specific to BS-DR genes, and not "Non-DR" genes, that are specific to resistant QTL haplotypes support the possibility of the presence of CRMs predicting the effectiveness of BS-DR gene regulation.

Characterization of CRM1 and CRM3 consensus sequences

The constituent motifs within both CRM1 and CRM3 exhibit position specificity across FA-DR genes and the entire promoterome (Figure 3-7, Figure 3-8). These patterns suggest that the rest of the nucleotides nested within the windows of CRM1 and CRM3 are conserved. A consensus sequence was generated for CRM1 and CRM3 by aligning every occurrence of the respective CRM across the promoterome of Nipponbare (Figure 3-11). The majority of nucleotides within occurrences of CRM1 and CRM3, respectively, show conservation. The motifs within CRM1 in all known DR genes, as seen in Figure 3-12a, also include an extra motif G downstream on the plus strand, complementary to the upstream motif G included in CRM1. This prompted the generation of a CRM1 consensus sequence that includes 184 nucleotides on the three-prime end to encompass the entire self-complementary structure, which is used in subsequent analyses as "CRM1-appended" (Figure 3-11). Many DR genes with CRM3 in their respective promoters contained a larger, palindromic motif profile as well with complementary flanking sites of Motif B as compared to CRM3 in *OsGLP8-6* (Figure 3-12b).

To identify what type of repetitive sequences CRM1 and CRM3 represent, the consensus sequences were aligned to both known miRNA stem loops from miRBase (www.mirbase.org) and the Rice Transposable Element (RiTE) database (www.genome.arizona.edu/cgi-bin/rite/index.cgi) (Figure 3-13). Both CRMs showed significant (E-value > 1e-10) alignment to segments of known miRNAs and Miniature Inverted-repeat Transposable Elements (MITEs). CRM1 aligned with the sorghum miRNA classified *miR6225*, whereas CRM3 aligned with many members from the rice *miR818* and *miR812* families. Additionally, CRM1 and CRM3 aligned with MITE superfamilies *PIF-Harbinger* and *Tc-Mariner*, respectively, which are both class II Terminal Inverted Repeat (TIR) transposons (Copetti et al., 2015; Lee & Kim, 2014). The

secondary structure of putatively transcribed RNA from CRM1 consensus, CRM3 consensus, and the occurrence of CRM3 within the *OsGLP8-6* promoter produces a structure that parallels an miRNA stem loop, as seen when compared to the known miRNA *miR818a* (Figure 3-14) (Sunkar et al., 2008). The *OsGLP8-6* CRM3 secondary structure exhibits less self-complementarity as the other stem loops shown.

Since CRM1 and CRM3 sequences show characteristics of a repeat element, rice expression data of small RNAs (sRNAs) at each CRM1 or CRM3 locus in known DR genes was extracted from an sRNA Illumina SBS sequencing database (Nakano et al., 2006). Across multiple experimental libraries, relative peaks of sRNA expression are found within each locus of CRM1 or CRM3 (Figure 3-15). These results also indicate the high level of repetitiveness at each CRM position.

To discover the sizes and possible epigenetic function of sRNA produced at the CRM loci, sRNA counts were examined from individual experiments in the sRNA database. In the majority of loci for both CRM1 and CRM3 in DR genes, sRNAs were 24 nucleotides in size and immunoprecipitated with the class of AGO4 proteins, and not with AGO1 proteins (Data extracted from Wu et al., 2009, 2010) (Table 3-4a). Some CRM loci produced both 21 and 24 nt sRNAs in abundance, such as loci for Phospholipase-D and *r11*. Data for total sRNA reads in processing mutants, Dicer-Like 3 (*dcl3*), Dicer-Like 1 (*dcl1*), and RNA-dependent RNA polymerase II (*rdr2*) were compared relative to the wild-type. In *dcl3* and *rdr2* mutants, 24 nt sRNAs were reduced compared to wild-type, but increased or unaffected in *dcl1* mutants. Conversely, the 21 nt sRNAs from a few loci generally decreased in *dcl1*, but increased in *rdr2* and *dcl3* mutants relative to wild-type (Table 3-4b). The production of 24 nt sRNA is dominant

across loci, and these results show an association with AGO1, DCL3 and RDR2 with CRM1 and CRM3 loci.

To shed light on possible epigenetic activity, methylation at each of the CRM1 and CRM3 loci was examined from publicly available rice bisulfite sequencing data (Figure 3-15) (mpss.danforthcenter.org) (Nakano et al., 2006). Within a 1 kb region surrounding each CRM locus, a relative increase or decrease of methylation was observed at the location of CRMs. For CRM1 loci, five out of the nine FA-DR genes that contain CRM1 in their promoters, *OsCERK1*, *OsWRKY53*, *OsRac1*, *Pia-RGA5*, and *nls1-1D*, show relative peaks in methylation (Figure 3-15a). One gene, *OsRAR1*, exhibits consistent methylation across the region around CRM1. The remaining three FA-DR loci, *OsAGO7*, *Phospholipase-D*, and *r11*, show no methylation, or a valley where methylation markedly decreases within CRM1 relative to the surrounding DNA. In CRM3 from FA-DR gene promoters, four of the six total loci, *Xb25*, *OsGLP8-6*, *Pia-RGA4*, and *Pi-ta*, show this valley trend. The gene, *spl28**, has a small peak in methylation at CRM3, and *OsNAC6* shows no methylation in the region. These combined results indicate a possible epigenetic regulatory aspect, through variable methylation/de-methylation, with CRM1 and CRM3 in DR genes.

Discussion

Defining and searching for BS-DR genes

A DR gene either promotes the DR, or mediates other cellular processes to increase plant fitness against multiple diseases (Boyd et al., 2013; Ke et al., 2017; Kou & Wang, 2010). These genes respond to pathways initiated by ETI and PTI. Thus, when considering the various infection mechanisms and physiologies of the broad pathogen spectrum, the plant defense

response exploits the plasticity of its entire genome in the structural and genotypic sense. DR genes that are co-expressed across different defense responses are considered BS-DR genes. Given the pivotal role of BS-DR genes in modulating the pathways involved in resistance, a well-orchestrated transcriptional response is necessary in strong basal resistance. In this study, we identified patterns in the promoters of BS-DR genes that provide clues as to how this action is carried out. The CRMs we identified may influence genome-wide transcription, since they are associated with known, functionally-validated DR genes and reside in many other genes throughout the genome.

BS-DR gene-associated short sequence motifs have functionality in known defense pathways

Many of the single short motifs (6-15 nt) identified here aligned with known *cis*-elements (Figure 3-5). The two motifs A and B matched the core DNA binding element of the WRKY transcription factor, and motif B is found in CRM3, CRM4, and CRM5. The presence of these *cis*-elements is essential for functional resistance to three major rice diseases, rice blast, sheath blight and bacterial blight (Hwang et al., 2016; Nakayama et al., 2013; Peng et al., 2016). Other motifs found in this study align with *cis*-elements involved in DR-related cellular processes. Lectins have been inferred to be involved in resistance through protein-carbohydrate binding during the defense response (Van Damme et al., 2004). Pertaining to nodule factor *cis*-elements, the communication and control of bacterial population between the host plant and *Rhizobium* bacteria is evolutionarily intertwined with the host-pathogen interaction (Cao et al., 2017). The major form of auxin, IAA, negatively regulates defense through induction of expansin-mediated cell wall loosening (Guilfoyle, 1999). Genes with promoters that harbor the auxin-responsive factor (motif L) may be contributing positively or negatively to this susceptibility response.

Ethylene acts as a regulator of host defense, and is found at higher levels in resistant interactions with *M. oryzae* (C. Yang et al., 2017).

Although small sequence motifs can be indicative of specific protein binding to alter transcription, *cis*-elements do not usually work alone. Often multiple *cis*-elements work in tandem to regulate genes involved the DR as well as abiotic stress and the auxin response (Berendzen et al., 2012; Deb & Kundu, 2015; Zou et al., 2011). To understand BS-DR gene regulation, the overall structure of promoters, in the form of CRMs, was addressed.

Interpreting the diverse characteristics of CRMs

Five different groupings of motifs were found to be conserved across promoters of BS-DR cluster genes, and show statistical enrichment in known DR genes and/or the BS-DR cluster (Table 3-3). The motifs found within each CRM infer broader implications for the conditional responsiveness of each gene containing the respective elements in their promoters.

- *CRM4 and CRM5 – W-box recurrence*

Motif B, the pattern found within both CRM4 and CRM5, is in perfect alignment with the core W-box. This WRKY binding site organizes into groupings of either two (CRM4) or three (CRM5) within a small window (Figure 3-6). Both CRMs are highly enriched in the BS-DR cluster of co-expressed genes, and are found in many FA-DR genes (Table 3-3). This suggests that W-boxes tend to work in tandem or repeated patterns. Indeed, some WRKY proteins form hetero- and homo-complexes when binding to promoters, and they do this where there are multiple, closely positioned W-boxes (Xu et al., 2006). For example, *OsWRKY4* is regulated by the binding of OsWRKY80 on the promoter to facilitate resistance to *R. solani*, and the W-box

sites in the promoter of *OsWRKY4* are in a close-proximity, duplicated pattern (Peng et al., 2016). In promoters of Arabidopsis genes involved in Systemic Acquired Resistance (SAR), W-boxes are in tandem, often as duplicates, and there are 4.3 W-boxes per promoter, relative to the expected 2.1 in the rest of the genome (Maleck et al 2000). While the W-box *cis*-element is in most promoters, its presence in duplicate or triplicate may be a more functional genotype, particularly when they are associated with regulation of BS-DR genes.

- *CRM2 – Two different motifs with interconnected roles*

The two motifs O and P align with functional *cis*-elements found in ethylene responsive and light responsive genes, respectively. CRM2 contains both elements, with two instances of motif O (Figure 3-6). Ethylene is a plant hormone known to play a key role in defense (C. Yang et al., 2017). Photosynthetic genes also play a role in the rice DR and are modulated during infection by *R. solani*, *X. oryzae* pv. *oryzae*, and other plant-pathogen interactions (Huot et al., 2014; Yu et al., 2014; Zhang et al., 2017). Interestingly, cross-talk exists between genes involved in the ethylene response and light-responsive genes, such as during the process of leaf greening and hypocotyl elongation (Yu & Huang, 2017; Zhong et al., 2009). Perhaps the grouping of *cis*-elements in CRM2 in BS-DR genes is indicative of the importance of regulating these two separate molecular pathways during infection. The identification of CRM2 is a step towards understanding how BS-DR genes with vastly different function intercommunicate, and helps to identify candidates that are involved.

- CRM1 and CRM3 – Signals for epigenetic modification

Since the *de novo* motif search method described in this study did not look for individual patterns greater than 15 nt due to computational restrictions of the motif-finding algorithms, any larger sequence similarities across promoters were overlooked. The constituent motif positions and strand specificities of CRM1 and CRM3 suggest more nucleotide conservation within CRMs, and indeed the consensus sequences illustrate the larger structural pattern that exists. Since CRM1 and CRM3 both align with segments of miRNAs as well as MITEs, they can generally be described as transposon-derived repeat elements, due to a lack of functional validation for involvement in any gene silencing pathways. The sRNAs that are transcribed in each CRM1 and CRM3 locus are hypothesized to be producing small-interfering RNAs (siRNA) and participating in *cis*-acting RNA-directed DNA methylation (RdDM) for a variety of reasons. First, the majority of reads are 24 nt in length and immunoprecipitated with AGO4 proteins, which is indicative of the siRNA processing pathway (Mallory & Vaucheret, 2010; Wu et al., 2009, 2010). Secondly, DCL1 is the protein involved in 21nt miRNA processing, and when mutated, *dcl1* lines do not show a decrease in the 24 nt sRNAs produced at the CRM loci (Kurihara, 2005; Wu et al., 2010). Mutation of RDR2 or DCL3, enzymes responsible for processing and amplification of siRNAs, results in a decrease of the amount of 24 nt reads relative to the wild-type, meaning these proteins are critical in the production of these reads. Methylation at CRM1 and CRM3 sites is highly variable across DR gene loci, with most CRM1 loci showing a peak in methylation, and many CRM3 loci conversely showing a methylation valley. The methylation data mined for this analysis is based on wild-type Nipponbare tissue under no stress; further experimentation is needed to determine if the active methylation/de-methylation of CRM1 and CRM3 loci are altered during pathogen infection.

The involvement of RdDM in plant disease resistance is well-studied in plants. A demethylating agent, 5-azadeoxycytidine, enhances resistance of rice to *X. oryzae* pv. *oryzae* (Akimoto et al., 2007). In *Arabidopsis*, active DNA de-methylation occurs during the resistant response to *P. syringae*, and many DR genes with repeat elements in their promoters are affected by epigenetic processes (Yu et al., 2013). One gene, *WRKY22*, is methylated in the promoter in the basal state, but actively demethylates during pathogen infection, allowing for transcription factors to bind and activate the gene (Yu et al., 2013). Conversely, active DNA methylation is important for defense against the tumor-inducing *Agrobacterium tumefaciens* and necrotrophic fungi (Gohlke et al., 2013; López, Ramírez, García-Andrade, Flors, & Vera, 2011). The identification of BS-DR-associated CRM1 and CRM3 reveals two separate patterns conserved across promoters, and more research is required to discern if these are key RdDM mechanisms for genomic regulation and disease resistance.

Polymorphisms in CRMs infer BS-DR gene functionality in disease resistance QTL

Our analysis shows that differences in the promoters of BS-DR genes between resistant and susceptible haplotypes include the mutation and/or structural change of CRMs (Figure 3-9). The ability for DR genes to modulate transcription during infection may be influenced by the regulatory mechanisms dictated by these genotypes. The two modules CRM4 and CRM5 contain polymorphisms in three BS-DR genes, *OsGLP8-7, OsOXO4*, and *OsPAL4*, and one gene not demonstrated to be involved in the DR, i.e., *OsOXO3* (Figure 3-9a-c). Since WRKY proteins show DR-functionality in duplicates and triplicates, mutations that break these CRMs may reduce transcriptional responsiveness of the associated gene, as seen in *OsGLP8-6, OsOXO4,* and *OsPAL4* promoters. In contrast, CRM5 is gained in the susceptible haplotype in the non-DR

OsOXO3, suggesting that an enhancement of expression of this locus due to CRM5 may increase susceptibility or play a role in different stresses or developmental stages. CRM2 is present in resistant haplotypes for the BS-DR *OsGLP8-9* and *OsPAL4* promoters. The lack of a complete CRM2 in susceptible promoters could leave the gene unresponsive to defense elicitors that induce ethylene and light-responsive activity. CRM1 and CRM3 are found in the promoter of *OsGLP8-6*, with CRM1 only found in the resistant haplotype (within the insertion). Both CRM1 and CRM3 are inferred to be involved in epigenetic modifications, thus the displacement of CRM3 and introduction of CRM1 in the resistant haplotype could be the reason for faster and higher expression of *OsGLP8*-6 (Davidson et al., 2010).

These examples of CRM polymorphisms in BS-DR QTL genes provide further evidence supporting the prospect that these CRMs have a predicative role in the contribution of other BS-DR genes to basal resistance. When breeding new varieties, the criteria for selection of parents could take advantage of this knowledge. Understanding the composition of their promoters, and how important that is for a strong basal resistance, will allow us to select varieties enriched in the broad-spectrum defense response across the genome.

Materials and Methods

Broad spectrum defense response transcriptome data

To identify rice defense response genes that are co-regulated across various diseases and treatments, we mined a collection of publicly available transcriptome data. Raw expression reads acquired from NCBI GEO (www.ncbi.nlm.nih.gov/geo/) and recent publications (Table S3-1: Supplemental), and were all treatment versus mock scenarios of resistant interactions. Affymetrix probe results files (CELs) were processed for each treatment-control study with the R

package, Affy, which utilizes Robust Multi-Array (RMA) background correction and quantile normalization as described by Irizarry, 2003. Affymetrix probes were matched to the MSU Nipponbare rice reference genome (rice.plantbiology.msu.edu/index.shtml) using the sequence alignment software, Vmatch (www.vmatch.de). Agilent microarray data was processed using the NCBI GEO software, GEO2R, which maps reads to rice loci and outputs an expression ratio based on resistant-mock data sets (www.ncbi.nlm.nih.gov/geo/geo2r/). Illumina RNA sequencing results in fastq format were checked for quality using FastQC (Babraham Bioinformatics). Sequence reads were preprocessed by removing the first 15 base adapter sequences, then trimming the ends with parameters of a minimum score of 20 and minimum read length of 20 using the fastx toolkit (hannonlab.cshl.edu/fastx_toolkit/index.html). The pre-processed fastQ files were aligned to the representative protein coding mRNAs from rice variety Nipponbare reference genome v7.0, and the resulting ".bam" file was assembled into transcripts and annotated. These alignment processes, along with calculating differential expression ratios was accomplished using TopHat, Cufflinks, Cuffdiff, and Cuffmerge, sequentially, from the Tuxedo analysis package (Trapnell et al., 2012).

To analyze the differential expression data sets across various platforms, each set was converted to a log base 2 scale fold change then normalized to make cross-comparison feasible. The 44 individual expression data sets were centered based on their respective medians within the 3*(interquartile range). Clipping of the data was done by removing expression values that were outside of 3*(interquartile range) from the first or third quartiles. Normalization was performed using a min-max linear transformation within the range of -1 to 1 (Figure 3-1).

Defense response gene co-expression analysis

Genes that had missing data for greater than 10% of the expression studies were excluded from the analysis. The rest of the missing data values were imputed using the median expression of the associated experiment. Pearson correlation, followed by average linkage hierarchical clustering of the distance measure, 1-correlation, was performed using the R packages "cor" and "hclust," respectively (www.R-project.org). Branch cutting of the resultant dendrogram was done using various parameter settings and algorithms of the dynamic tree cut method, an improvement from using a fixed cut height (Langfelder & Horvath, 2008). Specifically, the "cutreeHybrid" R-script was used with default parameters aside from an altered minimum cluster gap ranging from 0.15 to 0.40, or the "cutreeDynamic" R-script with "deepSplit" parameters, 0, 1, 2, 3, True, or False. Gene Ontology (GO) term and known DR gene enrichment analysis of co-expressed gene clusters utilized the Fisher exact test, with Benjamini-Hochberg correction and a False Discovery Rate (FDR) of 0.05 (Benjamini & Hochberg, 1995). Choosing the dendrogram branch cutting parameters to produce clusters for subsequent analysis involved finding the method which maximized Dunn Index and Connectivity, minimized Silhouette Width, and contained the most DR-GO and DR gene enriched clusters. GO terms were taken from the plant GOSlim database aligned to the rice genome (rice.plantbiology.msu.edu/index.shtml) (Table S3-2: Supplemental). A list of known defense genes (FA-DR genes) in the rice genome was compiled using recent literature, and publicly available databases, the Overview of functionally characterized Genes in Rice Online (OGRO) database and the Kansas State University rice defense gene collection (www.k-state.edu/ksudgc/) (Table S3-3: Supplemental) (Yamamoto et al., 2012).

Promoterome construction

The promoterome of the rice variety Nipponbare was constructed using the MSU v7.0 annotation and a python script. Only the representative gene model (contained furthest upstream 5' UTR) was chosen for each locus, and transposable element-encoding genes were omitted. Transcription start sites (TSSs) were assigned to the -1 position from the start of the genes. Promoters were limited to 2kb upstream of the assigned TSS, unless that segment overlapped with a neighboring gene sequence, in which case the 5' end of the promoter was truncated to the end of that respective gene. Promoters less than 50 bases long after this step were removed from the analysis due to constraints of the CRM-finding pipeline. Promoters for the rice variety IR64 were constructed using the genome sequence and annotation recently developed by Schatz et al 2017 (in prep.). To ensure accurate annotation of IR64 genes, gene orthologs were identified between IR64 and Nipponbare based on 90% similarity in full gene sequence. Accurate TSS positions in the IR64 orthologs were obtained by aligning Nipponbare gene sequences to IR64 contigs using BLAT on linux command line at default parameters aside from max Intron length of 600 (genome.ucsc.edu/cgi-bin/hgBlat?command=start). Matches with greater than 90% identity were labeled as IR64 orthologs to the corresponding Nipponbare gene, and the 5' end or 3' end (depending on + or – orientation, respectively) of the alignment was labeled the TSS in the IR64 contig. Any other IR64 annotated genes that did not have an orthologous Nipponbare gene were included in the IR64 gene set if they did not overlap with the previously identified orthologs. Promoters from the IR64 contigs were extracted, truncated, and filtered as described above. Consistency in promoter positioning and structure was verified using a graphical representation of the nucleotide tallies at each relative position in the promoterome for both varieties using an R script.

Sequence motif finding

Putative *cis*-elements *de novo* were identified using an ensemble approach with multiple motif-finding algorithms to search the promoteromes of the co-expressed BS-DR gene cluster derived from Nipponbare and IR64. The application, gimmemotifs, was modified to be used as a backbone for running component algorithms simultaneously (van Heeringen & Veenstra, 2011). Component algorithms were Mdmodule, MEME, Weeder, MotifSampler, trawler, Improbizer, BioProspector, AMD, Homer, and GADEM (Ao et al., 2004; Bailey et al., 2009; Conlon et al., 2003; Ettwiller et al., 2007; Heinz et al., 2010; Li, 2009; Liu et al., 2001; Pavesi et al., 2001; Shi et al., 2011; Thompson et al., 2003). Sequences of all promoters from each variety less those of the genes in the BS-DR cluster were used as background sequences when required per component algorithm. Default parameters for each algorithm were used aside from certain universal parameters of (1) size range between 6 and 15 base pairs, (2) search both plus and minus strand, and (3) output only top 15 motifs. Each motif Position Frequency Matrix (PFM) obtained from the multiple algorithm runs was tested for enrichment in the BS-DR cluster relative to the rest of the genome. A query sequence was counted if it held 85% similarity to a motif PFM. The enrichment test was done using a "with or without" score for each promoter in IR64 or Nipponbare, and a Fisher exact test with Bonferroni P-value correction. Many of the motifs deemed as enriched were highly similar and came from different motif-finding algorithms. Thus, matching motifs were combined into one using a Weighted Information Content (WIC) score and "iterative clustering" (van Heeringen & Veenstra, 2011). Clustered motifs were tested again for enrichment in the BS-DR cluster due to the changes imposed from motif merging (Figure 3-5). Comparison of these *de novo* motifs to known *cis*-elements was done using the program, TOMTOM, of the MEME suite (Gupta et al., 2007). A list of known

cis-elements was generated by combining the databases, PLACE (www.dna.affrc.go.jp/PLACE), atcisDB (arabidopsis.med.ohio-state.edu/AtcisDB), and TRANSFAC-plants (gene-regulation.com/pub/databases.html).

Searching for putative CRMs

CRMs were found using a window-based, pairwise alignment approach of motifs in the promoters of BS-DR cluster genes. This was accomplished with the Regulatory region Local Alignment tool (ReLA), which utilizes a Smith-Waterman algorithm on sequence motif profiles (González et al., 2012). The program was executed for both Nipponbare and IR64 BS-DR cluster promoters, with two separate size restrictions of the CRM search window: >50 nt, or >200 nt. Due to restrictions of the alignment method, ReLA uses one promoter as a reference to align to each other promoter, then outputs the promoters that match that reference promoter motif profile. Thus, ReLA was run 385 times, using a different BS-DR cluster promoter as the reference each time. From all runs of the program, the reference motif profiles that included the highest number of BS-DR promoters aligned with the test promoter (>10) were selected as putative CRMs. Enrichment of each independent CRM in promoters of BS-DR cluster genes or known DR genes was determined using a Fisher Exact test of the genes with CRMs in the BS-DR cluster versus the rest of the promoterome. Visualizations of CRMs and constituent motifs in promoters of QTL-based DR genes were done using a python script, and the PyX package (pyx.sourceforge.net). Motif densities in each CRM were calculated from the ratio of total occurrences of each motif start point for each nucleotide position in the CRM window, and plotted using an R script.

Classification of CRM1 and CRM3 sequences

For each respective CRM1 and CRM3, all occurrences across the promoterome of Nipponbare was extracted using a python script and aligned using a command line version of Clustal Omega (Sievers et al., 2014). The alignment was converted into an Hidden Markov Model (HMM) profile using *HMMer* (HMMer.org). The consensus sequences for CRM1 and CRM3 were then generated using percent occurrence values for each nucleotide at each position along the HMM profile using *HMMemit* (HMMer.org). Consensus sequences were searched using the BLAST utility in miRBase (www.mirbase.org) against all stem loop sequences, and in the RITE database (www.genome.arizona.edu/cgi-bin/rite/index.cgi) for all repeat elements of rice. Putative secondary structure of RNAs produced from CRM consensus sequences was accomplished using the RNAfold webserver (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi).

Using a publicly available Illumina SBS small RNA sequencing database, chromosomal coordinates of each CRM1 or CRM3 occurrence in known DR gene promoters were queried for total rice sRNA reads mapped to the regions (mpss.danforthcenter.org/) (Nakano et al., 2006). Study specific analyses were extracted to look at AGO immunoprecipitation and sRNA processing mutants from the database (Wu et al., 2009, 2010). Methylation results from bisulfite sequencing were extracted as total methylated reads from all experiments using Nipponbare in various tissue types (mpss.danforthcenter.org/~apps/DNA_Met/public/RICE_met/) (Nakano et al., 2006).

Table 3-1: Co-expression cluster validation measures. Scores and statistics on the co-expression clusters found from each attempt at dendrogram branch cutting using different methods and parameters. Chosen parameter set: "H_MG=0.25"

			Cluster Validation Measures			Functional Validation Measures (# of clusters)		
Set parameters	**# Clusters**	**Mean cluster size (genes)**	**Connectivity**	**Dunn index**	**Silhouette width**	**GO enriched**	**DR-GO enriched**	**DR-GO + DR gene enriched**
H_MG=0.15	116	127	21794	0.049	-0.011	46	17	0
H_MG=0.20	93	158	21315	0.057	-0.001	42	17	0
H_MG=0.25	65	226	19906	0.057	0.006	35	17	1
H_MG=0.30	37	397	18390	0.055	0.006	25	13	1
H_MG=0.35	22	668	17264	0.055	0.019	18	12	1
H_MG=0.40	14	1049	14840	0.054	0.027	13	13	1
H_DS=0	68	216	20166	0.057	0.006	35	17	1
H_DS=1	103	143	21476	0.055	-0.003	41	15	0
H_DS=2	123	119	22106	0.035	-0.016	47	19	0
H_DS=3	126	117	22271	0.035	-0.017	49	18	0
D_DS=F	38	387	20396	0.041	-0.114	30	19	0
D_DS=T	114	129	23719	0.031	-0.124	53	21	0

Table 3-2: Co-expressed gene clusters with enriched DR-GO terms. Twenty-eight total DR-GO terms are enriched, and are found across 17 clusters. Cluster "greenyellow" is the "BS-DR" cluster. Each cluster Is given, followed by a list of DR-GO terms enriched in the respective cluster, along with the GO term annotation.

ModuleColor	DR-GO #	Annotation
skyblue	3723	RNA binding (molecular_function)
	166	nucleotide binding (molecular_function)
brown	6810	transport (biological_process)
	9628	response to abiotic stimulus (biological_process)
turquoise	3723	RNA binding (molecular_function)
	3676	nucleic acid binding (molecular_function)
	166	nucleotide binding (molecular_function)
darkgreen	6810	transport (biological_process)
greenyellow	16301	kinase activity (molecular_function)
	6464	cellular protein modification process (biological_process)
	7165	signal transduction (biological_process)
	166	nucleotide binding (molecular_function)
	4871	signal transducer activity (molecular_function)
	5102	receptor binding (molecular_function)
lightyellow	3700	sequence-specific DNA binding transcription factor activity (molecular_function)
purple	6464	cellular protein modification process (biological_process)
	3677	DNA binding (molecular_function)
violet	3700	sequence-specific DNA binding transcription factor activity (molecular_function)
blue	3723	RNA binding (molecular_function)
bisque4	6810	transport (biological_process)
palevioletred3	6810	transport (biological_process)
orange	3700	sequence-specific DNA binding transcription factor activity (molecular_function)
coral1	16787	hydrolase activity (molecular_function)
midnightblue	7165	signal transduction (biological_process)
tan	6810	transport (biological_process)
plum1	6950	response to stress (biological_process)
	16301	kinase activity (molecular_function)
darkorange2	5777	peroxisome (cellular_component)

Table 3-3: Gene promoters with CRMs. Number of promoters with a given CRM for all genes in the two promoterome are given in the first column. Labeled "Orthologous" are the genes with a given CRM that that exist in both Nipponbare and IR64 varieties. The following column tallies the genes in the BS-DR cluster, and known DR genes. P-values are given in brackets from a test for enrichment in each given independent set relative to the rest of the promoterome (Fisher Exact Test).

Genes with CRMs in their promoters				Enrichment tests in DR gene sets	
CRM	**Variety**	**All genes**	**Orthologous**	**BS-DR cluster**	**FA-DR genes**
CRM1	IR64	796	678	11 [0.0777]	7 [0.0171]
	Nipponbare	750	670	13 [0.0595]	9 [0.0106]
CRM2	IR64	1164	947	19 [0.0053]	8 [0.0402]
	Nipponbare	1020	937	21 [0.0033]	12 [0.0063]
CRM3	IR64	866	714	12 [0.0665]	4 [0.3222]
	Nipponbare	846	751	16 [0.0200]	5 [0.3846]
CRM4	IR64	18112	13250	195 [6.577e-06]	66 [0.1725]
	Nipponbare	14787	12802	195 [5.658e-05]	76 [0.2424]
CRM5	IR64	5519	4011	74 [3.704e-05]	16 [0.7504]
	Nipponbare	4596	3948	76 [5.620e-05]	20 [0.7182]

Table 3-4: sRNA reads at DR gene CRM loci in sRNA pathway experiments. Reads mapped to CRM1 and CRM3 loci in DR gene promoters is shown for both 21 and 24nt sizes. Two separate studies are examined, AGO immunoprecipitation (a) and sRNA processing mutants (b).

Table 3-4a: AGO immunoprecipitation experiments. The total sRNA reads from Nipponbare tissue mapped to each CRM locus is in the first column. The following columns are sRNAs from immunoprecipitated AGO1 clade proteins (AGO1a, AGO1b, AGO1c), and AGO4 clade proteins (AGO4a, AGO4b, AGO16).

CRM1

Defense Gene CRM1 position	sRNA Read Size		AGO proteins immunoprecipitation					
		TOTAL	AGO1a	AGO1b	AGO1c	AGO4a	AGO4b	AGO16
OsACO7 (LOC_Os01g39860) [-829 to -682]	21	7	17	0	11	0	3	1
	24	66	0	0	0	103	42	59
OsCERK1 (LOC_Os08g42580) [-1183 to -1037]	21	32	21	0	30	4	3	5
	24	160	0	0	0	175	106	372
RAR1 (LOC_Os02g33180) [-756 to -613]	21	7	17	180	11	1	1	1
	24	49	0	0	0	139	85	274
OsRac1 (LOC_Os01g12900) [-455 to -306]	21	9	6	0	0	1	0	1
	24	32	0	0	0	15	17	38
Phospholipase-D (LOC_Os01g07760) [-1188 to -1040]	21	150	156	4	355	3	6	3
	24	158	0	0	0	202	223	181
OsWRKY53 (LOC_Os05g27730) [-1082 to -941]	21	15	5	0	0	3	3	2
	24	125	0	0	0	72	58	87
nls1-1D (LOC_Os11g14380) [-1192 to -1045]	21	26	11	0	0	6	2	3
	24	142	0	0	0	115	85	142
r11 (LOC_Os11g29920) [-505 to -363]	21	139	86	4	0	1	0	0
	24	45	0	0	0	28	18	32
Pia-RGA5 (LOC_Os11g11810) [-1177 to -1039]	21	10	0	0	0	1	1	1
	24	16	0	0	0	10	21	9
Pia-RGA5 (LOC_Os11g11810) [-401 to -257]	21	2	8	0	0	2	1	0
	24	29	0	0	0	33	26	37

CRM3

Defense Gene CRM3 position	sRNA Read Size		AGO proteins immunoprecipitation					
		TOTAL	AGO1a	AGO1b	AGO1c	AGO4a	AGO4b	AGO16
XB25 (LOC_Os09g33810) [-1233 to -995]	21	295	214	374	164	19	32	16
	24	1342	3	0	0	1339	1942	1129
*spl28** (LOC_Os01g50770) [-1274 to -1039]	21	39	83	212	19	14	8	2
	24	1547	0	0	0	2206	1163	1256
OsNAC6 (LOC_Os01g66120) [-1622 to -1350]	21	56	31	0	9	2	2	6
	24	195	0	0	0	111	113	117
Pi-ta (LOC_Os12g18360) [-1800 to -1579]	21	68	34	0	35	2	7	5
	24	259	0	0	0	172	176	153
OsGLP8-6 (LOC_Os08g09000) [-1344 to -1203]	21	1	0	0	0	0	0	0
	24	3	0	0	0	6	8	3
Pia-RGA4 (LOC_Os11g11790) [-1982 to -1747]	21	0	0	0	0	1	0	0
	24	6	0	0	0	11	16	17

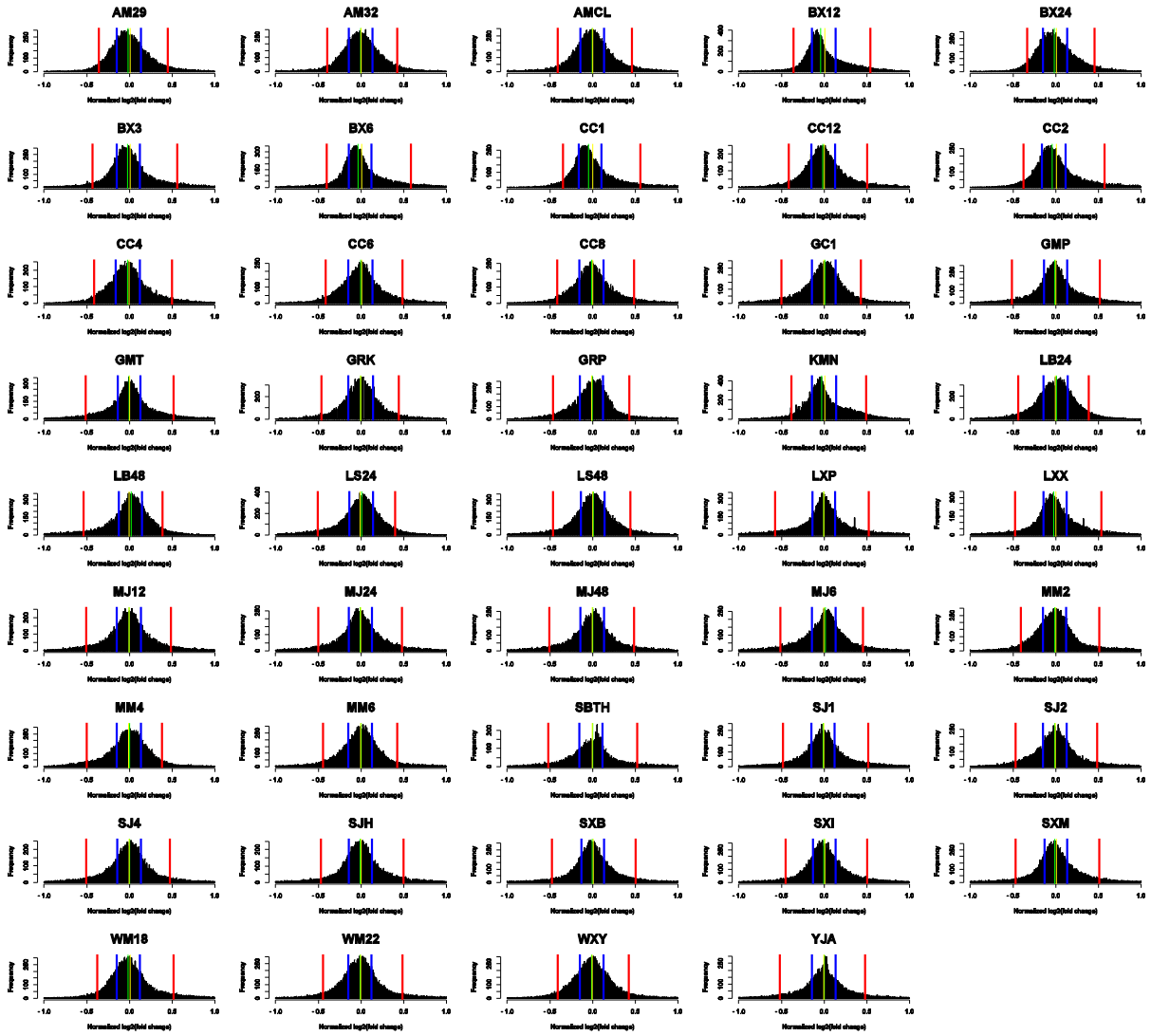
Table 3-4b: sRNA processing mutants. Wild-type Nipponbare tissue total sRNA reads mapped to each locus is in the first column. Reads of sRNAs in three different processing mutants, *Dicer-Like 1* (*dcl1*), *Dicer-like 3* (*dcl3*), and *RNA-Dependent RNA-polymerase II* (*rdr2*).

CRM1

Defense Gene CRM1 position	sRNA Read Size		sRNA processing mutants		
		WT	*dcl1*	*dcl3*	*rdr2*
OsACO7 (LOC_Os01g39860) [-829 to -682]	21	2	13	48	3
	24	38	56	21	37
OsCERK1 (LOC_Os08g42580) [-1183 to -1037]	21	49	31	77	42
	24	83	85	32	62
RAR1 (LOC_Os02g33180) [-756 to -613]	21	3	10	52	1
	24	29	49	21	21
OsRac1 (LOC_Os01g12900) [-455 to -306]	21	20	15	49	10
	24	19	22	5	9
Phospholipase-D (LOC_Os01g07760) [-1188 to -1040]	21	188	134	224	153
	24	121	192	55	90
OsWRKY53 (LOC_Os05g27730) [-1082 to -941]	21	18	14	55	7
	24	67	64	21	28
nls1-1D (LOC_Os11g14380) [-1192 to -1045]	21	36	39	64	27
	24	68	72	31	32
r11 (LOC_Os11g29920) [-505 to -363]	21	80	55	74	65
	24	35	34	7	32
Pia-RGA5 (LOC_Os11g11810) [-1177 to -1039]	21	3	6	9	1
	24	9	7	3	3
Pia-RGA5 (LOC_Os11g11810) [-401 to -257]	21	3	1	13	3
	24	15	26	7	9

CRM3

Defense Gene CRM3 position	sRNA Read Size		sRNA processing mutants		
		WT	*dcl1*	*dcl3*	*rdr2*
XB25 (LOC_Os09g33810) [-1233 to -995]	21	311	158	645	196
	24	868	854	297	377
*spl28** (LOC_Os01g50770) [-1274 to -1039]	21	43	42	150	20
	24	678	738	197	539
OsNAC6 (LOC_Os01g66120) [−1622 to -1350]	21	35	27	75	30
	24	231	216	56	124
Pi-ta (LOC_Os12g18360) [-1800 to -1579]	21	43	35	110	43
	24	271	240	76	149
OsGLP8-6 (LOC_Os08g09000) [-1344 to -1203]	21	0	3	2	0
	24	0	2	1	2
Pia-RGA4 (LOC_Os11g11790) [-1982 to -1747]	21	3	4	3	3
	24	8	6	2	5



$$X_{new} = \left(\frac{X - \min X}{\max X - \min X} \times 2\right) - 1$$

Figure 3-1: Post-processing expression data histograms. The distributions of scaled, centered, and normalized log base 2 fold change expression values for each experiment used in the co-expression analysis. Red lines show the outer 5% quantiles. Blue lines are marking the outer 25% quartiles. The median and mean are given as green and yellow lines, respectively. Normalization using the min-max linear transformation given below. Normalization was done separately for each individual experimental study. The vector X is the original scaled and centered expression data, X_{new} is the vector of expression values used for the analysis going forward.

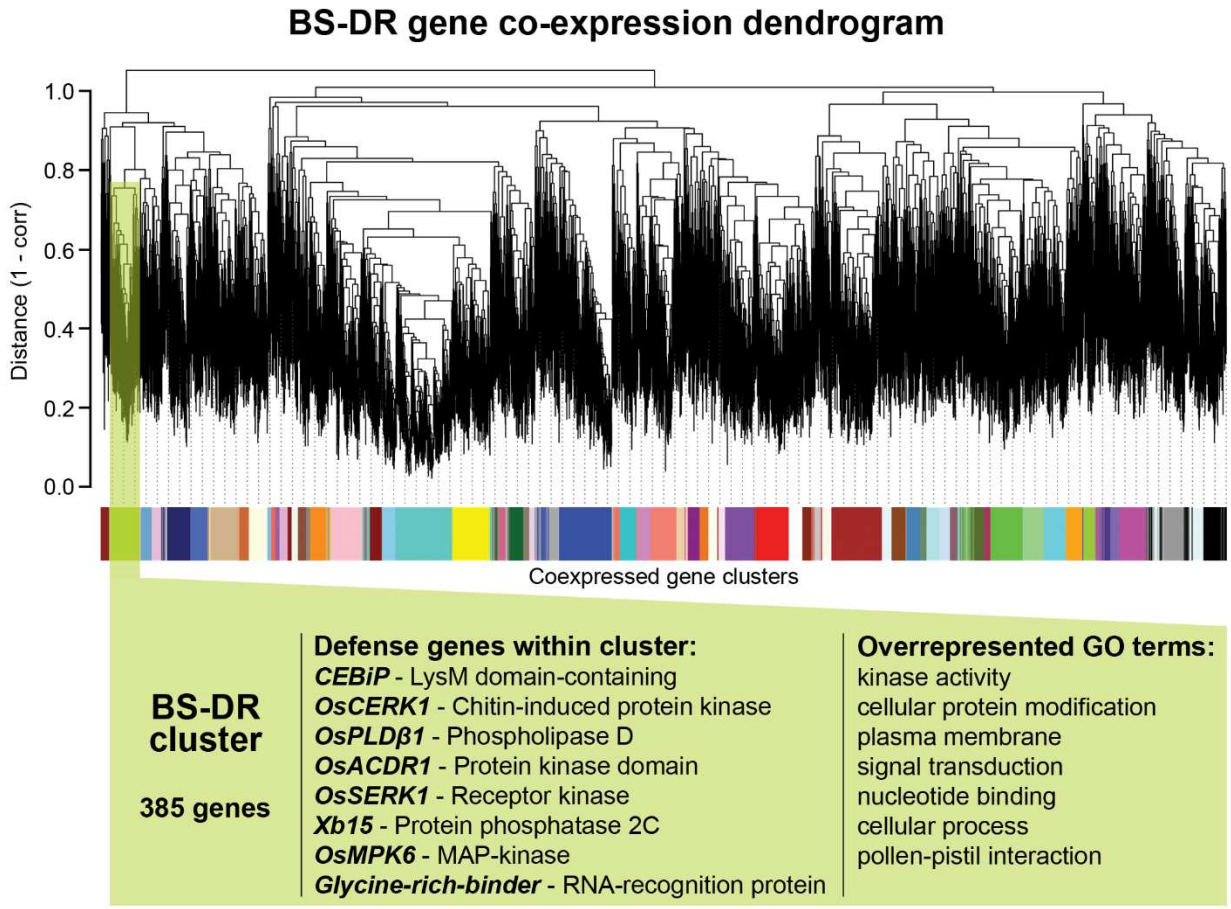


Figure 3-2: Broad Spectrum Defense Response (BS-DR) gene co-expression dendrogram and BS-DR specific gene cluster. Visual representation of the results of average linkage hierarchical clustering of distance measures, 1-corr (Pearson Correlation Coefficient, Y-axis). Colors along the base are distinct co-expression clusters amongst the entire set of genes included in the analysis (14688 genes) found from tree cut method "H_MG=0.25" (Figure 3-3, Table 3-1). Highlighted in green is the BS-DR gene co-expression cluster. Known FA-DR genes are listed and are enriched in the highlighted cluster (Fisher Test BH correction, P-value: 0.0014, Table S3-4: Supplemental). Gene Ontology terms that were significantly enriched in the highlighted cluster are also listed (Fisher exact test BH correction, Table S3-5: Supplemental, Table 3-7).

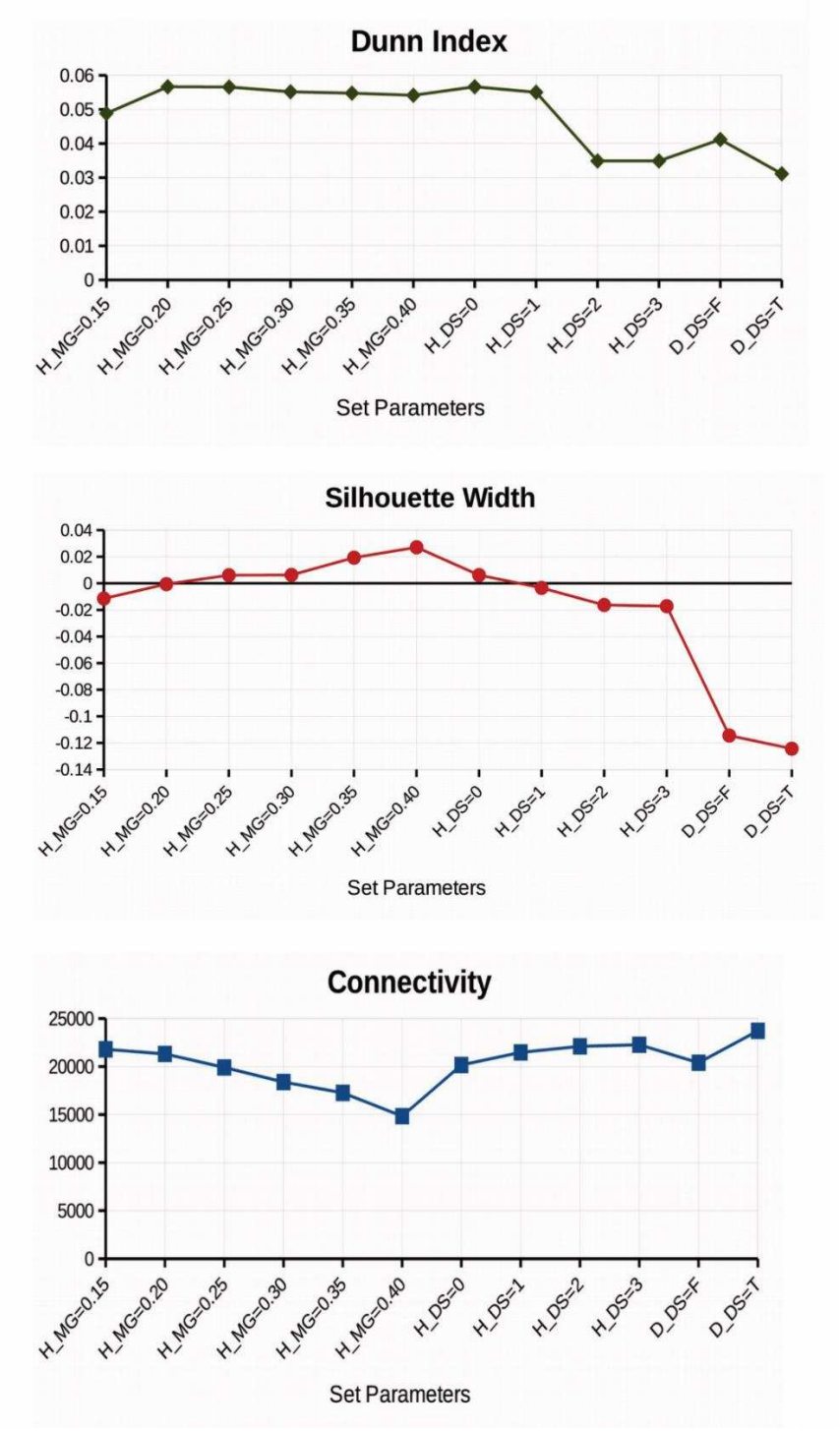


Figure 3-3: Cluster validation plots. Graphical representation of the different scoring measures for tree cutting methods of the co-expression dendrogram. Different scoring measures were the Dunn Index, Sillhouette Width, and Connectivity between clusters. All calculations were done as described in Xu et al 2009. Chosen parameter set: "H_MG=0.25"

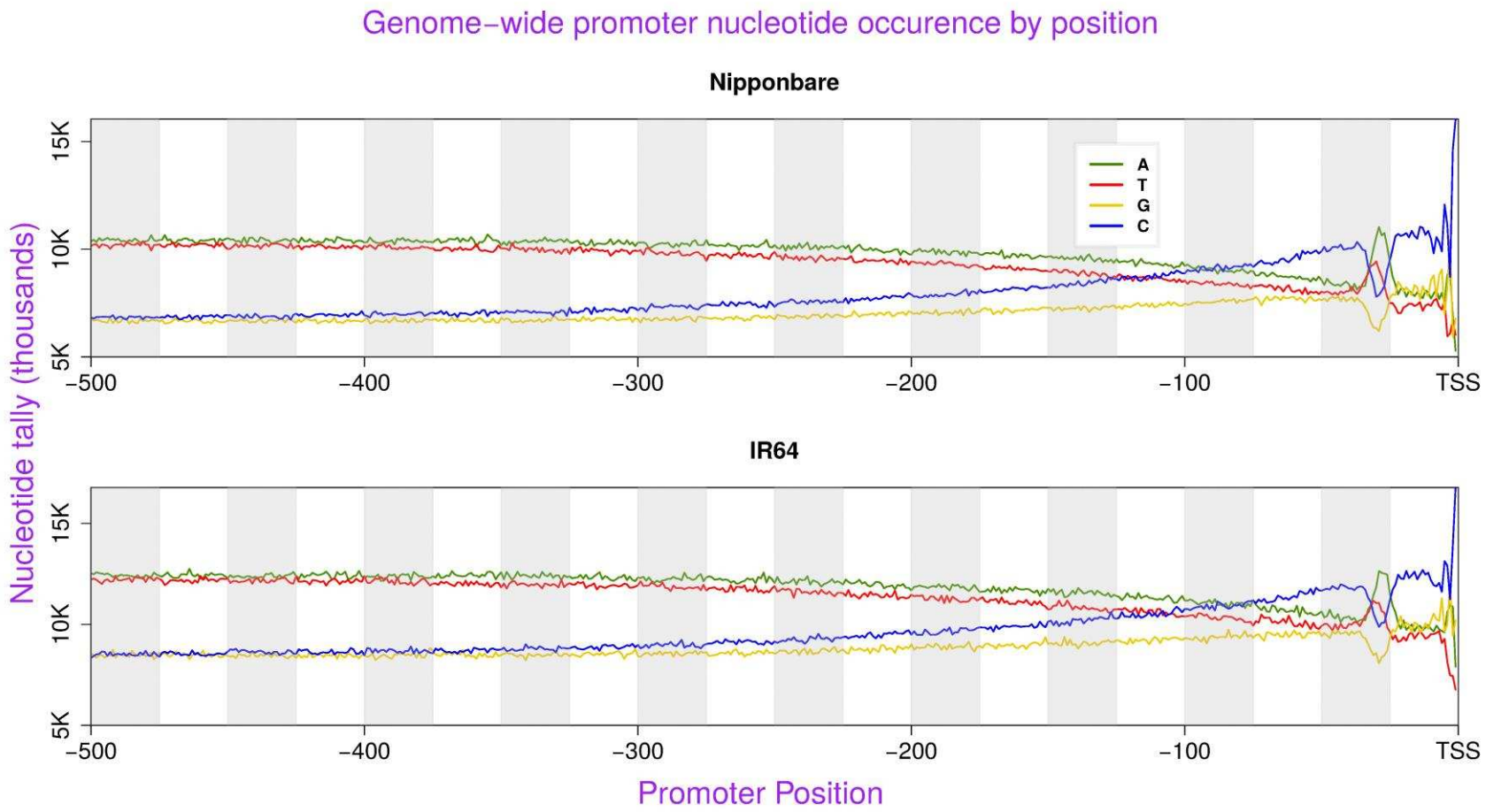


Figure 3-4: Genome-wide promoter nucleotide occurrence by position. Using the promoteromes from both IR64 and Nipponbare rice varieties, nucleotides were tallied for each position along the promoter from the Transcription Start Site to -500 bases upstream. There are a total of 36121 promoters in the Nipponbare set, and 43523 promoters in IR64. For each nucleotide positon along the X-axis, the number of promoters with A (green), T (red), G (yellow), or C (blue) is given on the Y-axis.

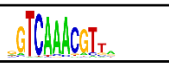
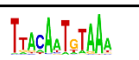
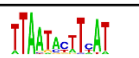
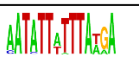
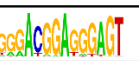
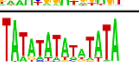
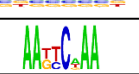
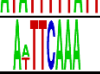
Motif ID	Discovered motif	*Cis*-element match	Annotation	DR Cluster genes		Rest of genome	
				N (384)	I (368)	N (35737)	I (43155)
A			W-box	187 [0.00156]	196 [1.66e-6]	14673	17710
B			W-box	366 [0.00025]	356 [0.00017]	32283	39718
C			DNA binding protein on soybean lectin promoter	272 ⊘ [0.01843]	269 [4.50e-5]	23468	27313
D		———	———	249 ⊘ [0.01648]	251 [4.64e-5]	21218	25080
E		———	———	307 ⊘ [0.00365]	304 [9.55e-5]	26415	32045
F		———	———	307 [0.00036]	298 [1.55e-5]	25832	30790
G			Binding site of soybean nodule factor	319 ⊘ [0.14737]	320 [0.00079]	28886	34760
H			Enhancer binding element Pea plastocyanin promoter	154 [0.00102]	155 [9.94e-6]	11596	13537
I		———	———	238 [3.84e-5]	230 [7.70e-5]	18500	22670
J		———	———	196 [0.00109]	187 ⊘ [0.00472]	15399	18950
K		*O.s.* miR1436	Mature miRNA	173 [0.00081]	164 [0.00016]	13232	15226
L			Element in promoter of auxin responsive gene	317 ⊘ [0.04265]	313 [0.00020]	28186	33442
M		———	———	42 [2.08e-7]	44 [7.38e-8]	1615	2112
N		———	———	24 [1.33e-5]	24 [7.07e-5]	813	1145
O			Light-responsive gene element	182 [0.00048]	187 [2.02e-7]	13912	16235
P			Ethylene response element	282 [0.00170]	280 [0.00011]	23704	28954

Figure 3-5: Short sequence motifs found in the DR gene cluster. Each row gives a different motif found using the ensemble algorithm method for finding short sequences overrepresented in the co-expressed BS-DR gene cluster. Motifs are given a letter identification for future reference. The *de novo* discovered motif is given in the second column, and the third column shows the verified known *cis*-element (if any) found using the program TOMTOM or a search of miRNA databases, followed by an annotation of that element in the fourth column. The number of gene promoters with the given motif is counted for both Nipponbare (N) and IR64 (I) in the BS-DR gene cluster or rest of the genome. Total number of genes for each set is given in parentheses next to the variety identifier. The BS-DR cluster gene numbers that are followed by an interdictory circle were not found to be significantly over-represented in the BS-DR cluster relative to the rest of the genome (P-values given in brackets, Fisher Exact Test, BH correction FDR: 0.05). Each motif has at least one variety in which it is enriched (overrepresented).

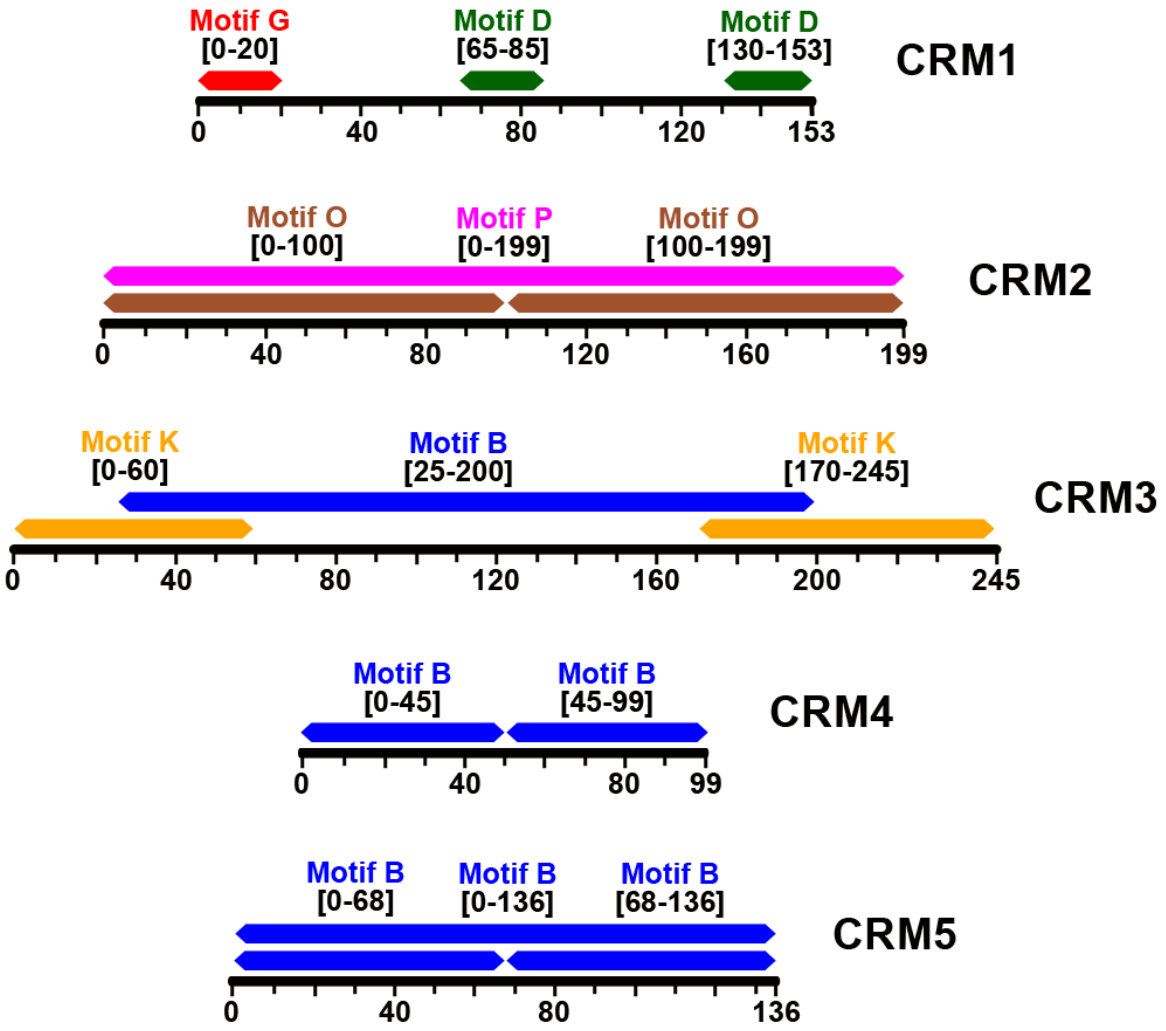
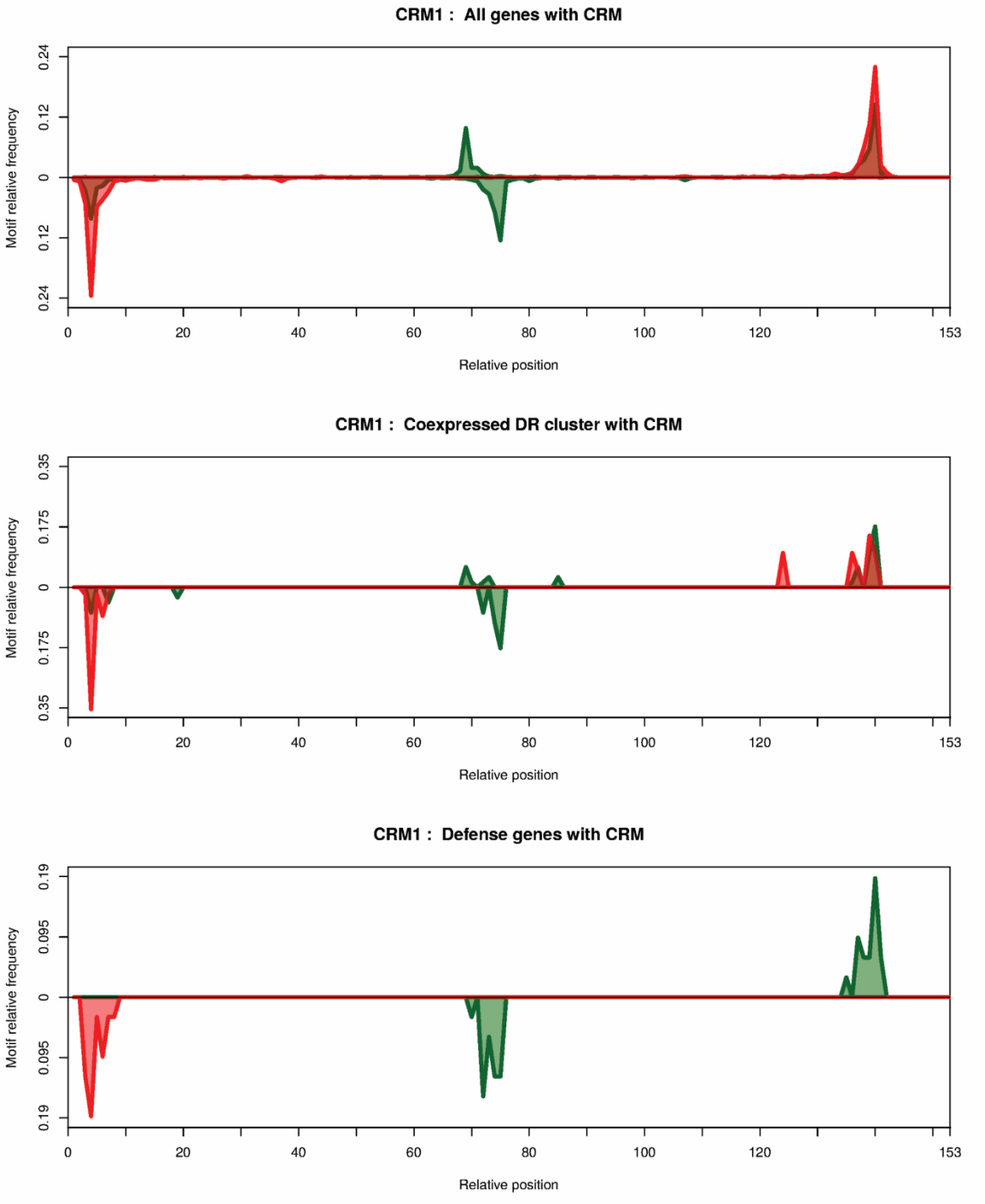
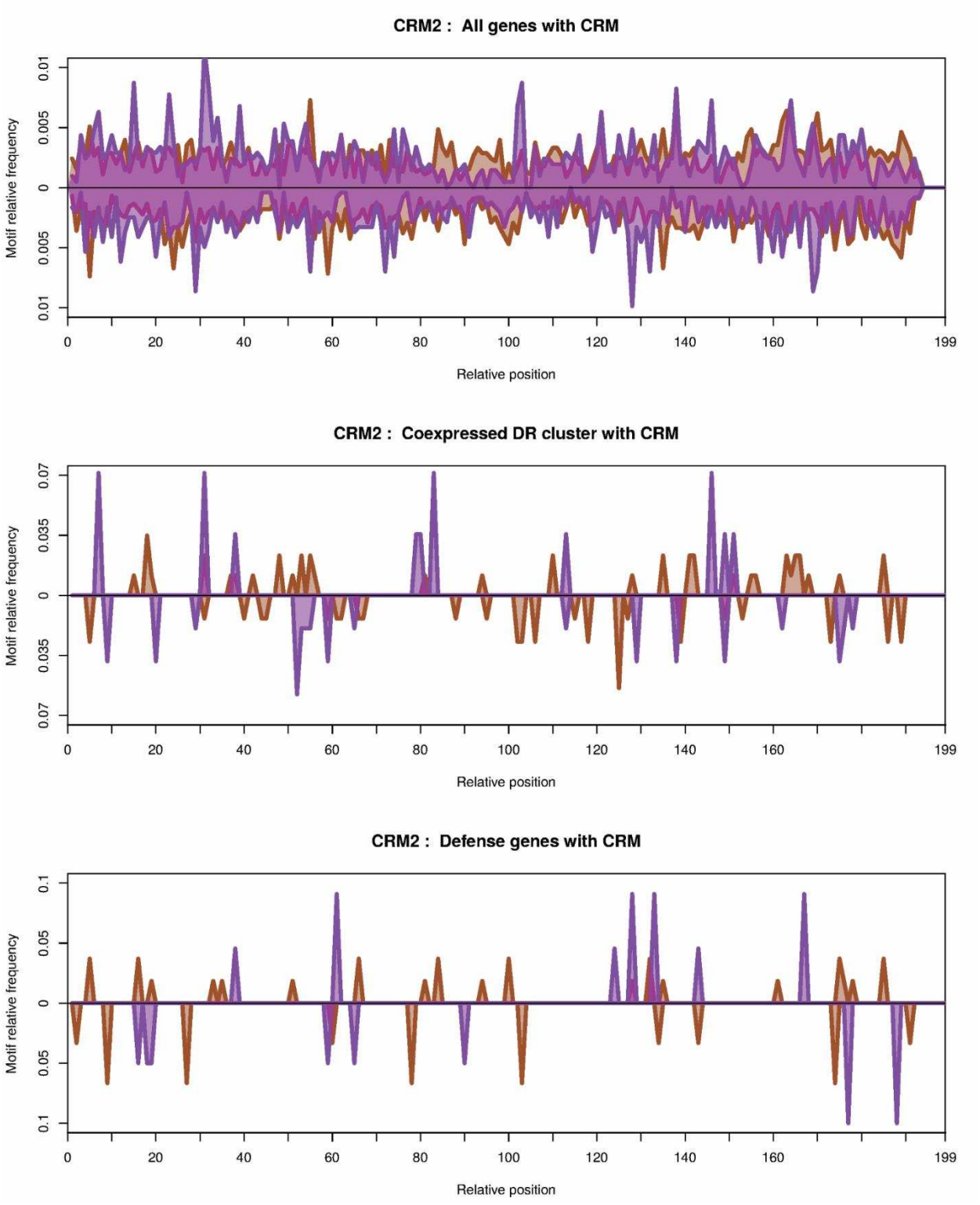


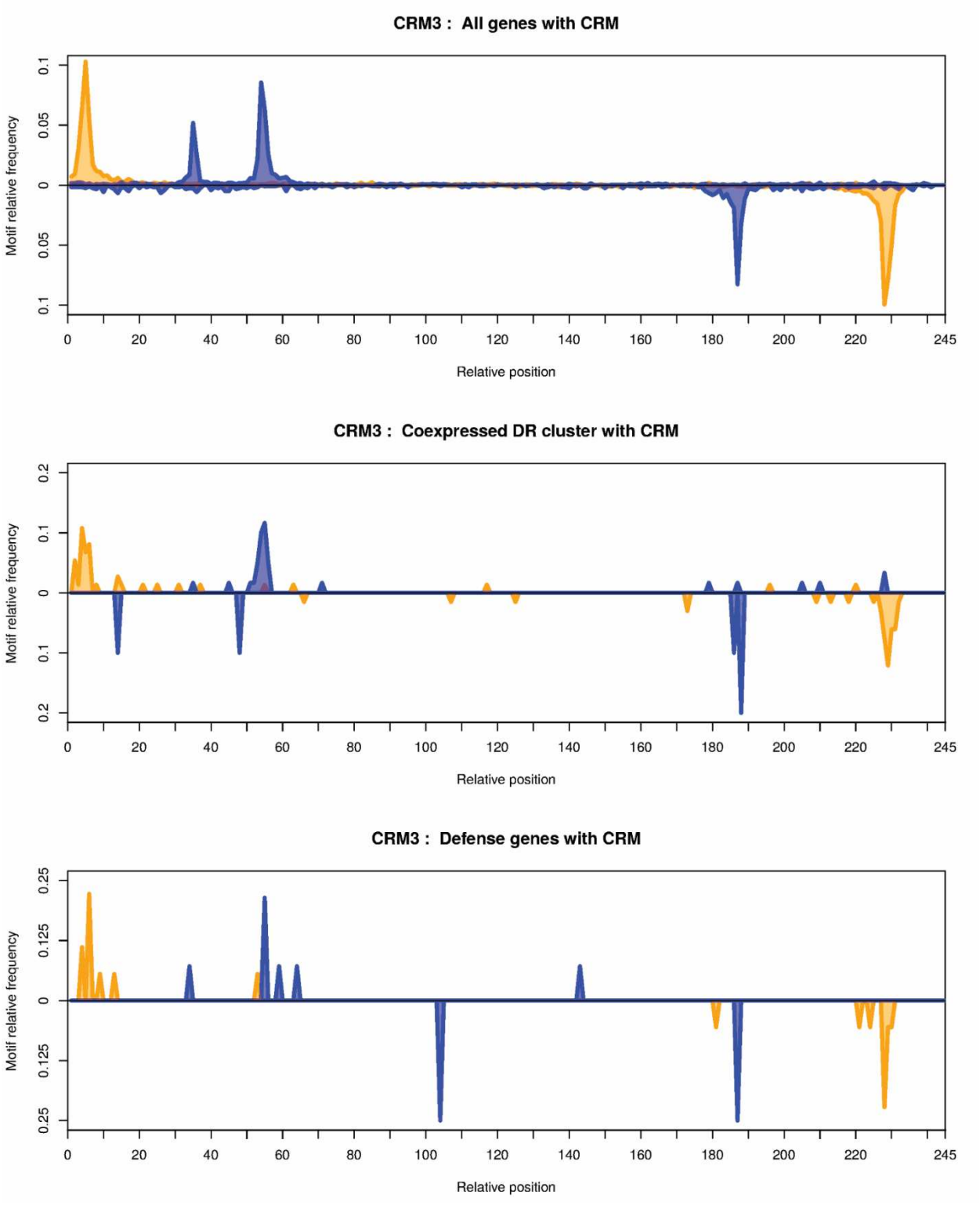
Figure 3-6: Visual of *Cis*-Regulatory Modules (CRMs). Each CRM found using a modification of the ReLA algorithm is shown, with constituent motifs given in distinct colors. Each range (X-axis) of the CRMs illustrates the window in which the constituent motifs were found along the promoters of BS-DR cluster genes. The sub-range for each constituent motif is also given as color-coded bars that span the region each motif is found. The CRMs, CRM2, CRM4, and CRM5 contain their respective constituent motifs in any position across the CRM window.


CRM1 : All genes with CRM
Motif relative frequency
0.24
0.12
0
0.12
0.24
0
20
40
60
80
100
120
153
Relative position
CRM1 : Coexpressed DR cluster with CRM
Motif relative frequency
0.35
0.175
0
0.175
0.35
0
20
40
60
80
100
120
153
Relative position
CRM1 : Defense genes with CRM
Motif relative frequency
0.19
0.095
0
0.095
0.19
0
20
40
60
80
100
120
153
Relative position

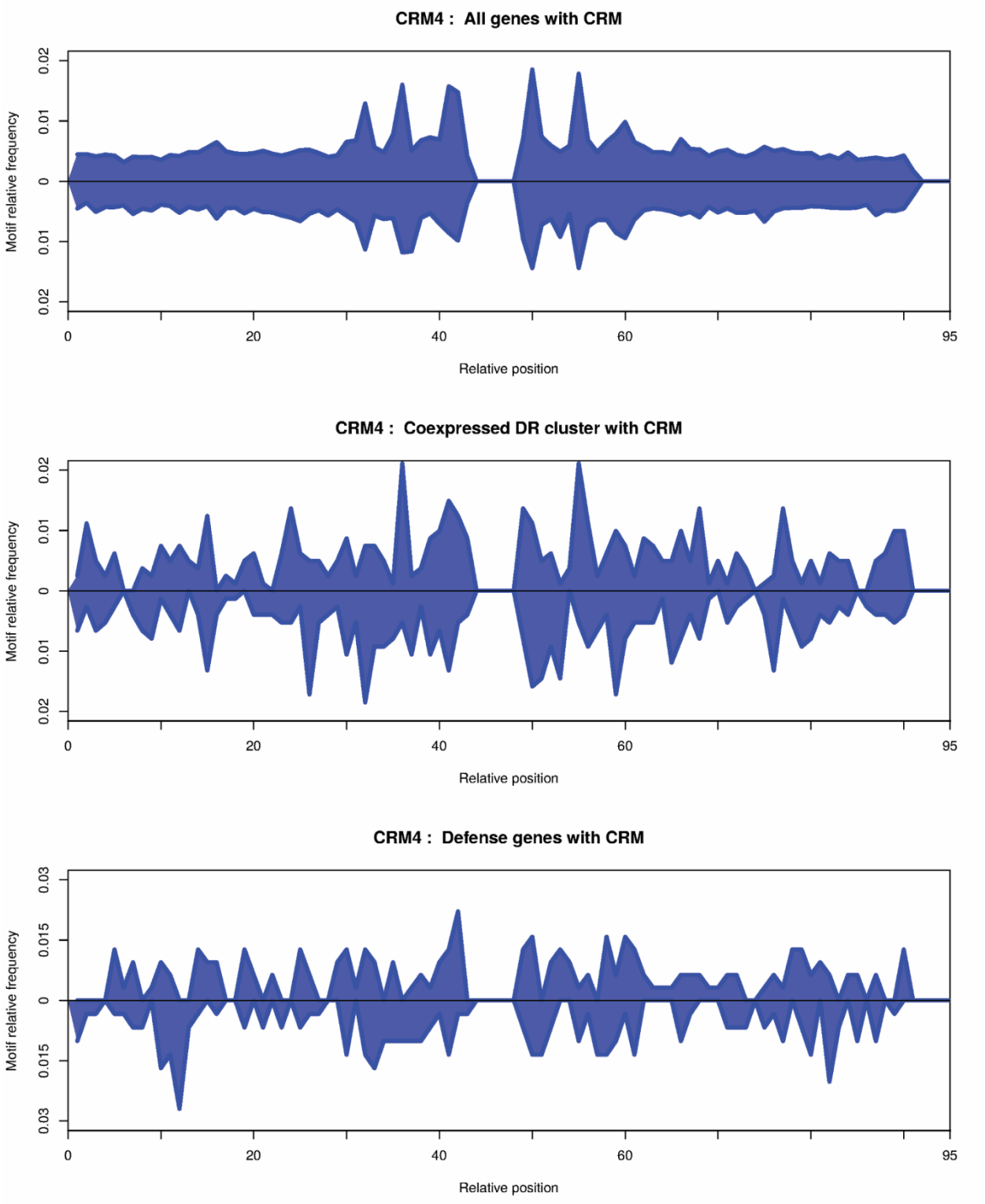
(A) CRM1


CRM2 : All genes with CRM
Motif relative frequency
0.01
0.005
0
0.005
0.01
0
20
40
60
80
100
120
140
160
199
Relative position
CRM2 : Coexpressed DR cluster with CRM
Motif relative frequency
0.07
0.035
0
0.035
0.07
0
20
40
60
80
100
120
140
160
199
Relative position
CRM2 : Defense genes with CRM
Motif relative frequency
0.1
0.05
0
0.05
0.1
0
20
40
60
80
100
120
140
160
199
Relative position

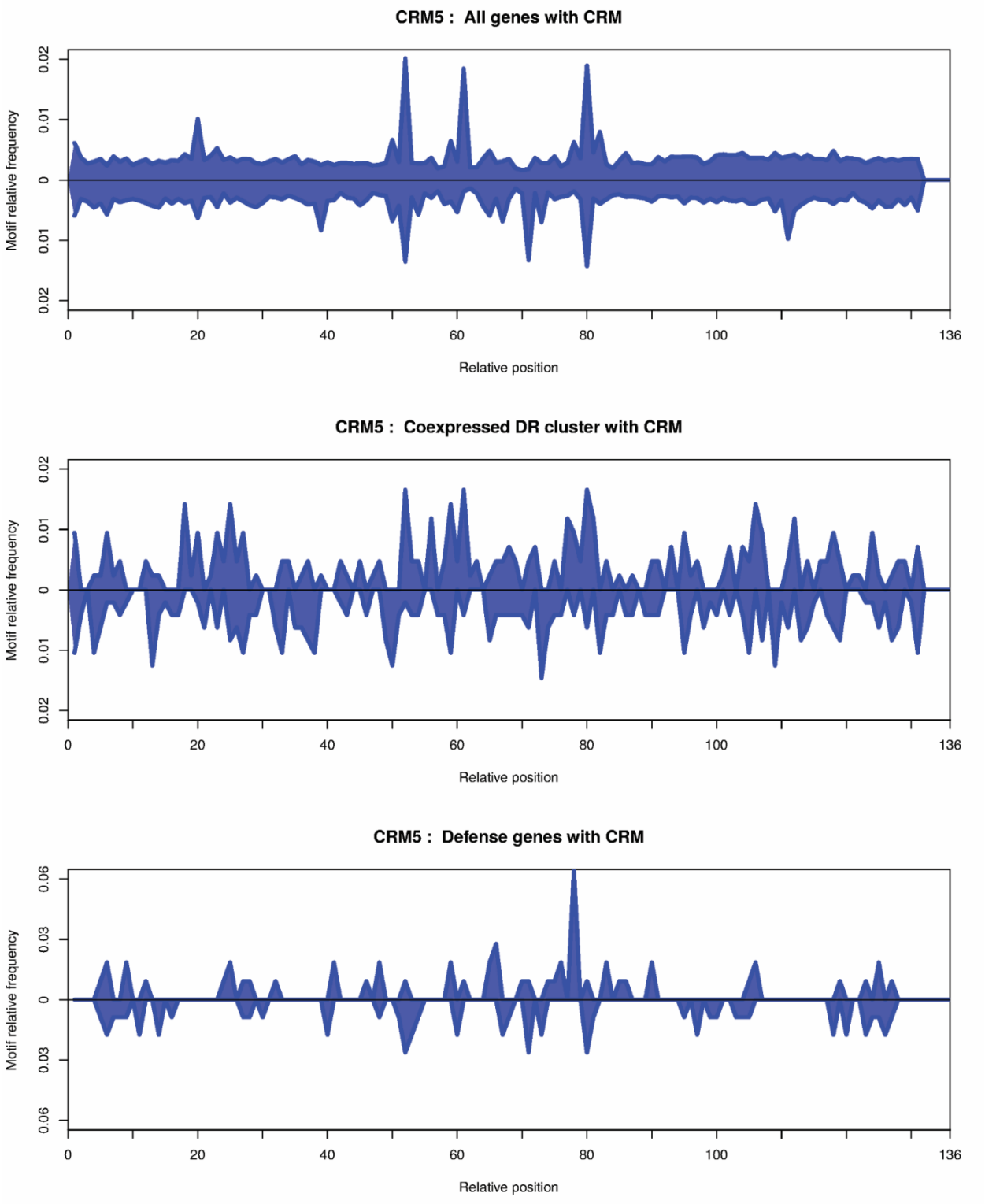
(B) CRM2


CRM3 : All genes with CRM
Motif relative frequency
0.1
0.05
0
0.05
0.1
0 20 40 60 80 100 120 140 160 180 200 220 245
Relative position
CRM3 : Coexpressed DR cluster with CRM
Motif relative frequency
0.2
0.1
0
0.1
0.2
0 20 40 60 80 100 120 140 160 180 200 220 245
Relative position
CRM3 : Defense genes with CRM
Motif relative frequency
0.25
0.125
0
0.125
0.25
0 20 40 60 80 100 120 140 160 180 200 220 245
Relative position

(C) CRM3


CRM4 : All genes with CRM
Motif relative frequency
0.02
0.01
0
0.01
0.02
0
20
40
60
95
Relative position
CRM4 : Coexpressed DR cluster with CRM
Motif relative frequency
0.02
0.01
0
0.01
0.02
0
20
40
60
95
Relative position
CRM4 : Defense genes with CRM
Motif relative frequency
0.03
0.015
0
0.015
0.03
0
20
40
60
95
Relative position

(D) CRM4



(E) CRM5

Figure 3-7: CRM motif densities. Frequency for each nucleotide position across the CRM1-5 (A-E) windows was calculated as a ratio of total occurrences of the constituent motif in the CRM found in all genes, BS-DR cluster genes, and FA-DR genes ("Defense Genes"). The strand which the motif is found is given as either below (- strand) or above (+ strand) the Y-axis origin.

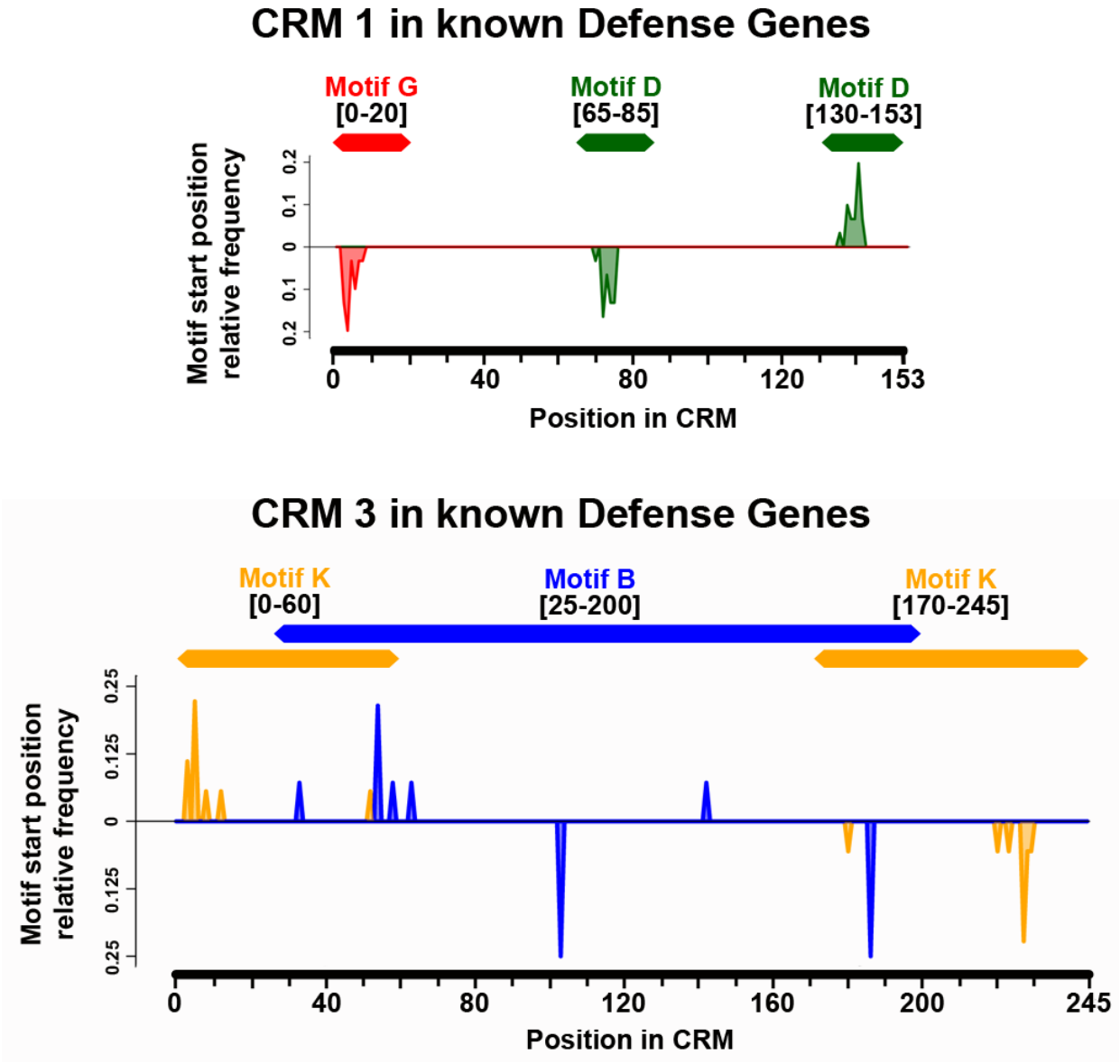


Figure 3-8: Constituent motif occurrence frequency within CRM1 and CRM3. The relative frequency of occurrence (Y-axis) is given for the constituent motifs for both CRMs. Frequency for each nucleotide position across the CRM window was calculated as a ratio of total occurrences of the constituent motif in the CRM found in all FA-DR genes ("Defense Genes"). The strand which the motif is found is given as either below (- strand) or above (+ strand) the Y-axis origin.

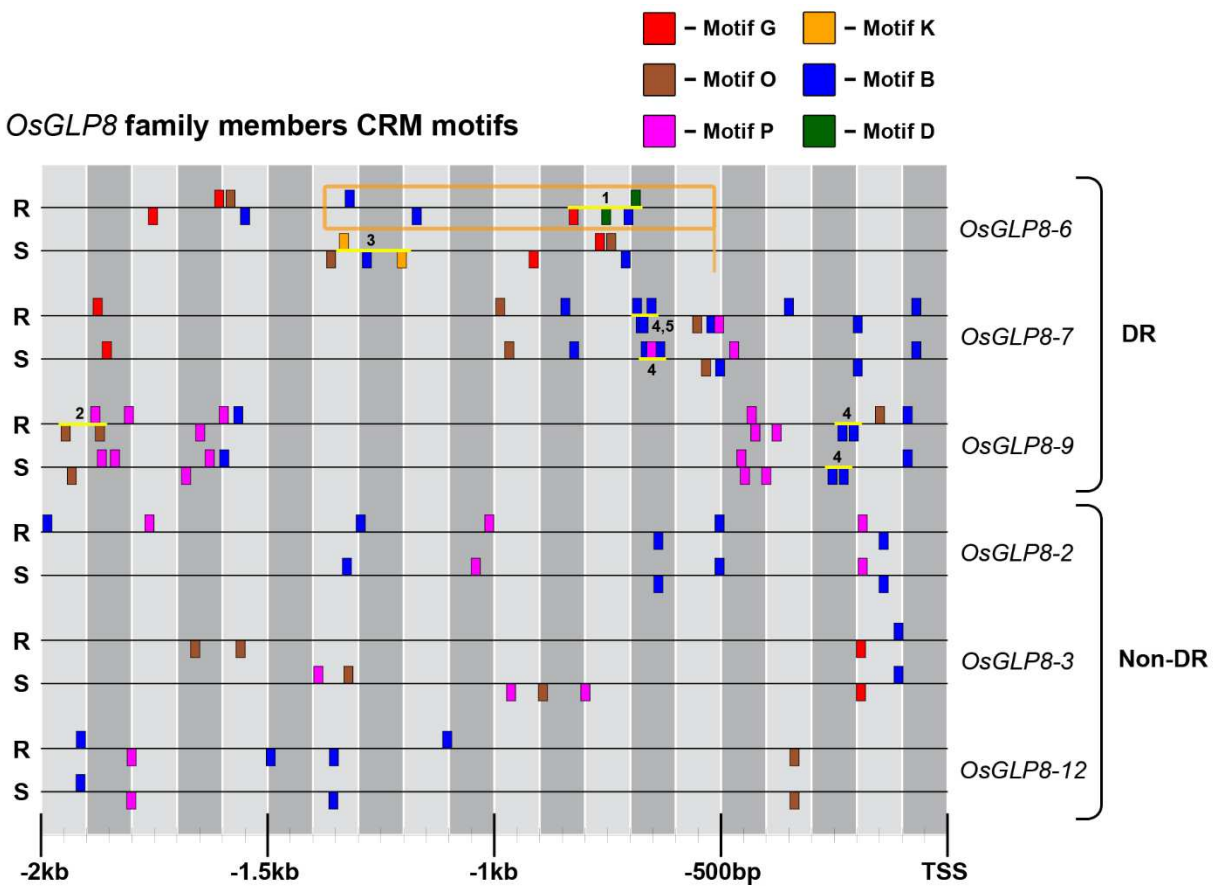


Figure 3-9a: CRM motifs in *OsGLP8* DR and non-DR genes.

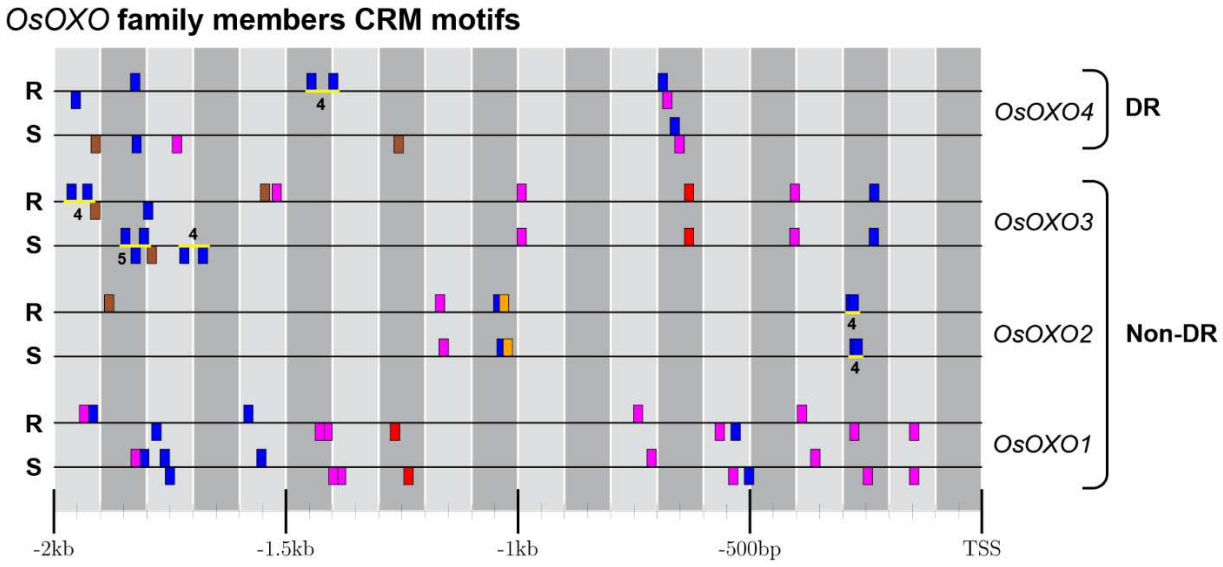


Figure 3-9b: CRM motifs in *OsOXO* DR and non-DR genes.

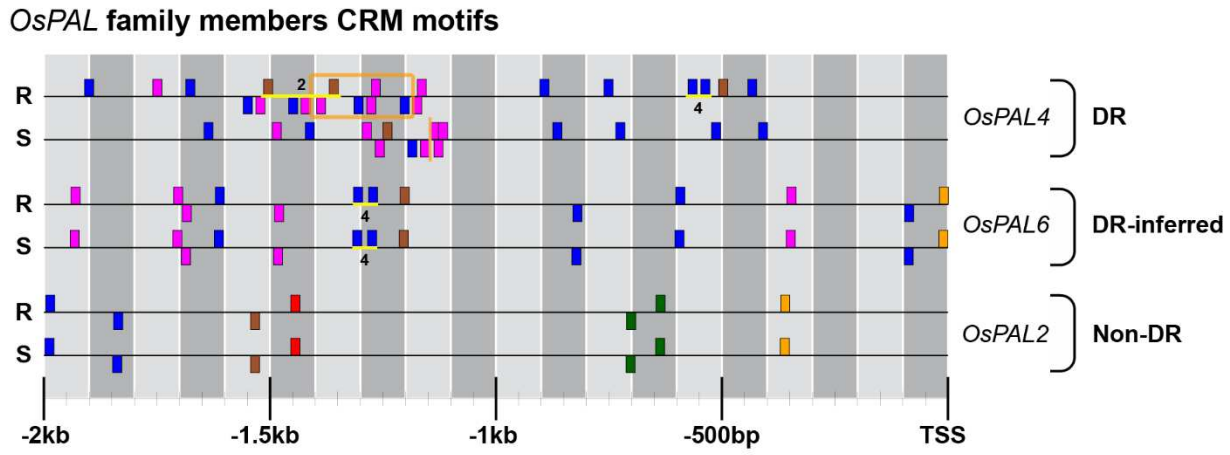


Figure 3-9c: CRM motifs in *OsPAL* DR and non-DR genes.

Figure 3-9: CRM differences between resistant and susceptible promoter haplotypes of DR gene family members. Each genes' promoter is shown as a horizontal black line starting from the left at -2kb, ending at the right at -1 from the TSS. Genes that have been found to be functionally related or non-related to the DR are labelled on the right. Constituent motifs from CRMs 1-5 are shown as colored boxes either above the line (+ strand) or below the line (- strand) in each promoter. CRMs found within the promoters are highlighted bright yellow and labeled by number. Indels of interest are boxed with orange color, and the site of indel in the opposing variety is shown as a vertical orange line. Sequences for specific varieties' promoters were obtained from the 3000 rice genomes project. The promoters of *OsGLP8* family members that reside in the rice blast disease resistance QTL for both resistant (SHZ-2) and susceptible (LTH) donors are given (a). The promoters of *OsOXO* family members that reside in the rice blast disease resistance QTL are shown for the susceptible (Azucena) donor. The resistant donor, Moroberekan, does not have available genome sequence, thus another japonica variety, Nipponbare, was used (b). The promoters of *OsPAL* family members that reside in the sheath blight disease resistance QTL for resistant (IR64) and susceptible (Azucena) are given (c).

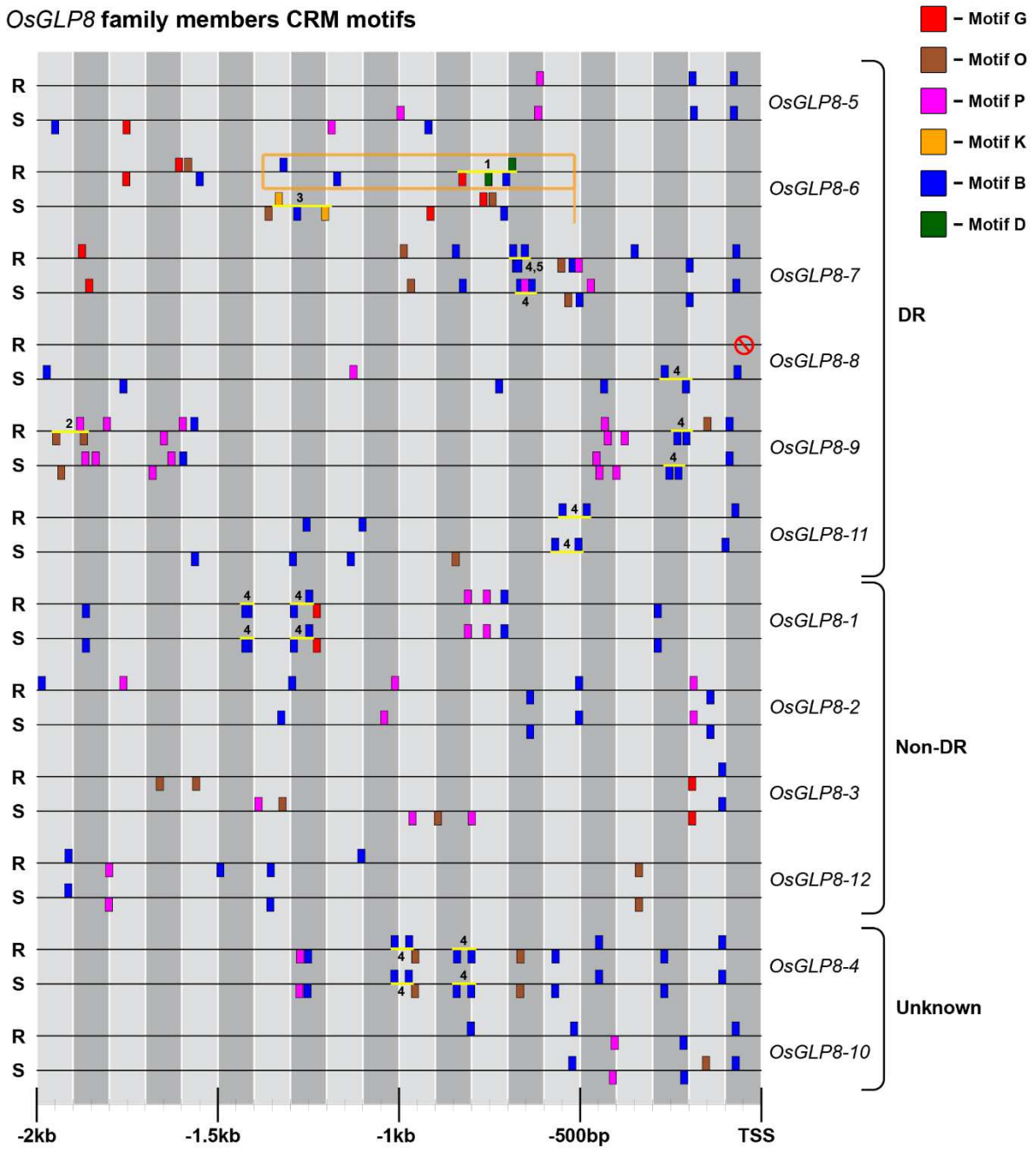


Figure 3-10a: Full OsGLP8 gene family CRM motif profiles. CRM motif profiles in resistant (SHZ-2) and susceptible (LTH) haplotypes.

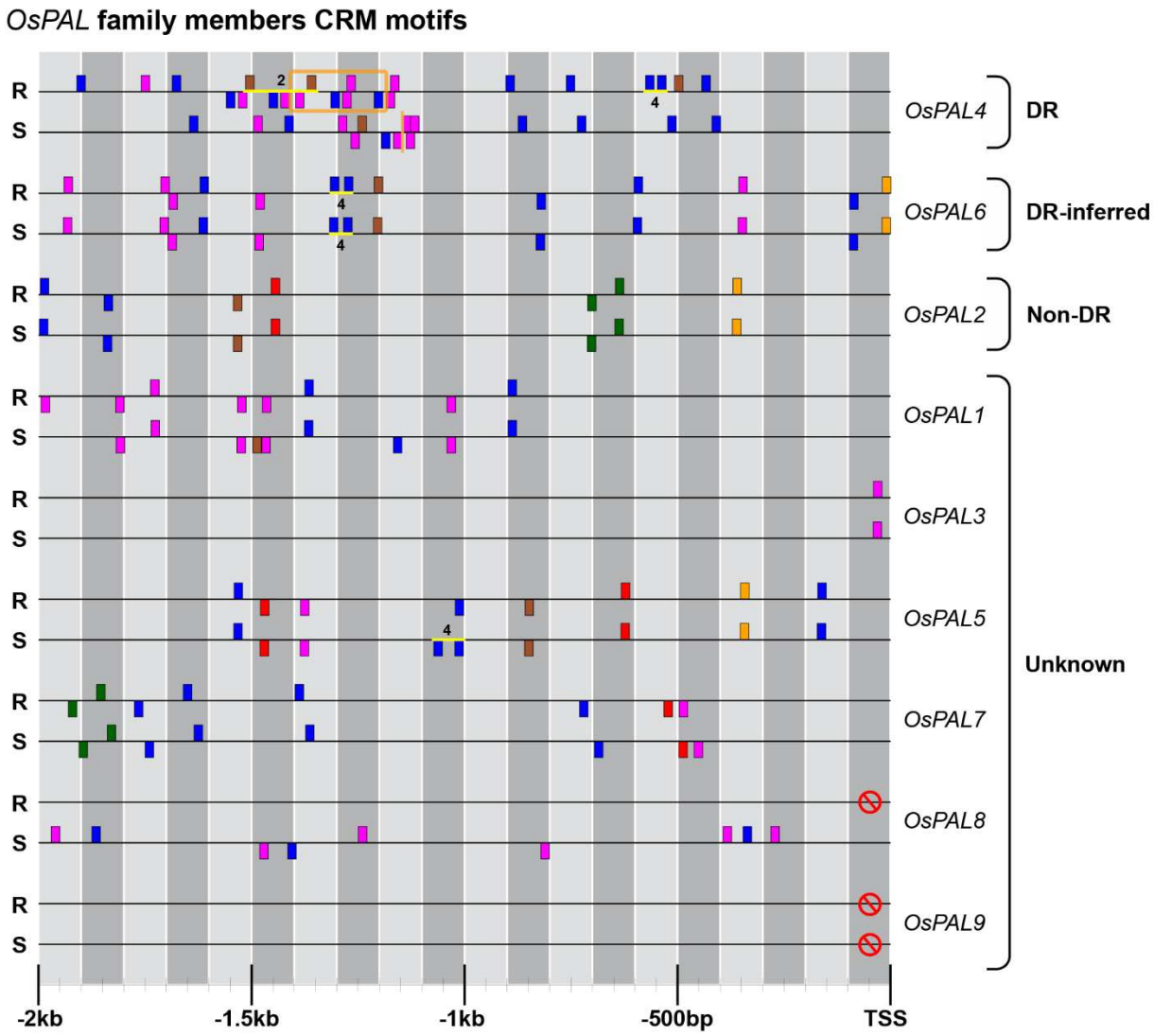


Figure 3-10b: Full OsPAL gene family CRM motif profiles. CRM motif profiles in resistant (IR64) and susceptible (Azucena) haplotypes.

Figure 3-10: Full gene family CRM motif profiles. The CRM motifs within each of the entire DR gene families are shown for *OsGLP8* (a) and *OsPAL* (b). These results include gene family members that have unknown functions in resistance positively or negatively. Red interdictory circles indicate a lack of promoter and/or gene sequence in the variety.

Figure 3-11: Consensus sequences of CRM1 and CRM3. The consensus sequence shown is a visual representation of the percentage of aligned nucleotides across every occurrence of each respective CRM in FASTA format. Each letter is the majority nucleotide for each position on the alignment. Capital letters represent nucleotides that are conserved in at least 50% of all sequences, and lowercase letters are less than 50%. The percentage shown on the right is based on the ratio of capital letters to lowercase letters. The length of each sequence is: CRM1 – 152 nt, CRM1-appended – 336 nt, CRM3 – 234 nt. Two CRM1 consensus sequences are shown. The first is generated based on the CRM1 window as seen in Figure 3-8. The second, "CRM1-appended," is generated from CRM1 occurrences with an additional 184 nt extended from the 3' end. This was due to the evidence of a larger "self-complementary" motif structure seen in FA-DR genes (Figure 3-12a). The CRM3 occurrence in the *OsGLP8-6* promoter is illustrated for comparison.

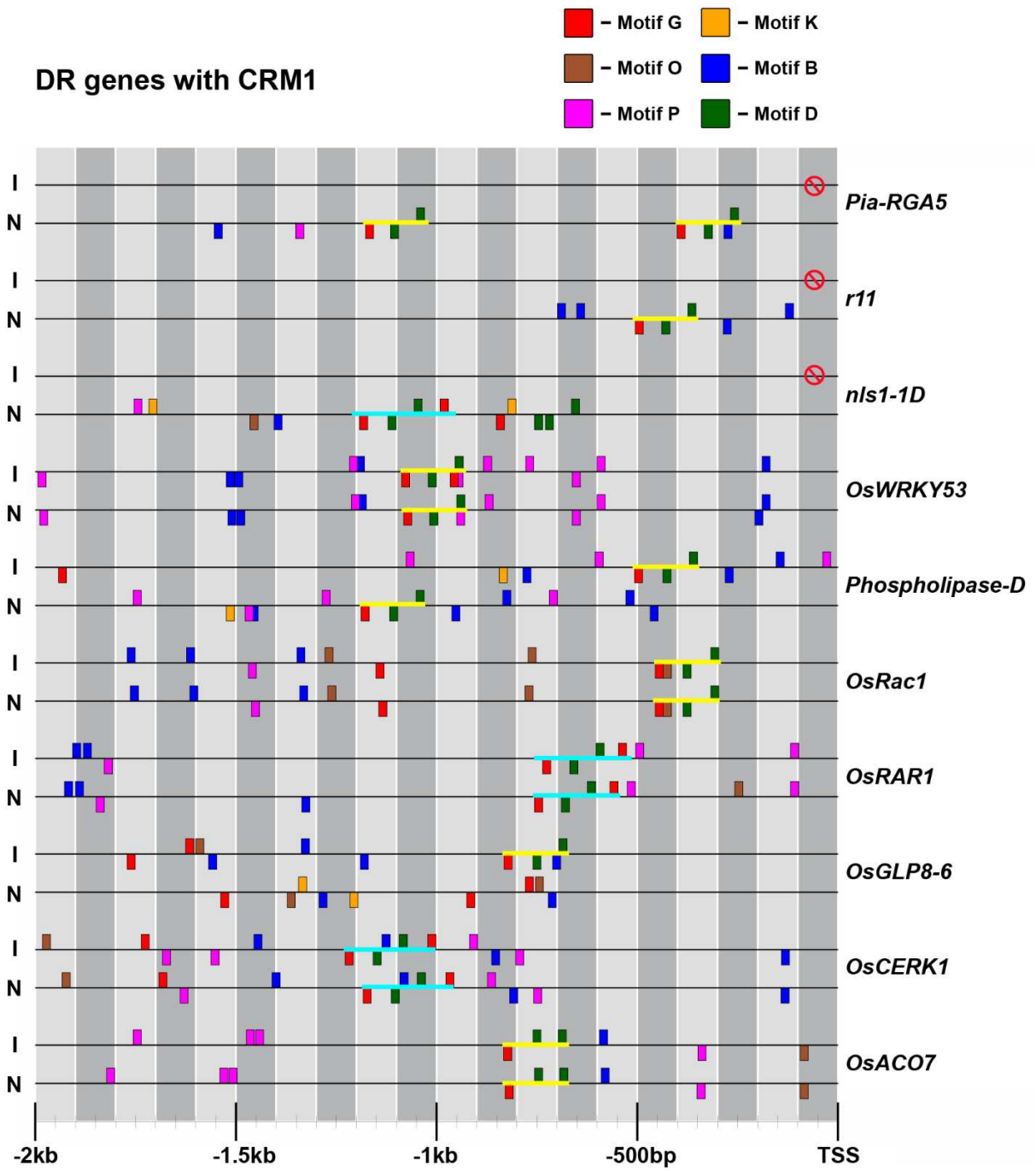


Figure 3-12a: CRM1 occurrences in DR genes.

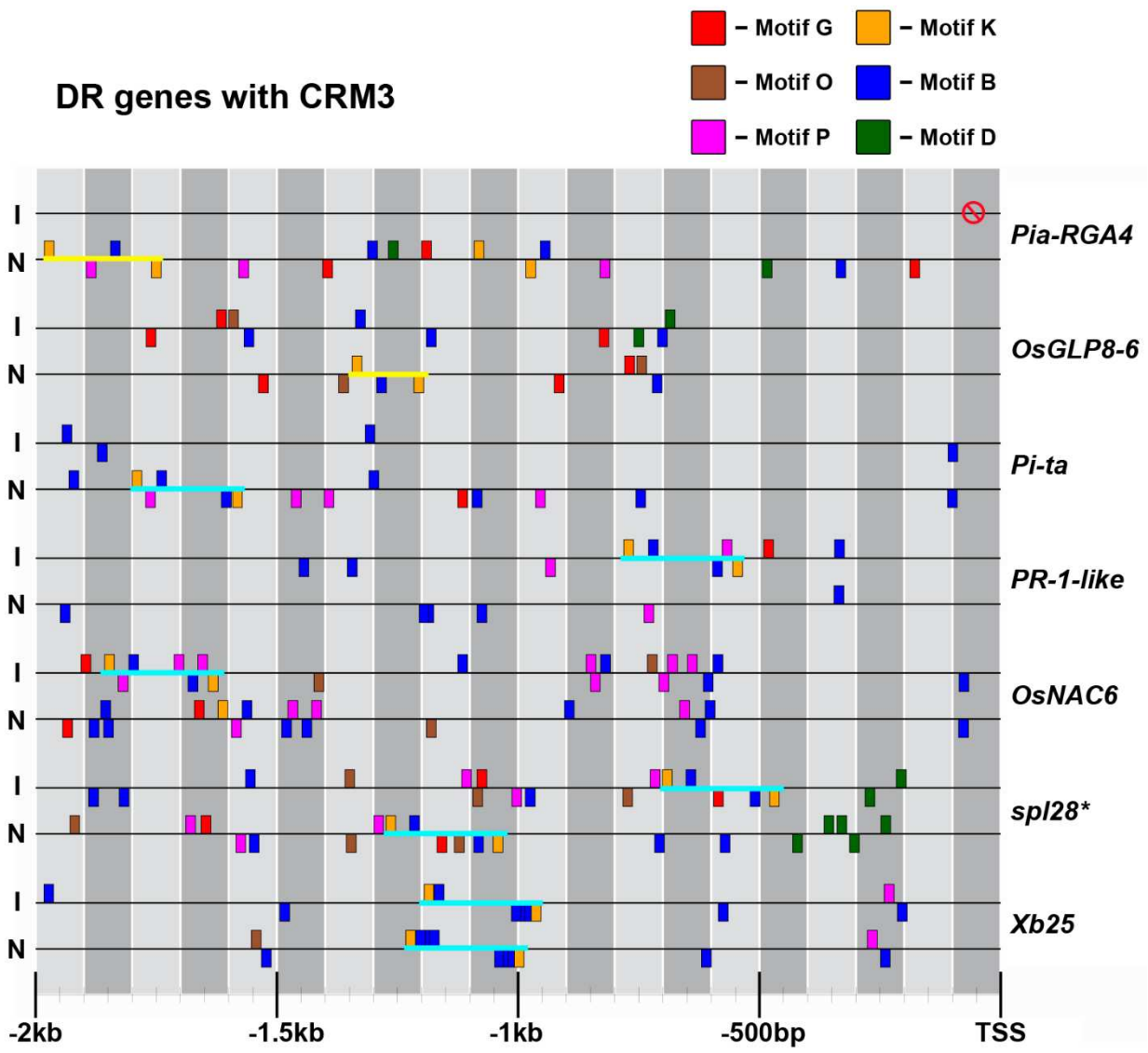


Figure 3-12b: CRM3 occurrences in DR genes.

Figure 3-12: CRM occurrences in known DR genes. The CRM motifs within each of the entire DR gene promoter for CRM1 (a) and CRM3 (b). Each DR gene promoter is from IR64 (I) or Nipponbare (N). Red interdictory circles indicate a lack of promoter and/or gene sequence in the variety. CRMs are highlighted in bright yellow or cyan. Cyan-highlighted regions are the larger, "self-complementary" version of the respective CRM. The cyan-highlighted versions of CRM1 are what prompted the generation of a "CRM1-appended" consensus sequence as seen in Figure 3-11. Cyan-highlighted CRM3 occurrences contain complementary Motif B sites.

(A) CRM1-appended / miR6225

```
    Query: 198-336
    sbi-MIR6225: 164-304
    score: 326
    evalue: 6e-20

UserSeq           198  uaaugauggauuaaauagauuuaaaagauucgucucgcaauuuacacgcaaauuuguguaauuaguuuuuuuuauuagucuacguuuaaauacuucauacaugugu--uacaauauuuaugaugugacacgcuaaaaaauuuu  336
                       |||| ||   |||| |||| ||||||||||||| |||||| |||||  |||||| | ||||||||||||| | ||  ||||  |||| |  || || |||||| |   | | |  |||||||| || |   | |||| ||||
sbi-MIR6225       304  uaauaauauauuuaauuagacuuaaaagauucguuucgcaaauuacagacaaauuauauaauuuaguuuuuauuuuucaucuauauuuuaguguuuaaugcaugugccguaagauucauuaugaugggaaaucuugaaaaguuuu  164


    Query: 198-299
    sbi-MIR6225: 1-102
    score: 312
    evalue: 9e-19

UserSeq           198  uaaugauggauuaaauagauuuaaaagauucgucucgcaauuuacacgcaaauuuguguaauuaguuuuuuuuauuagucuacguuuaaauacuucauacaugug  299
                       |||| ||| | ||| |||| | |||||||||| |||| | ||||| |   |||||||| |||||||||||| | || |||||  |||||| |||||  ||||||
sbi-MIR6225         1  uaaucauguacuaacuagacucaaaagauucaucucccgauuuauagauaaauuugugcaauuaguuuuuguuuuuugucuauauuuuaaugcuucacgcaugug  102


    Query: 2-94
    sbi-MIR6225: 2-94
    score: 258
    evalue: 3e-14

UserSeq             2  ugaaguauuaaaauuaaaauuaaaaaaaaaacaaauuacacaguuugcguguaaaauugcgagacgaaucuuuuaagccuaauuacuccaugauu  94
                       ||||| ||||||   | |  |||| |||||| |||| |||| |||   | ||||| | |||| |||||||| || ||| ||| | |||||||
sbi-MIR6225        94  ugaagcauuaaaauauagacaaaaacaaaaacuaauuugcacaauuuaucuauaaaucgggagaugaaucuuuugagucuaguuaguacaugauu  2


    Query: 2-94
    sbi-MIR6225: 211-303
    score: 249
    evalue: 1e-13

UserSeq             2  ugaaguauuaaaauuaaaauuaaaaaaaaaacaaauuacacaguuugcguguaaaauugcgagacgaaucuuuuaagccuaauuacuccaugauu  94
                       | ||  | ||||   | || |||| |||||| ||||| | | ||||  ||||| |||||| ||||||||||||| |||||||    || |||
sbi-MIR6225       211  uuaaacacuaaauauagaugaaaauaaaaacuaauuauauauauuugucuguaauuugcgaaacgaaucuuuuaagucuaauuaaauuauuauu  303
```

(B) CRM1-appended / Harb-MITE_0L024C_6443

```
>DTHm_#DNA/Harb-MITE_OLO24C_6443
          Length = 356

 Score =  186 bits (94), Expect = 4e-46
 Identities = 123/132 (93%), Gaps = 4/132 (3%)
 Strand = Plus / Minus


Query: 22  aaaaaaaaaacaaattacacagtttgcgtgtaaattgcgagacgaatcttttaagcctaa 81
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 263 aaaaaaaaaccaattacacagtttgcgtgtaaattgcgagacgaatcttttaagcctaa 204


Query: 82  ttactccatgatttgcaaaatgtgttgctacagtaaacatttgctaatcatgacggatta 141
           || | |||||||||| | |||||| |||||||||||||||||||||   |||||||||
Sbjct: 203 ttgcgccatgatttg-acaatgtggtgctacagtaaacatttgctaat---gacggatta 148


Query: 142 attaggcttaat 153
           ||||||||||||
Sbjct: 147 attaggcttaat 136


 Score =  119 bits (60), Expect = 8e-26
 Identities = 78/84 (92%)
 Strand = Plus / Plus


Query: 218 ttaaaagattcgtctcgcaatttacacgcaaattgtgtaattagtttttttattagtcta 277
           |||||||||||||||||||||||||||||||||| ||||||||| ||||||||  ||||
Sbjct: 210 ttaaaagattcgtctcgcaatttacacgcaaactgtgtaattggtttttttttttttcta 269


Query: 278 cgtttaatacttcatacatgtgtc 301
           | |||||||||||||||||||||
Sbjct: 270 catttaatacttcatacatgtgtc 293
```

Figure 3-13a: CRM1 alignment with miRNA and MITE sequences.

(A) CRM3 top alignments with known miRNAs

Accession	ID	Query Start	Query Stop	Subject Start	Subject Stop	Strand	Score	E-value
MI0005233	osa-MIR812a	47	222	33	197	-	307	2.00E-18
MI0019807	osa-MIR812r	23	222	25	213	-	292	3.00E-17
MI0019800	osa-MIR812q	23	261	2	225	-	287	8.00E-17
MI0005237	osa-MIR812e	41	221	2	169	-	264	6.00E-15
MI0005234	osa-MIR812b	51	222	11	169	+	255	3.00E-14
MI0023432	osa-MIR818f	35	260	47	258	+	253	5.00E-14
MI0019807	osa-MIR812r	33	211	17	184	+	241	5.00E-13
MI0017255	osa-MIR812o	41	222	3	171	-	233	2.00E-12
MI0019817	osa-MIR812t	38	220	29	199	-	228	6.00E-12
MI0019800	osa-MIR812q	82	261	67	241	+	224	1.00E-11
MI0005251	osa-MIR818e	34	243	9	202	-	222	2.00E-11
MI0005233	osa-MIR812a	33	216	25	197	+	212	1.00E-10
MI0017246	osa-MIR812l	38	226	34	209	-	209	2.00E-10
MI0017245	osa-MIR812k	56	207	44	183	-	208	3.00E-10

(B) CRM3 / Tc-Mariner_Chr1_5045258-5045503_36894

```
          Length = 246

 Score = 75.8 bits (38), Expect = 8e-13
 Identities = 41/42 (97%)
 Strand = Plus / Minus

Query: 43  ttttttttacaatgtttgaccgttcgtcttatttaaaaaattt 84
           |||||||| |||||||||||||||||||||||||||||||||
Sbjct: 210 ttttttttaaaatgtttgaccgttcgtcttatttaaaaaattt 169


 Score = 75.8 bits (38), Expect = 8e-13
 Identities = 41/42 (97%)
 Strand = Plus / Plus

Query: 179 aaatttttttaaataagacgaacggtcaaacattgtaaaaaaa 220
           |||||||||||||||||||||||||||||||| ||||||||
Sbjct: 169 aaatttttttaaataagacgaacggtcaaacattttaaaaaaa 210
```

Figure 3-13b: CRM3 alignment with miRNA and MITE sequences. A table of top alignments with miRNA stem loops is included due to the redundancy of matching results. CRM3 consensus sequences are aligned to miRNA sequences or MITE sequence, Tc-Mariner.

Figure 3-13: CRM1 and CRM3 consensus alignments with repeat elements. Top alignment results from CRM1-appended (a) and CRM3 (b) consensus sequences with miRNA stem loops from miRBase (www.mirbase.org) or repeat elements from RITE database (www.genome.arizona.edu/cgi-bin/rite/index.cgi).All alignment results are scored with E-values less than 1e-10.

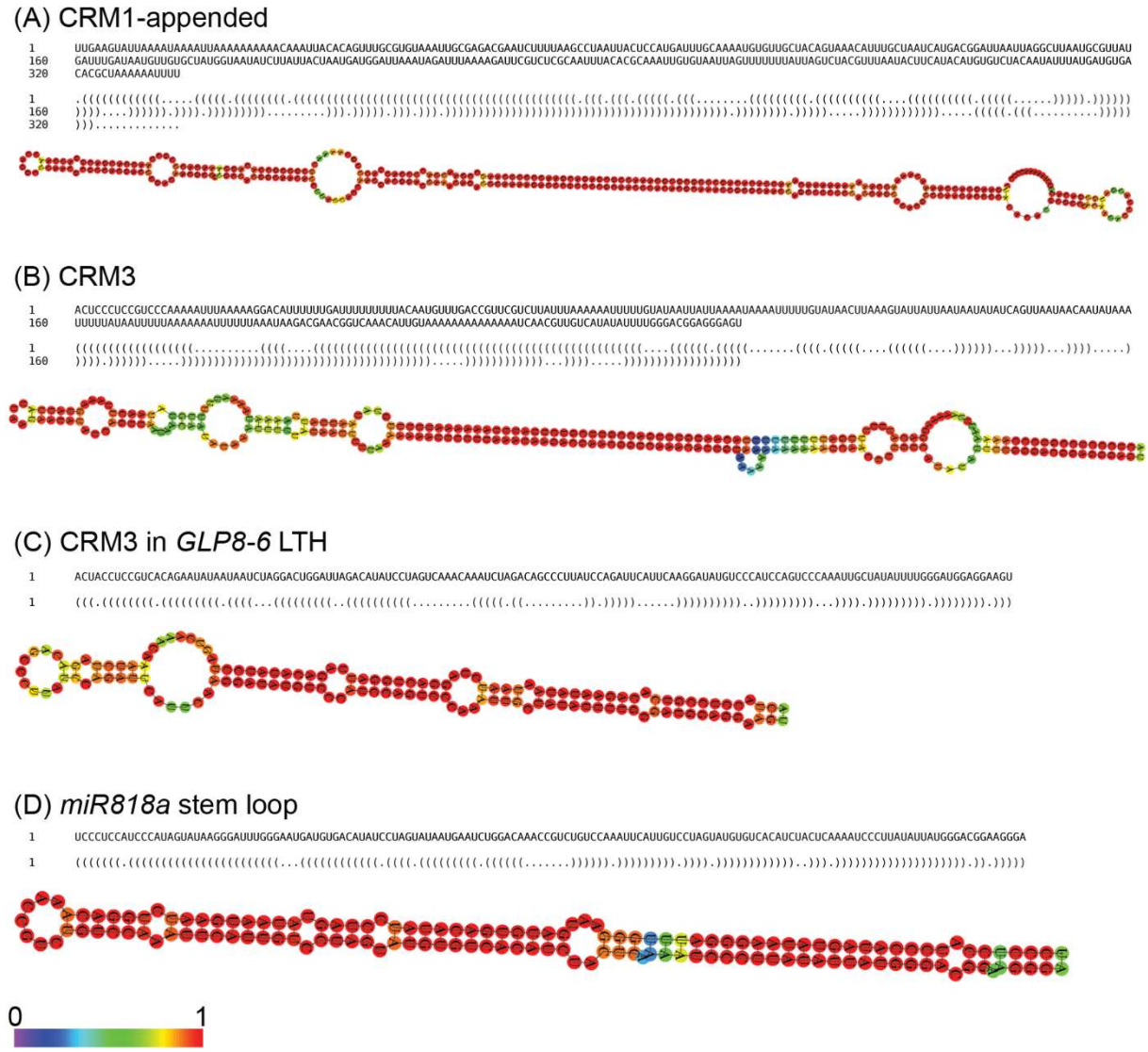
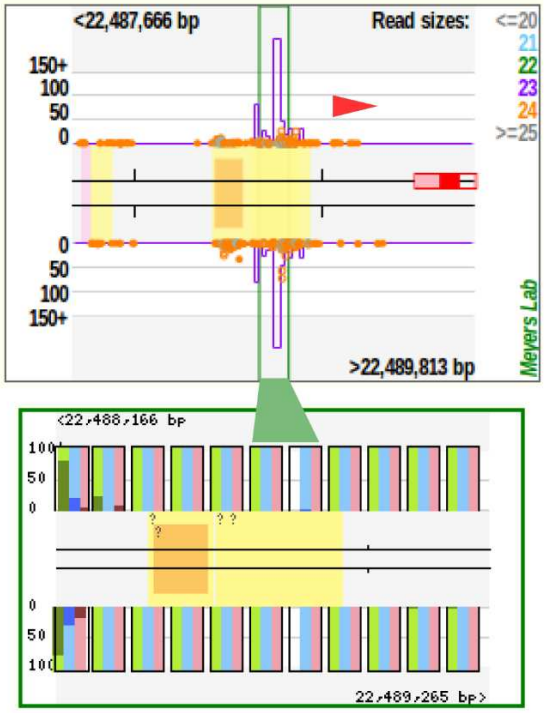


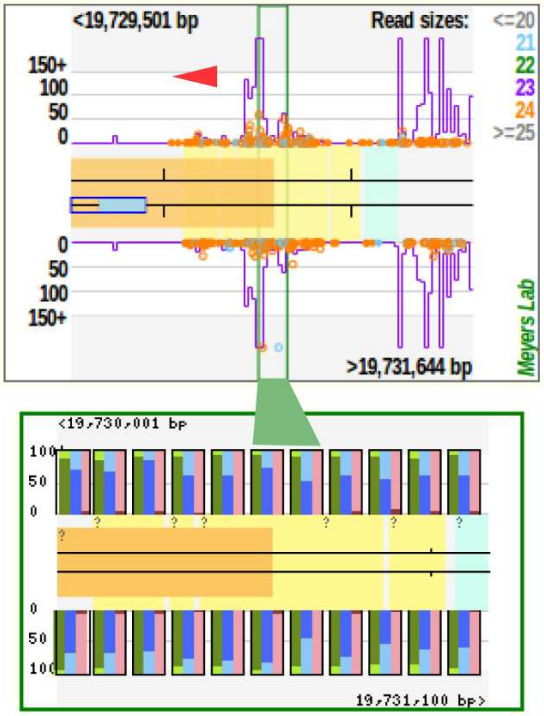
Figure 3-14: CRM1 and CRM3 secondary structure. Putative RNA secondary structure generated using the program RNAfold (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). The optimal secondary structure based on minimum free energy is illustrated in dot-bracket notation and graphically using base pair probabilities.

CRM1 total sRNA reads

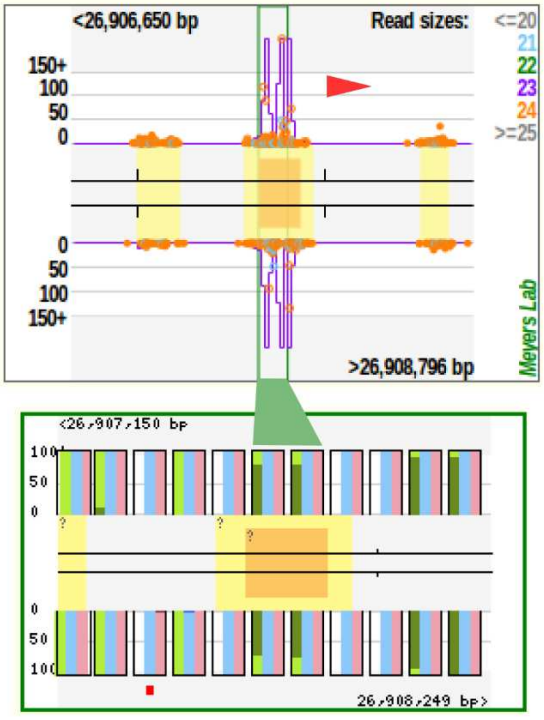
OsACO7 – LOC_Os01g39860
CRM1: -829 to -682



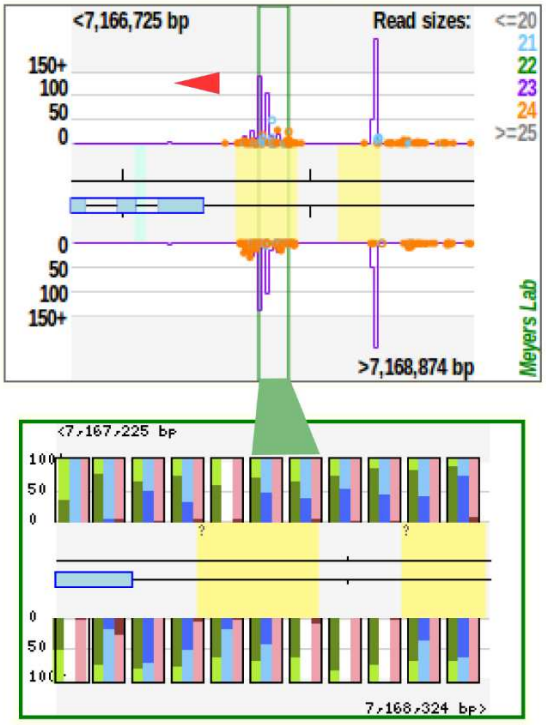
OsRAR1 – LOC_Os02g33180
CRM1: -756 to -613



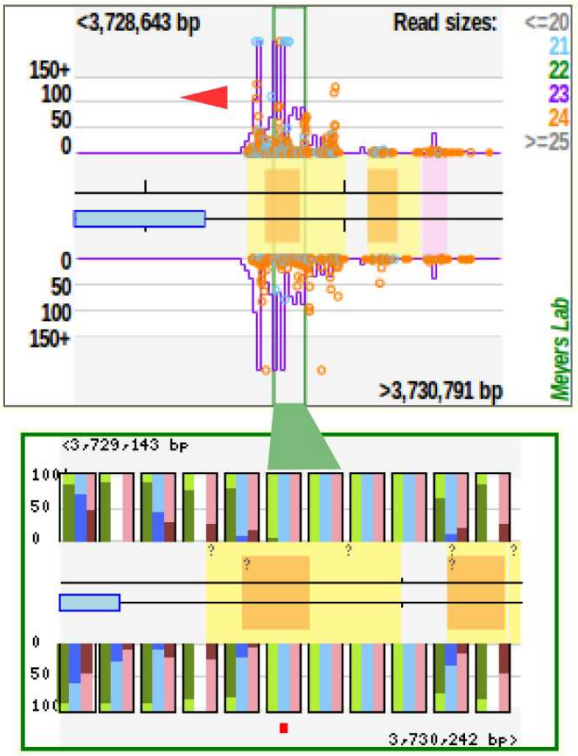
OsCERK1 – LOC_Os08g42580
CRM1: -1183 to -1037



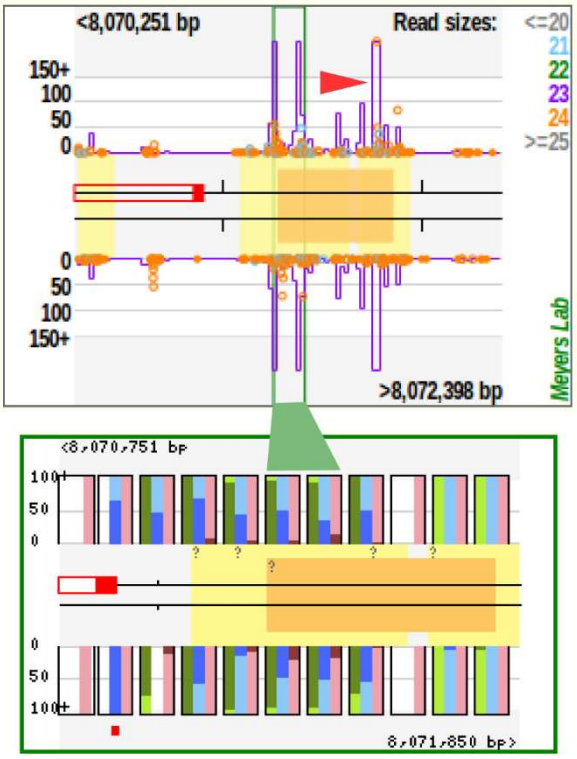
OsRac1 – LOC_Os01g12900
CRM1: -455 to -306



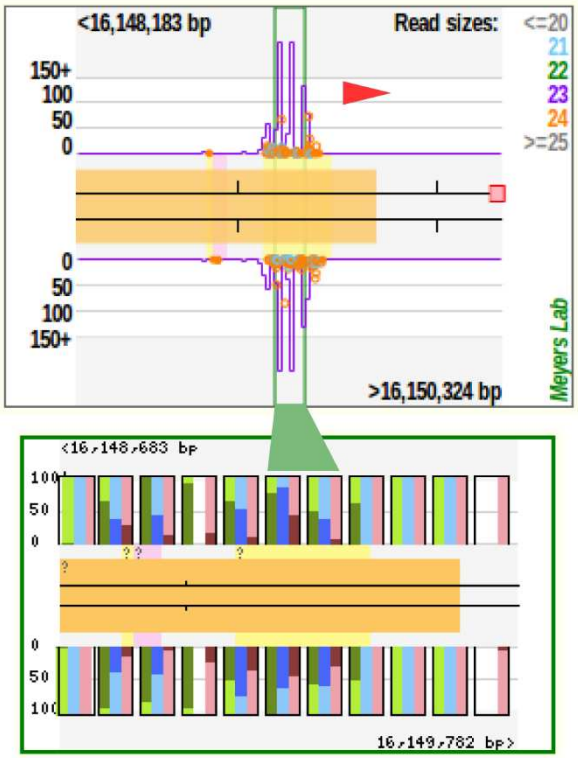
Phospholipase-D – LOC_Os01g07760
CRM1: -1188 to -1040



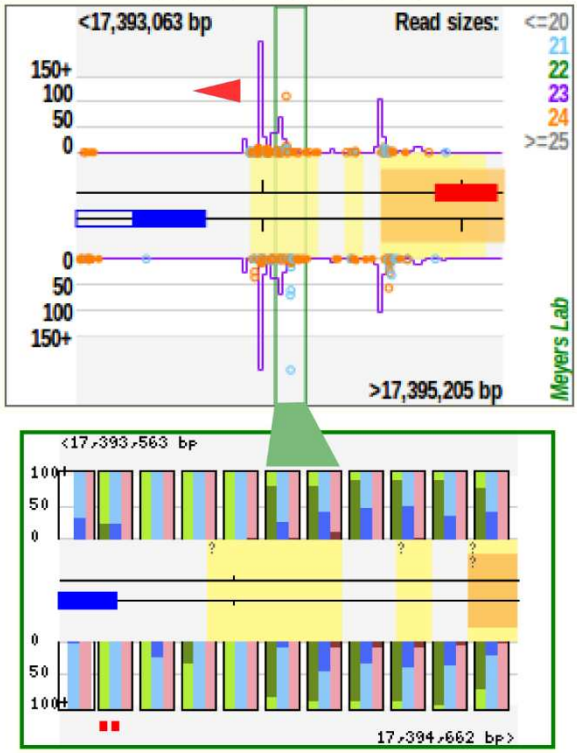
nls1-1D – LOC_Os11g14380
CRM1: -1192 to -1045



OsWRKY53 – LOC_Os05g27730
CRM1: -1082 to -941



r11 – LOC_Os11g29920
CRM1: -505 to -363



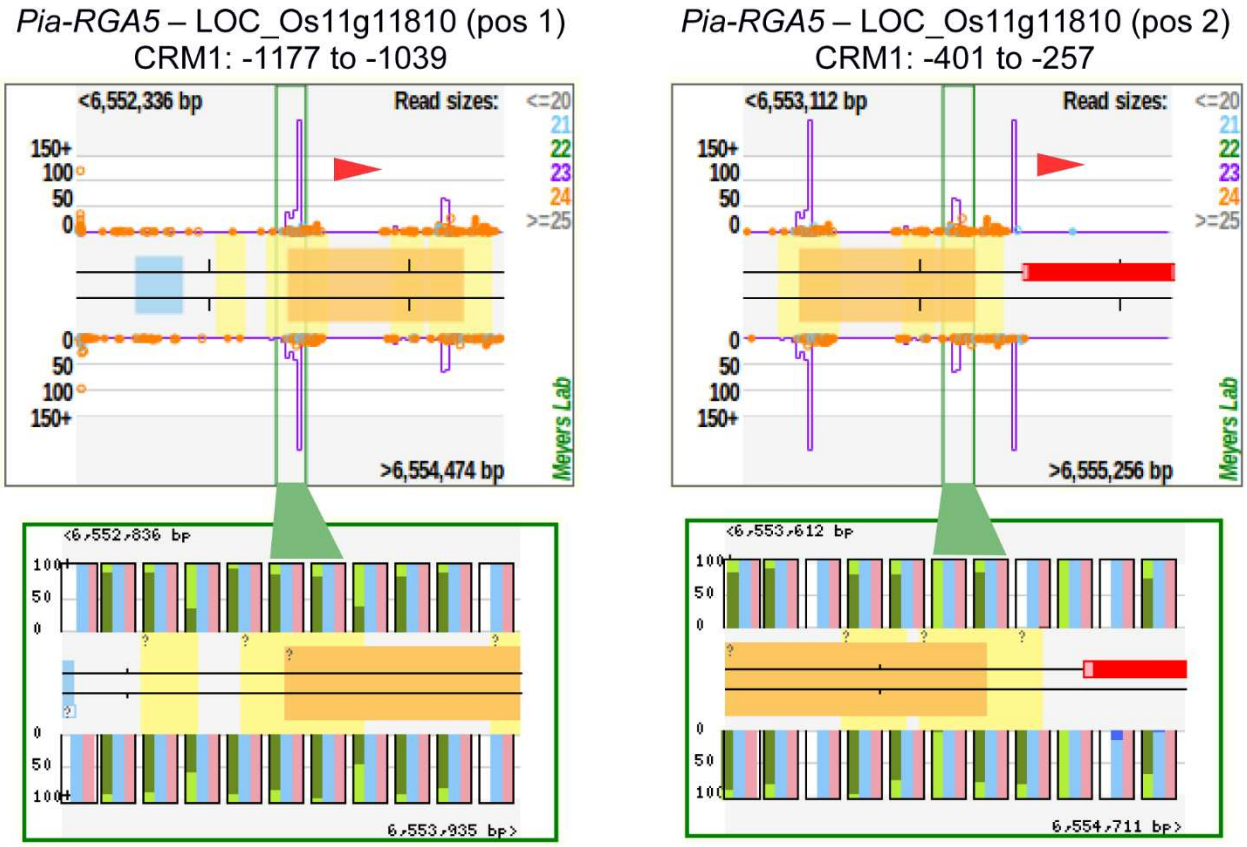
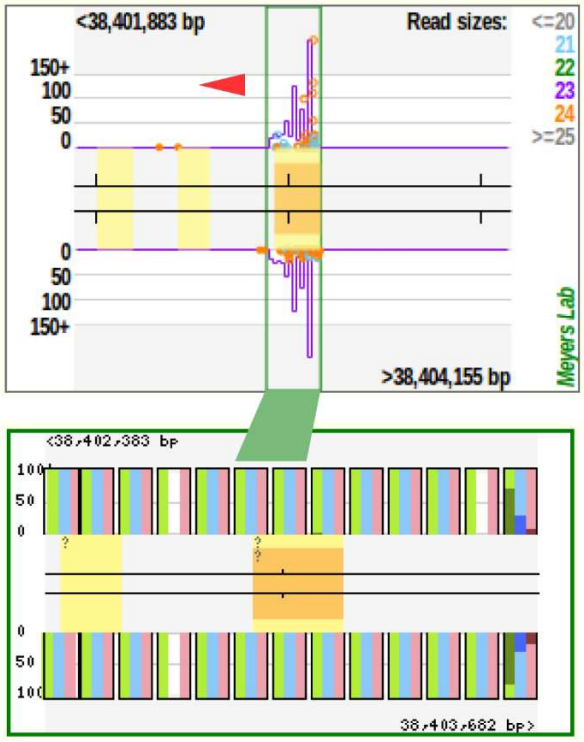


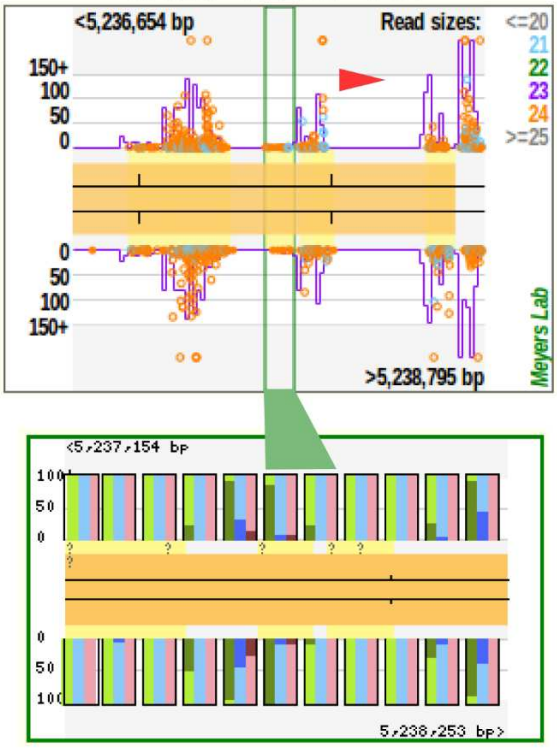
Figure 3-15a: CRM1 sRNA reads and methylation in known DR genes.

CRM3 total sRNA reads

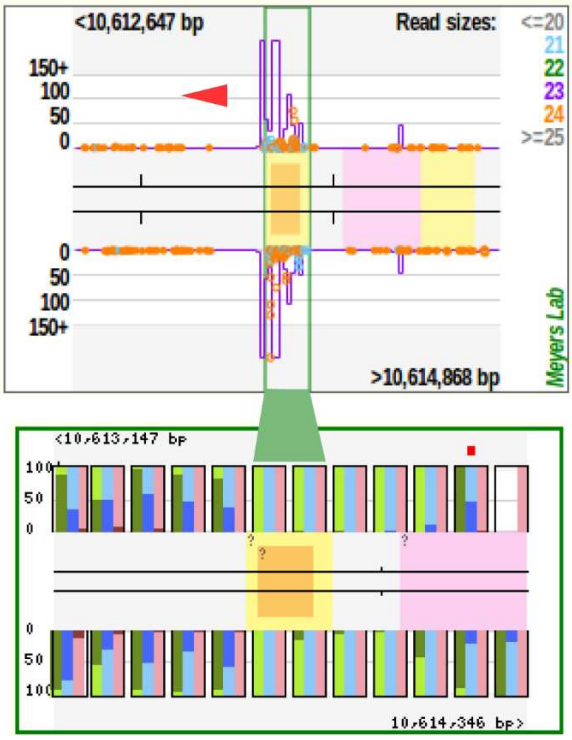
OsNAC6 – LOC_Os01g66120
CRM3: −1622 to -1350



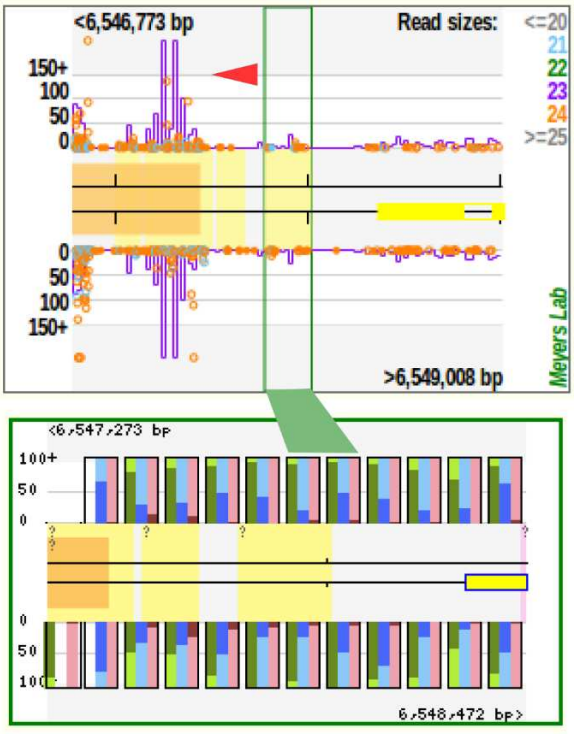
OsGLP8-6 – LOC_Os08g09000
CRM3: -1344 to -1203



Pi-ta – LOC_Os12g18360
CRM3: -1800 to -1579



Pia-RGA4- LOC_Os11g11790
CRM3: -1982 to -1747



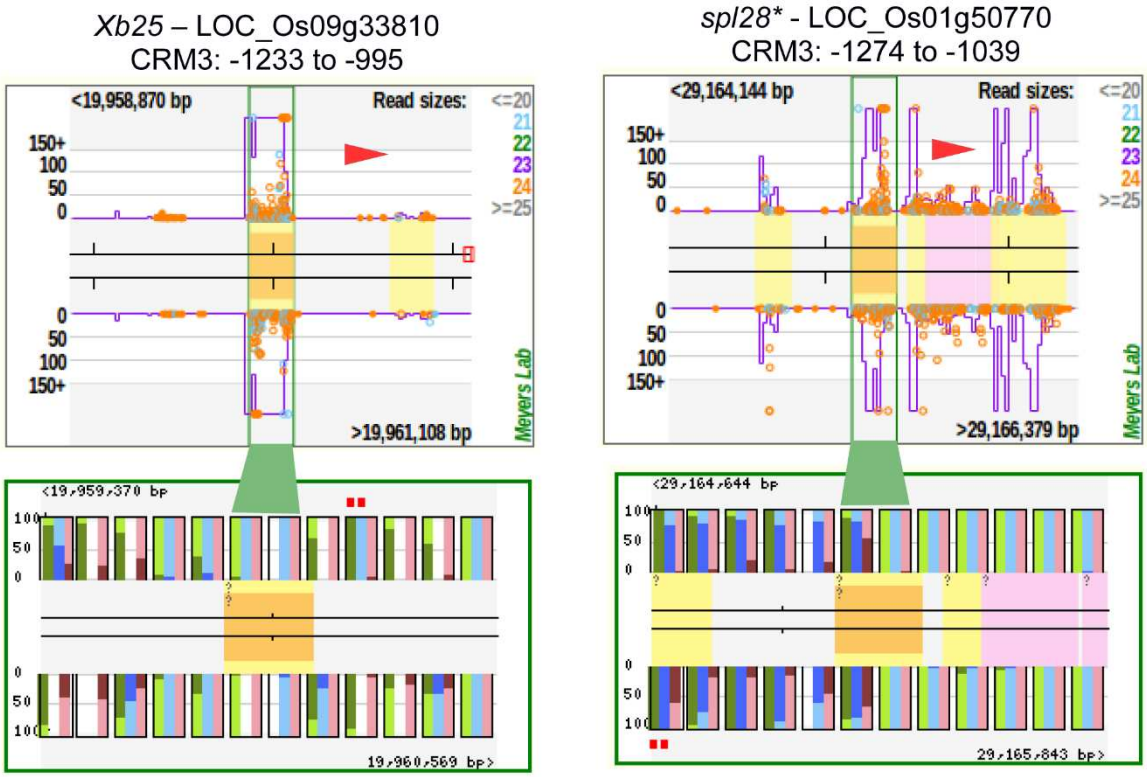


Figure 3-15b: CRM1 sRNA reads and methylation in known DR genes.

Figure 3-15: Total sRNA reads and methylation in CRM1 and CRM3 locations within known DR gene promoters. For each known DR gene containing a CRM1 (a) or CRM3 (b) a graphical representation of sRNA reads are shown, which was taken from the MPSS sRNA sequencing database (mpss.danforthcenter.org/dbs/index.php?SITE=rice_sRNA) (Nakano, 2006). Green boxes in the top plots (within a 2kb region) show the CRM position. Above and below the X-axis is plus or minus strand-aligned reads. The red arrow indicates the direction of the respective downstream DR gene. Orange dots are sRNA reads mapped to that location. The purple line graph shows the "K-mer" line, or degree of repetitiveness. A K-mer value of >50 indicates a lighly heterochromatic region, and known to produce sRNAs. The bottom graphs show total methylation within the 1kb surrounding the CRM location. Each partition along the X-axis represents a block of 100bp. Three separate methylations are represented: CG (green), CHG (blue), and CHH (red). Each colored bar is the relative percentage of methylated (darker color) to unmethylated (lighter color) reads. The green trapezoid from the top to bottom graphs shows where the CRM location resides in the methylation plot.

References

Akimoto, K., Katakami, H., Kim, H.-J., Ogawa, E., Sano, C. M., Wada, Y., & Sano, H. (2007). Epigenetic Inheritance in Rice Plants. *Annals of Botany*, *100*(2), 205–217. https://doi.org/10.1093/aob/mcm110

Ao, W., Gaudet, J., Kent, J., Muttumu, S., & Mango, S. E. (2004). Environmentally Induced Foregut Remodeling by PHA-4/FoxA and DAF-12/NHR. *Science*, *305*(5691), 1743–1746.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME S UITE : tools for motif discovery and searching, *37*(May), 202–208. https://doi.org/10.1093/nar/gkp335

Ballini, E., Morel, J.-B., Droc, G., Price, A., Courtois, B., Notteghem, J.-L., & Tharreau, D. (2008). A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. *Molecular Plant-Microbe Interactions : MPMI*, *21*(7), 859–68. https://doi.org/10.1094/MPMI-21-7-0859

Banniza, S., & Holderness, M. (2001). *Pathogen biology and diversity in major fungal diseases of rice: recent advances*. The Netherlands: Kluwer Academic Publishers.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*. https://doi.org/10.2307/2346101

Berendzen, K. W., Weiste, C., Wanke, D., Kilian, J., Harter, K., & Dröge-Laser, W. (2012). Bioinformatic cis-element analyses performed in Arabidopsis and rice disclose bZIP- and MYB-related binding sites as potential AuxRE-coupling elements in auxin-mediated transcription. *BMC Plant Biology*, *12*(1), 125. https://doi.org/10.1186/1471-2229-12-125

Boyd, L. a. (2006). Perspective: Can the durability of resistance be predicted? *Journal of the Science of Food and Agriculture*, *86*, 2523–2526. https://doi.org/10.1002/jsfa.2648

Boyd, L. a, Ridout, C., O'Sullivan, D. M., Leach, J. E., & Leung, H. (2013). Plant-pathogen interactions: disease resistance in modern agriculture. *Trends in Genetics : TIG*, *29*(4), 233–40. https://doi.org/10.1016/j.tig.2012.10.011

Cao, Y., Halane, M. K., Gassmann, W., & Stacey, G. (2017). The Role of Plant Innate Immunity in the Legume-Rhizobium Symbiosis. *Annual Review of Plant Biology*, *68*(1), 535–561. https://doi.org/10.1146/annurev-arplant-042916-041030

Carrillo, M. G. C., Goodwin, P. H., Leach, J. E., Leung, H., & Vera Cruz, C. M. (2009). Phylogenomic Relationships of Rice Oxalate Oxidases to the Cupin Superfamily and Their Association with Disease Resistance QTL. *Rice*, *2*(1), 67–79. https://doi.org/10.1007/s12284-009-9024-0

Chen, F., Srinivasa Reddy, M. S., Temple, S., Jackson, L., Shadle, G., & Dixon, R. a. (2006). Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wall-bound ferulic acid in alfalfa (Medicago sativa L.). *The Plant Journal*, *48*(1), 113–24. https://doi.org/10.1111/j.1365-313X.2006.02857.x

Civáň, P., & Švec, M. (2009). Genome-wide analysis of rice (Oryza sativa L. subsp. japonica) TATA box and Y Patch promoter elements. *Genome*, *52*(3), 294–297. https://doi.org/10.1139/G09-001

Conlon, E. M., Liu, X. S., Lieb, J. D., & Liu, J. S. (2003). Integrating regulatory motif discovery and genome-wide expression analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(6), 3339–3344. https://doi.org/10.1073/pnas.0630591100

Copetti, D., Zhang, J., El Baidouri, M., Gao, D., Wang, J., Barghini, E., … Wing, R. A. (2015). RiTE database: a resource database for genus-wide rice genomics and evolutionary biology. *BMC Genomics*, *16*(1), 538. https://doi.org/10.1186/s12864-015-1762-3

Davidson, R. M., Manosalva, P. M., Snelling, J., Bruce, M., Leung, H., & Leach, J. E. (2010). Rice Germin-Like Proteins: Allelic Diversity and Relationships to Early Stress Responses. *Rice*, *3*(1), 43–55. https://doi.org/10.1007/s12284-010-9038-7

Davidson, R. M., Reeves, P. a., Manosalva, P. M., & Leach, J. E. (2009). Germins: A diverse protein family important for crop improvement. *Plant Science*, *177*(6), 499–510. https://doi.org/10.1016/j.plantsci.2009.08.012

Deb, A., & Kundu, S. (2015). Deciphering cis-regulatory element mediated combinatorial regulation in rice under blast infected condition. *PLoS ONE*, *10*(9), 1–21. https://doi.org/10.1371/journal.pone.0137295

Deng, H., Liu, H., Li, X., Xiao, J., & Wang, S. (2012). A CCCH-Type Zinc Finger Nucleic Acid-Binding Protein Quantitatively Confers Resistance against Rice Bacterial Blight Disease. *Plant Physiology*, *158*(2), 876–889. https://doi.org/10.1104/pp.111.191379

Ettwiller, L., Paten, B., Ramialison, M., Birney, E., & Wittbrodt, J. (2007). Trawler: de novo regulatory motif discovery pipeline for chromatin immunoprecipitation. *Nature Methods*, *4*(7), 563–565. https://doi.org/10.1038/nmeth1061

Faris, J. D., Li, W. L., Liu, D. J., Chen, P. D., & Gill, B. S. (1999). Candidate gene analysis of quantitative disease resistance in wheat. *TAG Theoretical and Applied Genetics*, *98*(2), 219–225. https://doi.org/10.1007/s001220051061

Fu, J., Yu, H., Li, X., Xiao, J., & Wang, S. (2011). Rice GH3 gene family: Regulators of growth and development. *Plant Signaling & Behavior*, *6*(4), 570–574. https://doi.org/10.4161/psb.6.4.14947

Geffroy, V., Sévignac, M., De Oliveira, J. C. F., Fouilloux, G., Skroch, P., Thoquet, P., … Dron, M. (2000). Inheritance of Partial Resistance Against *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and Co-localization of Quantitative Trait Loci with Genes Involved in Specific Resistance. *Molecular Plant-Microbe Interactions*, *13*(3), 287–296. https://doi.org/10.1094/MPMI.2000.13.3.287

Gohlke, J., Scholz, C. J., Kneitz, S., Weber, D., Fuchs, J., Hedrich, R., & Deeken, R. (2013). DNA Methylation Mediated Control of Gene Expression Is Critical for Development of Crown Gall Tumors. *PLoS Genetics*, *9*(2). https://doi.org/10.1371/journal.pgen.1003267

González, S., Montserrat-sentís, B., Sánchez, F., Puiggròs, M., Blanco, E., Ramirez, A., & Torrents, D. (2012). ReLA, a local alignment search tool for the identification of distal and proximal gene regulatory regions and their conserved transcription factor binding sites. *Bioinformatics*, *28*(6), 763–770. https://doi.org/10.1093/bioinformatics/bts024

Goto, M. (1992). *Fundamentals of Bacterial Plant Pathology*. San Diego: Academic Press.

Guilfoyle, T. J. (1999). Auxin-regulated genes and promoters. In P. J. J. Hooykaas, M. A. Hall, & K. R. Libbenga (Eds.), *Biochemistry and Molecular Biology of Plant Hormones* (pp. 423–459). Amsterdam, The Netherlands: Elsevier.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. (2007). Quantifying similarity between motifs. *Genome Biology*, *8*(2), R24. https://doi.org/10.1186/gb-2007-8-2-r24

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., … Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.Simple

Hulsmans, S., Rodriguez, M., De Coninck, B., & Rolland, F. (2016). The SnRK1 Energy Sensor in Plant Biotic Interactions. *Trends in Plant Science*, *21*(8), 648–661. https://doi.org/10.1016/j.tplants.2016.04.008

Huot, B., Yao, J., Montgomery, B. L., & He, S. Y. (2014). Growth-defense tradeoffs in plants: A balancing act to optimize fitness. *Molecular Plant*, *7*(8), 1267–1287. https://doi.org/10.1093/mp/ssu049

Hwang, S. H., Kwon, S. Il, Jang, J. Y., Fang, I. L., Lee, H., Choi, C., … Hwang, D. J. (2016). OsWRKY51, a rice transcription factor, functions as a positive regulator in defense response against Xanthomonas oryzae pv. oryzae. *Plant Cell Reports*, *35*(9), 1975–1985. https://doi.org/10.1007/s00299-016-2012-0

Irizarry, R. A. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, *4*(2), 249–264. https://doi.org/10.1093/biostatistics/4.2.249

Jia, Y., Liu, G., Park, D., & Yang, Y. (2013). Inoculation and scoring methods for rice sheath blight disease. In Y. Yang (Ed.), *Rice Protocols* (Vol. 956, pp. 257–268). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-194-3

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*, 323–329. https://doi.org/10.1038/nature05286

K. Reddy, a. P. (1979). Relationship Between Nitrogen Fertilization, Bacterial Leaf Blight Severity, and Yield of Rice. *Phytopathology*, *69*(9), 970. https://doi.org/10.1094/Phyto-69-970

Karmakar, S., Molla, K. A., Chanda, P. K., Sarkar, S. N., Datta, S. K., & Datta, K. (2016). Green tissue-specific co-expression of chitinase and oxalate oxidase 4 genes in rice for enhanced resistance against sheath blight. *Planta*, *243*(1), 115–130. https://doi.org/10.1007/s00425-015-2398-x

Ke, Y., Deng, H., & Wang, S. (2017). Advances in understanding broad-spectrum resistance to pathogens in rice. *Plant Journal*, *90*(4), 738–748. https://doi.org/10.1111/tpj.13438

Kou, Y., & Wang, S. (2010). Broad-spectrum and durability: understanding of quantitative disease resistance. *Current Opinion in Plant Biology*, *13*(2), 181–5. https://doi.org/10.1016/j.pbi.2009.12.010

Kou, Y., & Wang, S. (2012). Toward an understanding of the molecular basis of quantitative disease resistance in rice. *Journal of Biotechnology*, *159*(4), 283–90. https://doi.org/10.1016/j.jbiotec.2011.07.002

Kurihara, Y. (2005). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna*, *12*(2), 206–212. https://doi.org/10.1261/rna.2146906

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559

Lee, S.-I., & Kim, N.-S. (2014). Transposable elements and genome size variations in plants. *Genomics & Informatics*, *12*(3), 87–97. https://doi.org/10.5808/GI.2014.12.3.87

Li, L. (2009). GADEM: A Genetic Algorithm Guided Formation of Spaced Dyads Coupled with an EM Algorithm for Motif Discovery. *Journal of Computational Biology*, *16*(2), 317–329. https://doi.org/10.1089/cmb.2008.16TT

Liu, B., Zhang, S., Zhu, X., Yang, Q., Wu, S., Mei, M., … Leung, H. (2004). Candidate defense genes as predictors of quantitative blast resistance in rice. *Molecular Plant-Microbe Interactions : MPMI*, *17*(10), 1146–52. https://doi.org/10.1094/MPMI.2004.17.10.1146

Liu, J., Wang, X., Mitchell, T., Hu, Y., Liu, X., Dai, L., & Wang, G. (2010). Recent progress and understanding of the molecular mechanisms of the rice – Magnaporthe oryzae interaction, *11*, 419–427. https://doi.org/10.1111/J.1364-3703.2009.00607.X

Liu, X., Brutlag, D. L., & Liu, J. S. (2001). BioProspector: Discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Symposium A Quarterly Journal In Modern Foreign Literatures*, *6*, 127–138. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11262934

López, A., Ramírez, V., García-Andrade, J., Flors, V., & Vera, P. (2011). The RNA silencing enzyme RNA polymerase V is required for plant immunity. *PLoS Genetics*, *7*(12). https://doi.org/10.1371/journal.pgen.1002434

Mallory, A., & Vaucheret, H. (2010). Form, Function, and Regulation of ARGONAUTE Proteins. *The Plant Cell*, *22*(12), 3879–3889. https://doi.org/10.1105/tpc.110.080671

Manosalva, P. M., Davidson, R. M., Liu, B., Zhu, X., Hulbert, S. H., Leung, H., & Leach, J. E. (2009). A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiology*, *149*, 286–96. https://doi.org/10.1104/pp.108.128348

Nakano, M., Nobuta, K., Kalyan, V., Shivakundan, S. T., Skogen, J. W., & Meyers, B. C. (2006). Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Research*, *34*(90001), D731–D735. https://doi.org/10.1093/nar/gkj077

Nakayama, A., Fukushima, S., Goto, S., Matsushita, A., Shimono, M., Sugano, S., … Takatsuji, H. (2013). Genome-wide identification of WRKY45-regulated genes that mediate benzothiadiazole-induced defense responses in rice. *BMC Plant Biology*, *13*(1), 150. https://doi.org/10.1186/1471-2229-13-150

Ou, S. H. (1985). *Rice Diseases* (2nd Editio). Kew, Surrey, England: Commonwealth Mycological Institute.

Pavesi, G., Mauri, G., & Pesole, G. (2001). An algorithm for finding signals of unknown length in DNA sequences. *Bioinformatics*, *17*(Suppl 1), S207–S214. https://doi.org/10.1093/bioinformatics/17.suppl_1.S207

Peng, X., Wang, H., Jang, J.-C., Xiao, T., He, H., Jiang, D., & Tang, X. (2016). OsWRKY80-OsWRKY4 Module as a Positive Regulatory Circuit in Rice Resistance Against Rhizoctonia solani. *Rice*, *9*(1), 63. https://doi.org/10.1186/s12284-016-0137-y

Ramalingam, J., Vera Cruz, C. M., Kukreja, K., Chittoor, J. M., Wu, J. L., Lee, S. W., … Leung, H. (2003). Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. *Molecular Plant-Microbe Interactions : MPMI*, *16*(1), 14–24. https://doi.org/10.1094/MPMI.2003.16.1.14

Shah, J. (2009). Plants under attack: systemic signals in defence. *Current Opinion in Plant Biology*, *12*(4), 459–464. https://doi.org/10.1016/j.pbi.2009.05.011

Sharma, T. R., Rai, A. K., Gupta, S. K., Vijayan, J., Devanna, B. N., & Ray, S. (2012). Rice Blast Management Through Host-Plant Resistance: Retrospect and Prospects. *Agricultural Research*, *1*(1), 37–52. https://doi.org/10.1007/s40003-011-0003-5

Shi, J., Yang, W., Chen, M., Du, Y., Zhang, J., & Wang, K. (2011). AMD, an automated motif discovery tool using stepwise refinement of gapped consensuses. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0024576

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539–539. https://doi.org/10.1038/msb.2011.75

Singh, R. P., Huerta-Espino, J., Bhavani, S., Herrera-Foessel, S. A., Singh, D., Singh, P. K., … Crossa, J. (2011). Race non-specific resistance to rust diseases in CIMMYT spring wheats. *Euphytica*, *179*(1), 175–186. https://doi.org/10.1007/s10681-010-0322-9

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W., & Zhu, J.-K. (2008). Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biology*, *8*(1), 25. https://doi.org/10.1186/1471-2229-8-25

Thompson, W., Rouchka, E. C., & Lawrence, C. E. (2003). Gibbs Recursive Sampler: Finding transcription factor binding sites. *Nucleic Acids Research*, *31*(13), 3580–3585. https://doi.org/10.1093/nar/gkg608

Tonnessen, B. W., Manosalva, P., Lang, J. M., Baraoidan, M., Bordeos, A., Mauleon, R., … Leach, J. E. (2014). Rice phenylalanine ammonia-lyase gene OsPAL4 is associated with broad spectrum disease resistance. *Plant Molecular Biology*, *87*(3), 273–286. https://doi.org/10.1007/s11103-014-0275-9

Torres, M. A., Jones, J. D. G., & Dangl, J. L. (2006). Reactive oxygen species signaling in response to pathogens. *Plant Physiology*, *141*(2), 373–378. https://doi.org/10.1104/pp.106.079467.conjunction

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578. https://doi.org/10.1038/nprot.2012.016

Troukhan, M., Tatarinova, T., Bouck, J., Flavell, R. B., & Alexandrov, N. N. (2009). Genome-wide discovery of cis-elements in promoter sequences using gene expression. *Omics : A Journal of Integrative Biology*, *13*(2), 139–51. https://doi.org/10.1089/omi.2008.0034

Vagndorf, N., Nielsen, N. H., Edriss, V., Andersen, J. R., Orabi, J., Jørgensen, L. N., & Jahoor, A. (2017). Genomewide association study reveals novel quantitative trait loci associated with resistance towards Septoria tritici blotch in North European winter wheat. *Plant Breeding*, (July 2016), 474–482. https://doi.org/10.1111/pbr.12490

Van Damme, E. J. M., Barre, A., Rougé, P., & Peumans, W. J. (2004). Cytoplasmic/nuclear plant lectins: A new story. *Trends in Plant Science*, *9*(10), 484–489. https://doi.org/10.1016/j.tplants.2004.08.003

van Heeringen, S. J., & Veenstra, G. J. C. (2011). GimmeMotifs: A de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics*, *27*(2), 270–271. https://doi.org/10.1093/bioinformatics/btq636

Wang, G., Ding, X., Yuan, M., Qiu, D., Li, X., Xu, C., & Wang, S. (2006). Dual function of rice OsDR8 gene in disease resistance and thiamine accumulation. *Plant Molecular Biology*, *60*(3), 437–449. https://doi.org/10.1007/s11103-005-4770-x

Wu, J. L., Sinha, P. K., Variar, M., Zheng, K. L., Leach, J. E., Courtois, B., & Leung, H. (2004). Association between molecular markers and blast resistance in an advanced backcross population of rice. *Theor Appl Genet*, *108*(6), 1024–1032.

Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., & Qi, Y. (2009). Rice MicroRNA effector complexes and targets. *The Plant Cell*, *21*(11), 3421–35. https://doi.org/10.1105/tpc.109.070938

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., & Qi, Y. (2010). DNA Methylation Mediated by a MicroRNA Pathway. *Molecular Cell*, *38*(3), 465–475. https://doi.org/10.1016/j.molcel.2010.03.008

Xu, X., Chen, C., Fan, B., & Chen, Z. (2006). Physical and Functional Interactions between and WRKY60 Transcription Factors. *The Plant Cell*, *18*(May), 1310–1326. https://doi.org/10.1105/tpc.105.037523.1

Xu, R., Wunsch, D., & IEEE Computational Intelligence Society. (2009). *Clustering* (IEEE series on computational intelligence). Hoboken, N.J. : Piscataway, NJ : IEEE Press: Wiley

Yamamoto, E., Yonemaru, J., Yamamoto, T., & Yano, M. (2012). OGRO: The Overview of functionally characterized Genes in Rice online database. *Rice*, *5*(1), 26. https://doi.org/10.1186/1939-8433-5-26

Yang, C., Li, W., Cao, J., Meng, F., Yu, Y., Huang, J., … Liu, J. (2017). Activation of ethylene signaling pathways enhances disease resistance by regulating ROS and phytoalexin production in rice. *Plant Journal*, *89*(2), 338–353. https://doi.org/10.1111/tpj.13388

Yu, A., Lepere, G., Jay, F., Wang, J., Bapaume, L., Wang, Y., … Navarro, L. (2013). Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proceedings of the National Academy of Sciences*, *110*(6), 2389–2394. https://doi.org/10.1073/pnas.1211757110

Yu, C., Chen, H., Tian, F., Leach, J. E., & He, C. (2014). Differentially-expressed genes in rice infected by Xanthomonas oryzae pv. oryzae relative to a flagellin-deficient mutant reveal potential functions of flagellin in host–pathogen interactions. *Rice*, *7*(1), 20. https://doi.org/10.1186/s12284-014-0020-7

Yu, Y., & Huang, R. (2017). Integration of Ethylene and Light Signaling Affects Hypocotyl Growth in Arabidopsis. *Frontiers in Plant Science*, *8*(January), 2012–2017. https://doi.org/10.3389/fpls.2017.00057

Zhang, H., & Wang, S. (2013). Rice versus Xanthomonas oryzae pv. oryzae: A unique pathosystem. *Current Opinion in Plant Biology*, *16*(2), 188–195. https://doi.org/10.1016/j.pbi.2013.02.008

Zhang, J., Chen, L., Fu, C., Wang, L., Liu, H., Cheng, Y., … Zheng, A. (2017). Comparative Transcriptome Analyses of Gene Expression Changes Triggered by Rhizoctonia solani AG1 IA Infection in Resistant and Susceptible Rice Varieties. *Frontiers in Plant Science*, *8*(August), 1–18. https://doi.org/10.3389/fpls.2017.01422

Zhong, S., Zhao, M., Shi, T., Shi, H., An, F., Zhao, Q., & Guo, H. (2009). EIN3/EIL1 cooperate with PIF1 to prevent photo-oxidation and to promote greening of Arabidopsis seedlings. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21431–21436. https://doi.org/10.1073/pnas.0907670106

Zipfel, C., & Robatzek, S. (2010). Pathogen-Associated Molecular Pattern-Triggered Immunity: Veni, Vidi...? *Future Perspectives in Plant Biology*, *154*, 551–554. https://doi.org/10.1104/pp.110.161547

Zou, C., Sun, K., Mackaluso, J. D., Seddon, A. E., Jin, R., Thomashow, M. F., & Shiu, S.-H. (2011). Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *108*(36), 14992–14997. https://doi.org/10.1073/pnas.1103202108

Chapter 4: Overall conclusions and outlook

Understanding broad spectrum resistance: broad to narrow, to broad

In the preceding chapters, the inquiry into the *O. sativa* Broad Spectrum Defense Response (BS-DR) encompassed single gene functionality as well as genome-wide regulation. At the genome level, multiple *cis*-Regulatory Modules (CRMs) were found to associate with BS-DR gene promoters. Some of these CRMs have been well-characterized in the literature, thus validating the analysis workflow. The tandem W-box motifs seen in CRM4 and CRM5 are recognized as functional elements in the activation of DR genes (Peng et al., 2016; Xu et al., 2006). Other CRMs are not as well understood, and possibly introduce a novel regulatory mechanism in BS-DR pathways. This is observed in CRM2, which putatively enables the cross-talk between ethylene and light-responsive genes, or the epigenetic mechanism that accompanies the presence of CRM1 and CRM3. The approach applied in this thesis yielded results that complement the research involving single BS-DR genes and DR gene families. The phenylalanine ammonia lyase gene, *OsPAL4*, illustrates the classical definition of a large-effect BS-DR gene. The encoded enzyme is a key player in the production of phenolic compounds, a staple in plant DR (Cho & Lee, 2015). A mutant of *OsPAL4* decreases resistance to diverse rice diseases. The experiments on the *ospal4* mutant also exhibited how other genes of the *OsPAL* family interact, such as the relative upregulation of *OsPAL2* in the mutant, which is in the same tandem duplication locus as *OsPAL4*. The increased expression of *OsPAL2* was observed in uninoculated tissue and it is unknown what caused the relative upregulation in basal levels of *OsPAL2*. Though, the increased levels of *OsPAL2* transcript did not prevent the susceptible mutant phenotype, inferring non-involvement with DR. These results led to the hypothesis that

promoter differences between these two family members effects their differential contribution to resistance. Indeed, the *OsPAL4* promoter containing CRM2 and CRM4 is present only in the resistant QTL donor, and no CRM exist in the promoter of *OsPAL2* in either the resistant or susceptible variety. Finding CRMs was accomplished in a blind, *de novo* fashion, thus it is very interesting and exciting to see the union play out between these two approaches of understanding BS-DR.

The importance of gaining knowledge on BS-DR at the gene and genome levels should stay in focus as new varieties are developed and released. As we pay attention to what is successful in durable BS-DR, the impacts of genetic background and polygenic responses emerge as paramount. During pathogen infection and disease spread, the response is throughout the organism, and various trade-offs need to occur for resources to be balanced between defense and plant development. Genes involved in the regulation of nutrient uptake, growth, ecological relationships, and environmental cues are all in transcriptional flux due to the stress of disease (Karasov et al., 2017). The genomic modifications from epigenetic mechanisms are an integral factor in balancing these different processes in the BS-DR (Deleris et al., 2016). For example, differential methylation/de-methylation of loci across the genome plays a role in the differentiation between a response to virulent pathogens or microbial symbionts (Zogli & Libault, 2017). Translating this knowledge to breeding is imperative, since epigenetic changes to DNA are inherited across generations (Holeski et al., 2012). Thus, the inheritance of BS-DR promoters containing CRM1 or CRM3 could be critical for transgenerational basal resistance. This work identifies new questions and potential mechanisms that are important to understanding durable BS-DR, and will hopefully influence variety development in the future.

Future experimental directions

The breadth of this work has focused on identifying candidates for future functional validation in BS-DR. For example, of the *OsPAL* gene family member within the disease resistance QTL, *OsPAL4* is likely to be the contributing gene, as is verified in the mutant experiments. The specific factor that distinguishes between a resistant and susceptible haplotype at the *OsPAL4* locus has yet to be discerned. The identification of putative BS-DR-associated CRMs in the *OsPAL4* promoter illuminates the possibility of these *cis*-elements being the key regulators in the resistant haplotype. Along with the *OsPAL4* candidate, many other genes involved in QTL-mediated resistance harbor CRM promoter polymorphisms such as *OsGLP8-6*, *OsGLP8-7*, *OsGLP8-9*, and *OsOXO4* (Figure 3-9). Future work could focus on other functionally-associated DR genes that contain a CRM (Table S4-1: Supplemental). Thus, there are multiple candidates for future studies on the functional validation of CRMs and BS-DR gene regulation. These examples show that building from a compilation of knowledge and data from the literature, bioinformatic analysis can refine our broad questions about BS-DR into concise, measurable hypotheses.

Experimental testing to determine the role of CRMs in BS-DR gene regulation will involve both observational and direct genome editing techniques. With recent editing technologies, such as TALEN and CRISPR/Cas9 systems, promoters can be altered to determine CRM function. The CRMs of interest could be deleted to determine if responsiveness to the DR is lost. More specifically, the CRM would be replaced with random nucleotides to preserve native promoter structure. A CRM may otherwise be introduced into a BS-DR gene promoter that lacks the CRM. Considering the BS-DR QTL genes mentioned above, the resistant and susceptible promoter haplotypes could be edited at the respective CRM loci. For example, the

OsPAL4 promoter contains CRM4 and CRM2 within the resistant allele, and mutating those regions may prove to alter *OsPAL4* gene expression during DR. Introduction of these CRMs in their respective positions in the susceptible haplotype promoter may "rescue" a resistance expression phenotype of *OsPAL4*, improving the DR of the susceptible host. One option for initial testing and validation of these promoter edits, before whole plant transformation, is through transient expression assays. Measuring the intensity of expression via luminescence of an edited or wild-type promoter fused to firefly luciferase, normalized by co-introduction of a ubiquitin promoter fused to *renilla* luciferase, could be tested in rice protoplasts or *Nicotiana benthamiana* leaves (Chen et al., 2006; Wang et al., 2017). Analyzing the differences in response to a defense elicitor, such as chitin, may show an effect between wild-type and edited BS-DR promoters. The context of these types of analyses must be considered, since suspension cultured rice cells will be under significant amounts of stress, and *N. benthamiana* may not have the correct genetic background to express the desired phenotype.

Ideally, when testing the functionality of promoter polymorphisms, it is important to analyze phenotypes of edited promoters within the native genetic background. For example, any mutations of resistant *OsPAL4* promoter CRMs would be ideally observed in IR64 to retain any regulatory mechanisms that act on this promoter in its wild-type state. Even if these experimental variables are well controlled, we must consider the unreconciled issues that come with genetic transformation. Methylation profiles of regenerated rice plants from tissue culture are altered significantly, and remain stable across generations (Stroud et al., 2013). The off-target binding of guide-RNAs from the CRISPR/Cas9 system can also be an issue, especially since the CRMs of interest are found in many genomic locations (Endo et al., 2015). Promoter editing is useful for initial validation of CRM function. Though as we move forward, the ideal

way to measure and understand BS-DR across the genome is through observational methods that do not disturb the native cellular and genomic structure.

Experiments to observe promoter activity and the effect of CRMs begins with first identifying varieties that are polymorphic at the loci of interest. We have identified those genotypes for the QTL-related BS-DR genes and some functionally associated DR genes which contain polymorphisms in CRMs between the reference varieties, IR64 and Nipponbare. In most cases of QTL analysis, however, the resistant and susceptible varieties are highly divergent, usually separated by sub-populations Indica and Japonica. This implies that the influence of genetic background is dissimilar among the two varieties. Ideally, using the 3000 rice genomes or other rice genomic resources, a variety can be found that is a member of the same subpopulation as the resistant variety, but contains a polymorphism in the target CRM. Thus, this would reduce the effect of genetic background as we analyze promoter activity.

One method of understanding the functionality of a CRM could utilize reverse Chromatin Immunoprecipitation (ChIP). Using a DNA probe containing the target CRM to bind the promoter segment of interest, or the promoter segment whose genotype has the CRM removed, the proteins bound to it can be pulled down. Mass spectrometric analysis can then discern the types of proteins bound to the CRM locus, and compare between promoter genotypes. There may exist a transcription factor or other protein associated with the CRM, and testing reverse ChIP both before and after pathogen infection may shed light on the involvement in DR. Another method for observing the BS-DR in a native genomic environment incorporates bisulfite treatment of genomic DNA and subsequent sequencing to measure degree of methylation across CRM loci. This type of analysis is especially valuable for surveying the role of CRM1 and CRM3 in methylation/de-methylation of BS-DR promoters. Bisulfite sequencing could be

administered to two varieties with polymorphic CRMs both before and after pathogen inoculation, and during a time course thereafter. Degree of methylation can be measured by mapping the reads back to the CRM loci of interest, thus hopefully discerning the mode of action of CRM1 or CRM3. The presence/absence of these CRMs could assist in epigenetic activity across many BS-DR genes across the genome, and other genes with the respective CRMs in their promoters could be measured as well to, in a way, predict their role in BS-DR. A very interesting candidate for this experiment is *OsGLP8-6*, which contains the large promoter insertion in variety SHZ-2, containing CRM1 in the 3' end (Figure 3-9a). Looking at the differences in methylation of SHZ-2 and LTH during infection may elucidate the hypothesized function of this CRM. The DR gene, *WRKY53*, is also a good candidate for this work, and has been shown to be involved in BS-DR (Chujo et al., 2007; Hu et al., 2016). The methylation in wild-type Nipponbare tissues is definitive in the CRM1 locus in this gene's promoter, and produces many sRNAs at this locus (Figures 3-12a, 3-15a). A critical piece of this analysis would be to accompany RNA-sequencing alongside bisulfite sequencing to elucidate the relationship between methylation and expression of BS-DR genes Since bisulfite sequencing is a genome-wide endeavor, there are many interesting questions that can be answered about the entire BS-DR, and the contribution of CRMs to this response. Analyzing the effects of different pathogens, including *X. oryzae* pv. *oryzae*, *M. oryzae*, and *R. solani* could shed light on trends that occur across these genes. Relating this data to the co-expressed BS-DR cluster of genes can validate this research even further.

As for future experimental endeavors in the bioinformatics realm, delving deeper into the hypothesis of CRM1 and CRM3 as putative genome-wide, regulatory repeat elements is a priority. The consensus sequence for both of these CRMs has been uncovered (Figure 3-11).

These sequences were found using small motifs initially, however, so a search back the other direction is the next step. There are other conserved elements in CRM1 and CRM3 along with the motifs seen in this work, and a full structure, evolutionary analysis, and physical map of these CMs across the genome is necessary. This will be especially helpful when looking back on results of bisulfite sequencing during the DR, to truly find a relationship of these CRMs with differentially methylated/regulated genes across the genome.

Looking forward: The importance of crossing

As we continue to strive for durable, broad spectrum resistance in new rice varieties, we will always face challenges of new disease outbreaks. The strategy to overcome this threat is attributed to an "arms race." This label rightfully expresses the relentless struggle. When a virulent pathogen arises, we introduce a new weapon, analogous to an R-gene, to restore resistance, but it is short-lived as the enemy builds its own arsenal, due to inevitable pathogen evolution (Ballini et al., 2008). Sadly, the cycle must continue. Since the introduction of a single R-gene cannot confidently provide durable resistance, a proposal to improve durability stacks multiple R-genes into one rice variety (Wulff et al., 2011). The transgenic approach to introducing R-gene stacks is seen as an easier method than crossing plants due to the speed and precision it provides (Kumari et al., 2017; Wulff et al., 2011). As newer gene editing technologies, such as CRISPR/Cas9 are perfected, this level of precision and efficiency will increase (Feng et al., 2014). Though, introducing an R-gene into a foreign genetic background can neglect the complex polygenic mechanism that resistance requires. For instance, some R-genes sense pathogen effector activity on another host protein, thus would require this other "guardee" protein to function (Mackey et al., 2002). Rice R-genes can often be large-effect DR

genes, not participating in ETI, hence they would require a strong BS-DR genetic background to effectively function (Ke et al., 2017). Furthermore, an R-gene's efficacy against *M. oryzae* was enhanced when another partial resistance QTL from the same genetic background was introduced (Barbary et al., 2016). This type of resistance would be lost if only single R-genes were introduced to a variety via transformation. Most strikingly, as mentioned above, epigenetic structure is affected by methods of genetic transformation (Stroud et al., 2013). Through the work presented here and others, we know the importance of epigenetics on disease resistance. Thus, an attempt to improve resistance with transgenic approaches may disrupt the inheritance of, or inhibit/alter DNA methylation mechanisms of BS-DR. Using classical breeding methods instead of transformation, along with marker assisted selection (MAS), has proven to be effective in introducing resistance to multiple pathogens (Das & Rao, 2015). Utilizing the natural mechanisms of recombination and sexual reproduction will preserve linkage groups and epigenetic inheritance that retain BS-DR in the population. As more BS-DR alleles are selected from diverse genetic backgrounds in breeding efforts, we could lessen the severity of the boom-bust cycle accredited to the arms race.

The key to understanding basal resistance cannot be simplified to a few loci. It rather lies in the many BS-DR genes, and the various alleles that arise across rice genomes. Recently, there has been a progressive decline in the genetic diversity and number of different rice varieties grown throughout the world (Heal et al., 2004; Singh et al., 2016). This has been attributed to breeding for specific stress tolerance traits while preserving a homogeneous phenotype in the population. Thus, the focus is on introgression of specific loci, and avoids other genetic contaminations from the donor parent. Since basal resistance can involve hundreds of genes

with small effect, these genomic "contaminations" may contain genetic treasures, such as unique BS-DR promoter alleles with CRMs.

The strategy used in this work combined multiple steps of different *computational analyses* to test the correlation between the BS-DR and a presence/absence of promoter CRMs. Associating CRMs with disease resistance was built on previous *biological validation* of several DR genes that function in broad spectrum disease resistance. Specifically, the promoters driving DR genes, *OsGLP8-6, OsOXO4,* and *OsPAL4*, had previously been shown to contribute to disease resistance. The results presented here show CRMs reside *exclusively* in each of their respective resistance alleles. Based on this information, we predict that a CRM signature within DR gene promoters can be used as a marker for future breeding practices to enrich for the most responsive and effective DR genes across the genome.

References

Ballini, E., Morel, J.-B., Droc, G., Price, A., Courtois, B., Notteghem, J.-L., & Tharreau, D. (2008). A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. *Molecular Plant-Microbe Interactions : MPMI*, *21*(7), 859–68. https://doi.org/10.1094/MPMI-21-7-0859

Barbary, A., Djian-Caporalino, C., Marteu, N., Fazari, A., Caromel, B., Castagnone-Sereno, P., & Palloix, A. (2016). Plant Genetic Background Increasing the Efficiency and Durability of Major Resistance Genes to Root-knot Nematodes Can Be Resolved into a Few Resistance QTLs. *Frontiers in Plant Science*, *7*(May), 1–9. https://doi.org/10.3389/fpls.2016.00632

Chen, S., Tao, L., Zeng, L., Vega-Sanchez, M. E., Umemura, K., & Wang, G. L. (2006). A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. *Molecular Plant Pathology*, *7*(5), 417–427. https://doi.org/10.1111/j.1364-3703.2006.00346.x

Cho, M. H., & Lee, S. W. (2015). Phenolic phytoalexins in rice: Biological functions and Biosynthesis. *International Journal of Molecular Sciences*, *16*(12), 29120–29133. https://doi.org/10.3390/ijms161226152

Chujo, T., Takai, R., Akimoto-Tomiyama, C., Ando, S., Minami, E., Nagamura, Y., … Yamane, H. (2007). Involvement of the elicitor-induced gene OsWRKY53 in the expression of defense-related genes in rice. *Biochimica et Biophysica Acta - Gene Structure and Expression*, *1769*(7–8), 497–505. https://doi.org/10.1016/j.bbaexp.2007.04.006

Das, G., & Rao, G. J. N. (2015). Molecular marker assisted gene stacking for biotic and abiotic stress resistance genes in an elite rice cultivar. *Frontiers in Plant Science*, *6*(September), 1–18. https://doi.org/10.3389/fpls.2015.00698

Deleris, A., Halter, T., & Navarro, L. (2016). DNA Methylation and Demethylation in Plant Immunity. *Annual Review of Phytopathology*, *54*(1), 579–603. https://doi.org/10.1146/annurev-phyto-080615-100308

Endo, M., Mikami, M., & Toki, S. (2015). Multigene knockout utilizing off-target mutations of the CRISPR/cas9 system in rice. *Plant and Cell Physiology*, *56*(1), 41–47. https://doi.org/10.1093/pcp/pcu154

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.-L., … Zhu, J.-K. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences*, *111*(12), 4632–4637. https://doi.org/10.1073/pnas.1400822111

Heal, G., Walker, B., Levin, S., Arrow, K., Dasgupta, P., Daily, G., … Starrett, D. (2004). Genetic diversity and interdependent crop choices in agriculture. *Resource and Energy Economics*, *26*(2), 175–184. https://doi.org/10.1016/j.reseneeco.2003.11.006

Holeski, L. M., Jander, G., & Agrawal, A. A. (2012). Transgenerational defense induction and epigenetic inheritance in plants. *Trends in Ecology and Evolution*, *27*(11), 618–626. https://doi.org/10.1016/j.tree.2012.07.011

Hu, L., Ye, M., Li, R., & Lou, Y. (2016). OsWRKY53, a versatile switch in regulating herbivore-induced defense responses in rice. *Plant Signaling and Behavior*, *11*(4), 1–4. https://doi.org/10.1080/15592324.2016.1169357

Karasov, T. L., Chae, E., Herman, J. J., & Bergelson, J. (2017). Mechanisms to Mitigate the Trade-Off between Growth and Defense. *The Plant Cell*, *29*(4), 666–680. https://doi.org/10.1105/tpc.16.00931

Ke, Y., Deng, H., & Wang, S. (2017). Advances in understanding broad-spectrum resistance to pathogens in rice. *Plant Journal*, *90*(4), 738–748. https://doi.org/10.1111/tpj.13438

Kumari, M., Rai, A. K., Devanna, B. N., Singh, P. K., Kapoor, R., Rajashekara, H., … Sharma, T. R. (2017). Co-transformation mediated stacking of blast resistance genes Pi54 and Pi54rh in rice provides broad spectrum resistance against Magnaporthe oryzae. *Plant Cell Reports*. https://doi.org/10.1007/s00299-017-2189-x

Mackey, D., Iii, B. F. H., Wiig, A., Dangl, J. L., Hill, C., & Carolina, N. (2002). RIN4 Interacts with Pseudomonas syringae Type III Effector Molecules and Is Required for RPM1-Mediated Resistance in Arabidopsis, *108*, 743–754.

Peng, X., Wang, H., Jang, J.-C., Xiao, T., He, H., Jiang, D., & Tang, X. (2016). OsWRKY80-OsWRKY4 Module as a Positive Regulatory Circuit in Rice Resistance Against Rhizoctonia solani. *Rice*, *9*(1), 63. https://doi.org/10.1186/s12284-016-0137-y

Singh, N., Choudhury, D. R., Tiwari, G., Singh, A. K., Kumar, S., Srinivasan, K., … Singh, R. (2016). Genetic diversity trend in Indian rice varieties: an analysis using SSR markers. *BMC Genetics*, *17*(1), 127. https://doi.org/10.1186/s12863-016-0437-7

Stroud, H., Ding, B., Simon, S. A., Feng, S., Bellizzi, M., Pellegrini, M., … Jacobsen, S. E. (2013). Plants regenerated from tissue culture contain stable epigenome changes in rice. *eLife*, *2013*(2), 1–14. https://doi.org/10.7554/eLife.00354

Wang, L., Rinaldi, F. C., Singh, P., Doyle, E. L., Dubrow, Z. E., Tran, T. T., … Bogdanove, A. J. (2017). TAL Effectors Drive Transcription Bidirectionally in Plants. *Molecular Plant*, *10*(2), 285–296. https://doi.org/10.1016/j.molp.2016.12.002

Wulff, B. B. H., Horvath, D. M., & Ward, E. R. (2011). Improving immunity in crops: New tactics in an old game. *Current Opinion in Plant Biology*, *14*(4), 468–476. https://doi.org/10.1016/j.pbi.2011.04.002

Xu, X., Chen, C., Fan, B., & Chen, Z. (2006). Physical and Functional Interactions between and WRKY60 Transcription Factors. *The Plant Cell*, *18*(May), 1310–1326. https://doi.org/10.1105/tpc.105.037523.1

Zogli, P., & Libault, M. (2017). Plant response to biotic stress: Is there a common epigenetic response during plant-pathogenic and symbiotic interactions? *Plant Science*, *263*(May), 89–93. https://doi.org/10.1016/j.plantsci.2017.07.008