THESIS

TALE-BOUND QTL: A COMPUTATIONAL INVESTIGATION OF BACTERIAL EFFECTOR ASSOCIATION WITH RESISTANCE QUANTITATIVE TRAIT LOCI IN *ORYZA SATIVA*

Submitted by

Jacob Emmett Sharkey

Department of Agricultural Biology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2022

Master's Committee:

Advisor: Jan E. Leach

Alejandra I. Huerta
Marc Nishimura
Robyn Roberts

ABSTRACT

TALE-BOUND QTL: A COMPUTATIONAL INVESTIGATION OF BACTERIAL EFFECTOR ASSOCIATION WITH RESISTANCE QUANTITATIVE TRAIT LOCI IN *ORYZA SATIVA*

Durable resistance to *Xanthomonas oryzae* pathovars *oryzae* (*Xoo*) and *oryzicola* (*Xoc*), which cause bacterial blight and bacterial leaf streak, respectively, is highly sought after in rice (*Oryza sativa*) due to the pathogens ability to impact maximum attainable yields. Regions of the rice genome associated with quantitative resistance to multiple strains of *Xoo* and *Xoc*, known as quantitative trait loci (QTL), were previously identified using a multi-parent advanced generation intercross (MAGIC) rice population and a combination of genome wide association studies and interval mapping. These QTL have been associated with decreased lesion lengths by *Xoc* and *Xoo* on rice. What remains unknown is the molecular basis for the induction of genes under these QTL during pathogen infection. Considering our biological question "what is the molecular basis for regulation of resistance QTL associated with *Xoo* and *Xoc*?", we predicted that part of the answer could be found by investigating the bacteria's direct interaction with the *O. sativa* genome. Upon infection, *Xoo* and *Xoc* injects the host with DNA-binding TALE (transcription activator-like effector) proteins. These effectors, when bound to their target plant gene promoter, induce gene transcription. We hypothesize that differential interactions of TALE with promoters of rice genes under the QTL lead to the resistant/susceptible phenotypes exhibited across varieties. To test this, we designed a pipeline that predicts TALE-regulated candidate genes involved in quantitative resistance. This pipeline identifies genes

that meet three criteria: (1) the presence of a binding site for an *X. oryzae* TALE in the gene's promoter, a strong correlation between binding site presence, and disease phenotypes and overlap of the gene with a resistance QTL. We used this pipeline with genomic and phenotypic data for the eight MAGIC founders to identify candidate genes involved in resistance against seven *Xoo* and *Xoc* strains. Candidate genes identified include ones encoding a patatin-like phospholipase and multiple NB-ARC containing proteins such as the Mla1 protein. Here, we exploit the abundant genomic data for the rice-*X. oryzae* systems and the ability to predict direct associations between bacterial proteins and plant genomes, to propose a method that could streamline the identification of genes involved in quantitative resistance to TALE-harboring *Xanthomonas*.

ACKNOWLEDGEMENTS

One hot summer day in 2019, I sat down for a job interview in Plant Sciences C023. That day I was interviewing for an undergraduate position which would entail greenhouse work and assisting with various lab duties. I vividly remember nervously walking across campus to meet with Emily Luna for this opportunity, pausing at the door because this truly was the opportunity I was looking for. Months prior I had left my job as a nurse aide and since was looking for my next step, a step into the world of research. The interview was standard, and many questions asked were ones that I had heard in other interviews. Something that stood out to me though was how highly Emily spoke of the lab environment, and the high regard for which she held her principal investigator, Professor Jan E. Leach. It wasn't until after joining the lab that I realized where this admiration came from. In Jan's lab I quickly felt welcome. This was a big deal to me because I often worried if I belonged in STEM as a person of color. So first, I say thank you to Emily Luna, for being the first point of contact, for seeing potential in me, and for bringing me into such an enriching space. After joining, Emily was always there as a mentor, a colleague, and a friend. The second person I met in the Leach lab was Federico Martin. I immediately knew that this guy meant business. Federico was quick to teach me the background and the methods of research and whether I was counting rice seed or running PCR products in an agarose gel, Fede always kept the spirits high. Federico, thank you for simultaneously teaching how much focus to detail research requires, but also how fun it can be. I will miss sitting across the bench pipetting to the sounds of Easy Star All-Stars. Another person who has largely impacted my career is Alvaro Perez-Quintero. Alvaro and I didn't start working together until 2020 when I began my master's, and before that we had only interacted a handful of times. Be that as it

may, over the past two years, Alvaro has been there every step of the way patiently teaching me computational biology. Thank you, Alvaro, for answering the immeasurable questions I have had over the past two years, for bringing me into the world of computer science, for superb book suggestions, and for all the pop-culture discussions. I also want to say thank you to the other members of the leach lab that I met along the way. Thank you, Casey Mazzotta, for showing me the ins and outs of the greenhouse. Thank you, Janet Hardin, for all our insightful discussions around the lab. Thank you, Jillian M. Lang, for all the career advice and tips about living abroad. Thank you, Jennifer Shipp for being there when I need to vent about the stresses of graduate school, and for all your advice on *Xanthomonas* inoculations. Finally, thank you Jan E. Leach. I have never met somebody so humble and so generous. Once, you told me your hope is that your students become smarter than you, and that your job was to propel our careers. Please know that this level of support has meant a world and a half to me. Thank you for believing in me, for imparting me with so many of your insights, and for setting the example of what it means to be a personified rising tide. To all in the Leach lab, I will carry you in my heart as I cross the Atlantic Ocean, and of course I'll be back to visit soon.

TABLE OF CONTENTS

1.1. INTRODUCTION

Plants constantly interact with a range of microbes. Some of these microbes are pathogenic to plants, and upon colonization of a plant, cause diseases with a range of phenotypes. Traditionally, the plant host can exhibit phenotypes that are summarized in resistance or susceptibility. Here we define resistance as successful plant defense against a pathogen that prevents the proliferation and spread of the microbe. Susceptibility would then be the successful spread of the pathogen and eventual destruction of the plant. Disease phenotypes, however, can range in severity rather than residing in a binary. Plant innate immunity historically has been grouped into two categories, PTI and ETI (PAMP-triggered immunity and effector-triggered immunity) (Jones & Dangl, 2006). It is the diverse combinations of signaling protein interactions and defense gene regulation that creates varied disease phenotypes and, while PTI and ETI have discerning properties, overlap between the two has been identified (Tsuda & Katagiri, 2010, Thomma *et al.*, 2011, Naveed *et al.*, 2020, Pruitt *et al.*, 2021, Ngou *et al.*, 2021).

During plant-pathogen interactions, PTI-associated defense responses are initiated by the recognition of the pathogen. This involves a range of extracellular plant proteins, known as pattern recognition receptors (PRR) that detect pathogen-associated molecular patterns and/or danger-associated molecular patterns (PAMPs, DAMPs) (Yuan *et al.*, 2021). PAMPs can be non-proteinaceous, like β-glucans of *Phytophthora* spp., or proteinaceous, like flagellin and elongation factors of bacterial pathogens (Naveed et al., 2020). DAMPS include biproducts of host damage by the pathogen such as cell wall fragments. Many PRRs are effective against a

range of microbes because they recognize molecules necessary for the pathogen's survival, such as flagellin which is required for bacterial motility (Sanguankiattichai *et al.*, 2022). Binding of the PRR to PAMPs or DAMPs initiates a cascade of defense signals that lead to oxidative bursts early in the disease cycle, as well as callose deposition and production of plant hormones. All these ultimately will assist the plant to ward off the pathogen. This phenomenon, however, places pressure on the pathogen, driving them to evolve evasive mechanisms of virulence. For example, one tactic used by the fungal pathogen *Magnaporthe oryzae* is to inject effector proteins that sequester chitin molecules into the plant's extracellular matrix, preventing pathogen detection and PTI (Mentlak *et al.*, 2012). Many pathogens are capable of injecting plant cells with effector proteins that inhibit various proteins in PTI signal pathways or manipulate the cell environment to the pathogens benefit. *Pseudomonas syringae* injects the effector, AvrRpt2, into the host, cleaving RPM1-interacting protein 4 (RIN4), a protein integral to stomatal opening, for degradation (Wang *et al.*, 2022). Over time plants have developed proteins that recognize effectors or effector biproducts (e.g., detection of RIN4 degradation in *Arabidopsis thaliana*), initiating effector-triggered immunity (ETI) (Wang *et al.*, 2022). Many effectors are recognized by nucleotide-binding (NB) and leucine-rich repeat (LRR) domain containing proteins. These proteins can bind effectors directly or indirectly, and once the effector is bound, a cascade of protein interactions and gene activation is initiated (Dodds & Rathjen, 2010). The proteins and genes involved in PTI and ETI are numerous, and the components of each vary across the plant-pathogen interactions. Research to understand these components is ongoing, and as we learn more about components that evoke a resistant phenotype, we gain more resources to produce more resilient plants.

Oryza sativa (rice) is one of the most cultivated crops worldwide, constituting nearly half of the calories consumed by 60% of the world's population (Khush, 1997). Bacterial blight (BB) and bacterial leaf streak (BLS) caused by *Xanthomonas oryzae* pathovars *oryzae* (*Xoo*) and *oryzicola* (*Xoc*) respectively, are two diseases of rice that greatly hamper crop production. While the two pathovars are closely related, *Xoo* and *Xoc* differ in lifestyles. *Xoo,* a vascular pathogen that enters the plant through wounds, growth cracks, or hydathodes, causes the leaves of susceptible plants to wilt and curl as the bacteria spreads through the xylem (Ou, 1985). *Xoc* is a non-vascular pathogen which enters the leaf through wounds or stomatal openings. In susceptible plants streaks of necrosis form parallel to leaf veins as the bacteria break down parenchyma cells (Ou, 1985).

Over 40 resistance genes (*R*-genes) have been identified in rice against *Xoo* (Hutin *et al.*, 2015). Some of these resistance genes, such as *Xa21*, code for PRRs that recognize the pathogen's complimenting effector avirulence gene product to initiate PTI (Boyd *et al.*, 2013). Others, such as *xa5* interfere with effectors of *Xoo* and their ability to manipulate the host-cell's environment to the pathogens' liking. These *R*-genes are single genes, or qualitative resistance genes, and while they are effective, the strong selective pressure they place on the bacteria reduces their durability. Thus, research has been conducted to identify quantitative resistance, or the summation of many defense response genes that are not race-specific (Boyd et al., 2013). Quantitative resistance would place less selective pressure on the pathogen, creating a more sustainable means of decreasing crop yield losses caused by BB and BLS.

A recently developed resource that expedited the identification of quantitative trait loci (QTL) in rice were the MAGIC (multiparent advanced generation inter-cross) populations

(Bandillo *et al.*, 2013). These populations comprise progeny advanced from eight elite founder (parent) varieties, and populations were constructed for japonica and indica accessions. We focus now on the indica MAGIC lines as they could be a new source of *R*-genes since most *R*-genes were identified in japonica varieties. In the MAGIC indica population, the founders (Fedearroz 50, IR45, IR46, IR77, PSBRc82, PSBRc152, Samba Mahsuri + sub1, and SHZ-2) were intercrossed, then single seed descent selfed for eight generations.

MAGIC populations allow for the resolution of individual genetic markers that could then be used in GWAS (genome-wide association studies) to identify trait-associated loci. For example, the MAGIC progeny were used to identify resistance QTL effective against bacterial blight disease using four *Xoo* strains from the Philippines (PXO61, PXO99, PXO86, and PXO341) (Boyd et al., 2013). Moreover, using a combination GWAS and IM (interval mapping) approach, they were also used to identify 11 BSR (broad-spectrum resistance) QTL and strain specific QTL effective against 20 African *Xoo* and *Xoc* strains as well as 12 QTL that are effective against *Xoo* strain PXO99A (Bossa-Castro *et al.*, 2018, Huerta *et al.*, 2021).

With resistance QTL identified, the question remains: What is the molecular basis for gene regulation within these genomic regions? Several prior studies had pointed to polymorphisms in promoters of rice defense response (DR) genes located under QTL as being responsible for DR gene responsiveness and enhanced resistance (Tonnessen *et al.*, 2015, Davidson *et al.*, 2010, Carrillo *et al.*, 2009, Li *et al.*, 2017, Fu *et al.*, 2011, Deng *et al.*, 2012). Thus, a potential explanation of gene regulation under QTL is the recruitment of TFs (transcription factors) to CREs (*cis*-regulatory elements) (for review, see Tonnessen *et al.*, 2021). For example, *de novo* analysis conducted by Tonnessen et al. (2015) identified a 229 bp deletion

in the *OsPAL4* promoter of a rice variety susceptible to multiple rice pathogens, including *M. oryzae*, *Rhizoctonia solani*, and *Xoo.* This deletion excluded CRMs (*cis*-regulatory motifs), or clusters of CREs, that may play a role in the defense gene regulation during infection (Tonnessen et al., 2015). CRMs were subsequently associated with promoters of defense genes underlying resistance QTL of rice (Tonnessen *et al.*, 2019, Tonnessen et al., 2021).

Another approach to addressing the question of gene regulation and QTL function, is by exploring how the bacteria interact directly with the rice genome. During *Xoo*/*Xoc* colonization of the plant tissue, the pathogen injects host cells with effectors called TALEs (transcription activator like effectors) (Perez-Quintero & Szurek, 2019). These proteins differ from other pathogen effectors in that they contain a domain that allows for localization of the protein to the nucleus. Once in the nucleus, TALEs slide along the host DNA and bind at a specific sequence known as the effector binding element (EBE). The sequence of the EBE is determined by the canonical central repeat (CCR) domain of the TALE, where residue 12 of each repeat determines binding specificity, and residue 13 stabilizes the effector binding to the DNA strand (Perez-Quintero & Szurek, 2019). Once bound, the activation domain of the TALE initiates gene transcription, for example, induction of susceptibility genes such as the *OsSWEET* genes (Perez-Quintero & Szurek, 2019, Streubel *et al.*, 2013). Polymorphisms in the EBE that prevent the TALE from binding lead to loss of that gene's activation and the associated phenotype. This concept and previously identified associations between Tal7b's EBE, a *Xoo* PXO86 TALE, with genes under resistance and susceptibility QTL (Huerta et al., 2021) inspired the hypothesis that the presence or absence of effectors binding EBEs could explain at least some of the resistance/susceptibility QTLs (Tonnessen et al., 2021). That is, EBE sequence variation

compared across rice varieties or lines leads to varied expression of plant defense genes. A difference in expression of these genes may give rise to varied disease phenotypes (Figure 1).

In this study, we take a step in addressing the role of EBE in rice promoters in quantitative resistance and susceptibility responses. Here, we describe a new pipeline that filters genes for ones that meet three criteria: they must overlap with QTL, contain an EBE in their promoters with sequence variation across the MAGIC founders, and this sequence variation must correlate with disease phenotypes (lesion lengths). Using this pipeline, we identified 54 genes that meet the above criteria for seven *Xoo* and *Xoc* strains. Because these 54 candidate genes meet our filtering criteria, we predict that they play a role in disease and/or resistance caused by these TALE-containing pathogens. *In vivo* regulation of these candidate genes, as well as other disease associated genes will be explored in the Appendix of this document.

1.2. EXPERIMENTAL PROCEDURES

Data acquisition

Genomic data files (in variant call format) for the eight MAGIC founders (Table **S1**) and four proof of concept varieties (Ejali, IR24, Khama, and SB) were collected from the 3k Rice Genome Project Database (3K.R.G.P., 2014). For the 3k Rice Genome Project, 3,000 rice varieties from 89 countries were sequenced with depth ranging from 4X to 60X and a minimum coverage of 92%. These Illumina libraries were mapped against the japonica reference genome (Nipponbare) and SNPs were identified. For our purposes, the reference sequence (Nipponbare version MSU7) (fasta) and annotation (gff) files were obtained from JGI Phytozome, which originally sourced these files from the Rice Genome Annotation Project (Kawahara *et al.*, 2013). The reference genome provided us with necessary gene coordinates and functional annotations of genes transcripts. *Xoo* and *Xoc* genomes were obtained from the NCBI database (fasta) (Table **S1**).

Pathogen aggressiveness (lesion length) data were collected for each MAGIC founder in response to inoculation with 22 *Xoo* and *Xoc* strains (Bandillo *et al.*, 2013, Bossa-Castro *et al.*, 2018, Huerta *et al.*, 2021) (Table **S2**). We validated phenotypes by inoculating the MAGIC founders with two strains (*Xoc* BAI5 and *Xoo* BAI3) (See "Phenotype validation methods" below). Lesion lengths for the four proof-of-concept varieties were derived from (Zaka *et al.*, 2018). Supporting interval coordinates for QTL were collected from (Bandillo *et al.*, 2013, Bossa-Castro *et al.*, 2018, Huerta *et al*., 2021, Zaka *et al.*, 2018).

Generation of the promoteromes

Gene coordinates were first extracted from the Nipponbare reference genome in Rstudio using the "rtracklayer" package (Lawrence *et al.*, 2009). Promoter coordinates were defined as 1000 bp upstream and 100 bp downstream of the translation start site (TSS). 100 bp downstream of the TSS was included to account for the multiple transcription start sites of each gene. In a Unix operating system, "samtools-faidx" was used to extract promoter subsequences from the reference genome. This was followed by the "vcf-consensus" command that applied SNP variants to the promoter sequences for each of the eight MAGIC founders (Figure **2a**) (Danecek *et al.*, 2011, Danecek *et al.*, 2021). Continuing, these are referred to as promoteromes. Promoteromes are stored in GitHub for public access.

TALE sequence extraction

In AnnoTALE, *Xoo* and *Xoc* (fasta) files were input, and nucleotide sequences were extracted for all TALEs in each strain's genome. Nucleotide sequences were then converted to amino acids sequences and repeat variable diresidues (RVDs) were extracted for each TALE (Figure **2b**) (Grau *et al.*, 2016). TALE repertoires for the 34 *Xoo* and *Xoc* strains are summarized in the Supplemental Data.

TALE targeting of gene promoters

Generated rice promoter sequences and *Xoo/Xoc* TALE RVD sequences were input into two prediction software, TALvez and PrediTALE (Perez-Quintero *et al.*, 2013, Erkes *et al.*, 2019). For a given strain's TALE repertoire, these software scan each MAGIC founder promoterome with each TALE RVD sequence to identify potential effector binding elements (EBE). Each plant

gene promoter's EBE then receives a score that represents likelihood of TALE binding (**Figure 2**). This score is then used to assign integer ranks to each promoter, denoting the top target, worst target, and every rank in between. On this scale, a rank of 1 is the most likely to be bound by a TALE when compared to all other promoters in the specified variety's genome. When running TALvez, the number of genes targeted by an individual TALE was limited to 2000. No limit was placed on PrediTALE predictions.

While the first steps of this pipeline included 34 different Xoo and Xoc strains, all required inputs (genomic sequence data, disease phenotypes, and mapped QTLs) were available for only seven strains (Xoc BAI5, Xoc BLS256, Xoo BAI3, Xoo MAI134, Xoo MAI145, Xoo PXO86, and Xoo PXO99A). Thus, the next steps included only these seven Xanthomonas strains.

Correlation between predicted TALE binding and disease phenotypes

Ranks from the previous step in the analysis were compiled into individual data sets for each of the seven *Xoo* and *Xoc* strains. In each data set the columns are the eight MAGIC founders and the rows are each gene transcript ID. Each cell then contains the lowest rank for the given gene (Figure **S1**). Only rows that had at least one rank less than or equal to 200 were moved forward. Ranks higher than 200 were considered unlikely targets. To filter for genes with sequence variation of EBEs across founders, the standard variation (SD) of ranks for each gene were calculated, and only genes with SDs of greater than 70 for TALVEZ and greater than 45 for PrediTALE were retained. For each gene, ranks were then correlated with lesion lengths measured from the Founders inoculated with *Xoo* and *Xoc* strains (Table **S2**; (Bandillo *et al.*,

2013, Bossa-Castro *et al.*, 2018, Huerta *et al.*, 2021). Pearson and Spearman correlations were used to account for linear and non-linear relationships between ranks and lesion lengths with a correlation-coefficient significance threshold set to an absolute value of 0.7 for both correlation methods (Nettleton, 2014). Only genes above these thresholds were analyzed further (Figure **2c**).

Gene coordinate overlaps with QTL

The list of TALE targeted genes with EBE variation across the MAGIC founders that correlated with lesion lengths was filtered further by identifying only those target genes that overlap with previously identified disease resistance QTL *(Bandillo et al., 2013, Bossa-Castro et al.*, 2018, Huerta *et al.*, 2021). QTL coordinates associated with *Xoo* and *Xoc* strains were input into Rstudio with the "GenomicRanges" package (Lawrence *et al.*, 2013). A subset of TALE target genes was identified with the "subsetByOverlaps" function for the seven strains for which QTL coordinates were available. Moving forward, the genes in the subset are referred to as candidate genes (Figure **2d**).

Sequence alignments and gene annotations

Variation in the EBE sequences was visualized and verified for the candidate genes using "ggmsa" (Zhou *et al.*, 2022). Gene function annotations were completed by extracting "Annotations" from the original MSU7 reference genome (gff) with the genomic ranges package (Lawrence *et al.*, 2009). Gene functions were appended to the dataset containing our candidate genes (Figure **2d**).

Plant materials and growing conditions

Seeds of eight MAGIC Indica founders (Fedearroz-50, IR45427-2B-2-2B-1-1, IR4630-22-2-5-1-3, IR77298-14-1-2-10, PSBRc82, PSBRc158, SHZ-2, Sambha Mahsuri + Sub1) (Table **S2**) were germinated in petri dishes on filter paper soaked with Maxim XL Fungicide at 25 ℃. Once the cotyledons emerged, plants were transplanted into soil filled pots in the greenhouse. The amount of time required for the cotyledon to emerge was different for each variety, so the start of germination was staggered to compensate for growth rate differences. Plants were grown in the greenhouse with a daytime temperature of 29 ℃ and a nighttime temperature of 23 ℃. Humidity was maintained at an average of 85%. At 2 weeks past germination, with three leaves emerged, plants were chelated with ferrous sulfate, and then fertilized twice a week with Peters Excel 15-5-15 Cal-Mag (Scotts, 300 mg/L).

Inoculation of MAGIC founders

For inoculum preparation, *Xoo* BAI3 and *Xoc* BAI5 strains were incubated on PSA (peptone sucrose agar) (Karganilla *et al.*, 1973) for 72 h at 28 ℃. Single colonies were taken from each and re-streaked onto new PSA plates, and the plates were incubated for 24-48 h. Inoculum was prepared by resuspending bacteria in sterile DI water and adjusting to an OD_{600} of 0.2 (10^8 CFU/mL).

Twenty-one-day old plants (four leaf stage) were inoculated with *Xoo* BAI3 and *Xoc* BAI5 suspensions, and an dH2O mock treatment. For *Xoc* BAI5, the second most extended leaf of four plants/founder was infiltrated with 0.1-0.3 uL of inoculum using a 1 mL, needle-less syringe (Reimers & Leach, 1991). The infiltration site was positioned directly over the leaf's xylem. For

Xoo BAI3, the tip of the second most extended leaf of four plants/founder was clipped using scissors dipped in inoculum (Kauffman *et al.*, 1973). For dH2O, the second most extended leaf of two plants were infiltrated using the same methods as above, while the second most extended leaf of the two remaining plants were clipped with scissors dipped in dH2O.

Lesion development and measurement

At 12 dpi (days post-inoculation) for infiltrated leaves, and 14 dpi for clipped leaves, leaves were collected and taped to transparencies. These were photographed and measurements of lesions were recorded using ImageJ (Schneider *et al.*, 2012). When calculating lesions extended from infiltration sites, 0.4 cm was subtracted from the total lesion length (0.4 cm = the diameter of the infiltration site).

1.3. RESULTS AND DISCUSSION

Proof-of-concept

The pipeline filters genes underlying QTL to identify candidate genes that have sequence variation in the EBEs of their promoters. Because EBE variation correlates strongly with disease phenotypic data, we predict that these genes are playing some role in response to *Xanthomonas*. To test the reliability of our pipeline to detect TALE-regulated candidate genes that function in disease susceptibility or resistance, we filtered for the known susceptibility gene *OsSWEET14*, targeted by a TALE (AvrXa7) in the *Xoo* strain PXO602 (Zaka *et al.*, 2018). Previous work identified sequence variation in *OsSWEET14*'s promoter across four indica varieties (Zaka *et al.*, 2018). This sequence variation was associated with induction of the *OsSWEET14* gene in two susceptible varieties (IR24 and SB), but not in resistant varieties (Ejali and Khama). It is important to note that we explored AvrXa7 targeting of *OsSWEET14* in these four indica varieties, as there is no variation in the *OsSWEET14* promoters of the japonica MAGIC Founders. To determine if our pipeline would identify the *OsSWEET14* promoter EBE as associated with phenotypes, these four varieties were applied to our pipeline. First, promoteromes were generated for Ejali, Khama, IR24, and SB. Then TALE targets were predicted for the TALE contained in *Xoo* PXO602 using TALvez and PrediTALE software (Perez-Quintero *et al.*, 2013, Erkes *et al.*, 2019). EBE ranks for each gene promoter were compiled into a data set containing all four varieties. We then tested for strong correlations between EBE ranks and lesion lengths induced by *Xoo* PXO602 as described above. Because the target of the TALE is known to be *OsSWEET14,* only genes that overlap with 18169150-18180250 bp on

chromosome 11 (the general coordinates of *OsSWEET14*) were screened (Kawahara *et al.*, 2013).

Using our pipeline, we found a strong correlation (Pearson = -0.996, Spearman = -0.736) between the EBE ranks and lesion lengths and detected a strong negative correlation between EBE rank and lesion length for susceptibility gene *OsSWEET14*, i.e., the EBE rank is lower for varieties with longer lesion lengths. The correlation is stronger when Pearson is used, therefore the relationship for this specific gene is more linear (Figure **4a-c**). A linear relationship observed between EBE rank and lesion length variables would indicate that as EBE rank increases lesion length decreases at a constant rate (if the correlation is negative, as in this case). The inverse would be true if the correlation is positive. Biologically, a constant rate of change of lesion length in relation to EBE rank could be associated with TALE binding. Thus, TALE associated expression levels of TALE bound genes live on a continuum rather than a binary (on/off). This continuum, however, was not observed by Zaka *et al.* (2018), as induction of *SWEET14* was only observed in the two susceptible varieties (IR24 and SB). In this case the linear relation observed between EBE rank and lesion length can be explained by our limited number of observations (correlation is calculated for four varieties). None-the-less, the results of this pipeline validation analysis indicate that (1) our promoter coordinate predictions are correct, (2) VCF files can be used to recreate promoter sequences, (3) AnnoTALE correctly extracted RVDs and predicted binding sites accurately (Grau *et al.*, 2016), and (4) correlation values can inform the relationships between genotype and phenotype. Overall, these results validate that our pipeline can identify candidate genes involved in TALE-activated susceptibility/resistance.

Disease phenotypes

Of the MAGIC founders, IR46 was most resistant to *Xoo* BAI3 followed by SHZ-2, both showing hypersensitive responses (HRs). Fedearroz 50, IR77, PSBRc158, PSBRc82, and Samba Mahsuri + sub1 were all moderately susceptible to *Xoo* BAI3, while IR45 was the most susceptible variety. For founders infiltrated with *Xoc* BAI5, Fedearroz 50, IR46, and SHZ-2 showed no spread of lesions, and an HR was observed as early as 5 dpi. IR77, Samba Mahsuri + sub1, PSBRc82, IR45, and PSBRc158 were increasingly susceptible in that order (Figure **5a-d**). For leaves infiltrated with dH2O, no lesions were observed. These results are concurrent with those observed by (Bossa-Castro *et al.*, 2018).

Identification of candidate genes by strain

The top 2,000 gene targets were predicted for each TALE from the repertoires of 34 *Xoo* and *Xoc* strains. We then assessed the correlation between EBE rank and lesion lengths and found 479 genes overall with EBE ranks that correlated with disease/resistance. Out of these, 49 genes overlap with resistance QTL associated with seven *Xoo* and *Xoc* strains (Figure **6a-f**). We extrapolated this candidate gene count further by looking at the direction of the correlation coefficients. Of the 49 genes, 23 have a positive correlation with lesion length and are predicted to play a role in resistance. The remaining 26 genes have a negative correlation with lesion length, potentially playing a role in susceptibility. Breaking this down by strain: One resistance gene and two susceptibility genes were predicted as targets of *Xoc* BAI5; three resistance genes and six susceptibility genes were predicted as targets of *Xoc* BLS256; five resistance genes were predicted for *Xoo* BAI3; one resistance gene was predicted for *Xoo* MAI134; two resistance

genes and one susceptibility gene were predicted for *Xoo* PXO86; and 11 resistance genes and 17 susceptibility genes were predicted for *Xoo* PXO99A. No genes under QTL associated with *Xoo* MAI145 were predicted as TALE targets.

Considering gene function, two of the four candidate gene targets for BAI5 are "unannotated" or "hypothetical" proteins (UHP). The 3rd is a potential resistance gene that codes for a SHR5, a receptor-like kinase (RLK). Downregulation of a similar gene in sugarcane was closely associated with the successful establishment of endophytic bacteria on their host (Vinagre *et al.*, 2006). For *Xoc* BLS256, one of the potential resistance genes predicted codes for a protein containing a pentatricopeptide repeat (PPR). Proteins containing PPRs often play a role in post-transcriptional regulation of RNA (Manna, 2015). Two of BLS256's targets code for NB-ARC domain containing proteins, and while NB-ARC proteins often bind pathogen effectors to elicit ETI, these two genes are predicted to play a role in susceptibility (Dodds & Rathjen, 2010). The remaining candidate genes for BLS256 are UHP. Three of the *Xoo* BAI3 targets are UHP and the other two are potential resistance genes. One of these genes' codes for the Mla1 protein, a nucleotide binding domain and leucine-rich repeat protein (NLR) that confers barley with resistance to powdery mildew (Lu *et al.*, 2016). The other codes for a patanin-like phospholipase (PLP). Activation of this gene may contribute to defense signaling as seen when the *Capsicum annuum* PLP gene is induced by *X. campestris* pv. *vesicatoria* in pepper leading to cell death (Kim *et al.*, 2014).

The single target predicted for *Xoo* MAI134 is a potential resistance gene that codes for a putative esterase. Predicted targets of *Xoo* PXO86 include genes that code for two UHPs potentially playing a role in resistance. The third candidate, predicted to play a role in

susceptibility to PXO86, is *phosphate-induced protein 1* (*OsPHI-1*). This gene is part of a family of *OsPHIs* that closely interacts with abscisic acid (ABA) in response to abiotic stresses (Quan *et al.*, 2019). Targets of note predicted to play a role in resistance to *Xoo* PXO99A include an NBS-LRR type resistance protein, and many UHPs. Interestingly, many of the potential susceptibility genes include ones that code for disease resistance protein RPM1 and an NB-ARC/LRR domain containing protein which traditionally confer plants with resistance (Dodds & Rathjen, 2010, Wang *et al.*, 2021). Additional targets of *Xoo* PXO99A can be found in the Supplemental Data section (Table **S4**).

The described methods explore resistance QTL associated with seven *Xoo* and *Xoc* strains for genes that meet specific criteria: ones that have EBEs in their promoters and sequence variation in their EBEs, and ones for which the sequence variation correlates with disease phenotypes of the given strain. With this pipeline we successfully identified a confirmed susceptibility gene (*OsSWEET14*), suggesting the pipeline has the potential to uncover new susceptibility or resistance genes. Subsequently, we predicted 49 candidate genes for the seven *Xoo* and *Xoc* strains.

1.4. CONCLUSION

Previous work identified resistance QTL associated with *Xoo* and *Xoc* strains (Bossa-Castro *et al.,* 2018, Huerta *et al.*, 2021, Bandillo *et al.*, 2013). However, there is no current knowledge as to what the molecular basis for gene regulation under these QTLs. We used computational analyses to explore the hypothesis that *Xoo* and *Xoc* TALE binding of gene promoters under QTL regulates quantitative resistance and susceptibility. Ultimately, our methods generate 49 hypotheses that must be confirmed *in vivo*. Upon inoculation of the MAGIC founders with one of the *Xoc* or *Xoo* for which candidate gene targets were predicted, we expect that candidate genes will be targeted by the predicted TALE and transcription will be induced. To confirm our new hypotheses, mRNA will be extracted from plants inoculated with *Xoo* or *Xoc* and raw sequencing data will be used to calculate differentially expressed genes. The *in vivo* analysis involving RNAseq experiments are underway (Appendix **B**). MAGIC founders were inoculated with *Xoo* BAI3 and *Xoc* BAI5 strains to compare differential gene expression when compared to a water control. RNAseq data collected will inform us whether our candidate genes are expressed in the presence of these two strains. If RNAseq data supports some of our 49 hypotheses while excluding others, we will use this knowledge to refine our pipeline's filtering power, for example, setting a higher threshold for correlation between EBE Rank and lesion length. Refining our pipeline and confirming its ability to predict novel resistance/susceptibility genes will provide evidence that resistance and susceptibility QTL for *Xoo* and *Xoc* are at least in part regulated by TALE. Confirmation of these candidate genes will

give us a better understanding of how *Xanthomonas*, prevalent on many species, achieves virulence, and will provide us with resistance markers to breed more resilient crops.

Limitations

The analysis outlined above required data inputs sourced from previous studies (3K.R.G.P., 2014, Kawahara *et al.*, 2013, Bandillo *et al.*, 2013, Bossa-Castro *et al.*, 2018, Huerta *et al.*, 2021). This means the analysis is first limited by any errors in the production of this primary data. For example, some error may arise from sourcing of SNP data. As stated above, the 3k rice genome project involved mapping short reads to the Nipponbare reference genome (3K.R.G.P., 2014). For many of the japonica varieties this isn't problematic. The MAGIC founders that we investigate here, however, are indica varieties. By mapping indica varieties to a japonica reference, we get accurate SNP variants for genes that successfully align. Rearrangements, or larger changes to the genome between japonica and indica varieties prevent full confidence of our promoter sequences for the 8 founders. One way to fix this is by using the secondary SNP data generated by mapping short reads to the IR64 reference genome (Mansueto *et al.*, 2017, Schatz *et al.*, 2014). This fix was attempted but proved more complicated than we anticipated. The biggest challenges experienced when using SNP data mapped to IR64 are that QTL coordinates were originally identified with the SNP data mapped to Nipponbare. The IR64 reference genome is also an older, more fragmented construct which requires joining of scaffolds to be comparable to the Nipponbare reference. Additional computational methods need exploring to address this limitation.

Our analysis showed a large variation in the number of targets predicted for *Xoc* strains/African *Xoo* strains and the *Xoo* strains from the Philippines (Table **S2**). This can partially be explained by the larger TALE repertoires of *Xoo* strains from the Philippines. Additionally, the number of candidate genes may be deflated due to the low number of targets predicted for PXO86. The QTL for PXO86 were determined using only GWAS (Bandillo *et al.*, 2013), whereas the QTL for the other six strains were resolved using both GWAS and IM (Bossa-Castro *et al.*, 2018, Huerta *et al.*, 2021). Recent studies have shown the benefit of using both GWAS and IM to identify QTL as each model can identify loci that the other does not (Raghavan *et al.*, 2017, Bossa-Castro *et al.*, 2018). This means the PXO86 QTL coordinates may not include all possible resistance QTL, and our "net" is capturing fewer genes at the QTL overlap step of the pipeline. This "net" size could be increased by coupling the GWAS with IM as with the other strains, however, we expect data gained from *in vivo* experiments to help us refine other steps of the pipeline and produce a more accurate candidate gene total for the strains analyzed.

Additionally, these methods only consider TALE-mediated, *cis*-regulation of genes underlying resistance QTL. There may be many genes involved in resistance and susceptibility to *Xoo* and *Xoc* strains that are regulated by other *cis*-regulatory elements (CREs), by trans-regulatory elements (TREs), and/or are outside of resistance QTL. The resistance or susceptibility conferred by the QTL might also be related to variation in gene function, for example non-synonymous or missense mutations in coding regions of some parents that alter protein function and subsequent defense responses (Tsuchiya & Eulgem, 2013, Zhang *et al.*, 2016). While this is a possibility, it was observed for maize and *Brassica napus* that trait- and disease- associated SNPs, respectively, were enriched in intergenic regions (Wallace *et al.*,

2014, Fikere *et al.*, 2020). Phenotypes may also be due to simultaneous interactions of genes on multiple QTLs. All these explanations cannot be validated with our current filtering methods. This pipeline does, however, have the potential to predict TALE-mediated, *cis*-regulation of genes that code for proteins involved in *trans*-regulation of genes or post-transcriptional modifications.

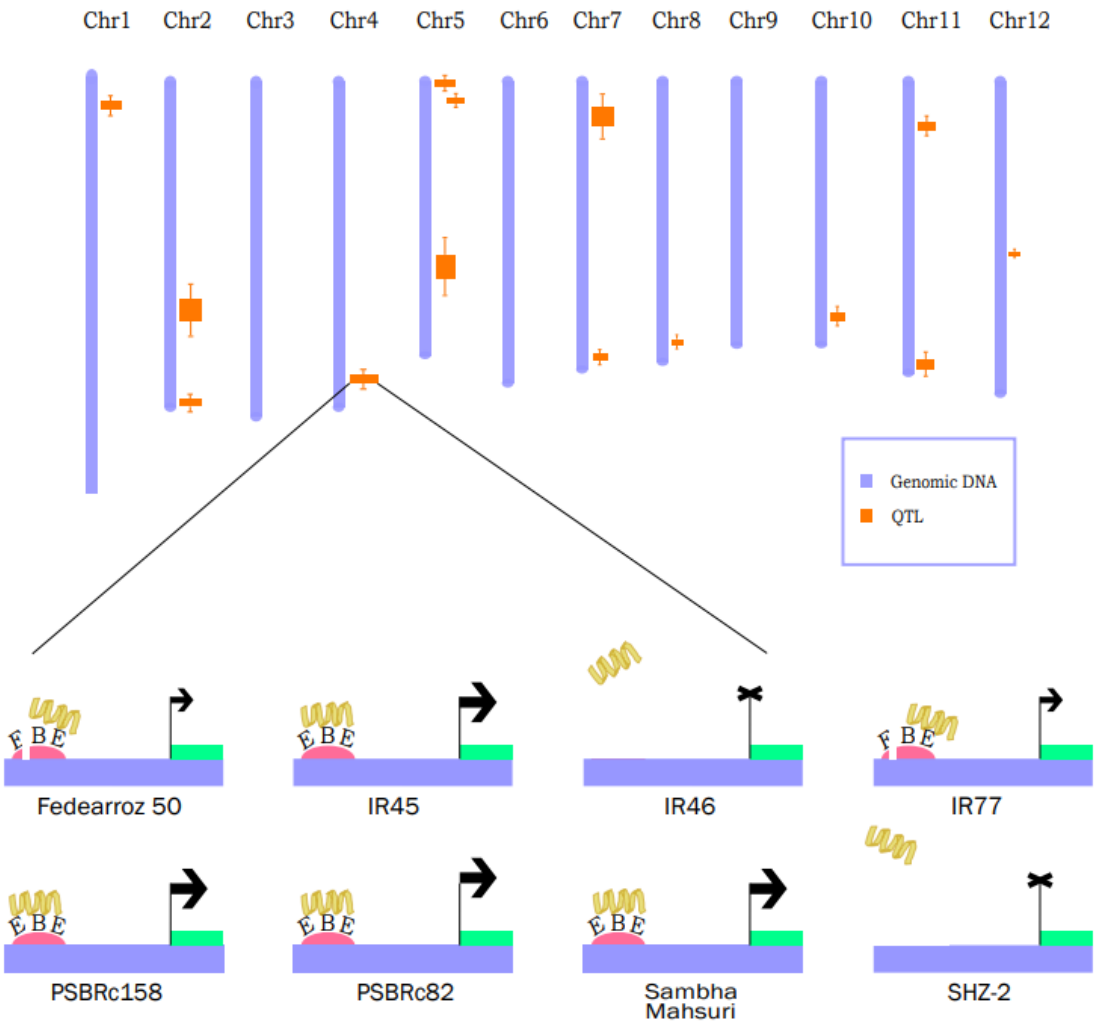

Figure 1. Representation of our proposed hypothesis: "*the presence or absence of effector binding EBEs could explain some of the resistance/susceptibility QTL.*" In this image genomic DNA is colored blue, QTL are orange bars, TALE are yellow curls, EBEs are pink, and coding regions are green. Under a QTL on chromosome 4, we zoom in on a hypothetical gene to compare EBE sequences across the MAGIC founders. The sequence variation that we see leads to varied binding of the TALE to the promoter, and thus, differential expression of the hypothetical gene. If activation of this gene leads to a susceptible phenotype, IR45 and Sambha Mahsuri would be the most susceptible and IR46 and SHZ-2 the least susceptible.

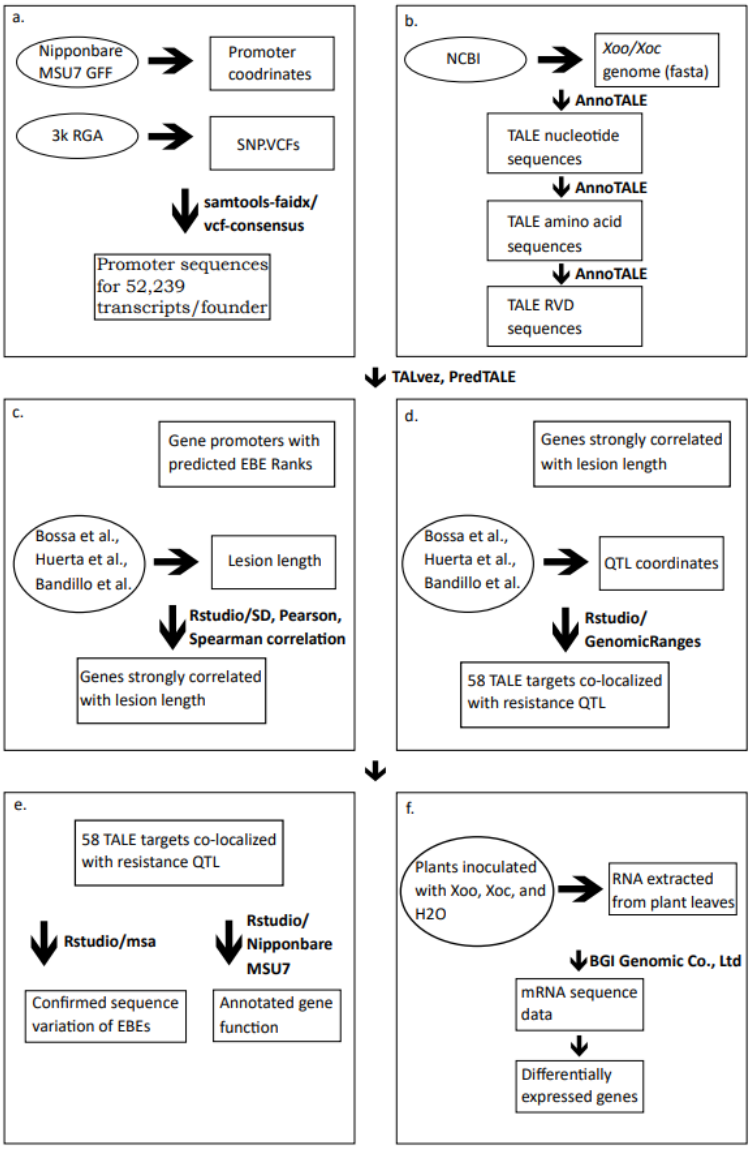

Figure 2. Schematic of the computational workflow. Ellipses represent sources of data, rectangles are products/inputs for the next step in the workflow, and the bolded text names software required for each step. **(a)** Promoter coordinates and SNP data are read against a reference genome to generate promoter sequences. **(b)** AnnoTALE is used to extract TALE sequences from *Xoo* and *Xoc* genomes, convert nucleotide sequences to amino acids, then extract RVDs. **(c)** The two products of **(a)** and **(b)** are inputs for TALvez and PrediTALE which predict TALE binding sites in the gene promoters. Once gene promoters are assigned EBE ranks, we use Pearson and Spearman to look for strong correlation with lesion length. **(d)** The genes strongly correlated with lesion lengths are filtered further for those that overlap with resistance QTL associated with *Xoo* and *Xoc.* **(e)** The TALE targets co-localized with QTL are then annotated with gene function using the Nipponbare reference genome, and sequence variation in their EBEs is confirmed with "msa" in Rstudio. **(e)** In **(f)**, we highlight future steps for this project to confirm candidate genes *in vivo*. These methods will be discussed more in the appendix.

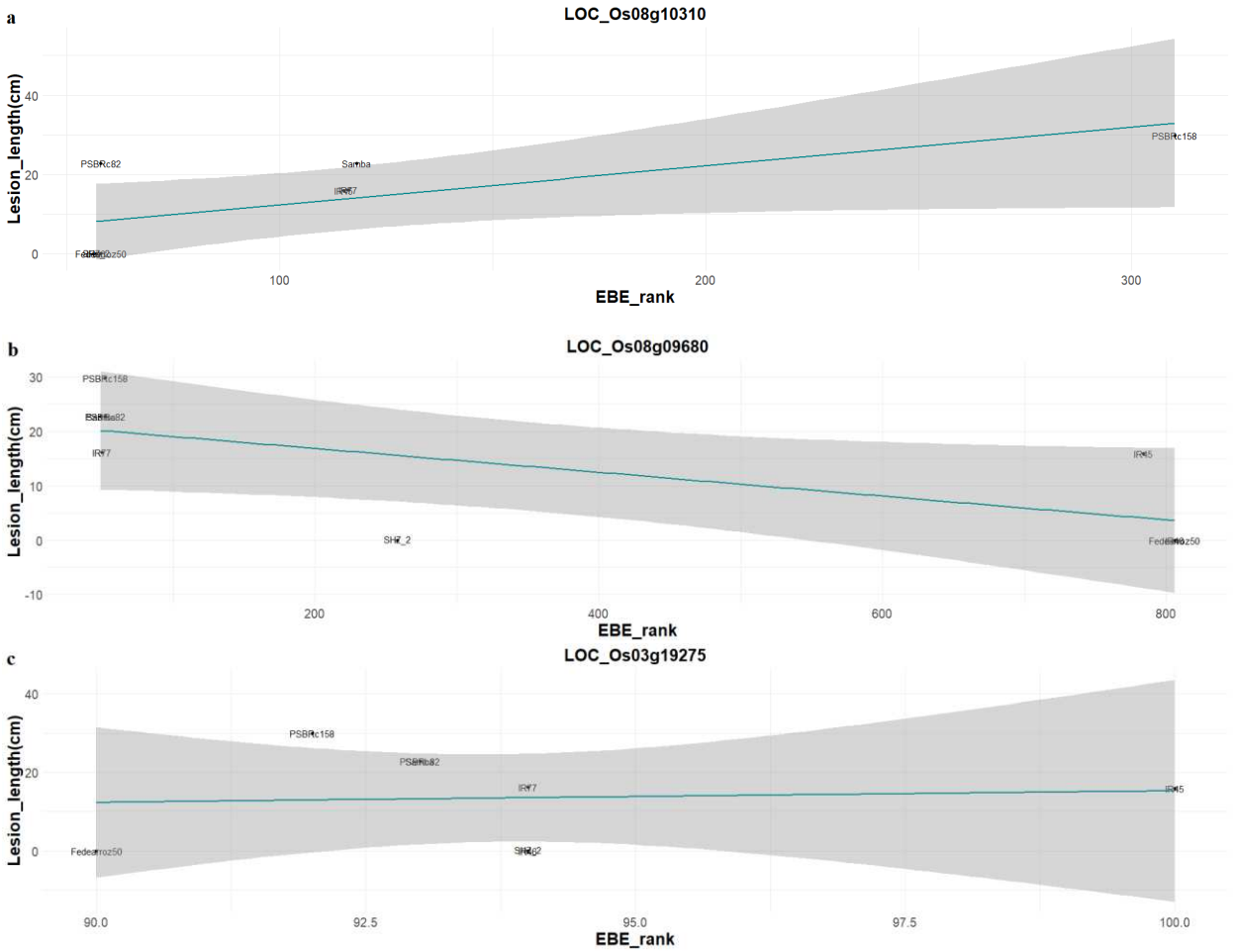

Figure 3. As we analyze correlation between EBE ranks of the MAGIC founders and lesion lengths, we consider strength of correlation ($|r| > 0.7$) and the direction of the correlation. Below are two candidate genes predicted for *Xoc* BAI5 and one gene with no correlation to phenotype: **(a)** a positive correlation occurs if lesion length increases as EBE rank increases. This indicates a gene is a predicted TALE target in resistant varieties and may be playing a role in resistance; (**b)** a negative correlation occurs if lesion length decreases as EBE rank increases. This indicates a gene is a predicted TALE target in susceptible varieties and may be playing a role in susceptibility; and **(c)** a gene for which EBE rank does not correlate with lesion length, indicating a gene for which TALE binding is not related to regulation of resistance or susceptibility.

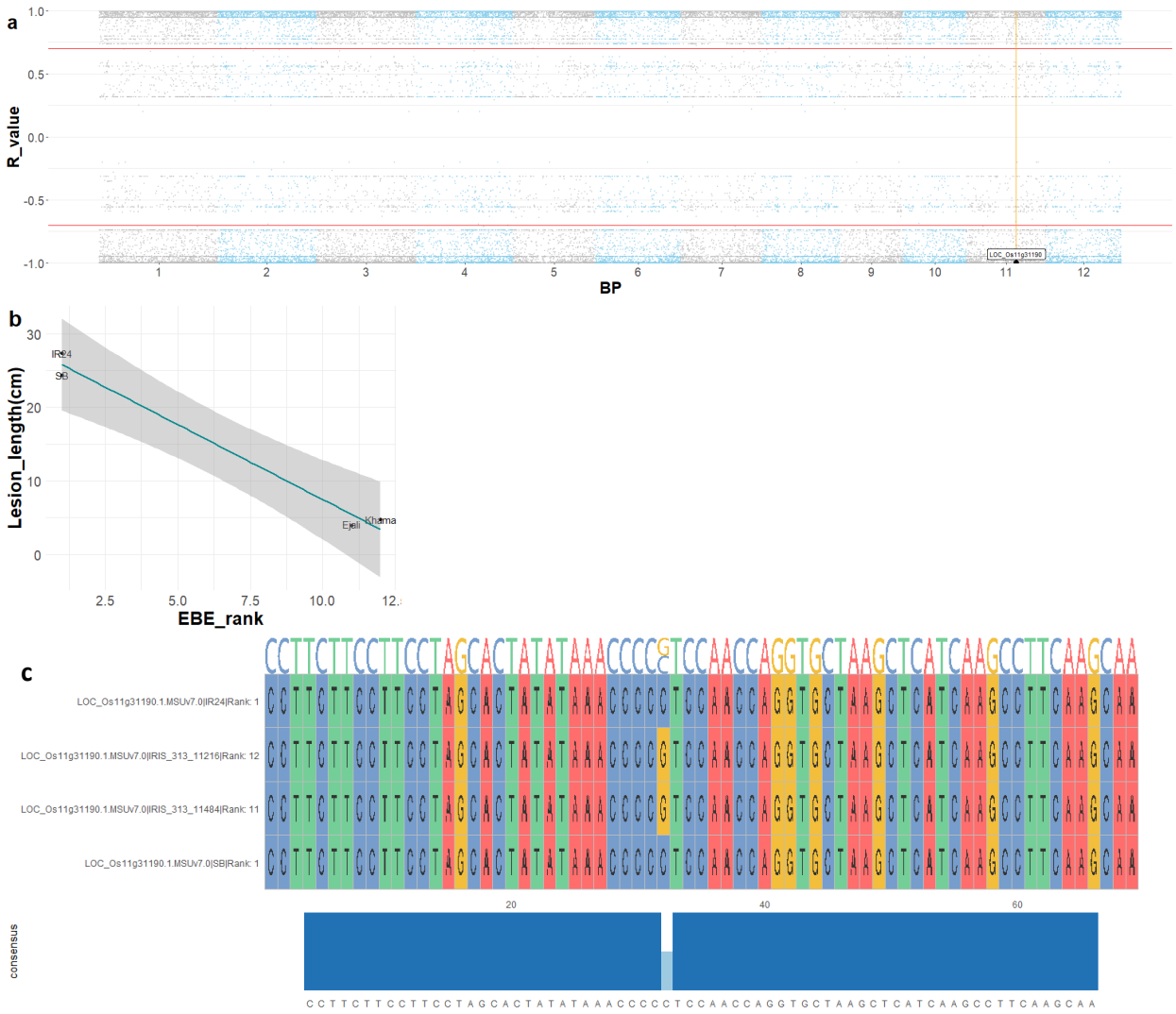

Figure 4. Pipeline validation. The proposed pipeline was used to identify *OsSWEET14*, a known susceptibility gene with EBE sequence variation in four varieties (IR24, Ejali, Khama, and SB) (Zaka *et al.*, 2018). This EBE variation leads to varied *OsSWEET14* gene induction upon inoculation with *Xoo* PXO602, a strain containing TALE (AvrXa7). **(a)** *OsSWEET14* is shown mapped to the rice genome, where the x-axis is bp position with chromosomes labeled. The y-axis is correlation value of the genes with a threshold for strong correlation set to 0.7 and -0.7. The orange line overlaying *OsSWEET14* are the coordinates used in place of QTL overlaps. **(b)** EBE rank plotted against lesion length to visualize the negative correlation and role in susceptibility of this gene. **(c)** EBEs across the four varieties. In the two resistant varieties (Ejali and Khama), there is a point mutation from "C" to "G."

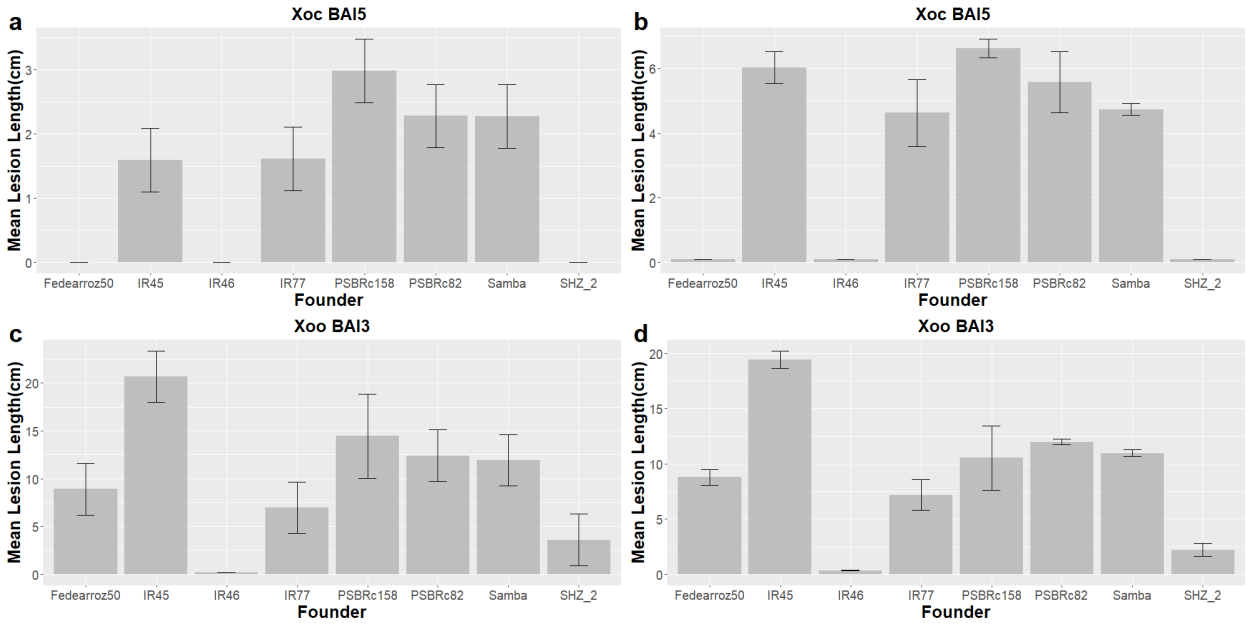

Figure 5. Phenotypic interactions of *Xoc* BAI5 and *Xoo* BAI3 with the indica MAGIC founders. Leaves were infiltrated with *Xoc* BAI5 and clipped with *Xoo* BAI3, and lesions were measured at 14 dpi. Mean lesion lengths were plotted on the y-axis, and the founder on the x-axis. (**a**) *Xoc* BAI5 lesion lengths (cm) from Bossa *et al.* (2018), (b) *Xoc* BAI5 lesion lengths (cm) from this study. (**c**) *Xoo* BAI3 lesion lengths (cm) recorded by Bossa *et al.* (2018), (b) *Xoo* BAI3 lesion lengths (cm) from this study. Error bars are the standard error calculated for the mean.

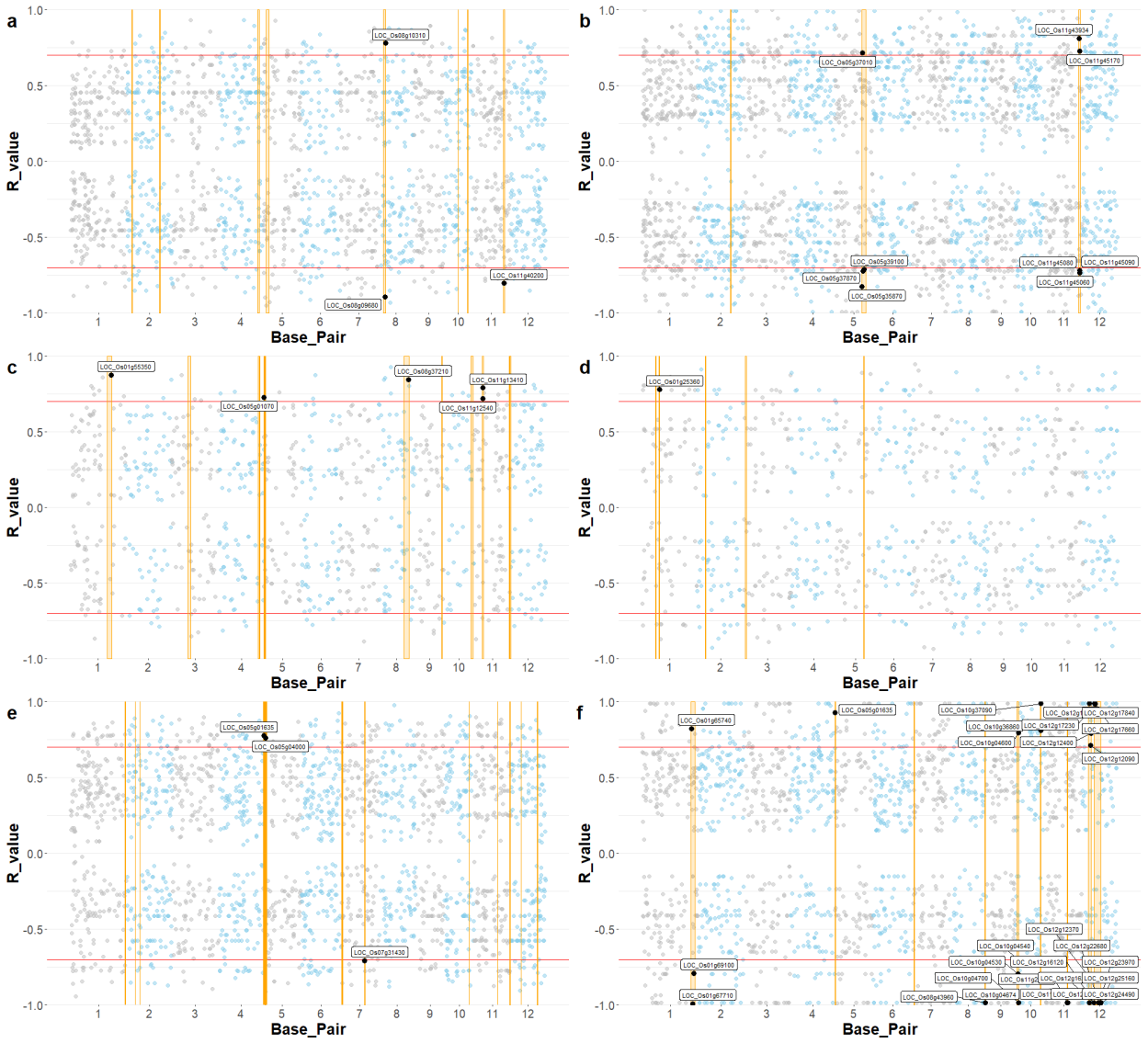

Figure 6. The 49 candidate genes mapped by strain to the 12 rice chromosomes. For each plot, base pair position within the rice genome (Nipponbare MSU7) is the x-axis, with chromosome numbers labeled (Kawahara *et al.*, 2013). The y-axis is the correlation value calculated for each gene. Each point then represents an individual gene. Thresholds of strong correlation have been set to 0.7 and -0.7 and are represented by the red horizontal lines. QTL are also mapped to each plot (orange boxes). **(a)** The three candidate genes identified for *Xoc* BAI5. **(b)** The nine candidate genes identified for *Xoc* BLS256. **(c)** The five candidate genes identified for *Xoo* BAI3. **(d)** The one candidate gene identified for *Xoo* MAI134. **(e)** The three candidate genes identified for *Xoo* PXO86. **(f)** The 28 candidate genes identified for *Xoc* PXO99A.

REFERENCES

3K.R.G.P. (2014) The 3,000 rice genomes project. *GigaScience,* **3,** 7.

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A., Lobina, I. T., Dilla-Ermita, C. J.*, et al.* (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice,* **6,** 11.

Bossa-Castro, A. M., Tekete, C., Raghavan, C., Delorean, E. E., Dereeper, A., Dagno, K.*, et al.* (2018) Allelic variation for broad-spectrum resistance and susceptibility to bacterial pathogens identified in a rice MAGIC population. *Plant Biotechnol J*.

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E. and Leung, H. (2013) Plant-pathogen interactions: disease resistance in modern agriculture. *Trends Genet,* **29,** 233-240.

Carrillo, M. G. C., Goodwin, P. H., Leach, J. E., Leung, H. and Cruz, C. M. V. (2009) Phylogenomic relationships of rice oxalate oxidases to the cupin superfamily and their association with disease resistance QTL. *Rice,* **2,** 67-79.

Cohen, S. P. and Leach, J. E. (2019) Abiotic and biotic stresses induce a core transcriptome response in rice. *Scientific Reports,* **9**.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A.*, et al.* (2011) The variant call format and VCFtools. *Bioinformatics,* **27,** 2156-2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O.*, et al.* (2021) Twelve years of SAMtools and BCFtools. *GigaScience,* **10**.

Davidson, R. M., Manosalva, P. M., Snelling, J., Bruce, M., Leung, H. and Leach, J. E. (2010) Rice germin-like proteins: Allelic diversity and relationships to early stress responses. *Rice,* **3,** 43-55.

Deng, H. Q., Liu, H. B., Li, X. H., Xiao, J. H. and Wang, S. P. (2012) A CCCH-type zinc finger nucleic acid-binding protein quantitatively confers resistance against rice bacterial blight disease. *Plant Physiol,* **158,** 876-889.

Dodds, P. N. and Rathjen, J. P. (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics,* **11,** 539-548.

Erkes, A., Mücke, S., Reschke, M., Boch, J. and Grau, J. (2019) PrediTALE: A novel model learned from quantitative data allows for new perspectives on TALE targeting. *PLOS Computational Biology,* **15,** e1007206.

Fikere, M., Barbulescu, D. M., Malmberg, M. M., Spangenberg, G. C., Cogan, N. O. I. and Daetwyler, H. D. (2020) Meta-analysis of GWAS in canola blackleg (Leptosphaeria maculans) disease traits demonstrates increased power from imputed whole-genome sequence. *Scientific Reports,* **10**.

Fu, J., Yu, H. H., Li, X. H., Xiao, J. H. and Wang, S. P. (2011) Rice GH3 gene family Regulators of growth and development. *Plant Signal Behav,* **6,** 570-574.

Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G.*, et al.* (2016) AnnoTALE: bioinformatics tools for identification, annotation and nomenclature of TALEs from Xanthomonas genomic sequences. *Scientific Reports,* **6,** 21077.

Huerta, A. I., Delorean, E. E., Bossa-Castro, A. M., Tonnessen, B. W., Raghavan, C., Corral, R.*, et al.* (2021) Resistance and susceptibility QTL identified in a rice MAGIC population by screening with a minor-effect virulence factor from *Xanthomonas oryzae* pv. *oryzae*. *Plant Biotechnol J,* **19,** 51-63.

Hutin, M., Sabot, F., Ghesquière, A., Koebnik, R. and Szurek, B. (2015) A knowledge-based molecular screen uncovers a broad-spectrum OsSWEET14 resistance allele to bacterial blight from wild rice. *Plant J,* **84,** 694-703.

Jones, J. D. G. and Dangl, J. L. (2006) The plant immune system. *Nature,* **444,** 323-329.

Karganilla, A., Natural, M. and Ou, S. (1973) A comparative study of culture media for *Xanthomonas oryzae*. *Phillip Agric,* **57,** 141-152.
Kauffman, H., Reddy, A., Hsiek, S. and Merca, S. (1973) An improved technique for evaluating resistance of rice varieties to *Xanthomonas oryzae*. *Plant Dis Rep,* **57,** 537-541.
Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S.*, et al.* (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice,* **6,** 4.
Khush, G. S. (1997) Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology,* **35,** 25-34.
Kim, D. S., Jeun, Y. and Hwang, B. K. (2014) The pepper patatin-like phospholipase CaPLP1 functions in plant cell death and defense signaling. *Plant Molecular Biology,* **84,** 329-344.
Lawrence, M., Gentleman, R. and Carey, V. (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics,* **25,** 1841-1842.
Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R.*, et al.* (2013) Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology,* **9,** e1003118.
Li, W. T., Zhu, Z. W., Chern, M. S., Yin, J. J., Yang, C., Ran, L.*, et al.* (2017) A natural allele of a transcription factor in rice confers broad-spectrum blast resistance. *Cell,* **170,** 114-+.
Lu, X., Kracher, B., Saur, I. M. L., Bauer, S., Ellwood, S. R., Wise, R.*, et al.* (2016) Allelic barley MLA immune receptors recognize sequence-unrelated avirulence effectors of the powdery mildew pathogen. *Proceedings of the National Academy of Sciences,* **113,** E6486-E6495.
Manna, S. (2015) An overview of pentatricopeptide repeat proteins and their applications. *Biochimie,* **113,** 93-99.
Mentlak, T. A., Kombrink, A., Shinya, T., Ryder, L. S., Otomo, I., Saitoh, H.*, et al.* (2012) Effector-Mediated Suppression of Chitin-Triggered Immunity by Magnaporthe oryzae Is Necessary for Rice Blast Disease *The Plant Cell,* **24,** 322-335.
Naveed, Z. A., Wei, X., Chen, J., Mubeen, H. and Ali, G. S. (2020) The PTI to ETI Continuum in Phytophthora-Plant Interactions. *Frontiers in Plant Science,* **11**.
Nettleton, D. (2014) Chapter 6 - Selection of Variables and Factor Derivation. In: *Commercial Data Mining.* (Nettleton, D., ed.). Boston: Morgan Kaufmann, pp. 79-104.
Ngou, B. P. M., Ahn, H.-K., Ding, P. and Jones, J. D. G. (2021) Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature,* **592,** 110-115.
Ou, S. H. (1985) *Rice Diseases*. Kew: Commonwealth Mycological Institute.
Perez-Quintero, A. L., Rodriguez, R. L., Dereeper, A., Lopez, C., Koebnik, R., Szurek, B.*, et al.* (2013) An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS One,* **8,** e68464.
Perez-Quintero, A. L. and Szurek, B. (2019) A Decade Decoded: Spies and Hackers in the History of TAL Effectors Research. *Annu Rev Phytopathol,* **57,** 459-481.
Pruitt, R. N., Gust, A. A. and Nürnberger, T. (2021) Plant immunity unified. *Nature Plants,* **7,** 382-383.
Quan, C., Bai, Z., Zheng, S., Zhou, J., Yu, Q., Xu, Z.*, et al.* (2019) Genome-wide analysis and environmental response profiling of phosphate-induced-1 family genes in rice (Oryza sativa). *Biotechnology & Biotechnological Equipment,* **33,** 627-638.
Raghavan, C., Mauleon, R., Lacorte, V., Jubay, M., Zaw, H., Bonifacio, J.*, et al.* (2017) Approaches in characterizing genetic structure and mapping in a rice multiparental population. *G3-Genes Genom Genet,* **7,** 1721-1730.
Reimers, P. J. and Leach, J. E. (1991) Race-specific resistance to X*anthomonas 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 P,* **38,** 39-55.
Sanguankiattichai, N., Buscaill, P. and Preston, G. M. (2022) How bacteria overcome flagellin pattern recognition in plants. *Current Opinion in Plant Biology,* **67**.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* **9,** 671-675.

Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J. and Szurek, B. (2013) Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol,* **200,** 808-819.

Thomma, B. P. H. J., Nürnberger, T. and Joosten, M. H. A. J. (2011) Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. *The Plant Cell,* **23,** 4-15.

Tonnessen, B. W., Bossa-Castro, A. M., Martin, F. and Leach, J. E. (2021) Intergenic spaces: a new frontier to improving plant health. *New Phytologist,* **232,** 1540-1548.

Tonnessen, B. W., Bossa-Castro, A. M., Mauleon, R., Alexandrov, N. and Leach, J. E. (2019) Shared cis-regulatory architecture identified across defense response genes is associated with broad-spectrum quantitative resistance in rice. *Scientific Reports,* **9**.

Tonnessen, B. W., Manosalva, P., Lang, J. M., Baraoidan, M., Bordeos, A., Mauleon, R.*, et al.* (2015) Rice phenylalanine ammonia-lyase gene *OsPAL4* is associated with broad spectrum disease resistance. *Plant Molecular Biology,* **87,** 273-286.

Tsuchiya, T. and Eulgem, T. (2013) An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *P Natl Acad Sci USA,* **110,** E3535-E3543.

Tsuda, K. and Katagiri, F. (2010) Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr Opin Plant Biol,* **13,** 459-465.

Vinagre, F., Vargas, C., Schwarcz, K., Cavalcante, J., Nogueira, E. M., Baldani, J. I.*, et al.* (2006) SHR5: a novel plant receptor kinase involved in plant-N2-fixing endophytic bacteria association. *J Exp Bot,* **57,** 559-569.

Wallace, J. G., Bradbury, P. J., Zhang, N. Y., Gibon, Y., Stitt, M. and Buckler, E. S. (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. *Plos Genet,* **10**.

Wang, A., Shu, X., Jing, X., Jiao, C., Chen, L., Zhang, J.*, et al.* (2021) Identification of rice (Oryza sativa L.) genes involved in sheath blight resistance via a genome-wide association study. *Plant Biotechnol J,* **19,** 1553-1566.

Wang, Y., Pruitt, R. N., Nürnberger, T. and Wang, Y. (2022) Evasion of plant immunity by microbial pathogens. *Nature Reviews Microbiology,* **20,** 449-464.

Yuan, M. H., Ngou, B. P. M., Ding, P. T. and Xiu-Fan, X. (2021) PTI-ETI crosstalk: an integrative view of plant immunity. *Current Opinion in Plant Biology,* **62**.

Zaka, A., Grande, G., Coronejo, T., Quibod, I. L., Chen, C.-W., Chang, S.-J.*, et al.* (2018) Natural variations in the promoter of OsSWEET13 and OsSWEET14 expand the range of resistance against Xanthomonas oryzae pv. oryzae. *PLOS ONE,* **13,** e0203711.

Zhang, H. T., Tao, Z., Hong, H. M., Chen, Z. H., Wu, C. Y., Li, X. H.*, et al.* (2016) Transposon-derived small RNA is responsible for modified function of WRKY45 locus. *Nature Plants,* **2**.

Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T. T., Wang, Q.*, et al.* (2022) ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Brief Bioinform*.

APPENDIX A

Supplementary material

Table S1. Agronomic traits of MAGIC indica founders (modified from (Bandillo et al., 2013, Bossa-Castro et al., 2018))

Designation	Germplasm/Variety	Origin	IRIS ID	GID†	Known *R* genes	Agronomic relevance
IR46	IR4630-22-2-5-1-3	IRRI	IRIS 313-15898	56023	*Xa4*	Salt tolerance
=	Fedearroz 50	Colombia	IRIS 313-15896	1846419	*Xa4*	Delayed senescence, quality traits, disease tolerance, breeding progenitor
IR77	IR77298-14-1-2-10	IRRI	IRIS 313-15901	2154106	*Xa4*	Drought tolerance, tungro resistance, IR64 background
SHZ-2	Shan-Huang Zhan-2 (SHZ-2)	China	IRIS 313-15897	402862	*Xa4*	Blast resistance, high yielding, breeding progenitor in China
=	PSBRc82 (IRRI123 or IR64633-87-2-2-3-3)	IRRI	CX358	94801	*Xa4, xa5*	High yielding, most popular variety of Philippines
Sambha	Sambha Mahsuri + Sub1 (IR 07F287)	IRRI	IRIS 313-15900	2254836		Mega variety, good grain quality, submergence tolerance
PSBRc158	PSBRc 158 (IRRI146 or IR77186-122-2-2-3)	IRRI	IRIS 313-15902	1111266	*Xa4*	High yielding in new plant type II (NPT) background
IR45	IR45427-2B-2-2B-1-1	IRRI	IRIS 313-15899	1935108	*Xa4*	Iron toxicity tolerance

"=" used if designation is the same as germplasm/variety, designations are used especially in figures. †GID, germplasm identification.

Table S2. Transcription activator-like effector repertoire for 34 *Xoc* and *Xoo* strains.

Pathovar	Strain	Number of TALE	Shared TALE
Xoo	AXO1947	9	-
Xoo	BAI3*	9	-
Xoo	KACC10331	12 + 1^{p}	-
Xoo	MAFF311018	12 + 1^{p}	-
Xoo	MAI1	9	-
Xoo	MAI68	9	-
Xoo	MAI73	9	MAI95 (all)
Xoo	MAI95	9	MAI73 (all)
Xoo	MAI99	9	-
Xoo	MAI106	9	MAI129 (all)
Xoo	MAI129	9	MAI106 (all)
Xoo	MAI134*	9	-
Xoo	MAI145*	9	-
Xoo	PXO71	15 + 5^{p}	-
Xoo	PXO83	16 + 2^{p}	PXO86 (all)
Xoo	PXO86*	16 + 2^{p}	PXO83 (all), PXO99A (1)
Xoo	PXO99A*	17 + 2^{p}	-
Xoo	PXO145	15 + 3^{p}	-
Xoo	PXO211	15 + 2^{p}	-
Xoo	PXO236	15 + 1^{p}	-
Xoo	PXO282	17 + 1^{p}	-
Xoo	PXO314	16 + 2^{p}	-
Xoo	PXO524	18 + 1^{p}	-

Xoo	PXO563	16 + 2^{p}	-
Xoo	PXO602	17 + 3^{p}	-
Xoc	B8-12	27 + 1^{p}	-
Xoc	BLS256*	27 + 1^{p}	-
Xoc	BLS279	26 + 1^{p}	-
Xoc	BXOR1	25 + 2^{p}	-
Xoc	CFBP2286	27 + 1^{p}	-
Xoc	CFBP7341 (BAI5)*	20 + 2^{p}	-
Xoc	CFBP7342	23 + 1^{p}	-
Xoc	L8	28 + 1^{p}	-
Xoc	RS105	23 + 1^{p}	-

"*" denotes strains for which genomic data, phenotypic data, and QTL coordinates were available for. "^{p}" are the amount of pseudo TALE genes identified for each strain.

	Fedearroz50	IR45	IR46	IR77	PSBRc82	PSBRc158	Samba	SHZ_2
LOC_Os01g01010.1.MSUv7.0 chr01:2449-3549	15	17	14	14	14	15	14	14
LOC_Os01g01010.2.MSUv7.0 chr01:2449-3549	16	18	15	15	15	16	15	15
LOC_Os01g01115.1.MSUv7.0 chr01:57658-58758	1	1	1	1	1	1	1	1
LOC_Os01g01140.1.MSUv7.0 chr01:68675-69775	2	2	2	2	2	2	2	1110
LOC_Os01g01369.1.MSUv7.0 Reversed: chr01:190131-191231	3	3	3	3	3	3	3	2
LOC_Os01g01380.1.MSUv7.0 chr01:192864-193964	4	4	4	4	4	4	4	3
LOC_Os01g01440.1.MSUv7.0 Reversed: chr01:220676-221776	87	80	84	78	77	76	79	83
LOC_Os01g01530.1.MSUv7.0 Reversed: chr01:277781-278881	5	5	5	5	5	5	5	4
LOC_Os01g01570.1.MSUv7.0 chr01:288696-289796	138	131	135	142	137	130	132	141
LOC_Os01g01620.1.MSUv7.0 Reversed: chr01:312964-314064	6	6	6	6	6	6	6	5

Figure S1. Example of the first 10 rows of EBE ranks predicted in the eight founders. The strain used in this example is *Xoc* BAI5. Row names include the gene ID, the chromosome of the gene and the gene coordinates. Each column is one of the MAGIC Founders. The cells represent EBE rank for each gene, organized by founder.

Table S3. Phenotypic data (mean lesion lengths) for MAGIC founders inoculated with *Xoc* BAI5, *Xoo* BAI3, and dH_2O. Lesion lengths are measured centimeters.

Cultivar	*Xoc* BAI5 (Bossa-Castro et al., 2018)	*Xoc* BAI5	*Xoo* BAI3 (Bossa-Castro et al., 2018)	*Xoo* BAI3
IR46	0.00	0.1	0.20	0.352
Fedearroz 50	0.00	0.1	8.90	7.610
IR77	1.613	4.627	6.98	7.191
SHZ-2	0.00	0.1	3.60	2.210
PSBRc82	2.280	5.571	12.41	11.988
Sambha	2.277	4.733	11.92	10.969
PSBRc158	2.987	6.622	14.45	10.533
IR45	1.590	6.025	20.66	19.422

Table S4. Candidate gene list organized by strain. The "Strain" column denotes which strain the candidate gene is a predicted target of. "Cultivar" lists the MAGIC founders for which there is an EBE present. "Gene-ID" gives the Nipponbare MSU7 gene name. "TALE" identifies which TALE of the given strain is targeting the candidate gene. "Gene Annotation" lists the expected function of the encoded protein based on recognizable domains.

Strain	Cultivar	Gene ID	TALE	Gene Annotation
Xoc BAI5	IR77, PSBRc82, PSBRc158, Samba	LOC_Os08g09680	TALE 14	conserved hypothetical protein
Xoc BAI5	Fedearroz 50, IR46, PSBRc82, SHZ-2	LOC_Os08g10310	TALE 15	SHR5-receptor-like kinase, putative, expressed
Xoc BAI5	Samba	LOC_Os11g40200	TALE 15	expressed protein
Xoc BLS256	IR45, IR46, IR77, PSBRc82, PSBRc158, Samba, SHZ-2	LOC_Os05g35870	TALE 21	expressed protein
Xoc BLS256	Fedearroz50, IR45, IR46, IR77, PSBRc82, SHZ-2	LOC_Os05g37010	TALE 22	expressed protein
Xoc BLS256	Fedearroz 50	LOC_Os05g37870	TALE 10	expressed protein
Xoc BLS256	IR45, PSBRc158, Sambha, SHZ-2	LOC_Os05g39100	TALE 16	hypothetical protein
Xoc BLS256	PSBRc82	LOC_Os11g43934	TALE 5	pentatricopeptide, putative, expressed
Xoc BLS256	Fedearroz 50, IR45, IR46, IR77, PSBRc158, Sambha, SHZ-2	LOC_Os11g45060	TALE22	NB-ARC domain containing protein, expressed
Xoc BLS256	IR45, IR46, IR77, PSBRc158, Sambha, SHZ-2	LOC_Os11g45080	TALE 21	expressed protein
Xoc BLS256	IR45, IR46, IR77, PSBRc158, Sambha, SHZ-2	LOC_Os11g45090	TALE 21	NB-ARC domain containing protein, expressed

Xoc BLS256	PSBRc82	LOC_Os11g45170	TALE 20	expressed protein
Xoo BAI3	IR46	LOC_Os01g55350	TALE 6	phosphoenolpyruvate carboxylase, putative, expressed
Xoo BAI3	Fedearroz 50, IR45, IR46, IR77, PSBRc82, Sambha, SHZ-2	LOC_Os05g01070	TALE 6	expressed protein
Xoo BAI3	IR46, IR77, SHZ-2	LOC_Os08g37210	TALE 2	patatin, putative, expressed
Xoo BAI3	Fedearroz 50, IR46, IR77, PSBRc82, PSBRc158, Sambha, SHZ-2	LOC_Os11g12540	TALE 3	expressed protein
Xoo BAI3	Fedearroz 50, IR46, IR77, PSBRc82, PSBRc158, Sambha, SHZ-2	LOC_Os11g13410	TALE 6	mla1, putative, expressed
Xoo MAI134	Sambha, SHZ-2	LOC_Os01g25360	TALE 1	esterase, putative, expressed
Xoo PXO86	PSBRc82	LOC_Os05g01635	TALE 6	expressed protein
Xoo PXO86	IR77, PSBRc82, PSBRc158,	LOC_Os05g04000	TALE 7	expressed protein
Xoo PXO86	Fedearroz 50, IR45, SHZ-2	LOC_Os07g31430	TALE 7	phosphate-induced protein 1 conserved region domain containing protein, expressed
Xoo PXO99A	SHZ-2	LOC_Os01g65740	TALE 15	plant-specific domain TIGR01615 family protein, expressed
Xoo PXO99A	IR45	LOC_Os01g67710	TALE 16	hypothetical protein
Xoo PXO99A	Fedearroz 50, IR45, IR46	LOC_Os01g69100	TALE 17	expressed protein

Xoo PXO99A	PSBRc82	LOC_Os05g01635	TALE 5	expressed protein
Xoo PXO99A	IR45, PSBRc158	LOC_Os08g43960	TALE 5	carrier, putative, expressed
Xoo PXO99A	Fedearroz 50, IR45, IR46, IR77, PSBRc158, Sambha	LOC_Os10g04530	TALE 15	expressed protein
Xoo PXO99A	Fedearroz 50, IR45, IR46, IR77, PSBRc158, Sambha	LOC_Os10g04540	TALE 15	expressed protein
Xoo PXO99A	IR46, PSBRc158, Sambha, SHZ-2	LOC_Os10g04600	TALE 10	OsFBX359 - F-box domain containing protein, expressed
Xoo PXO99A	IR45	LOC_Os10g04674	TALE 15	disease resistance protein RPM1, putative, expressed
Xoo PXO99A	IR45	LOC_Os10g04700	TALE 12	OsFBX361 - F-box domain containing protein, expressed
Xoo PXO99A	PSBRc158	LOC_Os10g36860	TALE 9	CRS1/YhbY domain containing protein, expressed
Xoo PXO99A	PSBRc158	LOC_Os10g37090	TALE 18	expressed protein
Xoo PXO99A	IR45	LOC_Os11g29760	TALE 18	hypothetical protein
Xoo PXO99A	IR45	LOC_Os11g29870	TALE 13	WRKY72, expressed
Xoo PXO99A	Fedearroz 50, IR46, IR77, PSBRc82, PSBRc158, Sambha, SHZ-2	LOC_Os12g10340	TALE 5 and 7 (Identical TALEs)	NBS-LRR type resistance protein, putative, expressed
Xoo PXO99A	IR45	LOC_Os12g10410	TALE 12	NB-ARC/LRR disease resistance protein, putative, expressed

Xoo PXO99A	IR46, IR77, PSBRc82, PSBRc158, Sambha, SHZ-2	LOC_Os12g12090	TALE 3	expressed protein
Xoo PXO99A	Fedarroz 50, IR45	LOC_Os12g12370	TALE 9	outer envelope protein, putative, expressed
Xoo PXO99A	SHZ-2	LOC_Os12g12400	TALE 2	retrotransposon protein, putative, unclassified, expressed
Xoo PXO99A	IR45	LOC_Os12g16120	TALE 14	expressed protein
Xoo PXO99A	IR45	LOC_Os12g16260	TALE 8	expressed protein
Xoo PXO99A	Fedearroz 50, IR46, IR77, PSBRc82, PSBRc158, Sambha, SHZ-2	LOC_Os12g17230	TALE 10	expressed protein
Xoo PXO99A	Fedearroz 50, IR46, IR77, PSBRc82, PSBRc158, Sambha, SHZ-2	LOC_Os12g17660	TALE 18	expressed protein
Xoo PXO99A	IR77, PSBRc82, PSBRc158, Sambha, SHZ-2	LOC_Os12g17840	TALE 15	ubiquitin family protein, expressed
Xoo PXO99A	IR45	LOC_Os12g22680	TALE 10	histone H3, putative, expressed
Xoo PXO99A	IR45	LOC_Os12g23970	TALE 15	expressed protein
Xoo PXO99A	IR45	LOC_Os12g24490	TALE 17	zinc finger, C3HC4 type domain containing protein, expressed
Xoo PXO99A	IR45	LOC_Os12g25160	TALE 14	MATE, putative, expressed

APPENDIX B: GENE EXPRESSION OF PLANTS INOCULATED WITH *XOO* AND *XOC* STRAINS

Introduction

In this thesis, I describe a pipeline that generates promoters sequences and relies on TALvez and PrediTALE to predict EBEs of inputted *Xoo* and *Xoc* TALE repertoires (Erkes *et al.*, 2019, Perez-Quintero *et al.*, 2013). EBE Ranks compared across multiple varieties are then correlated with lesion lengths caused by the *Xo* strain in question. Strongly correlated genes are filtered further by looking for overlaps with resistance QTL. This pipeline successfully identified 49 candidate genes for seven *Xoo* and *Xoc* strains (three for *Xoc* BAI5; nine for *Xoc* BLS256; five for *Xoo* BAI3; one for *Xoo* MAI134; three for *Xoo* PXO86; and 29 for *Xoo* PXO99A). Ultimately, these are just educated predictions that must be confirmed *in vivo*. In this section we introduce our strategy and initial progress towards expression analysis to validate candidate genes of *Xoo* BAI3 and *Xoc* BAI5.

Progress

Quality of the 72 RNA samples was tested using the Agilent RNA TapeStation, they were shipped to BGI Genomics Co., Ltd. for sequencing. BGI has since confirmed that all samples are 'A' quality and have moved forward with the sequencing. When RNAseq data becomes available, we will identify differentially expressed genes (DEGs) between dH2O and the *Xo* inoculated plants using methods similar to those used by (Cohen & Leach, 2019). DEGs will then be compared to our list of candidate genes. While this experiment will help us validate our pipeline, it will also generate additional data about the other genes involved in defense response to *Xoo* BAI3 and *Xoc* BAI5 (genes outside of QTL and unrelated to TALE targeting).

Ultimately, this will give us a better understanding of the disease cycles and potential markers for resistance to use when breeding new rice varieties.

Methods

Plant materials and growing conditions

Seeds of eight MAGIC Indica founders (Fedearroz-50, IR45427-2B-2-2B-1-1, IR4630-22-2-5-1-3, IR77298-14-1-2-10, PSBRc82, PSBRc158, SHZ-2, Sambha Mahsuri + Sub1) (Table **S2**) were germinated in petri dishes on filter paper soaked with Maxim XL Fungicide at 25 ℃. Once the cotyledon emerged, plants were transplanted into soil filled pots in the greenhouse. The amount of time required for the cotyledon to emerge was different for each variety, so the start of germination was staggered to compensate. Plants were grown in the greenhouse with a daytime temperature of 29 ℃ and a nighttime temperature of 23 ℃. Humidity was maintained at an average of 85%. At 2 weeks past germination, with three leaves emerged, plants were chelated with ferrous sulfate, and then fertilized twice a week with Peters Excel 15-5-15 Cal-Mag (Scotts, 300 mg/L).

Inoculation of MAGIC founders

For inoculum preparation, *Xoo* BAI3 and *Xoc* BAI5 strains were incubated on PSA (peptone sucrose agar) for 72 h at 28 ℃ (Karganilla *et al.*, 1973). Single colonies were taken from each and re-streaked onto new PSA plates, and the plates were incubated for 24-48 h.

Inoculum was prepared by resuspending bacteria in sterile DI water and adjusting to an OD_{600} of 0.2 (10^8 CFU/mL).

Twenty-one-day old plants (four leaf stage) were inoculated with *Xoo* BAI3 and *Xoc* BAI5 suspensions, and an dH2O mock treatment. For *Xoc* BAI5, the youngest most extended leaf of four plants/founder was infiltrated six times in the center with 0.1-0.3 uL of inoculum using a 1 mL, needle-less syringe (Reimers & Leach, 1991). At 6 hpi, leaf tissue infiltrated with BAI5 was collected and place immediately into liquid nitrogen. Because we want to associate DEGs with disease phenotypes, the second most extended leaf of four plants/founder was infiltrated once with the infiltration site was positioned directly over the leaf's xylem. For *Xoo* BAI3, the youngest most extended leaf of four plants/founder was infiltrated six times same as above. The tip of the second most extended leaf of four plants/founder was then clipped using scissors dipped in inoculum for phenotyping (Kauffman *et al.*, 1973). For dH2O, the youngest most extended leaf of four plants/founder was infiltrated six times same as the two treatments above. The second most extended leaf of two plants were infiltrated using the same methods as above, while the second most extended leaf of the two remaining plants were clipped with scissors dipped in dH2O.

Lesion development and measurement

At 12 dpi (days post-inoculation) for infiltrated leaves, and 14 dpi for clipped leaves, leaves were collected and taped to transparencies. These were photographed and measurements of lesions were recorded using ImageJ (Schneider *et al.*, 2012). When calculating

lesions extended from infiltration sites, 0.4 cm was subtracted from the total lesion length (0.4 cm = the diameter of the infiltration site). *Refer to Figure 5 in the main section of the thesis.*

RNA extraction

Frozen tissue was lysed with beads in Qiagen tissue lyser. Leaf powder was then stored at -80℃.

RNA was extracted from leaf tissue of the eight founders inoculated with *Xoo* BAI3, *Xoc* BAI5, or dH2O using the RNeasy Plant Mini Kit from Qiagen. Three biological replicates were generated for each treatment for a total of 72 RNA samples (Table **1**). 50 uL of nuclease free H_2O was added to the final column in the procedure to elute RNA. Concentration and quality of eluted RNA were checked before the RNA was stored at -20℃.

The minimum concentration of RNA extracted was 218 ng/uL with a max of 907 ng/uL and a mean of 447 ng/uL. 260/280 ratios of all RNA samples exceeded 2.0, however 16 samples had 260/230 ratios under 1.0 indicating the presence of impurities. These 16 samples were cleaned using the RNA Clean & Concentrator-5 kit from Zymo Research. After RNA-cleanup of these samples the minimum 260/230 ratio was 1.7. Quality of RNA was confirmed using the Agilent RNA TapeStation and all sample RINs were greater than 5.

Table **S5**

Below is the table submitted to BGI Genomics Co., Ltd for the sequencing of the 72 samples. This table includes information about the sample number and the corresponding sample name (variety and treatment). Additionally, the amount of RNA sent for sequencing and the 260/280 and 260/230 ratios for each sample are shown.

Sample Number	* Sample Name	* Species	* Quantity of Tubes	Concentration (ng/μl)	Volume (μl)	Total Quantity (μg)	OD260/280	OD260/230
1	IR45H2OA	Oryza sativa	1	80	25	2000	2.13	2
2	IR46H2OA	Oryza sativa	1	80	25	2000	2.06	2.33
3	IR77H2OA	Oryza sativa	1	80	25	2000	2.12	2
4	PSBRc82H2OA	Oryza sativa	1	80	25	2000	2.1	2.03
5	PSBRc158H2OA	Oryza sativa	1	80	25	2000	2.11	2.06
6	SambaH2OA	Oryza sativa	1	80	25	2000	2.07	2.22
7	SHZH2OA	Oryza sativa	1	80	25	2000	2.06	2.07
8	IR45H2OB	Oryza sativa	1	80	25	2000	2.09	2.09
9	IR46H2OB	Oryza sativa	1	80	25	2000	2.09	2.16
10	IR77H2OB	Oryza sativa	1	80	25	2000	2.08	1.74
11	PSBRc82H2OB	Oryza sativa	1	80	25	2000	2.09	2.4
12	PSBRc158H2OB	Oryza sativa	1	80	25	2000	2.05	2.11
13	SambaH2OB	Oryza sativa	1	80	25	2000	2.06	1.78
14	SHZH2OB	Oryza sativa	1	80	25	2000	2.07	2.33
15	IR45H2OC	Oryza sativa	1	80	25	2000	2.09	2.26
16	IR46H2OC	Oryza sativa	1	80	25	2000	2.06	2.26
17	IR77H2OC	Oryza sativa	1	80	25	2000	2.11	1.82
18	PSBRc82H2OC	Oryza sativa	1	80	25	2000	2.06	2.33
19	PSBRc158H2OC	Oryza sativa	1	80	25	2000	2.1	2.31
20	SambaH2OC	Oryza sativa	1	80	25	2000	2.07	1.87
21	SHZH2OC	Oryza sativa	1	80	25	2000	2.1	1.8
22	IR46BAI3A	Oryza sativa	1	80	25	2000	2.09	2.31
23	IR45BAI3A	Oryza sativa	1	80	25	2000	2.09	2.15
24	IR77BAI3A	Oryza sativa	1	80	25	2000	2.06	2.39
25	PSBRc158BAI3A	Oryza sativa	1	80	25	2000	2.09	2.2
26	PSBRc82BAI3A	Oryza sativa	1	80	25	2000	2.12	1.77
27	SambaBAI3A	Oryza sativa	1	80	25	2000	2.12	1.93
28	SHZBAI3A	Oryza sativa	1	80	25	2000	2.08	2.25
29	IR45BAI3B	Oryza sativa	1	80	25	2000	2.11	1.98
30	IR46BAI3B	Oryza sativa	1	80	25	2000	2.11	2.09
31	IR77BAI3B	Oryza sativa	1	80	25	2000	2.08	2.36
32	PSBRc82BAI3B	Oryza sativa	1	80	25	2000	2.1	2.35
33	PSBRc158BAI3B	Oryza sativa	1	80	25	2000	2.11	1.88
34	SambaBAI3B	Oryza sativa	1	80	25	2000	2.1	2.31
35	SHZBAI3B	Oryza sativa	1	80	25	2000	2.09	2.33
36	IR45BAI3C	Oryza sativa	1	80	25	2000	2.09	1.95
37	IR46BAI3C	Oryza sativa	1	80	25	2000	2.1	2.28
38	IR77BAI3C	Oryza sativa	1	80	25	2000	2.11	2.11
39	PSBRc82BAI3C	Oryza sativa	1	80	25	2000	2.08	2
40	PSBRc158BAI3C	Oryza sativa	1	80	25	2000	2.09	2.24
41	SambaBAI3C	Oryza sativa	1	80	25	2000	2.12	2.29
42	SHZBAI3C	Oryza sativa	1	80	25	2000	2.07	2.28
43	IR77BAI5A	Oryza sativa	1	80	25	2000	2.05	2.33
44	IR45BAI5A	Oryza sativa	1	80	25	2000	2.09	2.32
45	IR46BAI5A	Oryza sativa	1	80	25	2000	2.07	1.89
46	SHZBAI5A	Oryza sativa	1	80	25	2000	2.06	2.34
47	PSBRc82BAI5A	Oryza sativa	1	80	25	2000	2.12	1.9
48	PSBRc158BAI5A	Oryza sativa	1	80	25	2000	2.09	2.11
49	SambaBAI5A	Oryza sativa	1	80	25	2000	2.1	1.81
50	IR45BAI5B	Oryza sativa	1	80	25	2000	2.07	2.32
51	IR45BAI5B	Oryza sativa	1	80	25	2000	2.1	1.96
52	IR77BAI5B	Oryza sativa	1	80	25	2000	2.05	2.02
53	PSBRc82BAI5B	Oryza sativa	1	80	25	2000	2.07	2.24
54	PSBRc158BAI5B	Oryza sativa	1	80	25	2000	2.09	2.3
55	SambaBAI5B	Oryza sativa	1	80	25	2000	2.08	2.32
56	SHZBAI5B	Oryza sativa	1	80	25	2000	2.06	2.03
57	IR45BAI5C	Oryza sativa	1	80	25	2000	2.07	2.05
58	IR46BAI5C	Oryza sativa	1	80	25	2000	2.06	2.2
59	IR77BAI5C	Oryza sativa	1	80	25	2000	2.08	1.71
60	PSBRc82BAI5C	Oryza sativa	1	80	25	2000	2.03	2.07
61	PSBRc158BAI5C	Oryza sativa	1	80	25	2000	2.08	2.05
62	SambaBAI5C	Oryza sativa	1	80	25	2000	2.08	2.11
63	SHZBAI5C	Oryza sativa	1	80	25	2000	2.06	1.86
64	FEDH2OA	Oryza sativa	1	80	25	2000	2.08	2.35
65	FEDH2OB	Oryza sativa	1	80	25	2000	2.09	2.12
66	FEDH2OC	Oryza sativa	1	80	25	2000	2.09	2.31
67	FEDBAI3A	Oryza sativa	1	80	25	2000	2.11	1.82
68	FEDBAI3B	Oryza sativa	1	80	25	2000	2.08	2.34
69	FEDBAI3C	Oryza sativa	1	80	25	2000	2.1	1.73
70	FEDBAI5A	Oryza sativa	1	80	25	2000	2.09	2.32
71	FEDBAI5B	Oryza sativa	1	80	25	2000	2.09	2.29
72	FEDBAI5C	Oryza sativa	1	80	25	2000	2.09	2.3