EVOLUTION OF AN EFFECTOR FAMILY AND CHANGES IN HOST INTERACTION REPERTOIRES

A Thesis

By

NICHOLAS GREGORY FARMER

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Chair of Committee,	Daniel Ebbole
Committee Members,	Libo Shan
	Junqi Song
	Heath Blackmon
Head of Department,	Won Bo Shim

May 2023

Major Subject: Plant Pathology

ABSTRACT

Previously our lab described a 21-member gene family of the rice blast fungus *Pyricularia oryzae* that were named host-adapted genes (HAGs). Most AVR/effector genes of *P. oryzae* are either unique or have few paralogs. The presence of such a high number of paralogous HAG effectors suggests the potential for both redundancy and diversification in effector function. Redundancy may allow for the loss of some gene family members without loss of virulence activity. Redundancy may also allow for more precise regulation of effector expression or adaptation to allelic variation of host targets. On the other hand, divergence can allow for expansion of effector target repertoires that can also lead to increased fitness. Closely related *Pyricularia* species contain orthologous gene family members. However, in many cases the sequence divergence of orthologs is as great as is found between paralogs. One view is that orthologs would display conservation of host target interactions and paralogs would display diversification. Here I have tested these assumptions using Yeast Two-Hybrid assays to identify candidate rice target proteins that interact with members of the HAG effector family. Putative targets identified via Yeast Two-Hybrid were cross tested with the paralogous effectors from *P. oryzae* as well as orthologous effectors from other closely related *Pyricularia* species and species of *Colletotrichum*, allowing us to define overlap in the target repertoires of these effectors. The presence of the HAG family in *Colletotrichum* and only some species of *Pyricularia* supports the possibility of their arrival in *Pyricularia* via Horizontal Gene Transfer (HGT). We further characterized some of these interactions via split luciferase assays. Additional screening was conducted using chimeric HAG effectors made from shuffled exons to study this potential mode of effector evolution. Finally, a new family of putative effectors was characterized with similar indicators of HGT from *Colletotrichum* to *Pyricularia*.

DEDICATION

This thesis is dedicated to my beautiful, amazing fiancée Reagan. Without her support and encouragement, I could have never completed this work. I love you more than words could ever fully describe and cannot wait to spend the rest of our lives together.

ACKNOWLEDMENTS

I would like to thank my committee chair Dr. Daniel Ebbole for his guidance, patience, and support throughout my graduate studies as well as my committee members Dr. Libo Shan, Dr. Junqi Song, and Dr. Heath Blackmon. I would also like to thank the faculty and staff of the Department of Plant Pathology and Microbiology. In addition, I would like to extend a special thanks to all my friends and colleagues in the department for their support over the years, especially: Dr. Brendan Mormile, Dr. Joe Vasselli, Elek Nagy, and Oli Bedsole. Finally, I would like to thank all my friends at Duddley's Draw for helping me through my graduate experience one drink at a time.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of my advisor Dr. Daniel Ebbole of the Department of Plant Pathology and Microbiology and members Dr. Libo Shan of the Department of Biochemistry and Biophysics, Dr. Junqi Song of the Department of Plant Pathology and Microbiology, and Dr. Heath Blackmon of the Department of Biology.

Funding Sources

Graduate studies were supported by teaching and research assistantships provided by the Department of Plant Pathology and Microbiology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Pyricularia (Magnaporthe) oryzae is a fungus of the phylum Ascomycota that is known worldwide as the causal agent of rice blast disease, the most economically important disease of rice. With rice being one of the world's most important sources of nutrition especially in Southeast Asia, rice blast is a longstanding global concern (Wilson & Talbot, 2009). Additionally, with the recent emergence of a population of *P. oryzae* that parasitizes wheat to cause wheat blast, the fungus has become an even greater threat to world food supplies (Cruz & Valent, 2017). *P. oryzae* also has pathogenic populations infecting barley, finger millet, foxtail millet, and other wild grasses (Wilson & Talbot, 2009).

The lifecycle of *P. oryzae* takes place through six stages beginning when the asexual spore called a conidium lands on and attaches to the hydrophobic rice leaf surface. Then a germ tube grows out from the conidium, this is a narrow thread like tube. In the third stage the end of the conidium differentiates into a bulbous structure referred to as an appressorium. The appressorium has the unique feature of having its cell wall lined with melanin which allows the appressorium to generate very high levels of turgor pressure. In the next stage, because of this great pressure, a small peg develops at the base of the appressorium on the surface of the leaf. This peg, called the penetration peg, penetrates the rice leaf surface, and allows for growth inside of the host. Once into the plant host the next stage of the life cycle begins as the fungus spreads via hyphal growth within the first cell and then to further cells throughout the host. The invasive hyphae of the fungus grow surrounded by host plasma membrane and move to further cells via regions of the plant cell wall rich in plasmodesmata. Finally, after 72-96 hours, lesions develop on the surface of the host and, with sufficient humidity, these lesions sporulate producing conidia

which can then be dispersed to new hosts allowing the cycle to begin anew. These lesions can occur on all above ground surfaces of the host (Wilson & Talbot, 2009).

In addition to being one of the most important plant pathogens, *P. oryzae* is also known as an important model system for studying plant-pathogen interactions (Wilson & Talbot, 2009). Plants and pathogens interact through several levels of plant immune response. The primary mechanism in this plant immune system involves recognition of Pathogen Associated Molecular Patterns (PAMPs) by plant Pattern Recognition Receptors (PRRs). PAMPs are components that are common to different classes of microbes. For example, flg22 is a peptide component of bacterial flagellin and its presence can be recognized by PRRs on the plant surface. The recognition of PAMPs by PRRs provides the basal level of plant immune response and is what prevents most microbes from being pathogenic to a given host. This basal plant immunity is referred to as Pattern Triggered Immunity (PTI) and is carried out by cascades of downstream signals (Dodds & Rathjen, 2010). Plant pathogens have evolved mechanisms to interfere with PTI by interfering with steps in the cascade that trigger the PTI response. One such mechanism used by both bacterial and fungal pathogens is the secretion of small proteins called effectors into the host that can work to interfere with host proteins important to the PTI response and cause Effector Triggered Susceptibility (ETS) (Dodds & Rathjen, 2010). These effectors can be secreted into the host via a variety of mechanisms. In *P. oryzae*, they are secreted via a structure called the Biotrophic Interfacial Complex (BIC). The BIC is a plant membrane-rich structure that is associated with the invasive hyphae that form immediately after entering the plant cell as described above. It is here that the fungus can translocate the effectors, typically characterized by their small size and richness in cysteine residues, into the host cytoplasm causing inhibition of plant defenses in the first invaded cell and beyond to nearby cells to prime them for subsequent

invasion (Fernandez & Orth, 2018). Plants have evolved a further immune response called Effector Triggered Immunity (ETI) in which specific resistance (R) proteins can recognize these pathogen effectors and trigger an immune response. In these cases, these effectors can be called avirulence (AVR) proteins and this R-AVR protein interaction leads to strong and specific resistance against strains of the pathogen which have the gene to code for this AVR protein. In *P. oryzae* there are several examples of R-AVR protein interactions, the best characterized example is the interaction between the fungal AVRPiz-t protein and the rice resistance protein Piz-t (Park et al., 2012).

The work that will be detailed in this thesis is based on work with a family of 21 *P. oryzae* putative effector genes called Host Adapted Genes (HAG) and the HAG effector proteins they encode for. The first HAG gene was initially described in a study of *P. oryzae* genome sequences conducted by Zhong et al. published in 2018 (Zhong et al., 2018). The initial gene, MGG_17227 (later named HAG1), was recognized to be a gene deleted in all members of one of the three main clades of *P. oryzae* described in the study. Further investigation of the gene revealed that it coded for a putative effector gene that was secreted into the host cell and was able to suppress BAX-mediated cell death in tobacco leaf infiltration assays.

HAG1 is one of a 21 member gene family in *P. oryzae* that code for effectors similar to MGG_17227 and this large group of related effector genes became known as the HAG gene family (Ebbole et al., 2021). The genes were termed Host Adapted Genes due to allelic diversity of the genes being greater between populations of *P. oryzae* adapted to specific hosts of the pathogen than within each host-adapted population (Ebbole et al., 2021). Several conserved amino acid motifs present across the 21 effectors allow them to be recognized as an effector family. These motifs include a cysteine rich motif, a GCDx(n)GR motif (where “x(n)” represents

a variable number of non-conserved amino acids), and a C-terminal FIGCA motif (Figure 1.1, Ebbole et al. 2021). Additionally, it was determined using florescence microscopy that both HAG1 and HAG4 localized to the BIC where they are secreted into the host cells providing more evidence that these are in fact a family of effector genes.

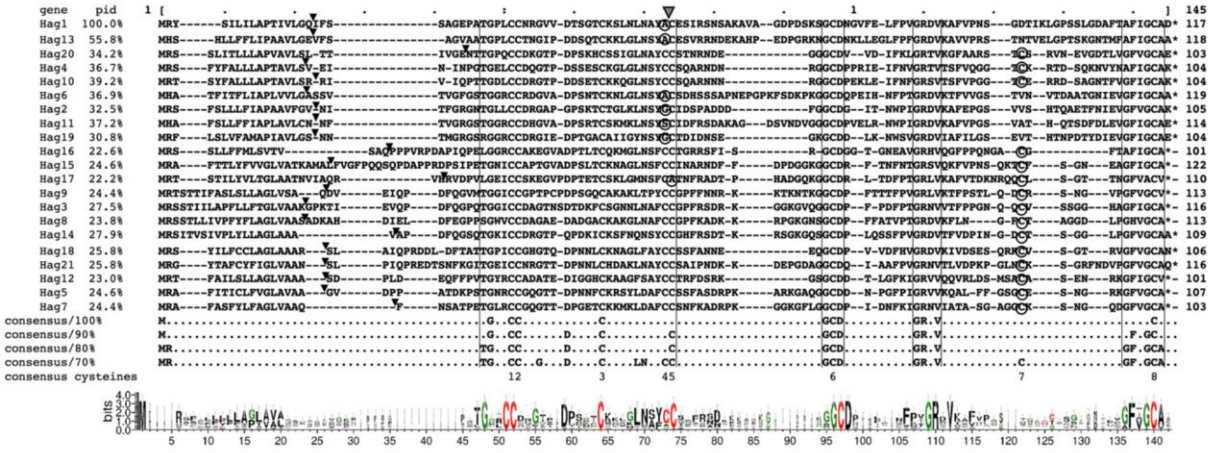


Fig. 1. Alignment of *Pyricularia oryzae* host-adapted gene (HAG) proteins. Sequence identity to Hag1p is shown (pid). Smaller arrowheads point to the predicted signal peptide cleavage sites. Indicated regions highlight a cysteine-rich, GCD tripeptide, GRxV, and C-terminal FI/VGCA motifs. Consensus sequences with 100 to 70% identity are indicated in the lines below the alignment with conserved cysteines numbered below. Circled residues highlight three types of proteins. The six-cysteine group has substitutions at position 4 (circled) and lack the cysteine at position 7 relative to those with eight cysteines (position 7 circled). Hag17p lacks cysteine 5 (circled alanine substitution) and has just seven cysteines. An intron in the genomic sequences lies precisely between the codons specifying cysteines 4 and 5 in all genes (large arrowhead above Hag1p sequence). A WebLogo graphic representation is at bottom.

As discussed above, plant pathogen effectors play important roles in parasitism, including countering plant immunity. However, investigation of the functional consequences of sequence divergence of fungal effectors during evolution is limited (Segretin et al., 2014). In addition to the 21 HAG family members in P. oryzae, many more family members found in other species of *Pyricularia* as well as cereal and grass pathogen species of *Colletotrichum* are found. The *Colletotrichum* members have yet to be described in published work and will only be discussed in this thesis in comparison of protein-protein interactions with *Pyricularia* HAG family members.

Most AVR/effector genes of *P. oryzae* are either unique or have few paralogs, making the HAG family interesting for the study of effector evolution. The presence of such a high number of

paralogous HAG effectors suggests the potential for both redundancy and diversification in effector function. Redundancy may allow for the loss of some gene family members without loss of virulence activity. This potential for redundancy is supported by presence/absence polymorphism of most members of the family in the rice infecting population. Redundancy may also allow for more precise regulation of effector expression or adaptation to allelic variation of host targets. Alternatively, divergence can allow for expansion of effector target repertoires that can also lead to increased fitness. With the large size of the family in *P. oryzae* and the presence of members in other species it is possible to robustly survey similarities/differences between orthologous and paralogous family members. One view is that orthologs would display conservation of host target interactions and paralogs would display diversification. We have tested these assumptions using Yeast Two-Hybrid assays to identify candidate rice target proteins that interact with members of the HAG effector family. Putative targets identified via Yeast Two-Hybrid were cross tested with the paralogous effectors from *P. oryzae* as well as orthologous effectors from other closely related Pyricularia species, allowing us to define overlap in the target repertoires of these effectors. Additionally, we have conducted split luciferase assays with some of these fungal effector- rice target interactions in tobacco to further verify the existence of these interactions in planta.

In Chapter 2, I further identified rice targets of HAG effectors by screening a rice yeast two-hybrid cDNA library. These findings provide further evidence that effectors typically interact with more than one host target protein. This also provides a larger set of host proteins to test for redundancy of protein-protein interaction among HAG effector orthologs and paralogs. This work provides an outline for future work to systematically investigate evolution of effector-target interactions.

In Chapter 3, I conducted a proof-of-concept study to show that interactions detected by yeast two-hybrid analysis could be validated by the split luciferase assay in planta. It was important to show that the detected yeast two-hybrid interactions were not specific to fusion proteins produced with GAL4 or any other aspect of the yeast system. By expressing proteins in tobacco, this demonstrated the ability to examine protein-protein interactions in plants and lays the groundwork for future studies to not only detect protein-protein interactions, but also examine the effect of the interaction on protein function. Specifically, I examined the protein kinase SnRK1 interaction with HAG proteins and this indicates it will be possible to examine the effect of SnRK1-HAG interactions on SnRK1 kinase activity in vivo.

In Chapter 4, I further examine evolution of HAGs on plant target interactions. HAG homologs of two *Colletotrichum* species and a *Cochliobolus* species were assessed for rice target interactions. Like *Pyricularia, Colletotrichum* species are hemibiotrophs, at first growing within living host tissue before evoking plant cell death after two to three days of host tissue colonization. I hypothesize HAG effectors play a similar role in parasitism and may have common host targets. The divergence of *Colletotrichum* and *Pyricularia* is thought to have occurred ~250-300 MYA (Hongsanan et al., 2017). *Cochliobolus heterostrophus* is known as a necrotroph, rapidly killing host cells, and then colonizing the dead/dying host tissue. Whether a HAG gene would be delivered into the host cell is unknown. *Cochliobolus* diverged from Pyricularia >300 MYA (Hongsanan et al., 2017). The sequence of the HAG gene from *C. heterostrophus* is quite poorly conserved with those of *Pyricularia* and *Colletotrichum*. Surprisingly, the *Cochliobolus* HAG gene does interact with at least one of the target proteins.

In Chapter 5, I examine the potential of creating chimeric HAG effector proteins made from two different members of the HAG family. This was of particular interest in line with one of the

purposes of this work being the study of effector evolution. In nature these types of chimeric effector proteins could be created via recombination or exon shuffling and give rise to novel effectors with different target specificities. This could be a driving factor in emerging pathogenicity.

In Chapter 6, I describe the findings of a project I led with undergraduate students to search for an additional protein family displaying the evolutionary characteristics of the HAG family. I hypothesized that since the HAG family was found exclusively in *Pyricularia* and *Colletotrichum*, but not in evolutionarily related saprophytic fungi, that other similar families might be found using a bioinformatics approach. This chapter therefore describes a new putative effector family that displays a discontinuous phylogenetic distribution, displays strong evidence of horizontal transfer, is primarily found in plant-associated fungi, and shows an in planta-specific gene expression pattern. This finding adds support to the hypothesis that the existence of these larger effector gene families may be more common than currently thought.

REFRENCES

Cruz CD, Valent B, 2017. Wheat blast disease: danger on the move. *Tropical Plant Pathology* **42**, 210-22.

Dodds PN, Rathjen JP, 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* **11**, 539-48.

Ebbole DJ, Chen M, Zhong Z, *et al.*, 2021. Evolution and Regulation of a Large Effector Family of Pyricularia oryzae. *Mol Plant Microbe Interact* **34**, 255-69.

Fernandez J, Orth K, 2018. Rise of a Cereal Killer: The Biology of Magnaporthe oryzae Biotrophic Growth. *Trends Microbiol* **26**, 582-97.

Hongsanan S, Maharachchikumbura SSN, Hyde KD, *et al.*, 2017. An updated phylogeny of Sordariomycetes based on phylogenetic and molecular clock evidence. *Fungal Diversity* **84**, 25-41.

Park CH, Chen S, Shirsekar G, *et al.*, 2012. The Magnaporthe oryzae effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* **24**, 4748-62.

Segretin ME, Pais M, Franceschetti M, *et al.*, 2014. Single amino acid mutations in the potato immune receptor R3a expand response to Phytophthora effectors. *Mol Plant Microbe Interact* **27**, 624-37.

Wilson RA, Talbot NJ, 2009. Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. *Nat Rev Microbiol* **7**, 185-95.

Zhong Z, Chen M, Lin L, *et al.*, 2018. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *ISME J* **12**, 1867-78.

CHAPTER II

IDENTIFYING HAG EFFECTOR INTERACTIONS WITH RICE TARGET PROTEINS VIA YEAST TWO HYBRID ASSAY

Introduction

Previous research on the HAG family revealed several characteristic features that support the hypothesis that the family functions as effectors (Ebbole et al., 2021, Chen et al., 2022). There are 21 family members in *P. oryzae*. A closely related species, *P*. sp. LS, a recently described *Pyricularia* species (Gomez Luciano et al., 2019), has orthologs of 20 of the 21 HAG genes found in *P. oryzae*. A slightly more distantly related species, *P. grisea*, also contains 20 HAG genes, however, only about half of them appear to be orthologs of the *P. oryzae/P*. sp. LS genes. The other *P. grisea* HAG homologs appear to have arisen by gene duplication and divergence (Chen et al., 2022). Interestingly, *P. grisea* has a HAG4 gene that is nearly identical to the *P. oryzae* HAG4 gene. *P. sp.* LS appears to lack HAG4, and it was proposed that *P. grisea* HAG4 was horizontally transferred to *P. oryzae* (Chen et al., 2022). Therefore, it is of interest to assess not only the interaction of *P. oryzae* HAG paralogs, but also to determine if orthologs found in other *Pyricularia* species share interaction with rice protein targets identified by the *P. oryzae* HAG protein interaction studies.

To further investigate the potential role of these effectors in contributing to pathogenesis in rice we identified putative interactions between the effectors and rice proteins. Yeast Two Hybrid (Y2H) analysis is a robust method used to study protein-protein interactions in yeast and we employed this method for our study (Bruckner et al., 2009, Wessling et al., 2014).

The basis for Y2H analysis is the two-domain structure of the GAL4 transcription factor. The two domains of this protein, an activation domain and a DNA binding domain, can be separated and when brought into close enough proximity still activate transcription of genes downstream of the GAL4 promoter. This ability to reconstitute the function to activate transcription allows for each domain to be attached to another protein and activate transcription if the two hybrid proteins interact with one another bringing the activation and binding domains back together. These hybrid proteins are referred to as bait (the fusion protein containing the GAL4 DNA binding domain) and prey (the fusion protein containing the GAL4 activation domain). The genes coding for these two hybrid proteins can be placed on separate vectors and then be delivered into yeast cells via transformation. Within the yeast strains, genes for steps in biosynthetic pathways for histidine and adenine have been deleted and new copies of the genes have been added downstream of a GAL4 promoter. A positive Y2H interaction is then detected via expression of the GAL4-inducible copies of the biosynthetic genes. In this chapter, I describe the further identification of putative HAG effector targets and characterization of the redundancy of interaction within the HAG family to develop hypotheses concerning the function of the HAG effectors in pathogenesis.

Materials and Methods

In this study we utilized the Matchmaker Yeast Two Hybrid system from Clonetech (Takara Bio USA) which provided the yeast lines, vectors, and recipes for selective media needed for the assay. Y2H experiments were carried out by two different protocols. The first using two different yeast lines AH109 and Y187, one with each yeast two hybrid vector transformed into it, that were mated to form diploid cells to bring the hybrid proteins together. In the second protocol one haploid yeast line, AH109, was used that was co-transformed with both necessary vectors. In all

other aspects these protocols were identical. The two yeast lines from Clonetech are described in Table 2.1

Table 2.1

Strain	Genotype	Reporters	Markers
AH109	*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2 : : GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3, MEL1 GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2, URA3::MEL1$_{UAS}$-MEL1$_{TATA}$-lacZ*	*HIS3, ADE2, lacZ, MEL1*	*trp1, leu2*
Y187	*MATα, ura3-52, his3-200, ade 2-101, trp 1-901, leu 2-3, 112, gal4Δ, met, gal80Δ, URA3 : : GAL1$_{UAS}$-GAL1$_{TATA}$-lacZ, MEL1*	*lacZ, MEL1*	*trp1, leu2*

The two vectors utilized in this study were pGADT7 (contains GAL4 activation domain) and pGBKT7 (contains GAL4 DNA binding domain). Their vector maps are shown in Figures 2.1 and 2.2 below.

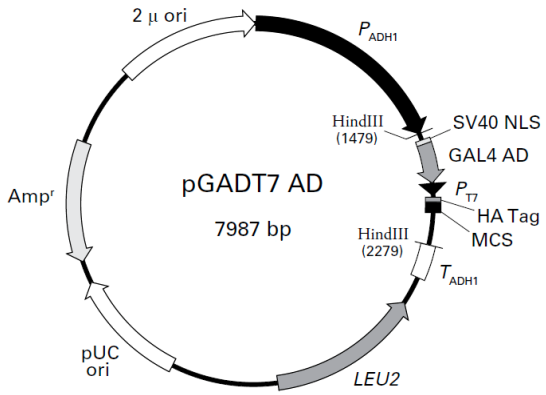


Figure 2.1 Plasmid map for pGADT7 AD vector Key: Amp^r - Ampicillin resistance marker, MCS - multiple cloning site, P_{T7} - T7 promoter, LEU2 - leucine selective marker, GAL4 AD - sequence for GAL4 activation domain (From manual for Matchmaker Yeast Two Hybrid, Clonetech)

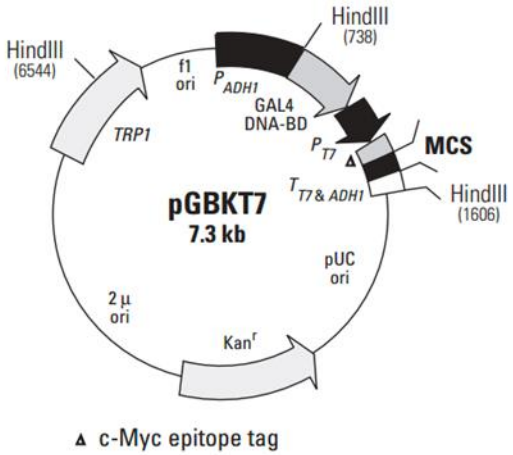


Figure 2.2 Plasmid map for vector pGBKT7 Key: Kan^r - Kanamycin resistance marker, MCS - multiple cloning site, P_{T7} - T7 promoter, TRP1 - tryptophan selective marker, GAL4 DNA-BD - sequence for GAL4 DNA binding domain (From manual for Matchmaker Yeast Two Hybrid, Clonetech)

The genotypes of the AH109 and Y187 yeast lines combined with these vectors allows us to use five types of media for selection purposes. For single vector transformations synthetic drop-out (SD) media lacking tryptophan, -trp, selects for presence of pGBKT7 that contains the TRP1 gene, and media lacking leucine, -leu, selects for presence of pGADT7 that contains the Leu2 gene. . SD media lacking both tryptophan and leucine (2-DO) was used to select for the presence of both vectors. Following growth on 2-DO media colonies were transferred to SD media lacking tryptophan, leucine, histidine, and adenine (4-DO) to select for positive protein-protein interactions between hybrid proteins that would activate transcription of the histidine and adenine biosynthetic genes driven by the GAL4 promoter. Yeast colonies that grew on 4-DO media were then replated onto SD media lacking tryptophan, leucine, and histidine, but with 30mM 3-amino-1,2,4-triazole (3-DO+30mM 3AT) which provides more stringent screening allowing us to focus on only the strongest positive interactions. These colonies were grown for 5 days unless otherwise specified. 3-amino-1,2,4-triazole is a competitive inhibitor of the product of the HIS3 gene imidazoleglycerol-phosphate dehydratase and inhibits biosynthesis of histidine, thus allowing only for those cells which are transcribing the constructed GAL4-inducible HIS3 gene at a high level (strong protein-protein interaction) to survive. All selective plate media was prepared fresh on the day of plating as I observed that yeast plated on media that had been stored long term at 4˚C were more variable in growth.

Presumptive positive rice targets that passed this screening were then replated to verify their growth and to obtain pure single colonies. These colonies were picked and used for yeast colony PCR. The picked colonies were added to a sterile 1.5 ml Eppendorf tube containing 500 ul of sterile water with 250 ul of 0.5 mm glass beads and disrupted using a bead beater for 1 minute to

lyse the yeast cells. Additionally, this lysate was then heated to 95 ˚C for 10 minutes to further lyse the cells. This crude cell lysate was used as the template for a PCR reaction with DNA primers for the T7 promoter sequence (forward) and the GAL4 activation domain (reverse). Reactions were run on a 1% agarose ethidium bromide gel to confirm the presence of a band prior to DNA sequencing (ETON Biosciences, Inc.). Returned nucleotide sequences of rice targets were analyzed using NCBI BLAST (Altschul et al., 1990) to identify protein sequence similarities and we built a catalog of these presumptive positive rice targets. Rice proteins that appeared in our catalog as multiple independent clones or from screenings with different HAG effectors were prioritized for further experiments. Additionally, identified rice target proteins that were already known to be important to pathogenesis based on prior studies in the literature were selected to be examined further.

To identify specific rice proteins that interact with a HAG effector being tested in Yeast Two Hybrid assays, a rice cDNA library in the pGADT7 vector was used (Chen et al., 2022). The cDNAs encoding rice prey proteins were inserted between the T7 promoter and the GAL4 activation domain sequence. A given HAG effector gene is inserted into pGBKT7 between the T7 promoter and the GAL4 DNA binding domain and enables the production of a HAG effector bait protein. This configuration of bait and prey was used for our initial screenings.

Individual yeast cell lines (prey only, bait only, and positive interactions with both bait and prey) were cultured and stored with 20% glycerol at -80˚C. Vector constructs were transformed into DH5α competent *E. coli* cells with selection using the Ampicillin (pGADT7) and Kanamycin (pGBKT7) resistance markers. DH5α cell lines were cultured and then stored in 20% glycerol at

-80˚C Vector construct stocks were harvested from cultures of these *E. coli* lines using the GeneJET Plasmid Purification Kit from Thermo Fisher Scientific and stocks were stored at -20˚C. Yeast cells were transformed with vector constructs via the lithium acetate-polyethylene glycol protocol (Agatep et al., 1998).

Later experiments used modified pGADT7 (pwy145) and pGBKT7 (pwy146) vectors obtained from the Song lab (Junqi Song, Texas A&M, personal communication) that are Gateway ready (Reece-Hoyes & Walhout, 2018). Full length sequences for selected rice target proteins were synthesized with Gateway compatible sequences facilitating transfer of these genes to the Y2H vectors described above. Additionally, some HAG effector genes were also synthesized for use with these Gateway vectors. Base Gateway ready vectors were maintained in DB.3.1 E. coli cells. Gateway reactions were conducted using the Gateway LR Clonease II Enzyme Mix (Invitrogen) following the manufacturer's instructions.

In order to screen our positive rice protein targets from the initial library screenings against other HAG effectors we employed a curing method to remove one bait vector from the yeast cells so that another could be added through transformation. This method simplified screening our targets with other HAG effectors. To cure the cells of the bait plasmid the yeast cells were grown overnight in liquid -leu media and then harvested and plated on -leu plate media. Colonies that turned a shade of pink/red were selected from these plates and spot plated on -leu and -trp plates. Colonies that grow on -leu plates but not -trp plates were cured of the bait plasmid and could be used for other screenings. This method takes advantage of the naturally occurring loss of the bait plasmid when the selective pressure for their presence (absence of tryptophan) was removed. The

pink/red color was used as an initial indicator of possible loss of the bait plasmid because it is caused by a buildup of an intermediate in the tryptophan biosynthetic pathway that would not occur if the *TRP1* gene from the bait plasmid was present.

The Y2H screen does have some potential for false positives if the bait protein is able to act similarly to the GAL4 activation domain in conjunction with the fused GAL4 DNA binding domain and cause activation of transcription of the selective genes without an interaction of two hybrid proteins. Previous work in our lab screened for this possibility by plating our yeast lines containing only our HAG effector bait constructs on 4-DO media. In doing this we found that HAG1, HAG18, and HAG20 bait vector yeast lines caused auto-activation and thus could not be used for screening against a prey library. Later experiments to assess HAGs that cause auto activation were conducted by swapping the effectors to the prey vector so that they are fused to the GAL4 activation domain instead of the GAL4 DNA binding domain. A second potential for false positives in this assay could be the fact that our prey library is a cDNA library and is made up of fragments of sequences and not complete coding regions. The interaction of these protein fragments with our HAG effectors may result in a positive but the full-length rice protein may not interact. To address this, we conducted later studies of the high priority targets using the full-length sequences. A third potential cause of false positives is that either the bait or the prey proteins interact with proteins non-specifically. Previous work in our lab identified that this was not the case for the HAG effectors, experiments testing for nonspecific interactions among our prey proteins are discussed later.

Results

The initial screenings of our rice library were conducted using two different HAG effectors, HAG4 and HAG10, as the bait proteins. We selected these two members of the HAG effector family for this screening because there was previous preliminary data for HAG4 (Ebbole et al., 2021) and because of the sequence similarity between the two effectors at the amino acid level. Rice target proteins identified from these screens and prioritized as described previously were cross screened with HAG4 from *Pyricularia grisea* (PgHAG4), HAG10 from a recently identified *Pyricularia* species isolated from *leersia* (referred to as *P.* spLS) (PlHAG10), HAG2, HAG7, and HAG13. These were selected for their varying levels of amino acid sequence divergence and because they represent both orthologs between species and paralogs in P. oryzae. This allows us to study conservation of rice target interactions across these two routes of evolution.

```
      %ID PoHAG4
PoHag2    43  MRSFSLLLFIAPAAVFGVNITFGRGNTGLLCCDRGAPGPSKTCTGLKLNSYGCIDSPADD  60
PoHag13   42  MHSH-LLFFLIPAAVLGEVFSAGVAATGPLCCTNGIPDDSQTCKKLGLNSYACESVRRND  59
PoHag7    39  MRAFASFYLFAGLVAAQFNS--ATPETGLRCCGQGTTDPGETCKKMKLDAFCCSNFKADR  58
PoHag4   100  MRSFYFALLLAPTAVLSVEI-NINPGTGELCCDQGTPDSSESCKGLGLNSYCCSQARNDN  59
PgHag4    88  MRSVYFALLLAPTAVLSYEI-TINPGTGELCCDHGTPDDSKTCKRLGLNSYCCSQARNNN  59
PoHag10   73  MRTSYFALLLAPTAVLSRRI-VIQPTTGDLCCDRGTPDDSETCKKQGLNSYCCSQARNNN  59
PlHag10   69  MRVSYFALLLAPTAVFSRRI-ELRPGTGDLCCDQGTPDASETCTKQGMHSFCCSQARNKN  59
                                          12          3        45

PoHag2        D---------FGGCDGI---TNWPIGRDVKAFEPGSVVSH--TQAETFNIEVGFVGCAK  105
PoHag13       EKAHPEDPGRKNGCDNKLLEGLFPFGRDVKAVVPRSTNTVELGPTSKGNTMFAFIGCAA  118
PoHag7        PK---GGKGFLGGCDPI---DNFKIGRNVIATASGAGGCK-------SNGQDGFVGCA-  103
PoHag4        ----------RGGCDPPRI-EIFNVGRTVTSFV-QGGTCK--RTDSQKNVYNAFIGCAK  104
PgHag4        ----------RGGCDPPRI-EIFNVGRTVTSYV-SGGTCK--RTDSEKNVYNAFIGCAK  104
PoHag10       ----------RGGCDPEKL-EIFNFGRSVTSFV-PGGTCE--RRDSAGNTFVGFIGCAK  104
PlHag10       ----------RGGCEPSKL-EIFNVGRTVTSFV-EGGTCE--RVDSAGNTFVGFVGCAK  104
                           6                        7                8
```

Figure 2.3 – Protein Sequence Alignment of HAG Effectors

Alignment of all HAG effectors used to screen rice target proteins. Red residues highlight conserved motifs, percent identity to HAG4 is listed after the name of the HAG effector, numbers below the alignment denote the conserved cysteine residues.

From the catalog of targets (Tables 2.2 and 2.3) we selected JMJ715, bZIP39, and ILR3 for further cross screening because of their multiple occurrences in the catalog. SnRK1 was selected for further cross screening because of both its multiple occurrences and its established relevance in pathogenesis in the literature.

Figure 2.4 depicts the results of this screen on the final selective media 3-DO+30mM 3AT. This screening found that the ILR3 and JMJ715 rice clones interacted with HAG4, HAG10, PgHAG4, PlHAG10, HAG2, HAG7, and HAG13 in yeast cells. We also found that the bZIP39 rice clone interacted with HAG4, HAG10, PgHAG4, PlHAG10, and HAG2, and the SnRK1 clone interacted with only HAG4, HAG10, PgHAG4, and PlHAG10. In this figure and all subsequent Y2H figures the protein listed on the left has been expressed as an AD-clone (vector pGADT7) and the protein listed on the right is expressed as a BD-clone (vector pGBKT7) unless otherwise specified. All Y2H experiments used the Lam and p53 as negative and positive controls interacting with the T-antigen plated at the same time with the same batch of media. Additional positive and negative controls were used as described.

Table 2.2 – Catalog of rice targets from HAG10 library screens

Best BLAST Hit	PCR Product Size
4-alpha-glucanotransferase DPE2-like	1277
bZIP transcription factor 39-like	1191
bZIP transcription factor 39-like	1021
bZIP transcription factor 39-like	1078
bZIP transcription factor 39-like	1247
calmodulin-binding transcription activator 3	927
chaperone protein dnaJ A7A, chloroplastic-like	1152
chaperone protein dnaJ A7A, chloroplastic-like	1147
chromatin structure-remodeling complex protein SYD	1292
chromatin structure-remodeling complex protein SYD	1262
chromatin structure-remodeling complex protein SYD	1271
chromatin structure-remodeling complex protein SYD	1238
F-box/kelch-repeat protein At1g55270	870
lysine-specific demethylase JMJ25	996
lysine-specific demethylase JMJ25	982
metallothionein 2b-like protein (MT2bL)	860
metallothionein 2b-like protein (MT2bL)	860
monothiol glutaredoxin-S12, chloroplastic-like	1095
OsMYC1 mRNA for bHLH-HALZ myc like protein	1057
OsMYC1 mRNA for bHLH-HALZ myc like protein	942
PITH domain-containing protein At3g04780	856
putative GEM-like protein 8	976
Rad21/Rec8-like protein mRNA	1267
serine/threonine protein kinase OSK1-like	1203
serine/threonine protein kinase OSK4-like	974
SUMO-activating enzyme subunit 1A	927
universal stress protein PHOS32	604

This table lists the identity of the best BLAST hit for all of the rice targets that we sequenced that were positives on initial 4-DO media screens with HAG10. It is important to note that the JMJ25 target represents JMJ715, the OsMYC targets represent ILR3, and OSK1-like represents SnRK1.

Table 2.3 – Catalog of rice targets from HAG4 library screens

Best BLAST Hit	PCR Product Size
AP-1 complex subunit gamma-2	1065
binding partner of ACD11 1	1211
bZIP transcription factor 39-like	1100
bZIP transcription factor 39-like	1219
F-box/kelch-repeat protein At1g55270	874
GTPase activating protein 1-like	976
high chlorophyll fluorescence phenotype 173	1226
lysine-specific demethylase JMJ25	1229
lysine-specific demethylase JMJ25	884
metal tolerance protein 3-like	784
metal tolerance protein 3-like	787
OsMYC1 mRNA for bHLH-HALZ myc like protein	872
phosphoglucomutase	709
probable protein phosphatase 2C 32	552
ROOT HAIR DEFECTIVE 3 homolog 2-like	708
RPM1-interacting protein 4	609
serine/threonine protein kinase OSK1-like	1183
ubiquitin-like domain-containing protein CIP73	1040
universal stress protein PHOS34	824
Wuyujing 3 HR-like lesion-inducer family protein	1161

This table lists the identity of the best BLAST hit for all the rice targets that we sequenced that were positives on initial 4-DO media screens with HAG4. It is important to note that the JMJ25 target represents JMJ715, the OsMYC targets represent ILR3, and OSK1-like represents SnRK1.

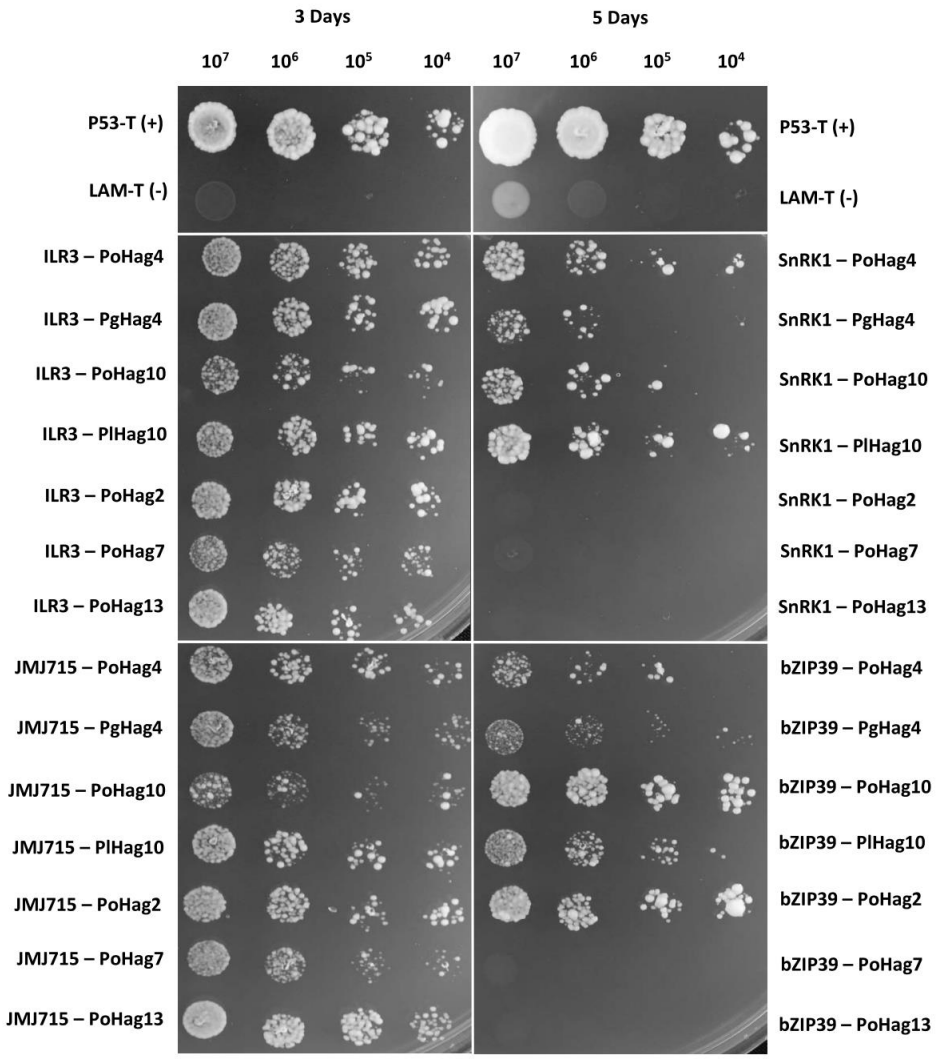


Figure 2.4 Interactions of Po paralogs and orthologs with identified rice proteins

Results for cross screening of interactions between the top four rice targets from our yeast two hybrid catalog created from initial library screening using HAG4 and HAG10 as bait. Further cross screening was conducted using orthologous and paralogous HAG effectors. p53-T and LAM-T are shown at the top representing the positive and negative control respectively. Spot plating was done on 3-DO+30mM 3AT media at dilutions of 10^7, 10^6,10^5, and 10^4 cells per mL.

In addition to cross screening this selection of HAG effectors with the clones we identified from the original library screening we also screened these HAG effectors against the full-length

versions of SnRK1 and JMJ715 from rice (Figure 2.5). These two targets were of particular interest for further experiments and as such we wanted to know if our initial Y2H results were replicable with full length proteins or if they might be artifacts of the initial screening using a library of protein fragments. The results for both SnRK1 and JMJ715 remained consistent when using the full-length proteins.

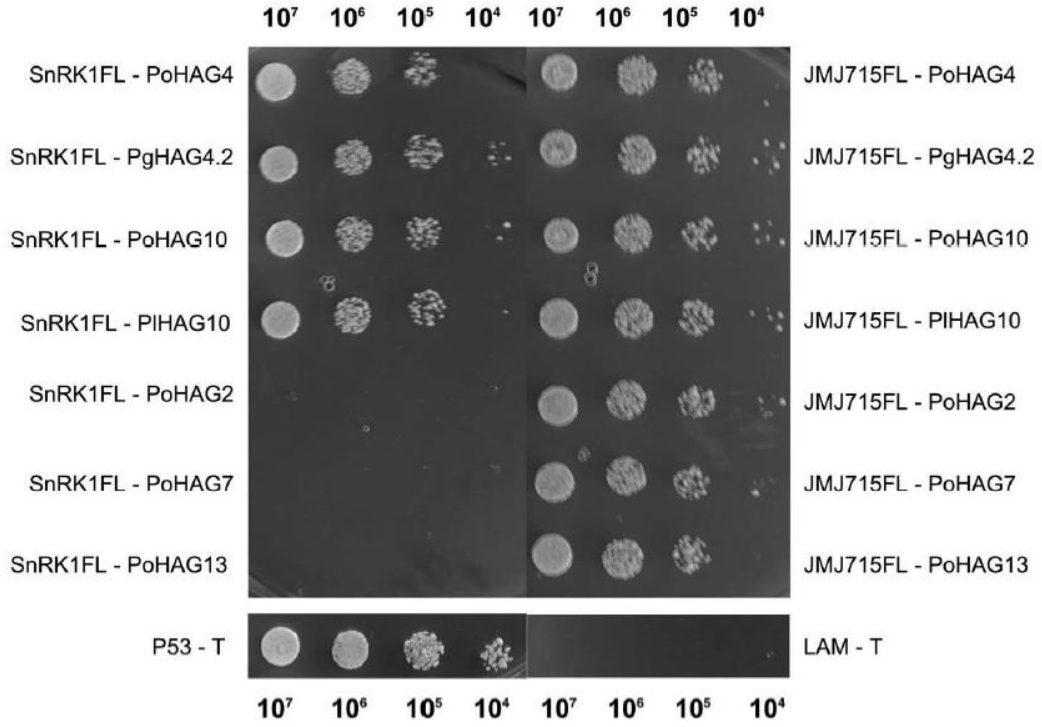


Figure 2.5 Testing of full length SnRK1 and JMJ715

Results from cross screening using synthesized versions of the full-length SnRK1 and JMJ715. All conditions are identical to experiments shown in Figure 1.3 otherwise.

A small screening of the library was conducted using HAG2 and we identified one target, Synamptotagmin-2, that we felt was of interest to use in cross screening with other HAGs due to its potential association with movement of effectors cell to cell during infection by some pathogens (Li et al., 2021, Liu et al., 2021). For this screening the same set of HAG effectors were used as in the previous two experiments. HAG4, HAG10, PgHAG4, PlHAG10, and HAG2

all interacted positively with Synaptotagmin-2 whereas HAG7 and HAG13 did not interact (Figure 2.6)

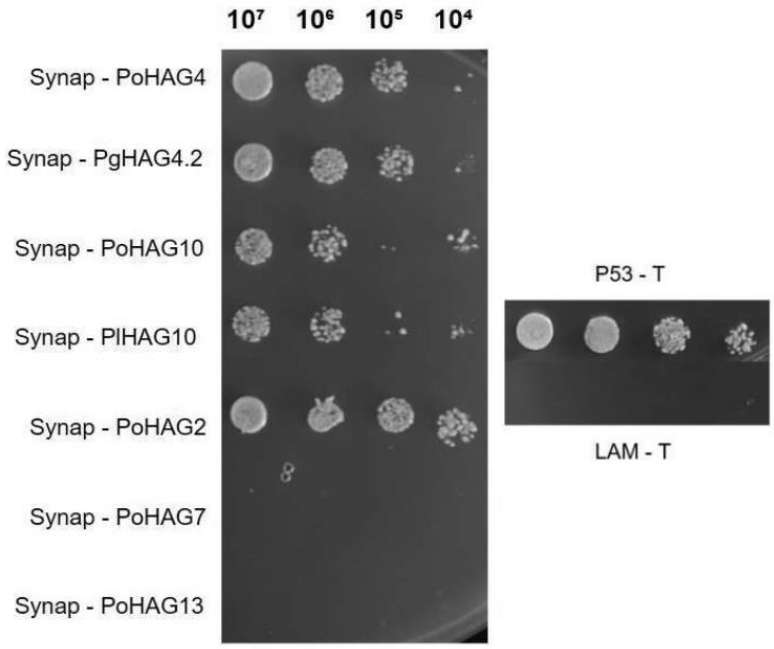


Figure 2.6 Synaptagmin-2 identified by HAG2, and interaction with other family members.

Results from screening Synaptotgmin-2 against the set of HAG effectors previously tested in Figures 2.4 and 2.5, all conditions were identical to these two previous screenings.

The finding of SnRK1 as a target for HAG interaction is of particular interest because reduction of SnRK1 transcript levels in rice by RNA interference leads to enhanced susceptibility to P. oryzae (Filipe et al., 2018). In this regard, SnRK1 is known to interact with a number of proteins in rice, including JMJ705, and JMJ705 also in implicated in plant defense against pathogens (Li et al., 2013). JMJ705 is a jumonji protein characterized by a jmjC domain that functions as a Histone lysine demethylase. This jumonji family of proteins containing jmjC domains also includes JMJ715, one of our initial rice targets. Figure 2.7 is an alignment of JMJ715 and JMJ705. These proteins are different classes of jmjC domain proteins with little primary sequence identity.

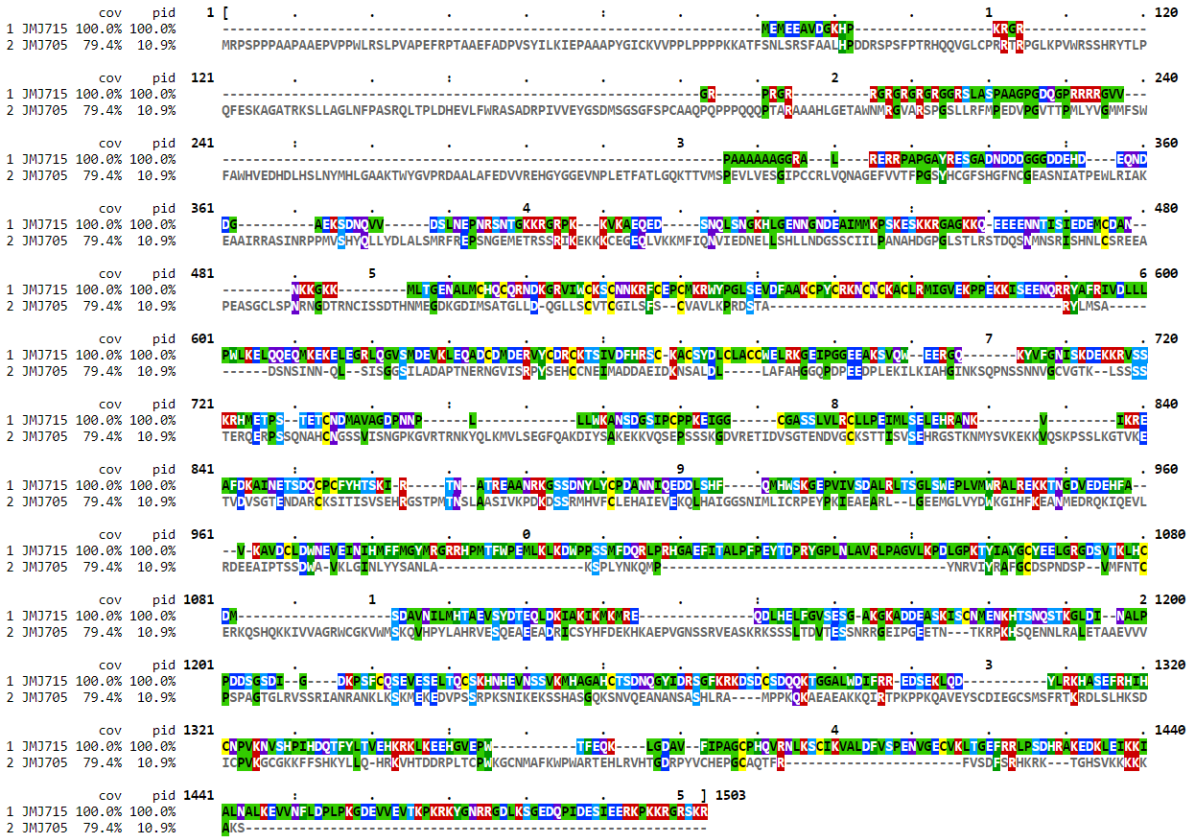


Figure 2.7 – Alignment of JMJ715 and JMJ705

A) This alignment of the rice target we obtained from the library screen JMJ715 and another member of the same protein family JMJ705 shows that JMJ705 only has 10.9% identity to JMJ715. B) The jmjC domains are boxed. The jmjC domains are aligned separately. Alignment generated in Mview (Brown et al., 1998)

Lysine methylation and demethylation control chromatin structure and interact with other proteins in complexes associated with chromatin. AvrPiz-t interacts with different rice proteins that themselves interact with each other (Tang et al., 2017). Therefore, we had a particular interest to screen the catalog for other proteins associated with histone lysine methylation. Di19 is an Arabidopsis protein associated with histone lysine methylation (Shen et al., 2021). A rice Di19 homolog, Di19-5 was preliminarily identified in a screen using HAG4 by our colleague in China. Therefore, I sought to validate this interaction with HAG4 and other HAG proteins and

with two other rice Di19 paralogs, Di19-6 and Di19-X. Figure 2.8 shows an alignment of the three Di19 proteins.

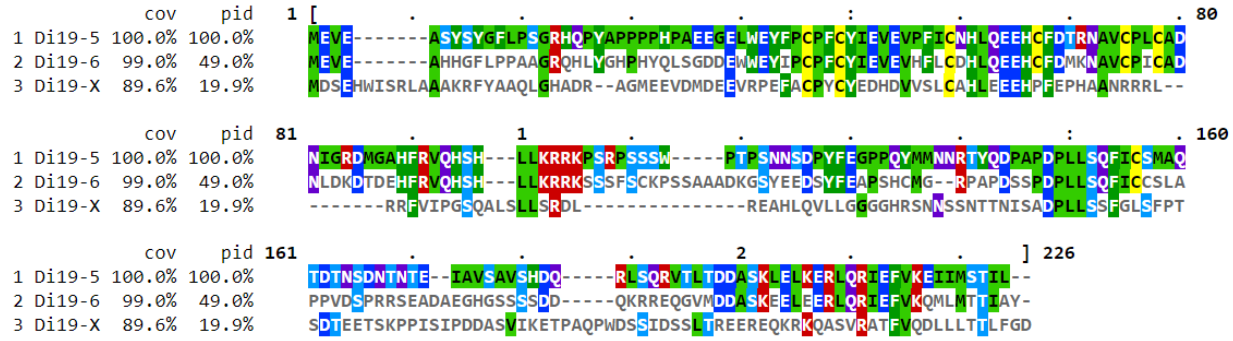


Figure 2.8 – Alignment of Di19-5, Di19-6, and Di19-X

This alignment of the three drought induced (Di) family proteins show that these Di19-5 and Di19-6 proteins only share ~20% sequence similarity with Di19-X whereas Di19-5 and Di19-6 share 49% sequence identity. Alignment generated in Mview (Brown et al., 1998)

I found that Di19-5 and Di19-6 interacted with HAG4, HAG10, PgHAG4, and PgHAG10 but not with HAG2, HAG7 or HAG13 (Figure 2.9). Our screening of Di19-X did not find any HAGs that interacted (Figure 2.10).

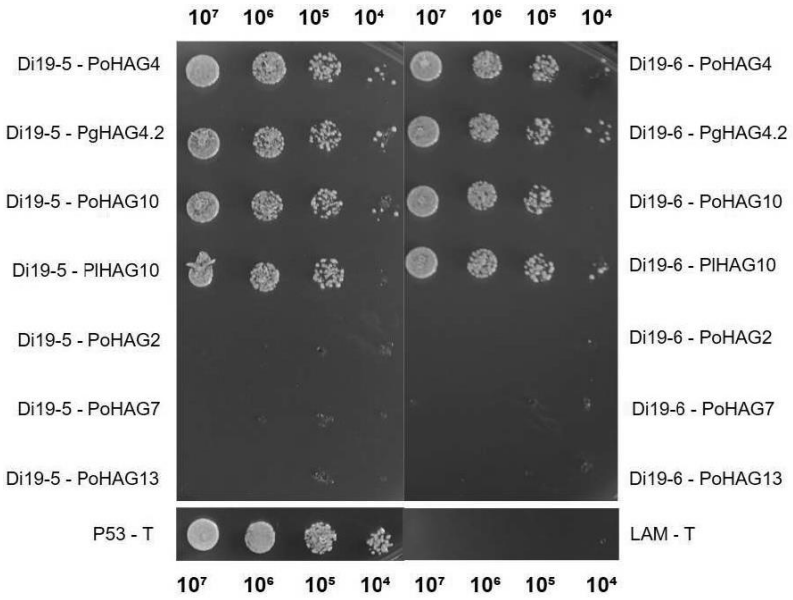


Figure 2.9 Testing additional targets related to histone methylation

Cross screening of the select subset of HAG effectors with Di19-5 and Di19-6, conditions were identical to all previously discussed cross screenings.

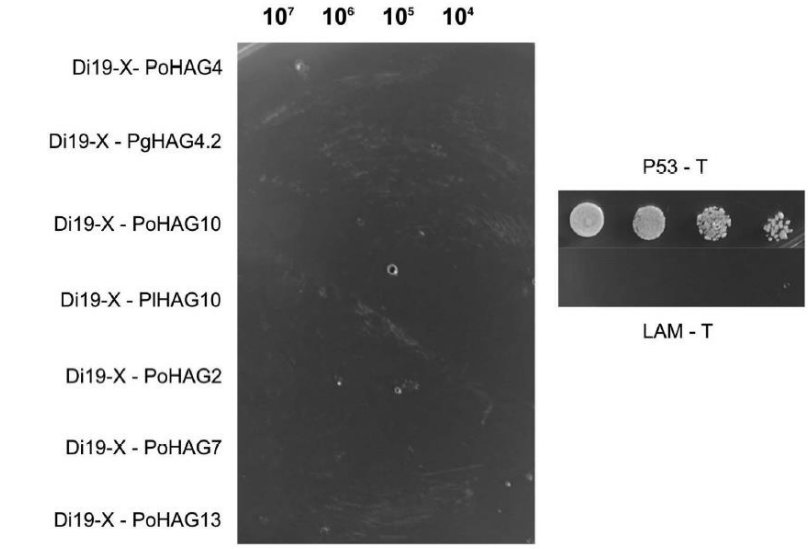


Figure 2.10 Defining family interaction with Di19-X

Cross screening of the select subset of HAG effectors with Di19-X, conditions were identical to all previously discussed cross screenings.

From the literature we learned that JMJ715 is a part of a large family jumonji (JmjC) domain containing proteins and that one such member, JMJ705, is known to interact with SnRK1 and this interaction is linked to defense response to other plant pathogens (Qian et al., 2019). As, such, we ordered a synthetic fragment of JMJ705 containing the JmjC domain (amino acids 18-392) and screened our subset of HAG effectors used in previous experiments. We did not find any positive interactions in this screening (Figure 2.11).

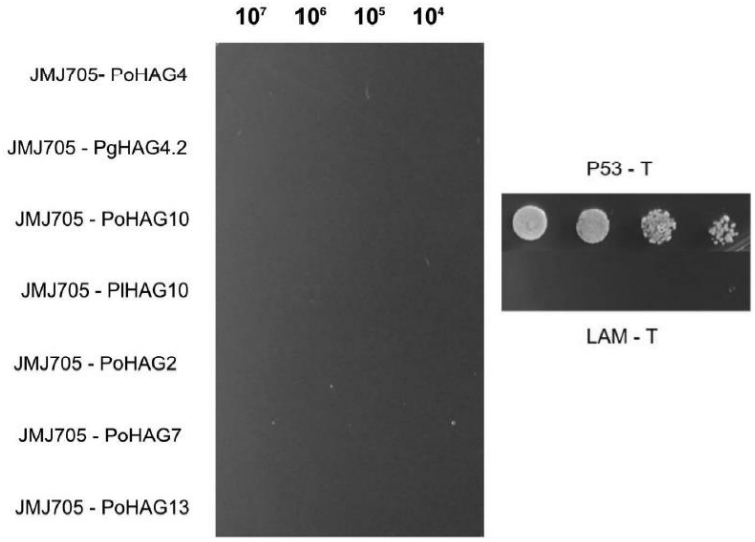


Figure 2.11 Defining HAG family interactions with JMJ705

Screening of JMJ705 fragment containing the JmjC domain against the subset of HAG effectors used previously, Conditions are identical to previous result platings.

As has been discussed previously, one potential pitfall of the yeast two hybrid assay is the potential for false positives. We screened our rice target proteins for non-specific protein interactions by screening them against another well studied effector from *P. oryzae* known as AVRPiz-t (figures 2.12 and 2.13). APIP5 is a known target of AVRPiz-t and was used for the positive control (Park et al., 2012, Zhang et al., 2022).

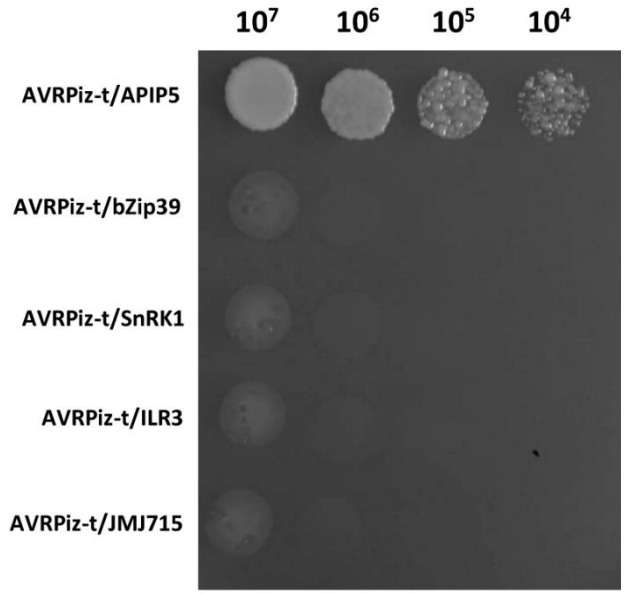


Figure 2.12 Screening the initial rice targets for non-specific interactions

Screening of the four rice target proteins from our initial screening for non-specific protein interactions with AVRPiz-t. The AVRPiz-t and APIP5 interaction was used as a positive control and 3DO+30mM 3AT media was used as in previous screenings with 5 days of growth.

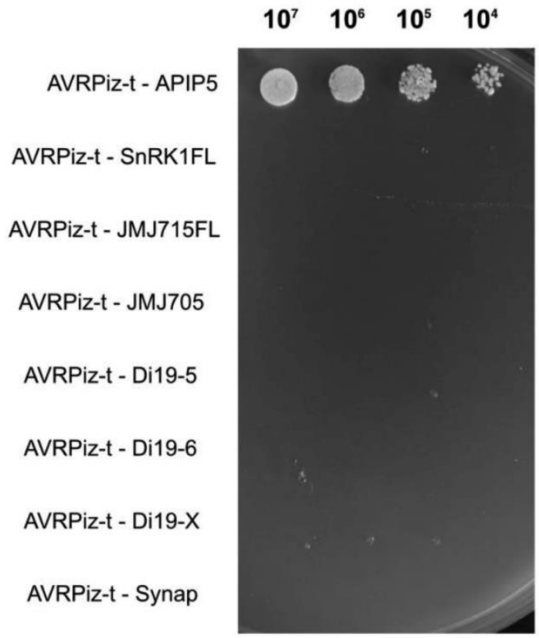


Figure 2.13 Screening of other rice targets for non-specific interactions

Screening of all synthetic genes used in previously discussed experiments and Synaptotagmin-2 from the small HAG2 library screening experiment for non-specific interaction with AVRPiz-t. The AVRPiz-t and APIP5 interaction was used as a positive control and 3DO+30mM 3AT media was used as in previous screenings.

Existing data from literature indicated that SnRK1 interacts with itself to form dimers and SnRK1 interacts with JMJ705 (Wang et al., 2021). As such we also screened all target proteins from the previous experiments against each other to determine if any of them interacted with each other, data that assisted us in forming a potential model for HAG interactions in rice and their potential to contribute to the diseased state in rice (Figures 2.14, 2.15, 2.16, 2.17, 2.18, 2.19, 2.20, 2.21, 2.22 and 2.23). From these experiments we learned following about yeast two hybrid interactions: 1) The SnRK1 fragment dimerizes in Y2H assays 2) The full-length SnRK1 dimerizes 3) The fragment and the full length SnRK1 dimerize 4) SnRK1 fragment and full-

length interact with JMJ715 fragment and JMJ715 full-length 5) SnRK1 fragment and full-length interact with JMJ705 jmjC domain , as would be expected from existing data indicating they interact and have a role in defense. All other screenings failed to detect SnRK1 interaction with other rice protein candidates.

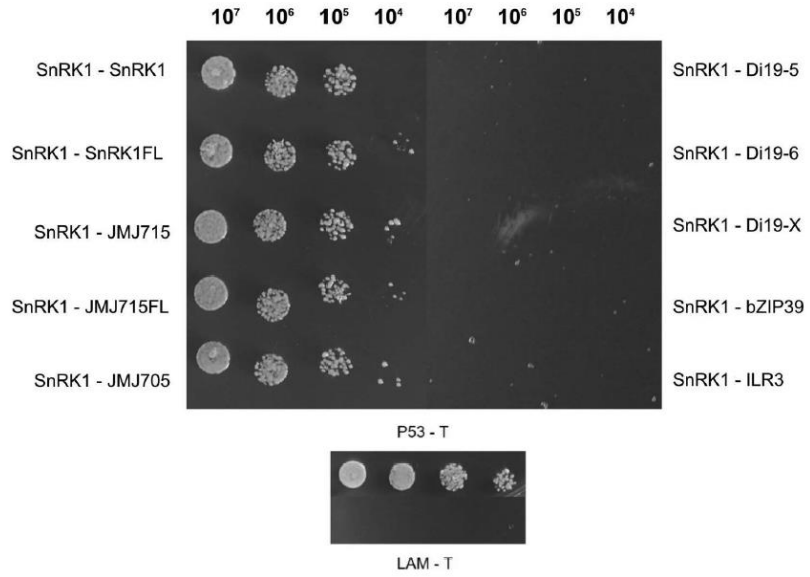


Figure 2.14 Screening of SnRK1 against other rice targets

Screening of SnRK1 fragment from initial library screenings against three other library targets and synthetic targets, plate media is 3-DO+30mM 3AT.

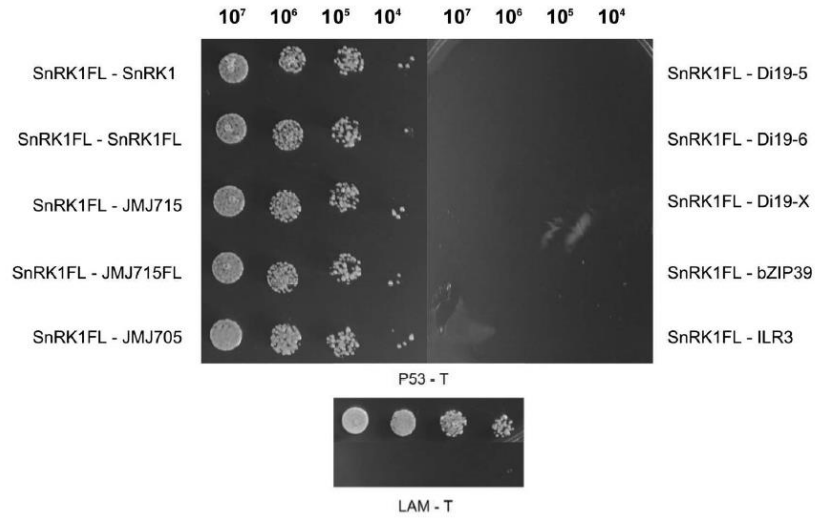


Figure 2.15 Screening of SnRK1 full-length protein against other rice targets

Screening of SnRK1 full-length protein against four library targets and synthetic targets, plate media is 3-DO+30mM 3AT.

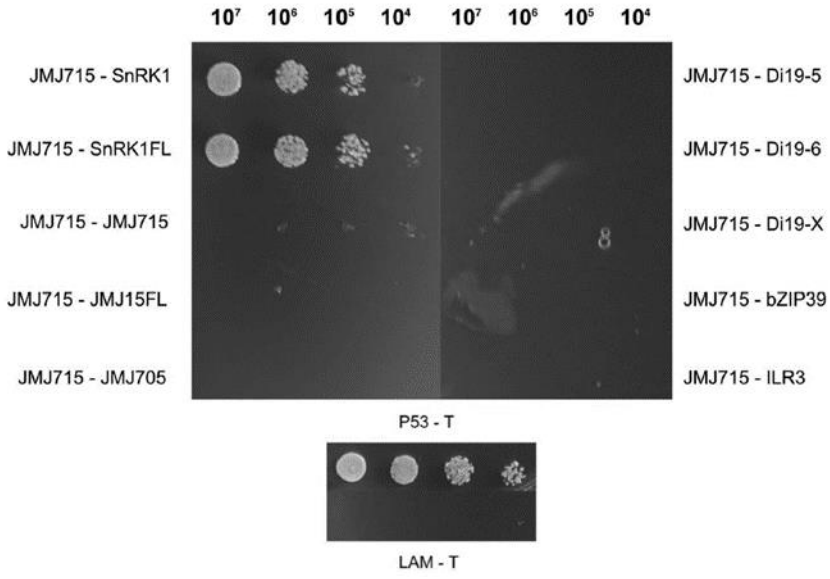


Figure 2.16 Screening of JMJ715 against other rice targets

Screening of JMJ715 fragment from initial library screenings against three other library targets and synthetic targets, plate media is 3-DO+30mM 3AT.

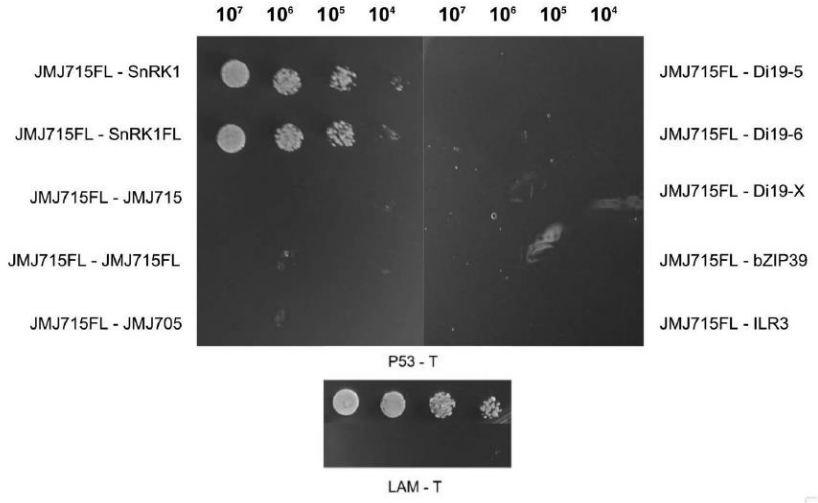


Figure 2.17 Screening of JMJ715 full-length protein against other rice targets

Screening of JMJ715 full-length protein against four library targets and synthetic targets, plate media is 3-DO+30mM 3AT.

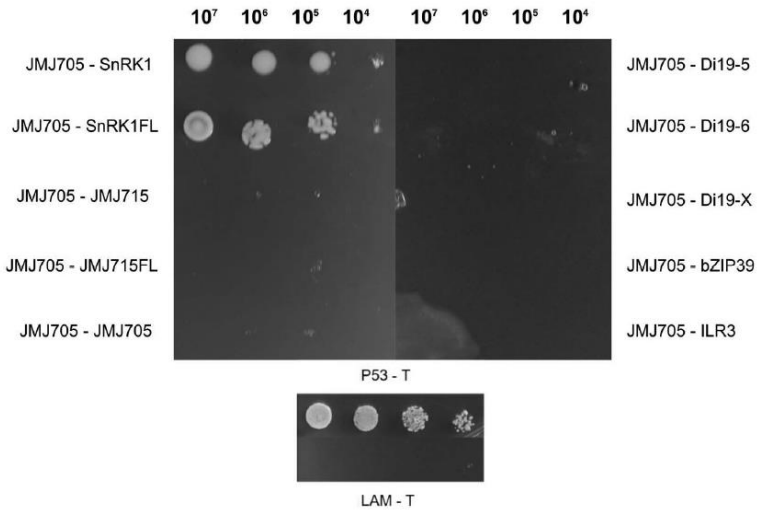


Figure 2.18 Screening of JMJ705 against other rice targets

Screening of JMJ705 fragment synthesized to contain the JmjC domain against seven targets from library screening and other synthetic targets, media is 3-DO+30mM 3AT.

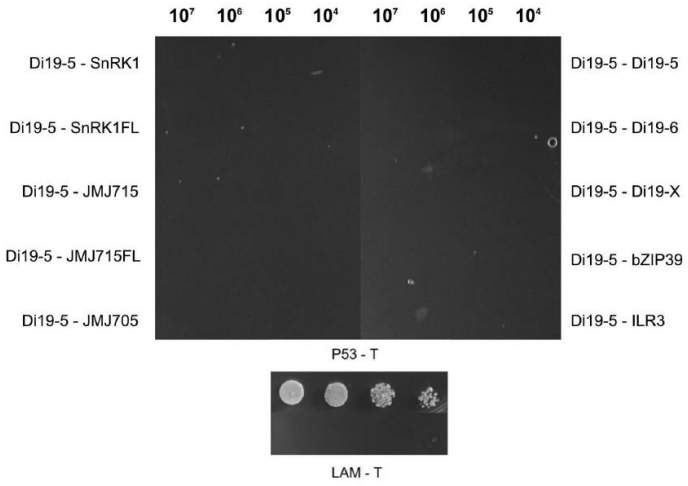


Figure 2.19 Screening of Di19-5 against other rice targets

Screening of Di19-5 full-length protein against four library targets and other synthetic targets, plate media is 3-DO+30mM 3AT.

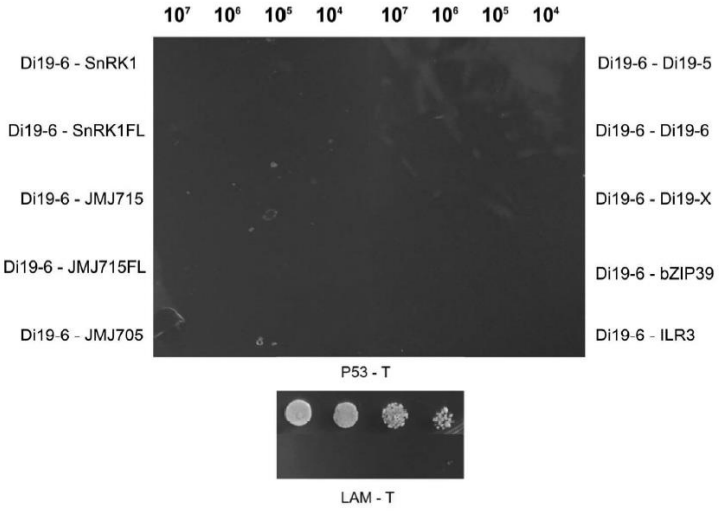


Figure 2.20 Screening of Di19-6 against other rice targets

Screening of Di19-6 full-length protein against four library targets and other synthetic targets, plate media is 3-DO+30mM 3AT.

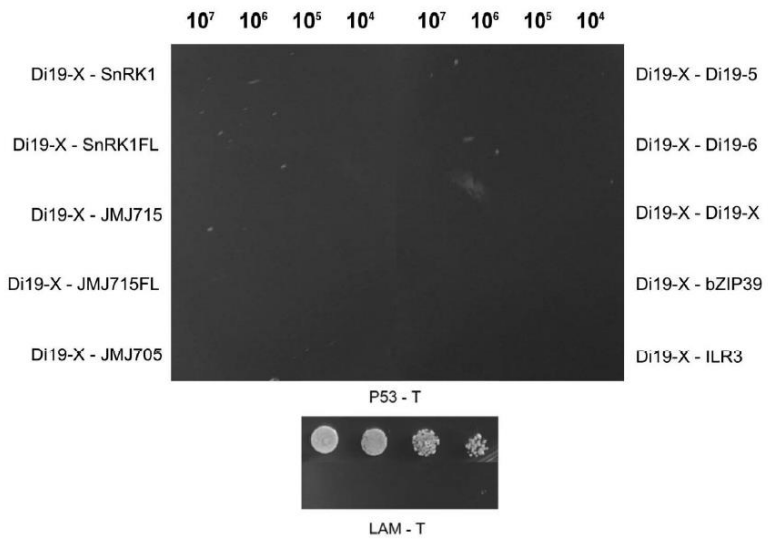


Figure 2.21 Screening of Di19-X against other rice targets

Screening of Di19-X full-length protein against four library targets and other synthetic targets, plate media is 3-DO+30mM 3AT.

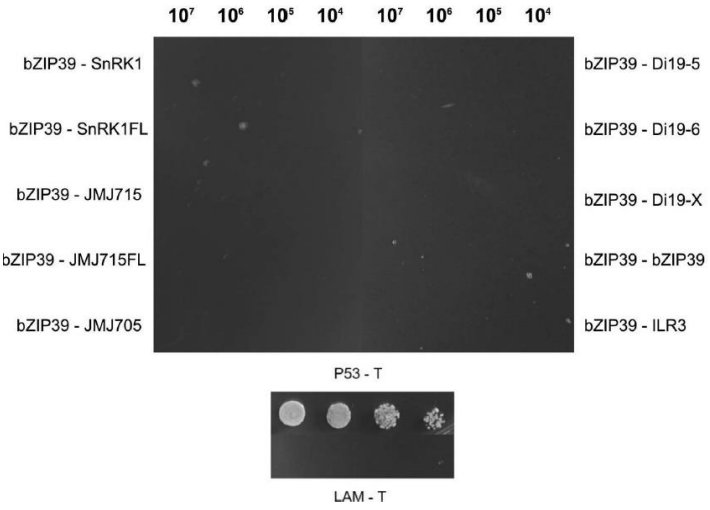


Figure 2.22 Screening of bZIP39 against other rice targets

Screening of bZIP39 fragment from library screening against three other library targets and synthetic targets, plate media is 3-DO+30mM 3AT.

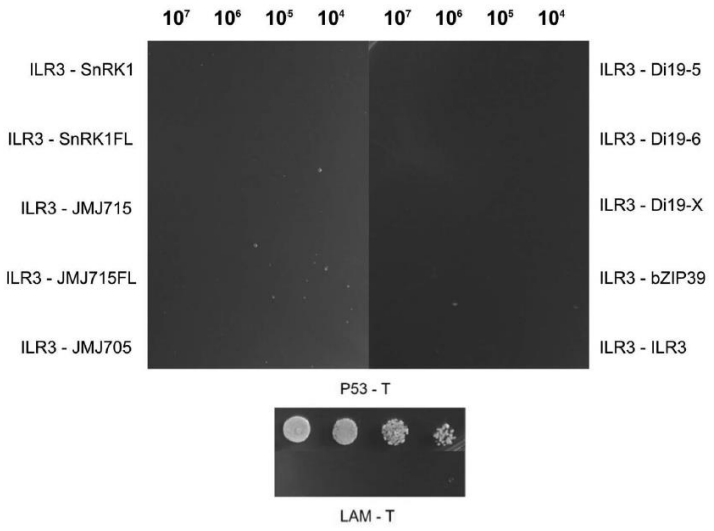


Figure 2.23 Screening of ILR3 against other rice targets

Screening of ILR3 fragment from library screening against three other library targets and synthetic targets, plate media is 3-DO+30mM 3AT.

Discussion

Initial screenings of a rice cDNA library with HAG4 and HAG10 led to an extensive catalog of potential targets (tables 2.2 and 2.3) from which four targets were selected for further study:

1) **SnRK1** (α catalytic subunit, but referred here to as simply SnRK1 as the β- and γ-subunits of the heterotrimer work to regulate substrate interaction only) – SnRK1 is a serine/threonine protein kinase that was identified by independent screenings with both HAG4 (1183 bp PCR product size) and HAG10 (1203 bp PCR product size). This protein is widely conserved throughout nature, known as SNF1 in yeast and AMP-activated protein kinase (AMPK) in mammals. It has roles in many cellular processes such as energy homeostasis, DNA repair, lipid metabolism, sugar metabolism, plant abiotic stress

response, and plant pathogen response among others (Herzig & Shaw, 2018, Crozet et al., 2014, Chen et al., 2020, Filipe et al., 2018) Additionally, a *Xanthomonas* effector AvrBsT interacts with a SnRK1 homolog directly to suppress the plant hypersensitive response (Szczesny et al., 2010). The multiple independent positive interactions from the library and its robust and central role in important cellular processes made SnRK1 our highest priority target of study.

2) **JMJ715** – JMJ715 is an inactive lysine demethylase with an amino acid substitution that makes it inactive in Arabidopsis. However, the zinc finger in JMJ715 is a RING finger and acts as an E3 ubiquitin ligase involved in rice resistance to insect herbivory (Zhang et al., 2021). JMJ715 is part of large family of histone demethylases that all contain jumonji (JmjC) domains (Qian et al., 2019) of which JMJ704 (Hou et al., 2015) and JMJ705 (Li et al., 2013) are known to be involved in rice response to *Xanthomonas* infection. We identified JMJ715 fragments from our initial library screenings twice by HAG4 (1229 and 884 bp PCR product sizes) and twice by HAG10 (996 and 982 bp PCR product sizes). Like SnRK1, the multiple independent identifications of JMJ715 from library screenings and established literature evidence for its (as well as other JMJ family proteins) role in plant defense made it our second high priority target for further study.
3) **bZIP39** – bZIP39 is a member of a large, diverse family of transcription factors that play a wide variety of roles in plant cell functions including plant response to pathogen stress (Ali et al., 2016, Alves et al., 2013, E et al., 2014). Specifically, bZIP39 has been demonstrated to be involved in ER stress response in rice which is a process directly involved with pathogen response (Takahashi et al., 2012). Earlier library screenings conducted in our laboratory identified bZIP39 as a positive target of HAG4. It was also

identified in this study using both HAG4 twice (1000 and 1219 bp PCR product sizes) and HAG10 four times (1191, 1021, 1078, and 1247 bp PCR product sizes). All the above factors made bZIP39 our third high priority target for further study.

4) **ILR3** – ILR3 is the name for this protein in *Arabidopsis* and its homolog in rice seems to be unnamed so we refer to it by the *Arabidopsis* name here. ILR3 is a bHLH transcription factor with known roles in iron homeostasis and pathogen response (Samira et al., 2018). We chose this as our final high priority target for study as it was identified from library screenings by HAG4 once (872 bp PCR fragment size) and by HAG10 twice (1057 and 942 bp PCR fragment sizes).

In addition to these four targets identified from library screenings we also selected four other protein targets because of their involvement with members of this four-target group in the literature:

1) **JMJ705** – As stated previously JMJ705 is a member of the large JMJ family of proteins that is made up of members with JmjC domains. JMJ705 was previously shown to play a role in plant defense. Another publication demonstrated an interaction between rice SnRK1 and JMJ705 that acts as a switch to control energy homeostasis (Wang et al., 2021). For this reason, we used a fragment of the rice JMJ705 protein containing the JmjC domain it shares with JMJ715 to add to our subsequent specific screening experiments beyond the general library screening. Knowing SnRK1 interacts with JMJ705 also lead us to the hypothesis that SnRK1 may also interact with JMJ715.
2) **Di19-5** – Di19 proteins have been demonstrated in the literature to be involved with protein complexes that establish histone lysine methylation, the antithetical process to the function of many JMJ family proteins (histone lysine demethylation) (Shen et al., 2021).

As such we selected Di19-5 (and two other Di19 protein homologs) as potential targets for HAG interactions.

3) **Di19-6** – Second Di19 protein homolog selected for study.
4) **Di19**-X **–** Third Di19 protein homolog selected for study.

We also conducted a screening with the seven HAG effectors against Synaptotagmin-2 which has been demonstrated to be involved in cell to cell movement and secretion of effectors in bacteria, oomycetes, and fungi (Li et al., 2021, Liu et al., 2021, Bozkurt et al., 2014, Lu et al., 2012). Seeing that Synaptotagmin-2 interacted with several of our HAG effectors indicated this target could be of interest for future research as we hypothesize HAGs may be further promoting colonization of plant tissue by the pathogen via interacting with Synaptotagmin via its role in forming ER membrane-plasma membrane junctions.

The data collected from our yeast two hybrid studies have supported hypotheses discussed above:

1) SnRK1 does interact with JMJ715 in yeast two hybrid as we had originally hypothesized, further experiments could tell us if this interaction is preserved in-planta and if it is a biologically relevant interaction like that of SnRK1 and JMJ705.
2) The full-length versions of SnRK1 and JMJ715 also interacted with the same set of HAG effectors lending support to these interactions being authentic and prompting us to explore further with in-planta experiments.
3) Di19-5 and Di19-6 interact with four of our HAG effectors and JMJ715 interacted with all seven, giving support to the idea that histone lysine methylation and demethylation may be a central hub of HAG regulation of plant defense responses. Di19-X did not interact with any HAG effectors tested but this is not unexpected as it is only ~20%

similar at the protein sequence level to Di19-5 and Di19-6 which are ~50% similar to each other at the protein sequence level (Figure 2.8).

This data also negated one of our other hypotheses:

1) With JMJ705 not interacting with any of the seven HAG effectors used in screening it seems less likely that HAG effectors target the JmjC domains generally. From figure 2.7 we can see that JMJ705 only shares 10.9% sequence identity at the protein level with JMJ715.

The results from our yeast two hybrid experiments above have also led us to form some new hypotheses regarding HAG effectors and their roles to be tested in experiments detailed in subsequent chapters:

1) We have a large set of orthologous and paralogous effectors that display a wide range of sequence divergence. The hypothesis is that orthologs will retain their targets even with large sequence divergence, whereas paralogs with the same level of sequence divergence will have different interactions. This hypothesis can be uniquely tested with simple yeast two hybrid experiments to study the changes in the interaction interfaces. Proof of concept for this type of study is provided in Chapter 4.
2) HAG effector proteins may evolve by duplication and divergence and by recombination between duplicated copies. We propose HAG proteins with new specificities could be produced from hybrid effectors produced by exon shuffling that may retain the targets of their constituent parental proteins. Proof of concept studies for this hypothesis are detailed in Chapter 6.

In summary, HAG protein interacts with apparent specificity with several different rice proteins as putative targets. This is consistent with the finding that most effectors interact with multiple

host targets (Wessling et al., 2014). The potential role for HAGs in the interaction with these targets and models for how these interactions relate to pathogenesis are discussed in Chapter 7.

REFERENCES

Agatep R, Kirkpatrick RD, Parchaliuk DL, Woods RA, Gietz RD, 1998. Transformation of Saccharomyces cerevisiae by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol. *Technical Tips Online* **3**, 133-7.

Ali Z, Sarwat SS, Karim I, Faridi R, Jaskani MJ, Khan AA, 2016. Functions of Plant's Bzip Transcription Factors. *Pakistan Journal of Agricultural Sciences* **53**, 303-14.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *J Mol Biol* **215**, 403-10.

Alves MS, Dadalto SP, Goncalves AB, De Souza GB, Barros VA, Fietto LG, 2013. Plant bZIP transcription factors responsive to pathogens: a review. *Int J Mol Sci* **14**, 7815-28.
Bozkurt TO, Richardson A, Dagdas YF, Mongrand S, Kamoun S, Raffaele S, 2014. The Plant Membrane-Associated REMORIN1.3 Accumulates in Discrete Perihaustorial Domains and Enhances Susceptibility to Phytophthora infestans. *Plant Physiol* **165**, 1005-18.

Brown NP, Leroy C, Sander C, 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* **14**, 380-1.

Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U, 2009. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* **10**, 2763-88.

Chen M, Farmer N, Zhong Z, *et al.*, 2022. HAG Effector Evolution in Pyricularia Species and Plant Cell Death Suppression by HAG4. *Mol Plant Microbe Interact* **35**, 694-705.

Chen Z, Wang C, Jain A, *et al.*, 2020. AMPK Interactome Reveals New Function in Non-homologous End Joining DNA Repair. *Mol Cell Proteomics* **19**, 467-77.

Crozet P, Margalha L, Confraria A, *et al.*, 2014. Mechanisms of regulation of SNF1/AMPK/SnRK1 protein kinases. *Front Plant Sci* **5**, 190.

E ZG, Zhang YP, Zhou JH, Wang L, 2014. Mini review roles of the bZIP gene family in rice. *Genet Mol Res* **13**, 3025-36.

Ebbole DJ, Chen M, Zhong Z, *et al.*, 2021. Evolution and Regulation of a Large Effector Family of Pyricularia oryzae. *Mol Plant Microbe Interact* **34**, 255-69.

Filipe O, De Vleesschauwer D, Haeck A, Demeestere K, Hofte M, 2018. The energy sensor OsSnRK1a confers broad-spectrum disease resistance in rice. *Sci Rep* **8**, 3864.

Gomez Luciano LB, Tsai IJ, Chuma I, *et al.*, 2019. Blast Fungal Genomes Show Frequent Chromosomal Changes, Gene Gains and Losses, and Effector Gene Turnover. *Mol Biol Evol* **36**, 1148-61.

Herzig S, Shaw RJ, 2018. AMPK: guardian of metabolism and mitochondrial homeostasis. *Nat Rev Mol Cell Biol* **19**, 121-35.

Hou Y, Wang L, Wang L, *et al.*, 2015. JMJ704 positively regulates rice defense response against Xanthomonas oryzae pv. oryzae infection via reducing H3K4me2/3 associated with negative disease resistance regulators. *BMC Plant Biol* **15**, 286.

Li T, Chen X, Zhong X, *et al.*, 2013. Jumonji C domain protein JMJ705-mediated removal of histone H3 lysine 27 trimethylation is involved in defense-related gene activation in rice. *Plant Cell* **25**, 4725-36.

Li Z, Variz H, Chen Y, Liu SL, Aung K, 2021. Plasmodesmata-Dependent Intercellular Movement of Bacterial Effectors. *Front Plant Sci* **12**, 640277.

Liu J, Zhang L, Yan D, 2021. Plasmodesmata-Involved Battle Against Pathogens and Potential Strategies for Strengthening Hosts. *Front Plant Sci* **12**, 644870.

Lu YJ, Schornack S, Spallek T, *et al.*, 2012. Patterns of plant subcellular responses to successful oomycete infections reveal differences in host cell reprogramming and endocytic trafficking. *Cell Microbiol* **14**, 682-97.

Park CH, Chen S, Shirsekar G, *et al.*, 2012. The Magnaporthe oryzae effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* **24**, 4748-62.

Qian Y, Chen C, Jiang L, Zhang J, Ren Q, 2019. Genome-wide identification, classification and expression analysis of the JmjC domain-containing histone demethylase gene family in maize. *BMC Genomics* **20**, 256.

Reece-Hoyes JS, Walhout AJM, 2018. Gateway Recombinational Cloning. *Cold Spring Harb Protoc* **2018**, pdb top094912.

Samira R, Li B, Kliebenstein D, *et al.*, 2018. The bHLH transcription factor ILR3 modulates multiple stress responses in Arabidopsis. *Plant Mol Biol* **97**, 297-309.

Shen Q, Lin Y, Li Y, Wang G, 2021. Dynamics of H3K27me3 Modification on Plant Adaptation to Environmental Cues. *Plants (Basel)* **10**.

Szczesny R, Buttner D, Escolar L, Schulze S, Seiferth A, Bonas U, 2010. Suppression of the AvrBs1-specific hypersensitive response by the YopJ effector homolog AvrBsT from Xanthomonas depends on a SNF1-related kinase. *New Phytol* **187**, 1058-74.

Takahashi H, Kawakatsu T, Wakasa Y, Hayashi S, Takaiwa F, 2012. A rice transmembrane bZIP transcription factor, OsbZIP39, regulates the endoplasmic reticulum stress response. *Plant Cell Physiol* **53**, 144-53.

Tang M, Ning Y, Shu X, *et al.*, 2017. The Nup98 Homolog APIP12 Targeted by the Effector AvrPiz-t is Involved in Rice Basal Resistance Against Magnaporthe oryzae. *Rice (N Y)* **10**, 5.

Wang W, Lu Y, Li J, *et al.*, 2021. SnRK1 stimulates the histone H3K27me3 demethylase JMJ705 to regulate a transcriptional switch to control energy homeostasis. *Plant Cell* **33**, 3721-42.

Wessling R, Epple P, Altmann S, *et al.*, 2014. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* **16**, 364-75.

Zhang F, Fang H, Wang M, *et al.*, 2022. APIP5 functions as a transcription factor and an RNA-binding protein to modulate cell death and immunity in rice. *Nucleic Acids Res* **50**, 5064-79.

Zhang Y, Chen M, Zhou S, Lou Y, Lu J, 2021. Silencing an E3 Ubiquitin Ligase Gene OsJMJ715 Enhances the Resistance of Rice to a Piercing-Sucking Herbivore by Activating ABA and JA Signaling Pathways. *Int J Mol Sci* **22**.

CHAPTER III

SLPIT LUCIFERASE *NICOTIANA BENTHAMIANA* ASSAYS

Introduction

The Y2H assays indicate protein-protein interactions that occur in yeast cells as fusions to GAL4 DNA binding or activation domains. Additional confidence in the validity of the interaction could be obtained in several ways. To demonstrate interactions in plant cells, I chose to employ a split luciferase assay system.

The split luciferase assay, like yeast two hybrid, makes use of a protein with two domains that can be separated into two fusion proteins and have its function restored when protein-protein interactions bring the two domains close together. In this case the protein used is firefly luciferase of which the separate n-terminal and c-terminal fragments are able to function when brought into close proximity (Azad et al., 2014).

Because SnRK1 was previously found to be targeted by a bacterial virulence factor in Arabidopsis, and inhibition of SnRK1 by the bacterial virulence factor led to reduced defense in the host (Filipe et al., 2018), I chose to assess interaction between two of the effectors that interact with rice SnRK1 in the Y2H assay, PgHAG4 and PlHAG10. A successful detection of the interaction would show that the interaction can occur in a plant cell (tobacco). Assays were performed to assess dimer formation by SnRK1, and heterodimer formation between SnRK1 and HAG proteins.

Materials and Methods

For these experiments we used two base vectors that were gateway ready, one containing the n-terminal luciferase sequence (pXJJ6) and one containing the c-terminal luciferase sequence (pJG014) (Junqi Song, Texas A&M University, personal communication). They were co-expressed in *Nicotiana benthamiana* leaves, delivered by agroinfiltration of both vectors in transformed GV3101 cells. Additionally, we used a third vector constitutively expressing luciferase as positive control (Libo Shan, Texas A&M University, personal communication) (Table 3.1).

Table 3.1 – Vector used for Split Luciferase assays.

Vector Name	Derived From	Modification	Selection	Origin
pLuc	pUC18	35s-Luciferase-nos	Ampicillin	Shan Lab
pXJJ6	pCambia1300	35s-nLUC-Gateway-nos	Kanamycin	Song Lab
pJG014	pCambia1300	35s-cLUC-Gateway-nos	Kanamycin	Song Lab

Gateway reactions were conducted using the Gateway LR Clonease II Enzyme Mix (Invitrogen) following the manufacturer's instructions. Agrobacterium competent cells were prepared, and transformations were performed using a freeze thaw method (Höfgen & Willmitzer, 1988). Agroinfiltration was performed following standard agroinfltration procedures for *N. benthamiana* (Norkunas et al., 2018).

Twenty-four hours following Agroinfiltration of *N. benthamiana* leaves regions of Agroinfiltration were excised with a hole punch to obtain small leaf disks to fit into 96-well plates for luminescence detection on a SpectraMax iD5 Multi-Mode Microplate Reader in

Luminescence Read Mode (Molecular Devices). Leaf disks were added to wells containing 100 uL of water. Once the plate was loaded it was sprayed uniformly with a solution of 0.2% Silwet and 0.2 uM luciferin. The plate was covered immediately with aluminum foil and placed subjected to vacuum infiltration for 1 minute to draw solution into the leaf disk. The plate was removed still covered and then left to quench luminescence induced by infiltration for 5-10 minutes after which the aluminum foil cover was removed. The plate was placed in the microplate reader, and luminescence was immediately measured. Luminescence was measured in Relative Light Units (RLUs) compared to reference non-Agroinfiltrated control leaf disks.

Results

Three assays were conducted to assess in planta protein protein interactions. It is known that SnRK1 forms a dimer in plant systems. Therefore, SnRK1 dimer formation was expected to serve as a positive control for function of the split luciferase. We further anticipated that SnRK1 interaction with PgHAG4 and PlHAG10 would be detected since these were interactions detected in yeast two hybrid assays. The specific assays were:

1) SnRK1 duplex – SnRK1 in pJG014 and SnRK1 in pXJJ6
2) SnRK1 in pJG014 and PgHAG4 in pXJJ6
3) SnRK1 in pJG014 and PlHAG10 in pXJJ66

Luminescence data from several rounds of split luciferase assays using these vector sets are detailed in Table 3.2 below in terms of RLUs and the average values for each of the four are charted in Figure 3.1.

Table 3.2 – Split Luciferase Luminescence Results in RLUs compared to negative control

Vectors Used			
Consituative Luciferase	SnRK1(pJG014)xSnRK1(pXJJ6)	SnRK1(pJG014)xPgHAG4.2(pXJJ6)	SnRK1(pJG014)XPspLSHAG10(pXJJ6)
12	28	10	32
45	14	8	27
39	5	22	3
55	32	5	26
42	26	15	18
52	23	15	10
65	32	21	24
57	19	9	28
25	29	16	5
54	13	24	18
54	24	22	24
42	33	8	24
	18	23	17
	26	20	30
	25	11	20
	32	13	28
	22	21	
	36	19	
	38	24	
	25	20	

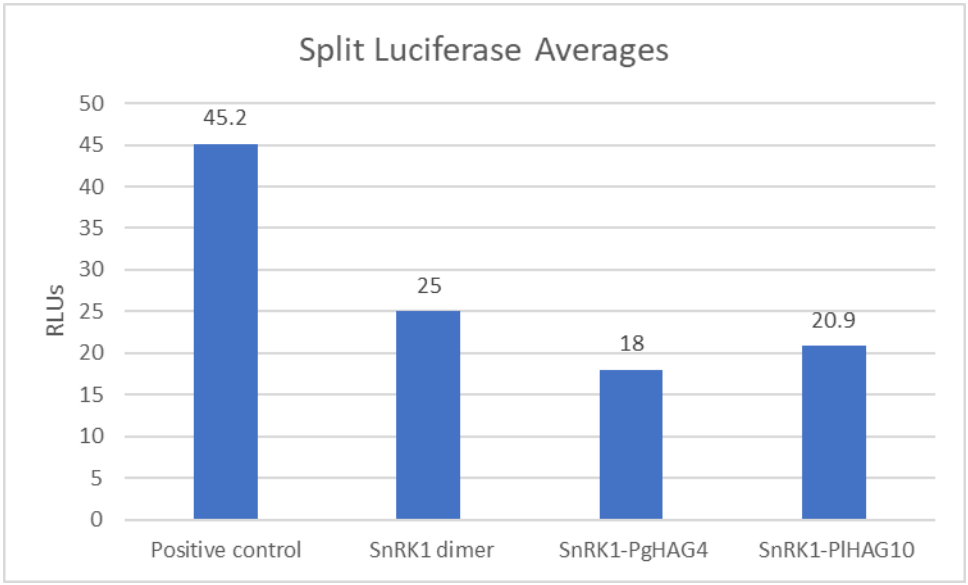


Figure 3.1 – Average RLUs for Split Luciferase Experiments.

These results from these split luciferase experiments can be summed up as follows:

1) Our positive control for this experiment functions as expected

2) SnRK1 forms dimers as reported in the literature and is a control for the split luciferase function
3) PgHAG4 and PlHAG10 interact with SnRK1 in planta

Discussion

The split luciferase experiments reported in this chapter have added support to the results from the chapter 2 yeast two hybrid experiments. The interactions of PgHAG4 and PlHAG10 with SnRK1 also observed in-planta. I was also able show SnRK1 dimerizes as has already been reported giving confidence to the results obtained for HAGs and SnRK1 in this assay.

Assuming that averaged over several experiments that the 35S promoter is equally active and equal amounts of plasmid are delivered by agroinfiltration for both SnRK1 constructs in the dimerization test, one would expect the potential to produce as many active luciferase enzyme (n-LUC and c-LUC) molecules as the control and therefore an equal amount of luminescence. However, since SnRK1 dimerizes, three types of SnRK1 dimer would be possible: n-lucSnRK1 with n-lucSnRK1, n-lucSnRK1 with c-lucSnRK1, c-lucSnRK1 with c-lucSnRK1. Only the heterodimer forms an active luciferase enzyme. The binding affinities and stability of these SnRK1 homodimers and heterodimers is unknown so it is not possible to accurately model the expected luminescence levels. However, if all affinities were equal, I would expect some homodimers would form (2/3 homodimers and 1/3 active heterodimer) and I would predict lower levels of luminescence than for the control plasmid, as was observed. Since we have not detected HAG dimer formation in Y2H experiments, all nLUC-HAG protein should be available to interact with cLUC-SnRK1 molecules. The amount of luciferase activity observed in the combinations of c-LUC-SnRK1 with either n-LUC-PgHAG4 or n-LUC-PlHAG10 are like the luminescence observed in the SnRK1 dimerization experiment. I would expect a level of

luminescence at least equal to that for the SnRK1 dimerization experiment since all the HAG protein is available for heterodimer formation. This suggests that HAG protein binding affinity for SnRK1 is lower than the SnRK1 dimerization affinity. Again, the assumptions made about expression and affinities for SnRK1 dimer formation, and that the positioning of the subunits in the dimers produce equal luciferase activity may not be correct. Future studies should be performed to analyze HAG-SnRK1 binding.

The ability to detect HAG-SnRK1 interactions shows that the system could be used as an in vivo test for inhibition of SnRK1 kinase activity. For example, the SAMS peptide is a highly specific target of SnRK1 enzyme (Avila et al., 2012). A SAMS peptide incubated with extracts of tobacco expressing SnRK1 or SnRK1 and HAGs could be used to measure SnRK1 in the two conditions. If inhibition of SnRK1 kinase activity by HAGs could be shown, this would directly show that HAGs expressed in planta could inhibit plant defense. Other approaches, such as construction of transgenic rice expressing HAGs and then testing for reduced SnRK1 kinase activity and defense against pathogens is another approach to assess the role of HAGs. Co-immunoprecipitation of SnRK1 with transgenic rice plants expressing an epitope tagged HAG protein would prove both in vivo binding and inhibition of defense.

This work indicates that the use of the split luciferase assay to further test HAG target interactions will be an important avenue for further characterization of interactions and the role of HAGs in the host-pathogen interaction.

REFERENCES

Avila J, Gregory OG, Su D, *et al.*, 2012. The beta-subunit of the SnRK1 complex is phosphorylated by the plant cell death suppressor Adi3. *Plant Physiol* **159**, 1277-90.

Azad T, Tashakor A, Hosseinkhani S, 2014. Split-luciferase complementary assay: applications, recent developments, and future perspectives. *Anal Bioanal Chem* **406**, 5541-60.

Filipe O, De Vleesschauwer D, Haeck A, Demeestere K, Hofte M, 2018. The energy sensor OsSnRK1a confers broad-spectrum disease resistance in rice. *Sci Rep* **8**, 3864.

Höfgen R, Willmitzer L, 1988. Storage of competent cells for Agrobacterium transformation. *Nucleic Acids Res* **16**, 9877.

Norkunas K, Harding R, Dale J, Dugdale B, 2018. Improving agroinfiltration-based transient gene expression in Nicotiana benthamiana. *Plant Methods* **14**, 71.

CHAPTER IV

HAG HOMOLOG EVOLUTION AND INTERACTIONS WITH RICE TARGET PROTEINS

Introduction

The phylogenetic distribution of HAGs led us to consider how evolution impacts the repertoire of interactions in paralogs and orthologs. The pattern of yeast two hybrid screenings discussed in Chapter 2, suggested that orthologs retain interactions and paralogs diverge further. To gain additional insight into the relationship between sequence divergence, orthology, paralogy, and interaction patterns, we conducted several smaller yeast two hybrid screenings with HAG homologs from *Pyricularia*, *Colletotrichum*, and the distantly related genus, *Cochliobolus*. HAG4 and HAG10 are very closely related paralogs in P. oryzae. The next most closely related paralog is HAG20. We tested HAG20 orthologs from *Pyricularia grisea* (PgHAG20), *P. pennisetigena* (PpHAG20), and *Pyricularia* sp. LS (PlHAG20) (Figure 4.1). Figure 4.2 presents a phylogenetic tree of *Pyricularia* species containing HAG genes. HAG genes are absent in *P. penniseti*. *P. penniseti* is thought to have diverged from the other Pyricularia species ~30 MYA (Gomez Luciano et al., 2019). As the HAG-containing species diverged from each other an estimated ~5 MYA, HAGs likely appeared in the lineage 5 to 30 MYA. Interestingly, C4 grasses arose ~25 - 35 MYA (Sage, 2004) and this lead to a large geographic expansion and many speciation events among the grasses. This may have led to increased encounters between pathogens of grasses and therefore more opportunities for HGT between formerly geographically isolated pathogens.

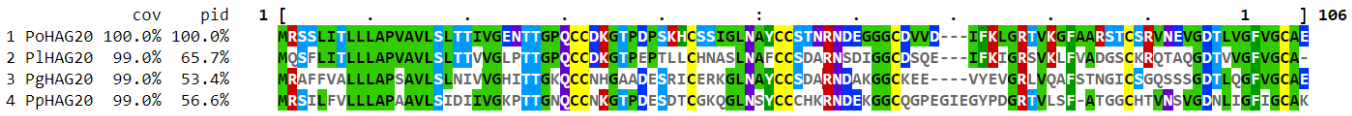


Figure 4.1 Alignment of *Pyricularia* HAG20s

Key: PoHAG20 = *P. oryzae* HAG20, PlHAG20 = *P.* sp. *leersia* HAG20, PgHAG20 = *P. grisea* HAG20, PpHAG20 = *P. pennisetagena* HAG20 alignment generated with Mview (Brown et al., 1998)

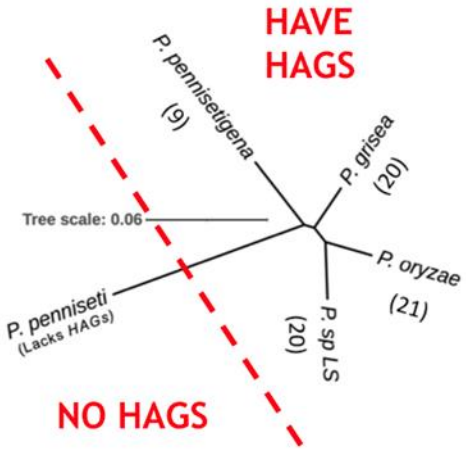


Figure 4.2 Phylogeny of HAG containing *Pyricularia* species

A depiction of the occurrence of HAG effectors in the genus *Pyricularia*. Values in parenthesis represent the number of HAG effectors identified from the available genome sequence data in *Pyricularia* species closesly related to *Pyricularia oryzae*. As we move further back evolutionarily, HAG genes stop appearing at the division between *P. pennisetigena* and *P. penniseti.*

HAG orthologs have been identified in species of Colletotrichum, a genus of Ascomycete fungi filled with well-studied plant pathogens (Cannon et al., 2012). This genus is thought to have diverged from Pyricularia ~250 MYA (Hongsanan et al., 2017). The gene structures between

Colletotrichum and Pyricularia HAGs are identical, including the occurrence of two classes of HAGs containing 6 or 8 cysteine residues. Strikingly, CgrHAG1 shares 65% sequence identity to *P. oryzae* HAG1. The absence of these classes of HAGs from other fungal species suggest the potential for HGT. Indeed, these HAG classes are only found in Colletotrichum species that are grass pathogens, and not in other Colletotrichum species. The divergence of clade of Colletotrichum grass pathogens from other Colletotrichum species is estimated to have occurred ~15 to ~28 MYA (Cannon et al., 2012). Since all Colletotrichum grass pathogens have HAGs, it seems more likely that HAGs were first acquired by Colletotrichum species soon after or concurrently with their adaptation to grasses, where they may have encountered other grass pathogens and acquired HAG genes. Whether they exchanged HAGs with Pyricularia is not proven but seems likely given the similarity of HAG protein sequences. Therefore, the question of common targets between Colletotrichum and Pyricularia HAGs is of particular interest. Two *Colletotrichum* HAG homologs: HAG1 from *Colletotrichum graminicola* (CgrHAG1) and HAG10 from *Colletotrichum falcatum* (CfaHAG10) were synthesized for Y2H assays. Additional, far more distantly related HAG sequences have also been discovered in other fungal species. In the distantly related corn pathogen *Cochliobolus heterostrophus* another HAG gene (ChetHAG1) was synthesized. Its greater sequence divergence from other HAGs (Fig 4.3), and the different lifestyle of the fungus (necrotroph) compared to Pyricularia and Colletotrichum (hemibiotroph) suggest that it might play a different role in the biology of the host-pathogen interaction and testing for interaction with the rice protein targets will be of interest.

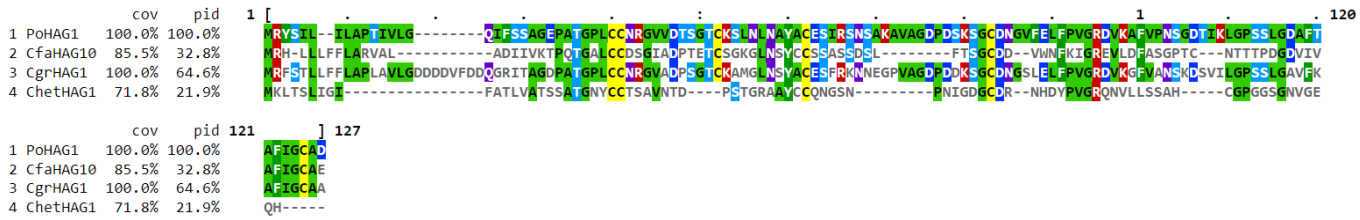


Figure 4.3 Alignment of *Colletotrichum* and *Cochliobolus* HAGs with *P. oryzae* HAG1

Key: PoHAG1 = *P. oryzae* HAG1, CfaHAG10 = *C. falcatum* HAG10, CgrHAG1 = *C. graminicola* HAG1, ChetHAG1 = *C. heterostrophus* HAG1. Alignment generated with Mview (Brown et al., 1998)

Materials and Methods

The protocols used in these experiments were identical to those used in previous Y2H screenings discussed in Chapter 2. The three HAG20 orthologs from other *Pyricularia* species and the three HAGs from outside the genus *Pyricularia* that were introduced in the previous section were synthesized with gateway sequences enabling them to quickly be moved into our previously discussed pwy146 (Junqui Song, Texas A&M University, personal communication) Y2H gateway vector. These Y2H constructs of PgHAG20, PlHAG20, PpHAG20, ChetHAG1, CfaHAG10, and CgrHAG1 were used in Y2H screens of our four rice targets SnRK1, bZIP39, ILR3, and JMJ715.

Results

Testing of autoactivating HAG1: Previous work in our lab found that HAG1 was auto activating, due to this finding all experiments in this chapter were conducted with the HAG effector expressed as an AD-clone and all target proteins as BD-clones. While we did not find the other three HAGs used in this assay to be auto-activating when expressed as BD-clones, this swapping

of domains made the assay uniform. Results of the most selective screening on 3-DO+30mM 3AT media showed that all four HAG orthologs interacted positively with all four rice targets identified from screenings in Chapter 1: SnRK1, JMJ715, bZIP39, and ILR3.

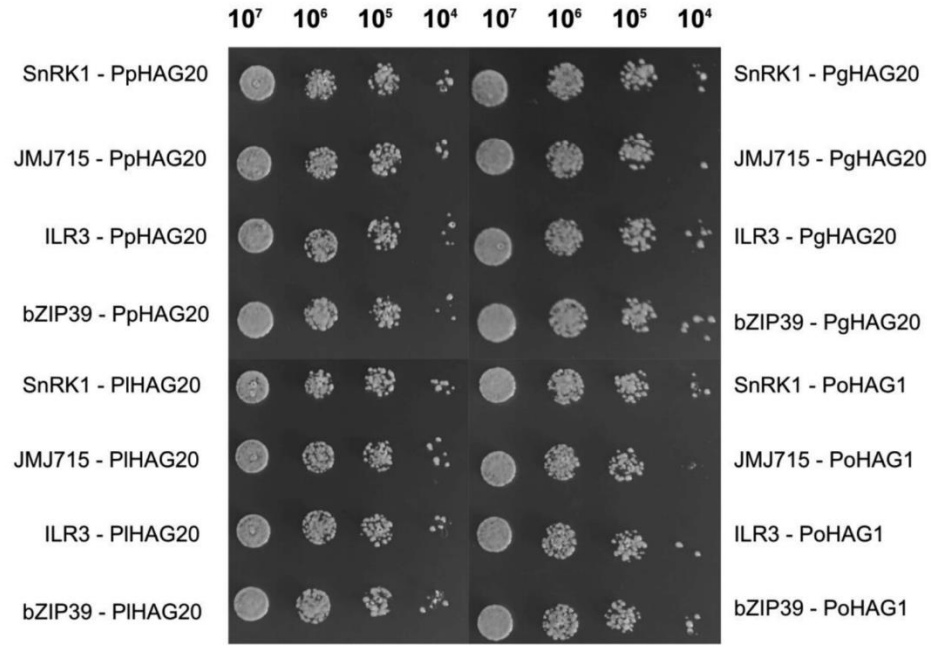


Figure 4.4

Yeast two hybrid screening results for four orthologous HAG effectors spanning the HAG containing species of *Pyricularia*. The plating shown was on 3-DO+30mM 3AT media as seen in the final screening results previously shown in Chapter 1. Here the rice target proteins were expressed as BD-clones and the HAG effectors were expressed as AD-clones.

Orthologs vs. Paralogs

Instead of screening a whole rice library, here we just wanted to screen the four targets previously identified in Chapter 1's library screen: SnRK1, JMJ715, bZIP39, and ILR3. We did not find ChetHAG1, CfaHAG10, or CgrHAG1 to autoactivate when expressed as BD-clones so

we used the same vector distributions as Chapter 1. Our screen found that CgrHAG1 interacted positively with all four rice targets while ChetHAG1 and CfaHAG10 interacted positively with JMJ715 and ILR3 and did not interact with bZIP39 or SnRK1 (Figure 4.5)

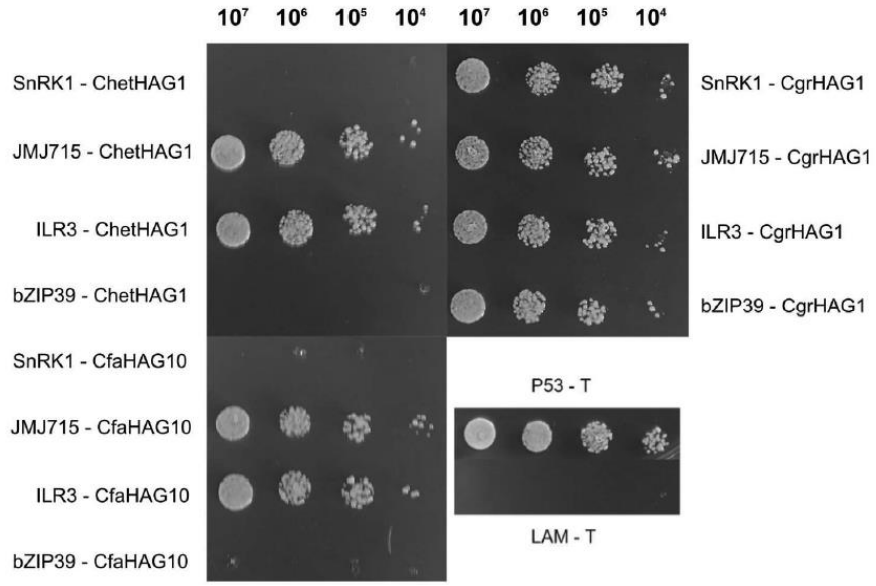


Figure 4.5

Yeast two hybrid screening results for three orthologous HAG effectors from three other Ascomycete plant pathogens. The plating shown was on 3-DO+30mM 3AT media as seen in the final screening results previously shown in Chapter 1.

Discussion

Previously in this thesis we posited the hypothesis that orthologous HAG effectors would maintain their target repertoires while paralogous HAG effectors with a similar level of amino acid divergence would be free to diversify their repertoires due to redundancy in the genome. The results of the screenings presented in this chapter lend support to this hypothesis as we saw all four orthologs here maintain all four target protein interactions. Clearly, a larger sample of

yeast two hybrid screenings with different HAG orthologs is needed to strengthen this hypothesis but these results point to those experiments being a worthwhile pursuit. With those facts in mind we can make use of the availability of many more HAGs in *Pyricularia* species shown in Figure 4.5.

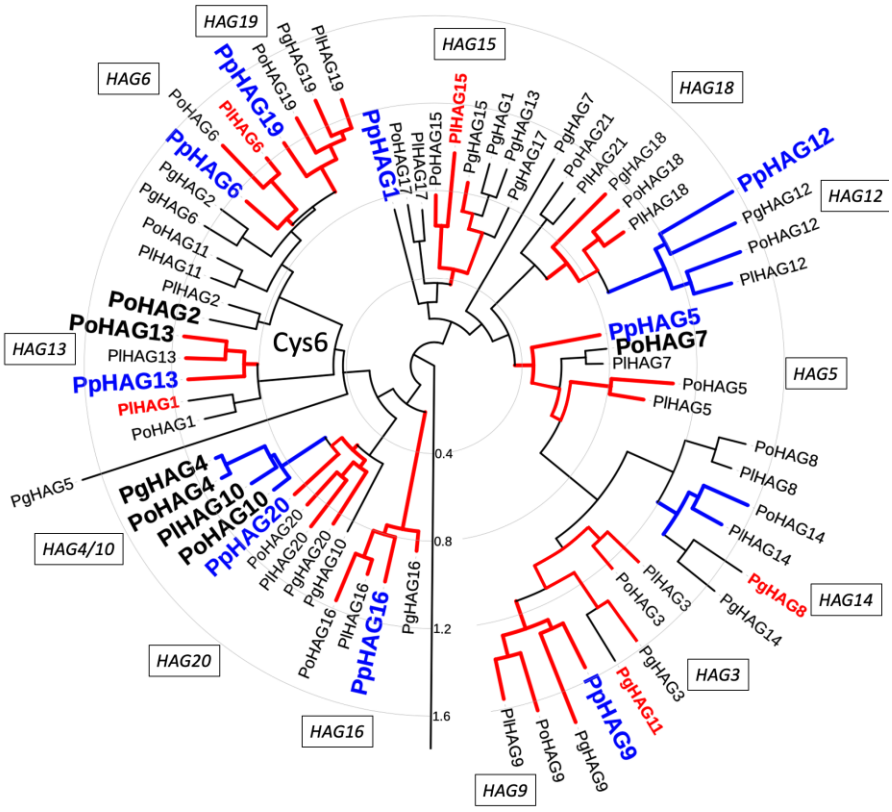


Figure 4.5 *Pyricularia* HAG tree

This is a DNA Maximum likelihood tree, the boxes on the outside highlight the 13 HAGs that are orthologs between *P. oryzae* and either *P. grisea* or *P. pennisettigena*. The *P. pennisittigena* HAG names are shown in Blue to give an idea of their distribution. The red branches denote the genes that have microsynteny to support the assignment of orthology and the blue branches are by best bi-directional BLAST hit. The 6 cysteine group forms a single clade. Some gene names can be seen in red, and these are pseudogenes, for example *P. grisea* HAG8 has the same frameshift mutation in all 9 *P. grisea* genome sequences available. Tree generated with ITOL (Letunic & Bork, 2021)

In this chapter we are presenting further yeast two hybrid experiments screening our core priority targets from rice library screens against three effectors in other species of *Pyricularia* and three HAG effectors from geneses outside of *Pyricularia*. It would be expected that effectors of grass/cereal infecting *Colletotrichum* species like *Colletotrichum graminicola* ,anthracnose leaf blight of corn (Ma et al., 2022), and *Colletotrichum falcatum* ,sugar cane red rot (Hossain et al., 2020), would interact with similar targets as their orthologs in *Pyricularia* which infect similar hosts and have a similar life cycle. What is incredibly interesting is that while *Cochliobolus heterostrophus (Bipolaris maydis)* also infects corn it is a necrotrophic pathogen (Manamgoda et al., 2011). This leads one to question if these interactions are biologically relevant or artifactual never occurring nature due to the hosts lifestyle. These HAG effectors in *C. heterostrophus* must certainly serve some purpose or they would not be maintained in the genome. Thus, we are left with two possibilities:

1) These HAG effectors have evolved a new purpose beyond secreted biotrophic effectors serving some yet unknown function.
2) Theses HAG effectors are still secreted into the host and act as effectors but only very early in infection.

Either of these possibilities would be a novel discovery if they could be supported by further experimental data.

REFERENCES

Brown NP, Leroy C, Sander C, 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* **14**, 380-1.

Cannon PF, Damm U, Johnston PR, Weir BS, 2012. Colletotrichum - current status and future directions. *Stud Mycol* **73**, 181-213.

Gomez Luciano LB, Tsai IJ, Chuma I, *et al.*, 2019. Blast Fungal Genomes Show Frequent Chromosomal Changes, Gene Gains and Losses, and Effector Gene Turnover. *Mol Biol Evol* **36**, 1148-61.

Hongsanan S, Maharachchikumbura SSN, Hyde KD, *et al.*, 2017. An updated phylogeny of Sordariomycetes based on phylogenetic and molecular clock evidence. *Fungal Diversity* **84**, 25-41.

Hossain MI, Ahmad K, Siddiqui Y, *et al.*, 2020. Current and Prospective Strategies on Detecting and Managing Colletotrichumfalcatum Causing Red Rot of Sugarcane. *Agronomy* **10**.

Letunic I, Bork P, 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W6.

Ma W, Gao X, Han T, *et al.*, 2022. Molecular Genetics of Anthracnose Resistance in Maize. *J Fungi (Basel)* **8**.

Manamgoda DS, Cai L, Bahkali AH, Chukeatirote E, Hyde KD, 2011. Cochliobolus: an overview and current status of species. *Fungal Diversity* **51**, 3-42.

Sage RF, 2004. The evolution of C(4) photosynthesis. *New Phytol* **161**, 341-70.

CHAPTER V

CHIMERIC HAG PROTEINS PRODUCE NOVEL PATTERNS OF INTERACTION WITH HOST PROTEINS

Introduction

Based on the remarkable ability of HAG effectors to maintain interactions with proteins across great sequence divergence I hypothesize that there is conservation of interaction domains in HAG proteins that are not obvious from simple sequence alignment. To test this idea further, chimeric HAGs that combine the first half (first exon) of one HAG effector with the second half (second exon) of a different HAG effector (exon shuffling) should retain some or all of the target interactions common to HAGs. For HAGs that differ in one or more target specificities the chimeric may retain or lose a unique target depending on the sequence required or the interaction.

Materials and Methods

The protocol for Y2H in this chapter is identical to the protocol that was used in Chapter 2. Synthetic HAG genes were produced and cloned in plasmid vectors (ETON Biosciences, Inc.) that I named HAG7-4 (first half HAG7, second half HAG4), HAG2-13 (first half HAG2, second half HAG13), HAG13-2 (first half HAG13, second half HAG2). HAG4-7 was also synthesized but clones of the gene could not be obtained. We assume the DNA fragment either inhibits plasmid replication or is lethal in *E. coli.* The three chimeric HAG genes were synthesized with Gateway compatibility and the Gateway protocol was used to move these genes into our desired vectors for Y2H. Figure 5.1 below is an alignment of our three chimeric HAG protein sequences along with the native sequences of their donor HAGs.

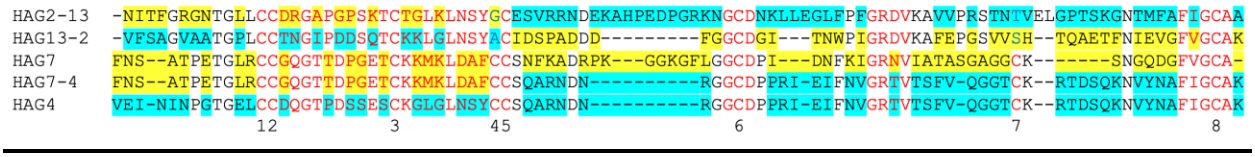


Figure 5.1 – Alignment of Chimeric Hag Protein Sequences with the Sequences of their Donor HAGs

HAG2-13 has the N-terminal region of HAG2 (yellow highlights for differences with HAG13) and the C-terminal region of HAG13 (blue sequences highlight differences with HAG2). For HAG chimerics the site of recombination is where the intron divides exon 1 from exon 2 directly between cysteine positions 4 and 5. For HAG7-4 the N-terminal region of HAG7 is highlighted in yellow and the C-terminal region is that of HAG4 is highlighted in blue.

Results

Results of this Y2H screening had HAG7-4 maintaining all HAG4 targets described in Chapter 2 with the exception of Synaptotagmin-2. HAG2-13 and HAG13-2 yielded identical results from the screening with both chimerics maintaining interaction with their common targets ILR3 and JMJ715 and HAG2 target synaptotagmin-2 both the chimerics were unable to interact with the HAG2 target bZIP39 (Figure 5.2).

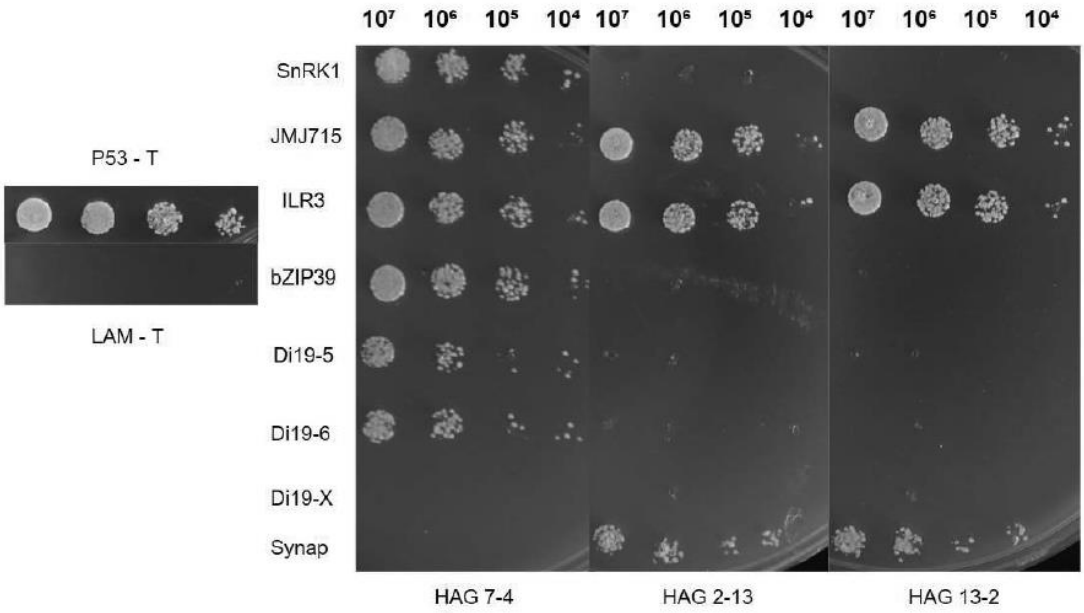


Figure 5.2 – Y2H Screen of Chimeric HAGs

The p53 – T antigen positive control and LAM – T antigen negative control are shown on the left. These were plated on a separate plate produced from the same media and inoculated on the same day. To the right are the Y2H screenings for HAG7-4, HAG2-13, HAG13-2 from left to right. Results shown are on 3-DO+30mM 3AT media. PoHAG7 interacts with ILR3 and JMJ715, but no other targets. HAG4 interacts with all targets except for Di19-X.

As we noted in Chapter 2, Di19-X differs from the other Di19 homologs significantly at the protein sequence level. Remarkably, the HAG7-4 chimeric protein with an N-terminal region of HAG7 and C-terminal region of HAG4 interacts with all the HAG4 targets except for Synaptotagmin.

Table 5.1 Interactions of HAGs and Chimeric HAGs with Rice Targets

	HAG4	HAG7	HAG7-4	HAG2	HAG13	HAG2-13	HAG13-2
SnRK1	+	-	+	-	-	-	-
JMJ715	+	+	+	+	+	+	+
Di19-5	+	-	+	-	-	-	-
Di19-6	+	-	+	-	-	-	-
Di19-X	-	-	-	-	-	-	-
ILR3	+	+	+	+	+	+	+
bZIP39	+	-	+	-	-	-	-
Synaptotagmin	+	-	-	+	-	+	+

Since the HAG7-4 gene interacts with all HAG4 targets except for the rice Synaptotagmin, we can conclude that either the binding sites for all of these interactions lie in the C-terminal region of HAG4 and only the Synaptotagmin site lies in the N-terminal region of HAG4. Alternatively, HAG7 may have residues that complete interaction interfaces with the HAG4 C-terminal region, but not bind to those targets in combination with its own C-terminus. Since HAG2 and three only differ in binding to Synaptotagmin, it is not surprising that both chimeric clones still bind to JMJ715 and ILR3. However, the finding that both HAG2-13 and HAG13-2 interact with Synaptotagmin but HAG13 does not is remarkable and further suggests that HAG13 contains some sequences that compensate for both the loss of the N- and C-terminal region.

Discussion

The results of these experiments support the hypothesis that chimeric HAG effectors can be created that maintain their protein interaction interface as evidenced by their ability to retain binding activity. The predicted structures of HAG7 and HAG4 (Figure 5.3) do not appear to be completely superimposable directly but share general structural features, such as the number of beta sheets and their relative relationship as well as the disulfide bridges. This result supports the idea that novel HAG effectors could be generated in the wild pathogen populations via exon shuffling or other recombination events, creating new specificities.

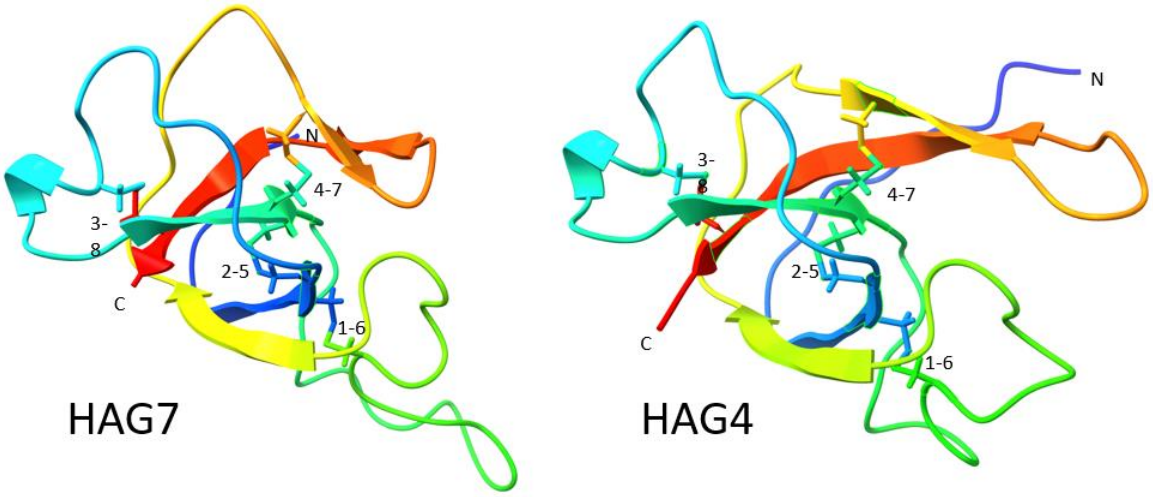


Figure 5.3 Comparison of HAG7 and HAG4 predicted structures

HAG7 and HAG4 alpha-fold models. The protein coded by the first exon and second exons are each rainbow colored along their length. The N-terminus (N) is blue and changes to red at the end (Cysteine 4) and begins as blue again (Cysteine 5) which in turn changes to red at the C-terminus (C). The positions of disulfide bonds between cysteines at positions 1 and 6, 2 and 5, 3 and 8, and 4 and 7 are labeled. Created using Alphafold (Jumper et al., 2021)

REFERENCES

Jumper J, Evans R, Pritzel A, *et al.*, 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-9.

CHAPTER VI

HORIZONTAL TRANSFER OF A PUTATIVE EFFECTOR GENE BETWEEN PLANT PATHOGENIC FUNGI

Introduction

Horizontal Gene Transfer (HGT) is a process in which genes are transferred from one organism to another, allowing for the quick acquisition of foreign genes between even distantly related organisms. This acquisition of foreign genetic material can be a powerful force among pathogenic species. HGT is often discussed as the cause behind problematic antibiotic resistance relevant to human disease, but it is an equally dangerous and challenging issue in the realm of plant pathogens. In the past HGT between plant pathogenic fungi was thought to be a rare occurrence but studies describing instances between eukaryotic organisms have become more and more common (Soanes & Richards, 2014). These instances have made it apparent that HGT plays a significant role in pathogen adaptation, especially between organisms that parasitize the same hosts or hosts that occupy the same environment. There are many examples where an HGT event provides a significant fitness advantage to a pathogen and is a pivotal factor in the pathogenicity of the organism (Friesen et al., 2006, Gardiner et al., 2012). HGT can occur via several different mechanisms such as the exchange of plasmids between two organisms, via transposon insertion, or even picked up from the environment during stress conditions (Nevoigt et al., 2000, Soanes & Richards, 2014). In plant pathogens HGT has been shown to allow for the transfer of toxins, cell wall degrading enzymes, or effector proteins all of which can rapidly allow a receiving organism that casually associates with a plant to become a disease-causing pathogen driving the fixation of the gene in a new population (Soanes & Richards, 2014, Friesen et al., 2006, Gardiner et al., 2012).

In the previous chapters I have described the HAG family of effectors present in *Pyricularia* species and grass pathogens in the genus *Colletotrichum.* My experiments detailed in the previous chapters found that some of these effectors may be relevant to pathogenesis, with some members of the HAG family showing signatures of HGT (Chen et al., 2022).

The work described in this chapter was initiated as part of an undergraduate research mentoring activity. I guided two students in the use of tools for identification of sequence homology and manual annotation of genes. Using available sequence data, we sought out other examples of putative effector genes in plant pathogens following a distribution indicative of their transfer via HGT events. Using the JGI Mycocosm MCL Clusters tool as a starting point, we searched for gene families that were present in *Colletotrichum* species as well as *Pyricularia oryzae* but not in species of fungi of intermediate phylogenetic distance to these two genera. We were able to identify another set of genes that seems to be preferentially associated with plant pathogenic fungi, primarily in *Colletotrichum* that are expressed during host infection. We named this family of putative effector genes the CWR genes due to the characteristic richness in Cysteine and Tryptophan residues in their protein sequences. Investigation of the region of HGT containing the CWR gene in *Pyricularia oryzae* revealed a 3.4kb region containing the CWR gene and a second gene that appeared to be co-acquired via HGT. Within *Colletotrichum* species there were many examples of a given genome having multiple CWR genes. These genes form three distinct classes we called CWR1, CWR2, and CWR3. Additionally, we found that there exist members of this gene family scattered across several other more distantly related species of fungi.

Materials and Methods

Bioinformatic analysis for genes unique to Pyricularia and Colletotrichum

We searched the Mycocosm database from JGI using the MCL Clusters tool (Grigoriev et al., 2014). We searched for clusters genes contained in both *Colletotrichum* and *Pyricularia* species but excluded clusters containing genes also found in several genera intermediate to these two (such as *Neurospora* and *Chaetomium*). Our selected criteria identified a cluster that was widespread in *Colletotrichum* and appeared in two species of *Pyricularia* (*Pyricularia oryzae* and *Pyricularia* sp. *leersia)* with most *Colletotrichum* species having multiple homologs. The sequences of these genes were downloaded from JGI and then further utilized to BLAST the entire JGI Mycocosm database as well as NCBI to collect any additional gene hits. Any new hits were then also used to BLAST these databases again to find any new hits until this iterative BLAST process returned no new hits. Our conditions for BLAST searches used the default settings and used the translated amino acid sequences after manual annotations.

Study of the protein sequences coded for by these genes revealed they coded for small, cysteine rich proteins. All annotated protein coding sequences were analyzed using Signal P which predicted all proteins to have secretion signals with >90% likelihood. Both factors support the hypothesis that this gene family codes for putative effector genes. Annotated protein sequences were aligned using Clustal Omega (Sievers et al., 2011) with alignments visualized via Mview (Brown et al., 1998), consensus sequences were generated with Weblogo3 (Crooks et al., 2004), and a protein distance neighbor joining tree was generated using Phyml (Guindon et al., 2010) and visualized with ITOL (Letunic & Bork, 2021).

RNA Seq data

We used the best available RNA Seq data *Pyricularia oryzae* (Jeon et al., 2020)from *Colletotrichum graminicola* (Buiate et al., 2017, O'Connell et al., 2012) where expression levels were analyzed for members of our gene family during host infection.

Results

Characterization of the CWR gene family

Our iterative BLAST search process yielded 93 gene family members across many species of *Colletotrichum* (Table 6.1).

Table 6.1 Occurrence of CWR genes in *Colletotrichum* species

Species	CWR1	CWR2	CWR3
C. abcissum^	Yes	Yes	No
*C. acutatum**	Yes	Yes	No
C. aenigma+	Yes	No	Yes
C. asianum+	Yes	No	Yes
C. camelliae+	Yes	No	Yes
*C. caudatum**	Yes	No	Yes
*C. cereale**	Yes	Yes	Yes
*C. chlorophyti**	Yes	No	No
*C. costaicense**	Yes	Yes	No
*C. cuscutae**	Yes^{p}	Yes	No
*C. eremocholae**	Yes	Yes	Yes
*C. falcatum**	Yes	Yes	Yes
C. filicis+	Yes	No	No
*C. fioriniae**	Yes	Yes	No
C. fruticola+	Yes	No	Yes
C. gleosporiodes+	Yes	Yes^{p}	No
*C. godetiae**	No	Yes	No
*C. graminicola**	No	Yes	No
C. higginsianum^	Yes	No	Yes
C. incatum^	Yes	No	Yes
C. lupini^	Yes	Yes	No
*C. melonis**	Yes	Yes	No
*C. navitas**	Yes	Yes	Yes
*C. nympaeae**	Yes	Yes	No
*C. orchidophilum**	Yes	No	Yes
C. orbiculare^	Yes	No	Yes
*C. paranaense**	Yes	Yes	No
*C. phormii**	Yes	Yes	No
C. salicis^	No	Yes	No
C. scovillei+	Yes	No	Yes
C. siamense+	Yes	No	Yes
C. sidae+	Yes	No	Yes
C. simmodsii^	Yes	Yes	No
C. shisoi+	Yes	Yes	No
*C. somersetensis**	Yes	Yes	Yes
*C. sp. CBS 101611**	Yes	Yes	No
C. spinosum+	Yes	No	Yes
C. sublineola^	Yes	Yes	Yes
C. tanaceti+	Yes	No	Yes
*C. tofieldiae**	Yes	No	No
C. trifolii+	Yes	Yes	No
C. truncatum+	Yes	Yes	No
C. viniferum+	Yes	No	Yes
*C. zoysiae**	Yes	Yes	No
*Glomerella acutata**	Yes	Yes	No
Glomerella cingulata^	Yes	No	Yes

Occurrence of CWR genes in *Colletotrichum* species "*" Denotes sequences found in JGI Mycocosm only, "+" denotes sequences found in NCBI only, "^" denotes that sequences were found in both. "^{P}" Denotes a pseudogene

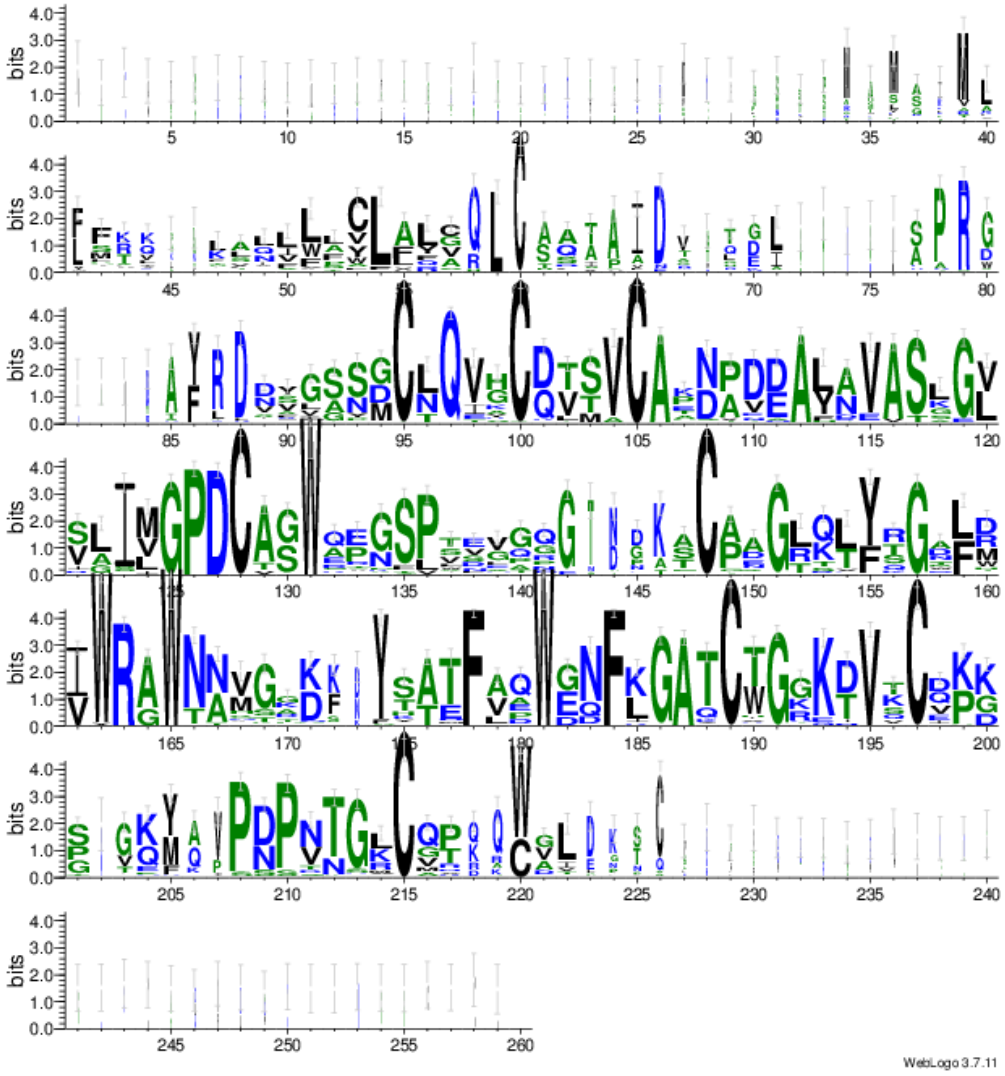
Alignment of the family of mature protein sequences revealed that the proteins were rich in Tryptophan leading us to name the family the Cysteine and Tryptophan Rich genes or CWR genes. Additionally, alignments and consensus sequences of the genes from *Colletotrichum* species show consistent presence of several other sequence motifs (Fig 6.1).

A)

<u>Figure 6.1</u> Alignment and Consensus Sequence of CWRs

A) Alignment of CWR1 of *Pyricularia oryzae* with CWR1, CWR2, and CWR3 from *Colletotrichum falcatum*. Conserved Cysteine and Tryptophan residues are highlighted.

B) Consensus sequence of CWR proteins of *Colletotrichum.* Even with nearly 100 different amino acid sequences and members from all 3 classes of the family, it is easy to identify the consistent presence of the namesake cysteine and tryptophan residues along with several other common motifs with eight conserved cysteines and five conserved tryptophans.



B)

We generated a neighbor joining tree of all CWR amino acid sequences and this revealed three distinct classes of CWR genes we named CWR1, CWR2, and CWR3. Most species have two or three of the classes although *C. graminicola*, *C. chlorophyti, C. godetiae,* and *C. tofieldiae* only have one. The three classes have characteristic c-terminal sequences that can be used to easily identify members of each class (Fig 6.2). Additionally, subgroups within each of the three CWR types seem to distinguish genes from pathogens of cereals and grasses and dicot pathogens (Figure 6.3).

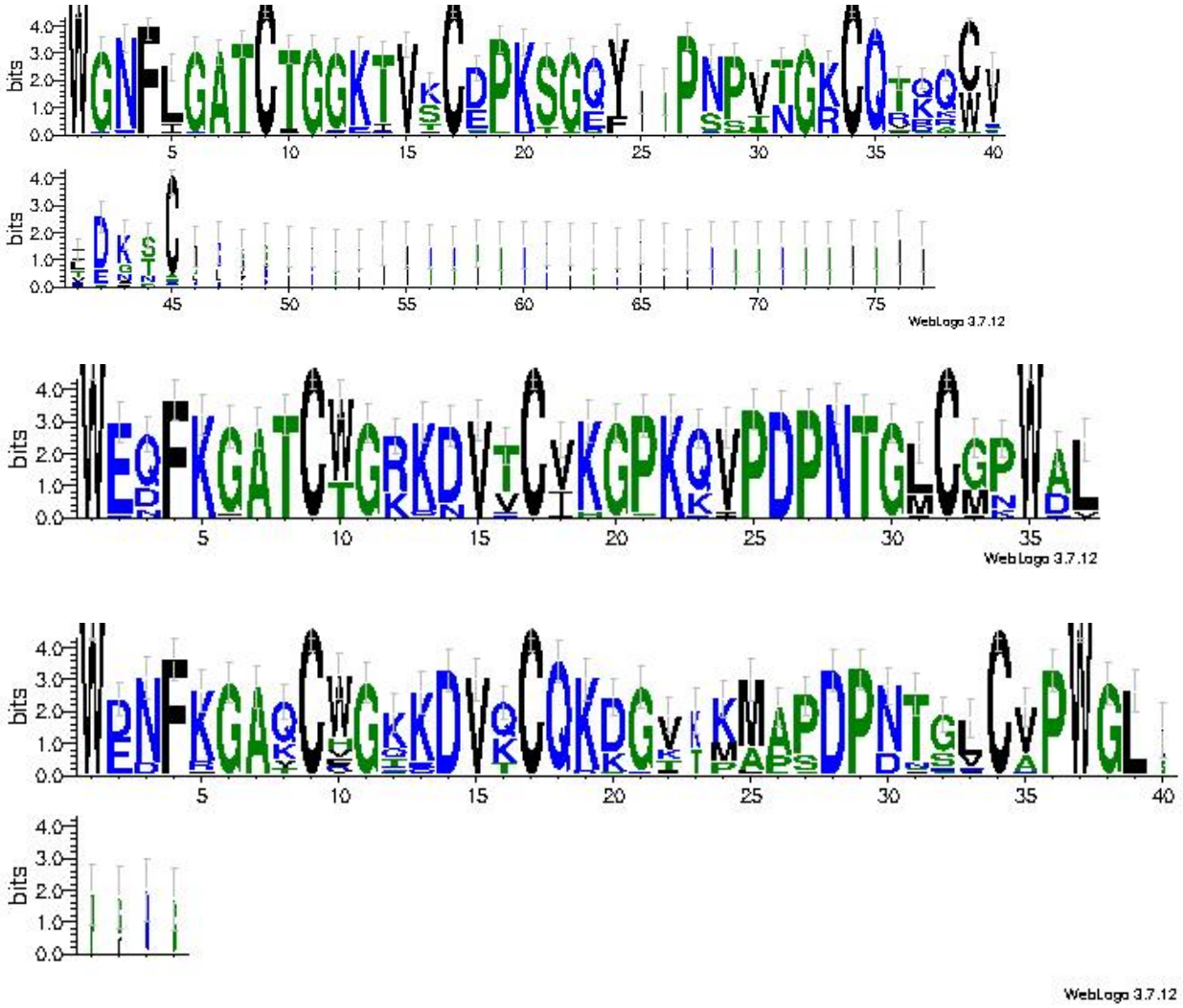


Figure 6.2 Consensus sequences for c-terminal characteristic sequences of CWRs

Comparison of Weblogo consensus sequences for the characteristic c-terminal sequences of the three CWR classes, from top to bottom the sequences come from CWR1, CWR2, and CWR3.

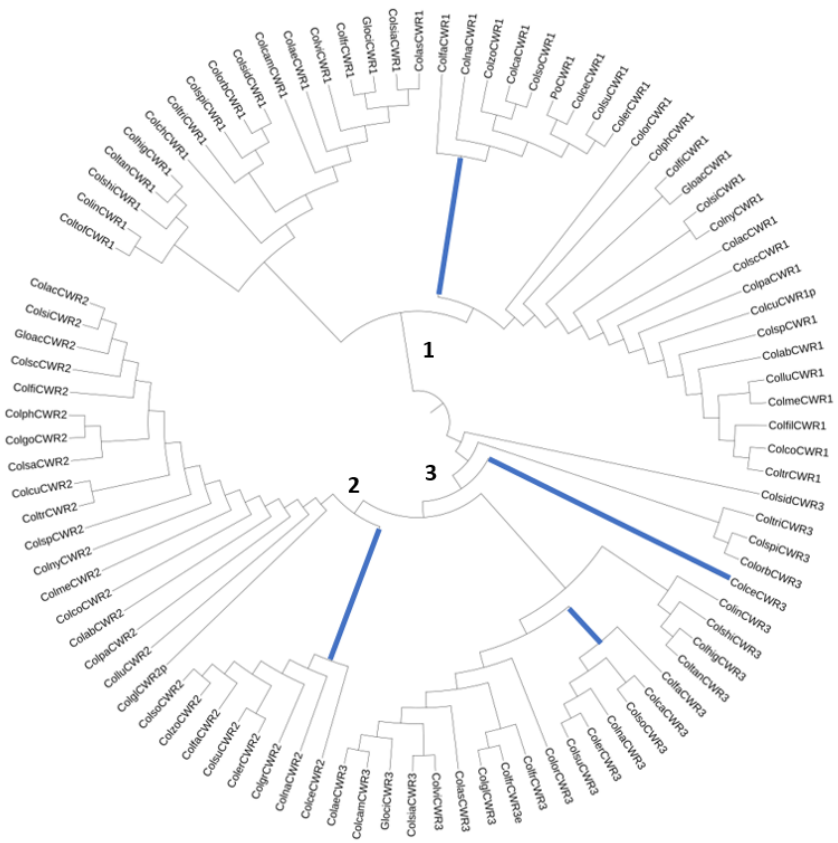


Figure 6.3 Tree of *Colletotrichum* CWR protein sequences

Neighbor joining tree of all *Colletotrichum* CWR amino acid sequences as well as the *P. oryzae* version. The branches for the three classes are labeled with the corresponding group number and each gene is labeled by species and gene class. Blue branches indicate pathogens of cereals and grasses. Tree generated with ITOL (Letunic & Bork, 2021)

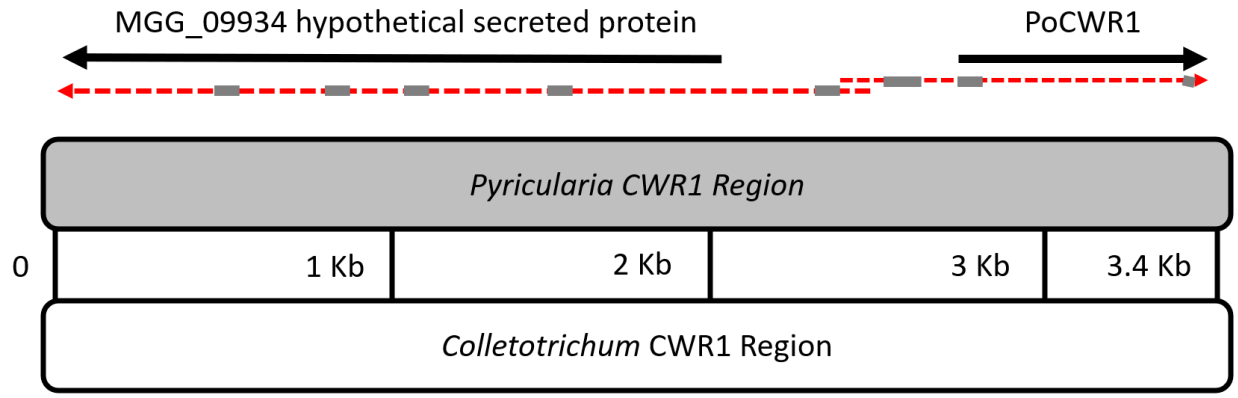
CWR genes in Pyricularia and other scattered species of Ascomycete fungi

In addition to the expansive family of CWR genes present in various species of *Colletotrichum* we discovered that a single CWR gene was present in *Pyricularia oryzae* and the Pyricularia sp. *Leersia* but not in the next most closely related species of *Pyricularia, P. grisea* or *P. pennisetigena*. The gene is identical in both *Pyricularia* species reflecting their close phylogenic relationship. The two species were estimated to have diverged 0.7 MYA(Gomez Luciano et al.,

2019). Only the gene from *P. oryzae* has been included in analysis here. There are no identifiable CWR genes present in species spanning the phylogenetic gap between the genera *Colletotrichum* and *Pyricularia*. This is consistent with an HGT event between an unknown species of *Colletotrichum* and the ancestor of the two species of *Pyricularia*.

Notably the potential region of HGT included a second gene found adjacent to CWR1 genes in *Colletotrichum* as well as in these *Pyricularia* species (Figure 6.4). This gene has been previously annotated as a dUTPase but this annotation appears erroneous and the function is unknown. Although the two genes are adjacent in the CWR1 genes, homologs of the adjacent gene are not found near CWR2 and CWR3 genes.

A)



B)

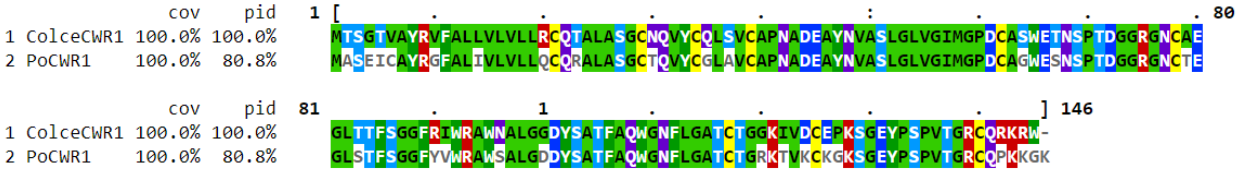


Figure 6.4 Region of HGT with CWR gene and alignment of most similar Colletotrichum protein sequence

A) Diagram of the region of Horizontal Gene Transfer between *Colletotrichum* species and *P. oryzae*. Black lines show region containing coding sequence. Red dotted lines indicate transcript RNA and grey boxes indicates positions of introns based on alignment with mRNA seq data

B) Alignment of *P. oryzae* CWR1 with CWR1 from *C. cereale*, the most closely related sequence on the protein neighbor joining tree. Alignment generated with Mview (Brown et al., 1998)

Interestingly, we discovered that there were several members of the CWR gene family scattered discontinuously throughout the Ascomycete phylogenetic tree. There does not seem to be a pattern to this distribution phylogenetically or in relation to the lifestyles of these fungi. In all,

we identified additional CWR genes in the following species: *Hyaloscypha finlandica*, *Hyaloscypha bicolor*, *Coleophoma cylindrospora*, *Coleophoma crateriformis*, *Protomyces lactucae-debilis*, and *Cenococcum geophilum*. All these CWRs are relatively dissimilar in the different genera and the addition of them to the CWR neighbor joining tree places them all far outside those genes from *Colletotrichum* and *Pyricularia* with percent identities ranging from 11% to 42% sequence identity to CWR1 from *C. cereale* (Figure 6.5)

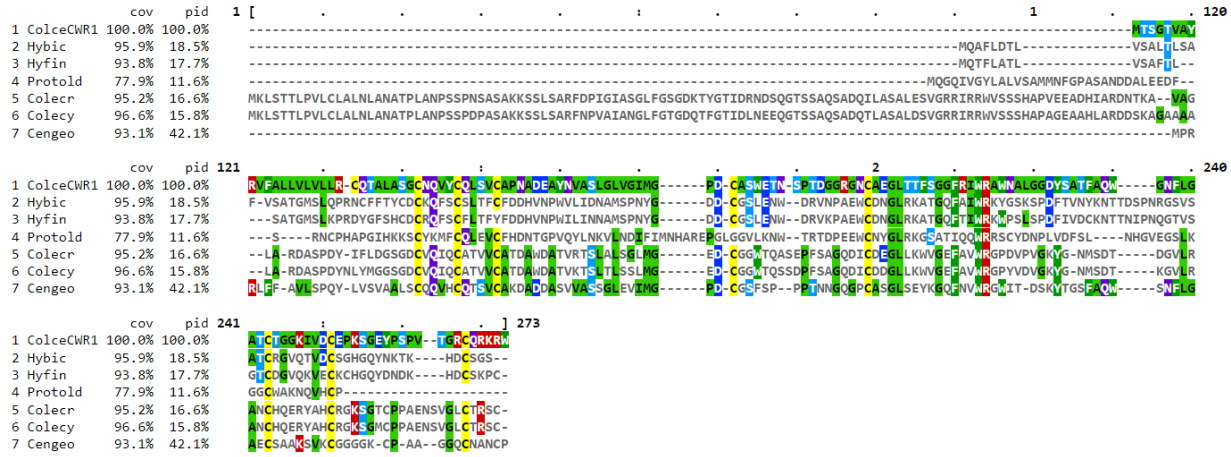


Figure 6.5 Alignment of discontinuously distributed CWRs with CWR1 from *C. cereale*

Key: ColceCWR1 = CWR1 from *Colletotrichum cereale*, Hybic = sequence from *h. bicolor*, Hyfin = sequence from *H. finlandica*, Protold = sequence from *P. lactucae-debilis*, Colecr = sequence from *C. crateriformis*, Colecy = sequence from *C. cylindrospora*, Cengeo = sequence from *C. geophilium.* Alignment generated with Mview *(Brown et al., 1998)*

Expression of CWR genes during infection

In the study of this new family of putative effectors we were interested in determining if any previous studies were conducted involving RNA expression data related to host infection where reads were generated for any of our genes. In searching the literature, we were able to find two such studies, in *Pyricularia oryzae* (Jeon et al., 2020*)* and in *Colletotrichum graminicola*

(O'Connell et al., 2012, Buiate et al., 2017) where expression levels of a CWR gene was measured in culture as well as during host infection. In the case of the *Pyricularia* study the expression of the co-transferred gene we identified was also found to be regulated during infection. This data is represented in Tables 6.2 and 6.3 below relative to 1000 actin fragments/kb. Table 6.2 shows RNA seq data for *P. oryzae* that shows expression levels of CWR1 and co-transferred gene MGG_11719 relative to 1000 actin fragments/kb, both genes are coordinately expressed at 36 hpi, during biotrophic growth (Jeon et al., 2020). Table 6.3 shows RNA seq data for *Colletotrichum graminicola* which only has CWR2 and shows low expression in mycelial grown on glucose but strong induction at 24 hpi (appressoria initiating penetration) and 36 hpi (growth within first invaded cell) (O'Connell et al., 2012, Buiate et al., 2017).

Table 6.2 ***P. oryzae*** **CWR gene expression**

Gene Name	Broad ID	Mycelium	18hpi	27hpi	36hpi	45hpi
none	MGG_09934	2	5	6	120	19
PoCWR1	MGG_11719	0	7	9	121	29
Actin	MGG_O3982	1000	1000	1000	1000	1000

Table 6.3 ***C. graminicola*** **CWR gene expression**

Gene	Gene ID	Glucose	24 hpi	36 hpi	60 hpi
ColgrCWR2	GLRG_11583	1.5	277	288	51
Actin	GLRG_03056	1000	1000	1000	1000

Discussion

For this study we set out to identify genes that had patterns of appearance like HAG genes and that had indications they may have been acquired via Horizontal Gene Transfer between species. We were able to identify a family with a wide and discontinuous phylogenetic distribution like HAG genes. The ease with which we identified the CWR genes suggests this phenomenon may be common.

In this study we identified that the CWR family of proteins had several features common to effectors: small, cysteine rich, secreted proteins, and their expression is upregulated during stages of host infection in *Colletotrichum* and *Pyricularia*. Yeast two hybrid studies probing various libraries from important hosts of the pathogens possessing these CWR genes could shed light on common targets of CWR effectors. Even though they are secreted we do not know if the CWR proteins are secreted into the host cytoplasm like the HAG effectors. If so, it will be interesting to see if CWR effectors from different *Colletotrichum* species that parasitize different hosts (such as grasses/cereals vs dicots) target the same or similar host proteins. The three distinct classes of CWR proteins appear to have originated due to duplication and divergence, as such it will be of interest to determine if they have evolved distinct roles as they have diverged.

The presence of CWR family members in widely distributed species of Ascomycete fungi beyond *Colletotrichum* and *Pyricularia* is also a matter of interest that warrants more study. When looking at the lifestyles of these other fungi they are mostly endophytic and non-pathogenic although all of them have been identified as having some association with plants. *Cenococcum geopilium* is unique among these fungi as it has a near ubiquitous global distribution as a fungal endophyte. The role for the CWR gene of *C. geopilium may* play a potentially different role from the CWRs of *Colletotrichum* plant pathogens One possibility is that, at some point in history, these genes were acquired via HGT from *Colletotrichum* ancestors like we have proposed for the *Pyricularia* species. These genes are much more divergent in sequence so these potential HGT events must have occurred long ago if this is the case. There is a robust availability of genome sequences throughout the Ascomycete fungi and only a few genera have these CWR genes making it seem less likely that the CWR genes were lost from all other Ascomycetes.

In this study we made use publicly available genome sequences to search for patterns of gene occurrence associated uniquely with *Colletotrichum* and *Pyricularia* based on the hypothesis that if HAG genes were horizontally transferred between these genera, there could be other horizontally transferred genes. Further, such genes might relate to a role in pathogenesis, and if so, would display expression patterns consistent with this idea. We propose the CWR genes are effectors that may contribute to disease in a wide variety of plant hosts of economic significance parasitized by *Colletotrichum* species. This kind of analysis could potentially be adapted to automated searches to identify all such patterns resembling HGT and cross-screening against published infection-specific gene expression data. This study is an example of how the massive amount of available genome sequence data could be studied to further our understanding of effector evolution.

REFERENCES

Brown NP, Leroy C, Sander C, 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* **14**, 380-1.

Buiate EaS, Xavier KV, Moore N, *et al.*, 2017. A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species Colletotrichum graminicola and Colletotrichum sublineola. *BMC Genomics* **18**, 67.

Chen M, Farmer N, Zhong Z, *et al.*, 2022. HAG Effector Evolution in Pyricularia Species and Plant Cell Death Suppression by HAG4. *Mol Plant Microbe Interact* **35**, 694-705.

Crooks GE, Hon G, Chandonia JM, Brenner SE, 2004. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-90.

Friesen TL, Stukenbrock EH, Liu Z, *et al.*, 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet* **38**, 953-6.

Gardiner DM, Mcdonald MC, Covarelli L, *et al.*, 2012. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathog* **8**, e1002952.

Gomez Luciano LB, Tsai IJ, Chuma I, *et al.*, 2019. Blast Fungal Genomes Show Frequent Chromosomal Changes, Gene Gains and Losses, and Effector Gene Turnover. *Mol Biol Evol* **36**, 1148-61.

Grigoriev IV, Nikitin R, Haridas S, *et al.*, 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* **42**, D699-704.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O, 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-21.

Jeon J, Lee GW, Kim KT, *et al.*, 2020. Transcriptome Profiling of the Rice Blast Fungus Magnaporthe oryzae and Its Host Oryza sativa During Infection. *Mol Plant Microbe Interact* **33**, 141-4.

Letunic I, Bork P, 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W6.

Nevoigt E, Fassbender A, Stahl U, 2000. Cells of the yeast Saccharomyces cerevisiae are transformable by DNA under non-artificial conditions. *Yeast* **16**, 1107-10.

O'connell RJ, Thon MR, Hacquard S, *et al.*, 2012. Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. *Nat Genet* **44**, 1060-5.

Sievers F, Wilm A, Dineen D, *et al.*, 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

Soanes D, Richards TA, 2014. Horizontal gene transfer in eukaryotic plant pathogens. *Annu Rev Phytopathol* **52**, 583-614.

CHAPTER VII

CONCLUSION

Our understanding of fungal effectors, their emergence, their evolution, and their target repertoires is a rapidly changing field of study. It has become clear that effectors play diverse and multiple roles in contribution to pathogenic states in plant hosts and the work discussed in this thesis has added support to this view.

In Chapter 2 I presented the results of a variety of yeast two hybrid experiments that used five HAG family paralogs from the rice blast fungus *Pyricularia oryzae* (HAG4, HAG10, HAG2, HAG7 and HAG13) and two HAG family orthologs from *Pyricularia grisea* (PgHAG4) and *Pyricularia* sp. *leersia* (PlHAG10) to screen rice target proteins for interactions. I identified four rice proteins from our rice cDNA library (SnRK1, bZIP39, JMJ715, and ILR3) that were screened against this subgroup of HAG effectors. Additionally, I also screened this group of HAG effectors with four rice proteins (Di19-5, Di19-6, Di19-X, and JMJ705) involved with histone lysine methylation/demethylation, the process centrally involving our two most promising targets SnRK1 and JMJ715. This process of methylation/demethylation is known to be associated with regulation of expression/repression of defense genes in response to pathogen challenge (Roy et al., 2018, Margueron & Reinberg, 2011, Laugesen et al., 2019, Shen et al., 2021, Hou et al., 2015, Dutta et al., 2017, Li et al., 2013). My results indicated that several of these HAG effectors interact with rice proteins involved with both histone lysine methylation and histone lysine demethylation. In Chapter 3 my results indicate that at least some of these interactions occur in planta in tobacco, making it likely that these interactions really do have a role in pathogenesis. These results have led me to construct a potential model for HAG effector action in rice which can be seen in figure 7.1.

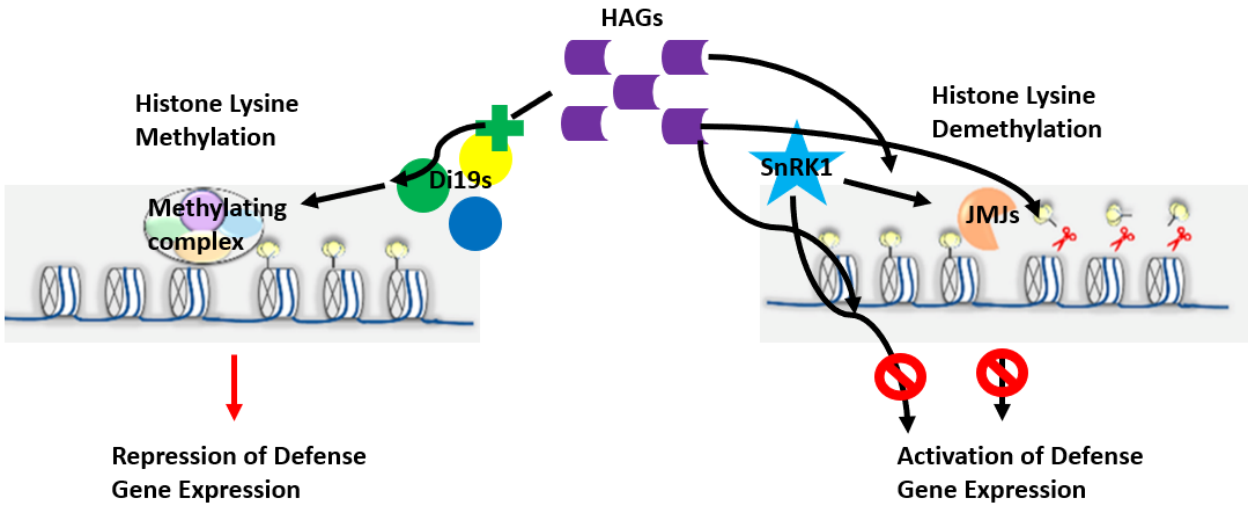


Figure 7.1 Potential model for HAG effector action in histone lysine methylation/demethylation

From the literature it is known that Di19 proteins act in trans with a PRC complex that works to methylate lysines on histones leading to the inaccessibility of defense genes for expression. Conversely, it is known that JMJ family proteins demethylate these lysines on histones and open up defense related genes for expression during pathogen challenge. It is also known that SnRK1 acts in expression of these defense genes both through interaction with JMJ proteins and directly. I propose that HAG effectors work with some Di19 proteins to recruit them to the methylation complex leading to increased repression of defense gene expression while some HAG genes interact with SnRK1 and JMJ proteins to inhibit activation of defense gene expression. This potential model places HAG genes at the center of this histone lysine methylation/demethylation process and would allow the fungus to impact defense responses from both sides. Future experiments studying histone lysine methylation during rice infection with *P. oryzae* may be able to measure expression levels of specific defense associated genes in infections with *P. oryzae*

strains with all HAGs compared to those missing several HAGs and give a clearer view on HAG impact on rice defense responses.

In addition to this potential role for HAGs in regulation of histone lysine methylation I also examined HAG effector interactions with Synaptotagmin-2 via Y2H analysis. Existing literature indicates that Synaptotagmin-2 plays a role in delivery of effectors to initially infected plant cells and cells infected subsequently (Liu et al., 2021, Li et al., 2021). I was able to determine that some of these same HAG effectors interact with Synaptotagmin-2. Although, I did not explore this interaction further to determine their biological relevance it will be interesting to see if these interactions hold true in further experiments in planta. This may be a second mechanism by which HAG effectors prime the plant cells for infection by the fungus.

In Chapter 4 I conducted Y2H experiments with HAG family members from other species of *Pyricularia*, species of *Colletotrichum*, and a necrotrophic pathogen *Cochliobolous heterostrophus*. These orthologous HAGs maintained some of the same targets identified to be involved in these important defense pathways. It will be of interest to see if these targets are in fact biologically relevant and if they too have an impact on plant defense response to these other pathogens, especially so with *Cochliobolous heterostrophus* which has a different lifestyle than the other pathogens discussed. Future experiments could determine if fungal strains lacking these HAGs are less successful at pathogenesis than those that possess theses HAGs.

In Chapter 5 I explored Y2H interactions with our identified targets and chimeric HAG effectors created by swapping exons of the wild-type HAGs. These chimeric HAGs maintained many of the same interactions as their whole counterparts. If, in future studies, these interactions are determined to be preserved biologically than we may have demonstrated novel mechanisms for

effector evolution that could maintain virulence on existing hosts and potentially create new pathogenic interactions on novel hosts.

My results from Chapter 6 add support to the idea that there exist many yet to be studied fungal effector families that have been exchanged between pathogens that occupy similar ecological niches and exist in the wild on similar host species.

REFERENCES

Dutta A, Choudhary P, Caruana J, Raina R, 2017. JMJ27, an Arabidopsis H3K9 histone demethylase, modulates defense against Pseudomonas syringae and flowering time. *Plant J* **91**, 1015-28.

Hou Y, Wang L, Wang L, *et al.*, 2015. JMJ704 positively regulates rice defense response against Xanthomonas oryzae pv. oryzae infection via reducing H3K4me2/3 associated with negative disease resistance regulators. *BMC Plant Biol* **15**, 286.

Laugesen A, Hojfeldt JW, Helin K, 2019. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol Cell* **74**, 8-18.

Li T, Chen X, Zhong X, *et al.*, 2013. Jumonji C domain protein JMJ705-mediated removal of histone H3 lysine 27 trimethylation is involved in defense-related gene activation in rice. *Plant Cell* **25**, 4725-36.

Li Z, Variz H, Chen Y, Liu SL, Aung K, 2021. Plasmodesmata-Dependent Intercellular Movement of Bacterial Effectors. *Front Plant Sci* **12**, 640277.

Liu J, Zhang L, Yan D, 2021. Plasmodesmata-Involved Battle Against Pathogens and Potential Strategies for Strengthening Hosts. *Front Plant Sci* **12**, 644870.

Margueron R, Reinberg D, 2011. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-9.

Roy S, Gupta P, Rajabhoj MP, Maruthachalam R, Nandi AK, 2018. The Polycomb-Group Repressor MEDEA Attenuates Pathogen Defense. *Plant Physiol* **177**, 1728-42.

Shen Q, Lin Y, Li Y, Wang G, 2021. Dynamics of H3K27me3 Modification on Plant Adaptation to Environmental Cues. *Plants (Basel)* **10**.