PROJECT DESCRIPTION

The role of transposable elements and DNA methylation on immunity gene regulation and diversification in maize and the model monocot *Setaria viridis*

Introduction

Cereal crops including maize, wheat, barley, and rice make up greater than 40% of the world's caloric intake and are important foundations for global food security (Savary et al., 2019). Despite management strategies, 15-20% of US wheat and maize crops are lost to pests and pathogens annually (Savary et al., 2019). Control of pathogens is often attained through application of pesticides; however, this is often not economically feasible for growers. In addition to the economic impact of pesticide use, there can also be negative environmental and health impacts associated with the use of pesticides (Zhang, 2018). An alternative approach is growing cultivars that encode genetic resistance. Discovery of resistance often comes from screening large domestic and wild germplasm collections (Lenne and Wood, 1991). These screens have been made more efficient by basic plant immunity research which has provided foundational knowledge on the types of genes and molecular mechanisms often associated with disease resistance (Kourelis and van der Hoorn, 2018). Understanding which genomic regions to focus on accelerates resistance gene discovery and creates opportunities to engineer synthetic resistance genes.

An emerging area in the study of plant disease resistance is the role of dynamic DNA methylation on immunity gene expression. In plants DNA methylation occurs at cytosine residues in any context (CG, CH, or CHG where H is any nucleotide other than guanine). Studies in *Arabidopsis* have shown that methylation is deposited by DOMAINS REARRANGED METHYLASE 2 (DRM2) with the help of DECREASE IN DNA METHYLATION (DDM1) and maintained by METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLASE2 and 3 (CMT2 and CMT3) (Zhang et al., 2018). Small RNAs guide methylation machinery to selectively methylate genomic regions, often repetitive sequences including transposable elements (TEs). Active demethylation occurs through excision and repair carried out by REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DME-LIKE2 (DML2), and DML3 (Agius et al., 2006; Mok et al., 2010; Zhang et al., 2018).

The methylation status of genic and intergenic DNA sequences changes when plant cells undergo stress (Dowen et al., 2012; Deleris et al., 2016; Kawakatsu and Ecker, 2019). TEs, generally highly methylated which keeps their expression levels low, are primary targets of these methylation changes. Demethylation of TEs can lead to increased expression of both the TE and nearby genes (Lisch, 2009). Plant immunity genes are often located in TE-rich regions of the genome and differential TE methylation is believed to play an important role in regulating the expression of these genes (van Wersch et al., 2020). Studies linking methylation and plant immunity have shown that decreased methylation is associated with increased resistance to biotrophic bacterial and oomycete pathogens (Dowen et al., 2012; Luna and Ton, 2012; Gohlke et al., 2013). Demethylation deficient *Arabidopsis* mutants are more susceptible to bacterial and fungal pathogens (Le et al., 2014). Little is known about demethylation mechanism and specificity, but ~80% of regions targeted by demethylases are near annotated genes (Penterman et al., 2007). Demethylation of plant immunity gene promoters may be a mechanism for rapid expression changes necessary for effective disease resistance.

Immunity genes from the nucleotide-binding and leucine-rich repeat (NLR) class are among the largest gene families in plants (Clark et al., 2007; Ossowski et al., 2008) and, along with TEs, exhibit some of the highest diversity even within a given species (Van de Weyer et al., 2019). Recent sequencing efforts have focused on generating pan-genomes for plant species rather than relying on a single reference genome (Bayer et al., 2020). In maize, the long-read sequence assembly of the maize nested association mapping (NAM) lines (McMullen et al., 2009) is an exciting resource for examining NLR immunity gene and TE diversity. RNAseq data for all NAM lines and whole genome bisulfite data for a sub-set of lines have been generated, allowing intersection of methylation and expression data with gene and TE diversity. These analyses will allow us to observe correlative relationships (**Objective I**), however exploring the mechanisms involved in maize is difficult due to lethality of methylation pathway mutants and the long timelines associated with transformation in general. An appropriate model system for maize

would facilitate testing hypotheses generated from the analysis of the wealth of information generated by the NAM sequencing project.

In the field of plant immunity much of our knowledge comes from work on the model dicot *Arabidopsis thaliana (Agrawal, 2018)*. The importance of *A. thaliana* as a model cannot be questioned, however a more closely related plant would serve as a better model for maize. *Setaria viridis* is a diploid C4 grass that grows in diverse niches around the globe. *S. viridis* is being developed as a monocot model due to its close phylogenetic relationship to important cereal crops, small genome size, short lifespan, and ease of growth. Importantly, *S. viridis* transformation protocols have been established and seed can be collected from transformants in only eight to ten weeks (Brutnell et al., 2010; Li and Brutnell, 2011). CRISPR/Cas generated disruptions of the methylation and demethylation pathway in *Setaria* would provide a means to determine the impacts of hypo- and hyper-methylation during normal growing conditions and when plants are experiencing biotic stress (**Objective II**).

To date, all plant studies looking at dynamic DNA methylation and gene expression changes in response to biotic stress have used homogenized whole tissues. This approach has been sufficient to detect genome-wide changes in DNA methylation and differentially expressed genes, however plant cells that are directly interacting with a pathogen respond differently than cells that are distal to the infection. The ability to parse these pools of cells would allow finer resolution and detection of signal that may go undetected in whole tissue homogenate. The isolation of nuclei from tagged specific cell types (INTACT) system utilizes a transgenic plant with a cell-type or condition-specific promoter driving a reporter protein fused to a nuclear membrane targeting motif (Deal and Henikoff, 2010). First developed for use in *Arabidopsis*, the INTACT system has been modified for use in tomato and the monocot rice (Ron et al., 2014; Reynoso et al., 2018). Tagged nuclei are collected with a crude nuclear extraction protocol and sorted via magnetic bead-based separation. The engineering of a reporter driven by a biotic stress inducible gene promoter to drive the INTACT system would allow the sorting of nuclei from cells that are directly experiencing stress for downstream DNA and RNA sequencing. This would improve the sensitivity for detecting differentially methylated regions and localized gene expression changes (**Objective III**).

The objectives proposed in this application require the development of skills related to transposon biology, DNA methylation, and the management of large complex datasets while building on a strong foundation in plant-microbe interactions to get a wholistic view of the complex relationships between TEs, DNA methylation, and immunity gene expression.

Research Objectives

Objective I: Determine the impact of DNA methylation and proximity to TEs on maize immunity gene expression in the diverse NAM lines. The repetitive nature of both TEs and plant immunity genes make them particularly challenging to sequence and annotate prior to the advent of long-read sequencing technologies. NLR immunity genes are often found in TE-rich regions of the *Arabidopsis* genome (Lai et al., 2020). The recent sequencing and high-quality annotation of the 25 NAM maize lines (maizegdb.org) provide an opportunity to examine the relationship between TEs and immunity genes in an important monocot crop species.

Using the NAM data I will first generate a pan-maize NLR-ome by identifying and extracting immunity gene sequences from each of the 25 genomes. Phylogenetic relationships will be determined by alignment of the NLR central conserved nucleotide-binding region. NLRs will be classified as common (found in most lines) or rare (found in one or only a few lines). Next, I will determine the number of immunity genes that contain TEs either within gene sequences or in the five kilobase region up and downstream of the gene. These data will be compared with a control 'non-NLR' gene set to determine if, as has been reported in *Arabidopsis*, maize NLRs are significantly more likely to colocalize with TEs. Existing RNA sequencing data will be used to calculate expression levels for genes and TEs in the NAM lines.

DNA methylation data has been generated for a sub-set of the maize NAM lines. With this data, I will examine DNA methylation of immunity genes in these sequenced lines. A recent study has created a

framework for classifying genes as either unmethylated, having gene body-like or TE-like methylation patterns (Kenchanmane Raju et al., 2020). I will modify this framework to classify maize genes into these categories to determine if NLRs are unique in their methylation patterns when compared to a control 'non-NLR' gene set. Taken together, the immunity gene sequence diversity, DNA methylation data, and expression levels will allow us to begin to make hypotheses about how these factors interact to generate the diverse disease resistance spectra within this species. Due to the difficulty of generating methylation mutants in maize, a model system is desirable for testing these hypotheses.
Expected Outcomes: 1) A pan-maize NLR-ome and NLR phylogeny. 2) Expression patterns, TE association, and methylation status for all NAM NLR immunity genes. 3) Classification of all maize genes into distinct methylation type categories.

Objective II: Monitor DNA methylation and gene expression changes in *Setaria viridis* methylation and demethylation pathway mutants. CRISPR/Cas mutants of the MET1, DRM, CMT, and DDM methylation gene families are currently in the transformation pipeline at UMN. Transgene free knockouts of all gene family members for MET1 and DRM have been acquired and there is evidence for successful editing at the other targets (Table 1). Tissue from transgene free Δdrm mutant progeny has been sampled and submitted for whole genome bisulfite and RNA sequencing. Preliminary analysis of the bisulfite sequencing data indicates a global reduction in CHH methylation across genes and TEs (Fig. 1A and B). RNAseq analysis identified 690 genes that are differentially expressed in the Δdrm mutant when compared with wildtype plants grown under normal conditions. Visualization of CHH methylation of differentially expressed genes indicates that, at least in some cases, CHH islands at TE sequences upstream of genes have changed drastically in the Δdrm mutant (Fig. 1C). These data are promising, and it will be interesting to see how they compare to data for other methylation pathway mutants. It is also important to include mutants in the demethylation pathway. *Setaria* homologs of the *Arabidopsis* demethylation proteins ROS1 and DME have been identified and I am currently designing CRISPR/Cas constructs to generate these targeted gene disruptions (Table 1).

		A	B
Methylation	Met1	Sevir.2G06230	Sevir.9G05350
	DDM	Sevir.2G02780	Sevir.9G11100
	DRM	Sevir.9G57480	Sevir.9G49620
	CMT	Sevir.9G30620	Sevir.9G48290
De-Meth	ROS1	Sevir.2g158400	NA
	DME	Sevir.3g228900	NA

Table 1: *Setaria* homologs of methylation and demethylation genes. There are two copies of each of the methylation genes, A and B. Grey rows indicate targets for which gene edited *Setaria* plants have been generated

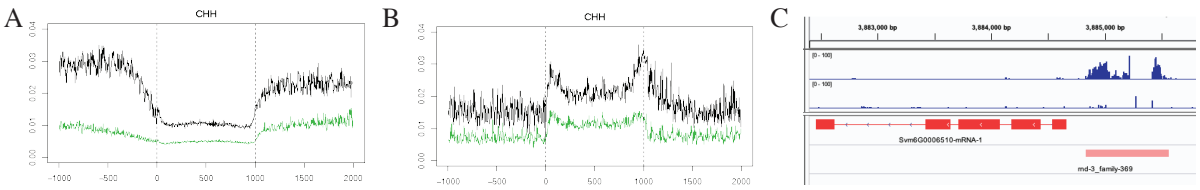

Figure 1: Metaplots showing average % CHH methylation in wildtype (black) and a Δdrm mutant (green) across (**A**) genes and (**B**) TEs in the *Setaria* genome. (**C**) CHH methylation in the wildtype (top) and Δdrm mutant (bottom) overlaps a transposable element upstream of differentially expressed gene Svm6G0006510

Following a primary assay carried out to compare all mutant and wildtype *Setaria* expression and methylomes under normal growth conditions, mutants will be challenged with the bacterial pathogen *Xanthomons vasicola* pv. holcicola (Xvh) and the oomycete pathogen *Sclerospora graminicola* (Sg). Inoculation methods for both Xvh and Sg on closely related monocots have been established (Kulkarni et al., 2016; Lang et al., 2017). Disease severity will be scored, and tissue will be sampled for bisulfite and RNA sequencing. Data will be analyzed following established best practices currently used by the Springer group. Genome-wide identification of differentially methylated regions (DMRs) will be carried out with a focus on DMRs located within 2 kb of annotated immunity genes. Differential expression analyses of TEs and immunity genes will be performed based on the RNAseq data. Together these data will indicate whether *Setaria* methylation mutants are more or less susceptible and what methylation and gene expression changes underlie these differences in susceptibility
Expected Outcomes: 1) generation of a panel of methylation/demethylation mutants for the *Setaria* research community 2) The first transcriptomic and methylation study of *Setaria*-pathogen interactions 3) Association of differentially methylated genomic regions with immunity gene expression 4) Expertise in bisulfite sequencing and analysis.

Objective III: Generate stress reporter *Setaria viridis* lines to sort nuclei from infected and uninfected cells prior to methylation and RNA sequencing. While the experiments described in Objective II will provide valuable insights into the mechanisms of biotic stress response and the role of DNA methylation in mediating this response, these experiments utilize crude plant tissue homogenate. Plant infection by pathogens often occurs at distinct foci, and cells that are distal to these foci are likely experiencing very different stresses from those cells in direct contact with pathogens. Meaningful insights into the responses of plants to pathogens will likely rely upon increased understanding of cell-specific responses. Objective III will focus on developing key resources to enable these types of studies.

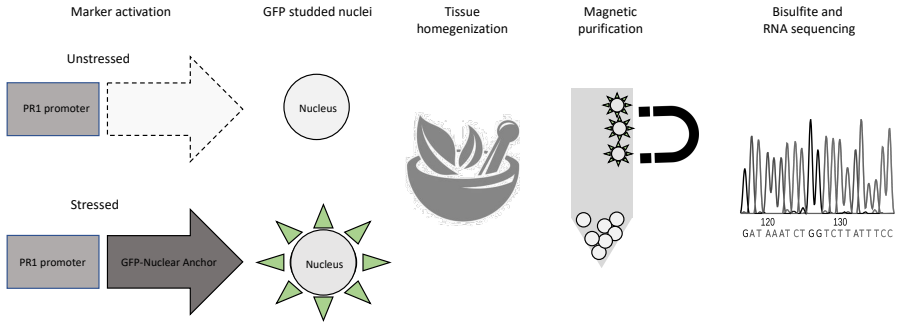

Figure 2: INTACT workflow. Transgenic reporter plants express nuclear anchored GFP when activated by stress. Whole tissues are homogenized. Magnetic antiGFP beads are used to purify nuclei only from cells that are expressing the reporter. Stressed and unstressed cells can be sequenced separately.

The INTACT system utilizes a genetically encoded, nuclear bound, reporter to allow sorting of nucei from crude plant homogenate (Fig 2). Initially, two *Setaria* reporter lines will be designed to respond to two different types of biotic stresses. The first will use the promoter of the *Setaria* homolog of the salicylic acid responsive PATHOGENESIS-RELATED 1 (PR1) gene which is known to be induced by biotrophic pathogens. A second reporter will encode the jasmonic acid and necrotrophic pathogen induced promoter of the ALLENE SYNTHASE OXIDE 2-LIKE (AOS) (Kouzai et al., 2016). These reporter lines will allow sorting nuclei from cells that are directly experiencing stress from cells distal to the infection site. Following sorting, RNA and DNA will be extracted from pooled nuclei for bisulfite and RNA sequencing. Results of INTACT sorting will be compared with whole tissue analysis to determine if additional signal for dynamic methylation or gene expression is detected.

Expected Outcomes: 1) Generation of biotic stress reporter *Setaria* lines that will be of value to the plant-microbe interaction community and a foundation for my future independent research program 2) High resolution analysis of differentially methylated regions and gene expression in cells sorted from tissue homogenate 3) Comparison of DNA methylation and expression data from whole tissue and sorted nuclear preparations.

Training Objectives

In order to transition to career independence, it is important that I develop both technical and leadership skills during my postdoctoral training. The development of technical skills related to DNA methylation and gene expression will be accomplished by interaction with Dr. Nathan Springer, his lab members, and other members of the UMN community. In addition, through bi-weekly meetings with Dr. Springer, we will critically evaluate my technical and career development progress. Technical and professional development workshops (virtual or in-person) will be identified by Dr. Springer and myself and attended when appropriate.

Dr. Springer's expertise in DNA methylation combined with the resources available through the University of Minnesota will provide a supportive environment for me to acquire the skills necessary for success in the next step of my career. The generation of fluorescent reporter *Setaria italica* lines will generate a valuable tool to launch my independent research group. These lines represent a novel resource to address plant-biotic interactions that will be critical for several research projects to investigate the basis of plant disease resistance. For example, plants that encode multiple reporters could be used to sort nuclei from cells experiencing different or overlapping stresses, such as heat and disease. Dr. Springer is supportive of me developing these materials and using them as the foundation of my research program when I depart his group.

Choice of sponsoring scientist and host institution

Dr. Springer has expertise in DNA methylation and gene expression and has worked in several monocot systems. Members of the Springer lab have established bioinformatics pipelines for large-scale methylome, transcriptome and genome variation studies. Interactions with Dr. Springer and members of his research group will provide valuable opportunities as I develop my bioinformatics skillset and perspectives for –omic studies of crop plants. The Springer group works closely with the group of Dr. Candice Schmidt at UMN. Dr. Schmidt and her group will be a valuable resource in the analysis of the NAM maize data. Dr. Springer is involved in collaborative projects with Dr. Feng Zhang to develop genome edited *S. italica* lines. I will work with Dr. Zhang to develop materials needed for Objectives II and III. I have also reached out to UMN plant pathology department professor Dr. Brian Steffenson who has agreed to help with sourcing pathogens and the development of disease assays. This multi-department team will expand my professional network and facilitate input from a diverse team of experienced researchers.

I will have biweekly one-on-one meetings with Dr. Springer at which professional development and scientific progress will be discussed. At these meetings, Dr. Springer will help me identify development opportunities, provide feedback on experimental design, and help troubleshoot any technical issues I encounter. As mentioned above, I have developed relationships with additional faculty at UMN including Drs. Candice Hirsch, Feng Zhang, and Brian Steffenson, each of whom will provide additional access to resources and guidance during the course of the fellowship and beyond.

Public Outreach

The University of Minnesota Twin Cities campuses are unique among public universities with large agriculture programs in that they are located in major metropolitan areas in Minneapolis and Saint Paul. This proximity to urban areas provides an opportunity to expose young urban Minnesotans to agriculture and plant science as potential academic and career paths. Despite these opportunities to interface with the local community, many of the state's lowest performing schools are in the Minneapolis

and Saint Paul districts. I'm looking forward to contributing to outreach efforts specifically aimed at exposing students from these demographics to plant science.

A group of graduate students and postdocs at UMN has created an outreach program called Market Science (www.marketsci.org). This group carries out outreach events at three Twin Cities farmer's markets to connect with adults and youth in the community. In addition to weekly markets, Market Science also has a presence at several community festivals throughout the year. I have joined the group and I'm excited about participating in events in the coming season. In my first year I will participate in existing programs and encourage the group to expand their presence to additional markets and events in under-represented communities such as the Hmong Farmer's Market. In the second year of the fellowship I will design and run an outreach module using the phenotypic diversity of corn ears as a gateway to plant genetics. This program will be scalable for audiences of any age or background. In the final year of the fellowship I will continue my activities with Market Science while training new graduate students to continue developing and deploying the maize plant genetics outreach activity.

Broader Impacts

Transposable elements and DNA methylation are largely understudied in the plant immunity community despite the evidence that they play integral roles in plant immunity gene diversification and regulation. In addition to contributing knowledge of how these genomic features interact in plant defense, each of objective generates a useful resource for the plant science community. Objective I generates a maize pan-NLRome, useful for those working in maize disease resistance and immunity gene evolution. Objective II generates a series of *Setaria* methylation and demethylation mutants that will have many uses outside the scope of this proposal. The stress reporter lines generated in Objective III will provide a blueprint for generating reporters for additional stresses such as heat or salinity. As described above, the stress reporter lines will be a foundational part of my future research. In addition to being disseminated at domestic and international scientific meetings such as PAG and the American Society of Plant Biologists, findings from this project will be posted as preprints and submitted to appropriate open-access peer reviewed scientific journals. Biological materials and data generated in this project will be shared with the research community (see Data Management Plan).

Timeline

I joined the Springer Lab in September of 2020, nine months prior to the start of the fellowship. Since joining, I have been training on the TE annotation, RNAseq, and DNA methylation sequencing pipelines. I carried out the initial analysis of the Δdrm mutant and will be continuing this analysis for the next several months. This lead time will allow me to begin productive research at the onset of the fellowship.

Efforts on **Objective I through III** will be staggered over the duration of the fellowship. All resources are currently available for the completion of **Objective I**, so this will be prioritized with the goal of finishing analysis and submitting a manuscript before the end of year 1. **Objectives II and III** require the generation of genome edited or transgenic *Setaria* plants. It will be important that transformation materials are designed, built, and delivered to Dr. Zhang's facility for transformation by the middle of year 1. Work on **Objective II** will begin before year 2. Ideally the demethylation mutants will be available at this time, but if not, work will start with available methylation mutants and the remaining mutants will be assayed as they become available. The biotic reporter lines required for **Objective III** are traditional transgenic plants and should be easier to generate and screen than the methylation/demethylation pathway knockouts. Experiments outlined in **Objective III** will begin in mid-year 2 and continue through the end of the fellowship, though lab work will ramp down in the last quarter of year 3 to allow adequate time for data analysis and documentation.

Upon completion of the fellowship we will have three submitted manuscripts, one from each objective. Additionally, we will have generated resources for the research community including *Setaria* methylation/demethylation pathway mutants, and *Setaria* stress reporters.