The role of histone marks in genome organization and histone cross-talk in *Magnaporthe oryzae*

by

David Rowe

B.S. Plant Sciences, University of Maryland, College Park, 2020
B.S. Cell Biology and Genetics, University of Maryland, College Park, 2020

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Dr. David E Cook

Copyright

Abstract

Magnaporthe oryzae is the filamentous fungus that causes rice blast disease. The organization of genes and genetic variation in plant pathogenic fungal genomes is an important area of research to better understand the evolution of pathogens and genetic diversity of fungal populations. The organization of fungal genomes has been described in different ways. Some descriptions separate the genome into gene-dense and gene-sparse regions that differ in the abundance of effector genes and DNA variation, called the two-speed genome. Other fungal genomes have variable regions without differences in gene density. As such, it is unclear what description of genome organization is most relevant to understanding pathogen biology and evolution. Additionally, it remains unknown what factors drive particular descriptions of genome organization. One hypothesis is that the epigenome dictates genome organization described through the lens of variability and evolution. Under this model, active and repressive epigenetic histone modifications contribute to genome organization of stable and variable genomic regions, respectively, that have different evolutionary trajectories. To test this hypothesis, we completed four high-quality genome assemblies and histone modification maps to enable high-resolution analysis of DNA variation and the epigenome present in multiple reference strains, against a population of rice-infecting *M. oryzae* strains. Overall, the genomes of all four strains showed similar patterns of organization, defined by their epigenetic states, the presence of genes versus transposable elements (TEs) and their type of DNA variation. We did not find evidence of the genomes being organized into gene-dense and gene-sparse regions. The results showed that the repressive histone marks, H3K27me3 and H3K9me3, were associated with higher single nucleotide polymorphism frequencies in genes, while H3K27ac, a mark of active transcription, is associated with higher insertion and deletion mutation frequencies. The results support the

hypothesis that the tested epigenetic marks are associated with specific types of genome variation and contribute to the organization of the genome.

In addition to the comparative genomics research, we used a direct genetic approach to examine histone cross-talk and its role in transcription and genome organization. These experiments focused on tri-methylation of H3 Lysine 36 (H3K36me3) and H3 Lysine 27 (H3K27me3). This is because previous research in the lab showed that disruption of Polycomb Repressive Complex (PRC2), the enzyme complex responsible for methylating H3K27, also impacted the enrichment of H3K36me3, a mark typically associated with actively transcribed genes. Two lysine methyltransferases, Su(var)3-9, Enhancer of zeste, Trithorax-2, SET2, and absent, small, or homeotic-1, ASH1, are responsible for H3K36me3 in filamentous fungi, and single and double deletion strains for the two enzymes, Δ*set2* and Δ*ash1*, were constructed. Phenotypic analysis showed that both enzymes are required for normal growth, development and host infection, indicating the broad importance of H3K36me3 in *M. oryzae*. Results from histone chromatin immunoprecipitation sequencing (ChIP) showed that ASH1 and SET2 function in separate genomic regions, consistent with results from other Ascomycetes. In *M. oryzae*, ASH1-mediated H3K36me3 largely overlapped regions of PRC2-mediated H3K27me3, and interestingly, the Δ*ash1* strain lost H3K27me3 in 56.5% of regions co-marked by ASH1 and PRC2. The results indicate that ASH1-mediated H3K36me3, but not SET2, can co-locate and sometimes be required for H3K27me3 deposition. The mechanism of this interaction is not known. The proposed function of H3K36me3 is to repress DNA transcription, and this could explain why the one mark is found in disparate regions of the genome. We propose that SET2-mediated H3K36me3 in actively transcribed genes slows transcription to prevent erroneous

transcripts whereas ASH1-mediated H3K36me3 that frequently co-occurs in H3K27me3 marked regions of the genome helps contribute to stronger repression of genes in heterochromatin.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I want to acknowledge and thank my major professor, Dr. David Cook. His guidance, expertise, and patience have greatly improved my knowledge and ability to do science. He has challenged me to do the best work possible since my first day at K-State and I would not have accomplished so much without him.

I would also like to thank the many wonderful lab members I have had the pleasure to learn from and work with. Jun Huang and Wei Zhang have taught me so many of the experimental techniques used in my research, and I truly appreciate all that they have taught me. I would also like to thank the other members of the lab who have helped contribute to ideas and discussions through lab meetings and conversations: Sandeep Marla, Divya Mishra, Veerendra Sharma, Joseph Fenoglio, Pratima Subedi, and Tom McAnany.

I want to thank all of the friends I have made in the department – my time at K-State would not be the same without you. Thanks for all the good times together.

Finally, I want to give a special thanks to my family and friends from home. You all have kept me going through the highs and lows of grad school and this achievement was only possible with your love and support.

Chapter 1 - General Introduction

Epigenetics Overview

Epigenetics is the study of phenotypic changes that occur without alterations to the underlying DNA sequence. One of the major mechanisms underlying epigenetic regulation is the addition or removal of chemical modifications to the histone proteins that package DNA into chromatin (Jenuwein and Allis 2001). Epigenetics also encompasses other forms of regulation including direct methylation of the DNA itself as well as non-coding RNAs regulating transcription (Bender 2004; Costa 2008), however the focus of this research will only be on the role of histones. Histone modifications, known as histone marks, can either promote or inhibit gene expression by altering the structure of chromatin and influencing the accessibility of DNA to transcription factors and other regulatory proteins as well as serve as anchors for other interacting proteins to bind to nucleosomes (Lai et al. 2022; Harr et al. 2015). Two widely studied histone marks that are associated with open and accessible regions of the genome called euchromatin and actively transcribed genes are H3 lysine 27 acetylation (H3K27ac) and H3 lysine 4 tri-methylation (H3K4me3) (Miao and Natarajan 2005). The histone marks H3K27ac is associated with active gene expression and it is often found at the promoters of actively transcribed genes (Creyghton et al. 2010). One of the ways lysine acetylation functions is to change the positively charged lysine amino acid into a neutrally charged residue which lessens the interaction between the histone protein and the negatively charged DNA molecule, however, this typically is not sufficient alone to regulate transcription (Wolffe and Hayes 1999). Another way that histone marks function is through recognition by reader proteins. For example, in yeast the chromatin remodeling protein chromodomain-helicase DNA-binding 1 (CHD1) has a binding site that recognizes H3K4me3 and is functionally involved in gene expression and transcriptional

elongation (Pray-Grant et al. 2005). Histone marks also contribute to the formation of heterochromatin, a chromatin state where genes are generally transcriptionally inactive. H3K27me3 and H3K9me3 are two well-known histone marks that contribute to heterochromatin (Lai et al. 2022). In fungi, these two marks tend to form broad domains that coincide with repressed genes (Jamieson et al. 2013). These marks also repress highly repetitive transposable elements (TEs) which can cause mutations if allowed to move around in the genome (Basenko et al. 2015).

ASH1 and SET2 deposit H3K36me3 in different regions of the genome

While the terms euchromatin and heterochromatin generally define two states of genome activity, biology is often not this simple. For example, some histone modifications are present at both euchromatic and heterochromatic regions, such as H3K36me3 (Connolly et al. 2013). In fungi, H3K36me3 is deposited by two proteins Su(var)3-9, Enhancer of zeste, Trithorax-2 (SET2) and absent, small, or homeotic-1 (ASH1) (Strahl et al. 2002; Tanaka et al. 2007). SET domain containing proteins are found across eukaryotes and SET2 and ASH1 are both H3K36-specific methyltransferases that function in separate, discrete regions of the genome (i.e. eu- versus heterochromatin) (Freitag 2017; Bicocca et al. 2018; Janevska et al. 2018). SET2-mediated H3K36me3 is associated with euchromatin in part because the SET2 enzyme has an RNA polymerase II interacting domain, Set2 Rpb1 Interacting (SRI) allowing it to mark histones during active DNA transcription (Kizer et al. 2005). On the other hand, ASH1-mediated H3K36me3 is associated with repressed genes and is reported to coincide with H3K27me3 marked regions of the genome (Bicocca et al. 2018). It is thought that H3K36me3 functions as a repressive mark that helps prevent defects of transcription elongation which can happen when

RNA polymerases collide or become stuck (Carrozza et al. 2005). From these findings, H3K36me3 appears to be a repressive mark that is often found in euchromatin.

Cross-talk between histone marks

Histone marks can also interact with one another to influence chromatin structure and gene expression, termed histone cross-talk (Zhang et al. 2015; Schmitges et al. 2011). For example, the presence of H3K4me3 and H3K36me3 can help recruit chromatin remodelers and transcriptional activators (Lee et al. 2017). The histone modification H4K20me3, regulates the enrichment of other histone marks, such as H3K27me3 and ASH1-mediated H3K36me3 (Möller et al. 2023). In *Magnaporthe oryzae*, loss of H3K27me3 resulted in the loss of H3K36me3 in previously H3K27me3 marked regions, but the gain of H3K36me3 in other regions (Zhang et al. 2021). The different interactions of H3K36me3 to the loss of H3K27me3 may be due to the different enzymes SET2 and ASH1. These findings illustrate the potential for cross-talk between histone marks to both direct or suppress other histone modifications, but there are many details of histone-cross talk that remain unknown, especially in filamentous fungi. Overall, chemical histone modifications function in concert to modulate chromatin structure and gene expression.

Different descriptions of fungal genome organization

In addition to the role of chromatin in gene transcriptional availability, there is a growing body of evidence that euchromatin and heterochromatin may be involved in organizing the genome into variable and stable regions. Genome organization is important in plant pathogenic fungi because effectors, secreted proteins that aid in pathogenesis and infection, are often grouped within genomes (Torres et al. 2020). Some pathogen genomes are described as ‘two-speed genomes’ that contain gene-dense and gene-sparse regions (Raffaele et al. 2010). In *Phytophthora infestans*, genes localized in gene-sparse regions and displayed higher rates of

non-synonymous substitutions (Raffaele et al. 2010). However, other filamentous plant pathogens do not display gene-dense and sparse regions (Frantzeskakis et al. 2018). In fungal pathogens such as *Verticillium dahliae*, the causal agent of vascular wilt diseases (Cook et al. 2020), and *Blumeria graminis*, the causal agent of powdery mildew on barely (Frantzeskakis et al. 2018), there is not evidence of gene-dense or sparse regions. Instead, genome organization in *V. dahliae* is defined by widely shared and conserved core regions and lineage specific regions that are divergent between distantly related strains (Jonge et al. 2013; Faino et al. 2016). In *B. graminis*, there are not regions that are more variable than others which has been termed a one-speed genome in contrast to the two-speed genome hypothesis.

Magnaporthe oryzae overview

Magnaporthe oryzae is a filamentous fungus that is the causal agent of Rice Blast disease (Asibi et al. 2019) which destroys 30% of rice production globally (Nalley et al. 2016). The pathogen can infect leaves, panicles, and other above ground tissues (Talbot 2003). It does this through the formation of an appressorium, a highly melanized structure that can build up an enormous amount of turgor pressure (Wilson and Talbot 2009; Osés-Ruiz et al. 2021). This pressure is directed through a penetration peg to enter the host cell wall (Wilson and Talbot 2009). Once inside the cell, invasive hyphae grow and a biotrophic interfacial complex (BIC) develops (Zhang and Xu 2014; Khang et al. 2010). The BIC is a center for cytoplasmic effector secretion and translocation into the host cell (Mosquera et al. 2009) that is partially made of host membranes to facilitate translocation into the host (Zhang and Xu 2014). Other effectors are secreted directly from the invasive hyphae into the apoplast between the fungal cell wall and the host plant membrane (Mosquera et al. 2009). The success of infection varies with the presence of effector genes as well as their expression depending on the strain of *M. oryzae* (Kang et al. 2001;

Valent et al. 1991). Rice infecting strains of *M. oryzae* are globally distributed, but form four distinct lineages, one which displays characteristics of sexually recombining strains, and three clonal lineages (Thierry et al. 2022; Latorre et al. 2020). Across these four lineages is a diversity of supernumerary mini-chromosomes (Orbach et al. 1996). These mini-chromosomes are dispensable, vary in number between strains, and contain repeat-rich sequences and many transposable elements (Chuma et al. 2003). Additionally, it has been demonstrated that mini-chromosomes can undergo large-scale rearrangements during and after meiosis (Chuma et al. 2003). Strains that have mini-chromosomes often will have some effector genes located on the mini-chromosomes (Peng et al. 2019; Langner et al. 2021), however it is still unclear how the mini-chromosomes might contribute to virulence or the evolution of effector genes.

Understanding the potential role of the epigenome in organizing variable and stable regions of fungal genomes is an active area of research with important implications in plant pathology and disease management. Knowledge of which effector genes are rapidly evolving can help inform plant breeders of which resistance genes in rice would be the most useful in providing long-term resistance against rice blast. The associations between variable regions of the genome and other genomic and epigenomic features can help enable the prediction of these important effector genes, as well as inform how these pathogens have evolved to infect their specific host. The interactions between epigenetic histone marks is also an important area of research as it has implications in the finely tuned regulation of genes which may play a role in the complex interactions between pathogen and host during infection.

Chapter 2 - Natural genomic variation in rice blast genomes is associated with specific histone modifications

Note: This chapter was done in collaboration with Jun Huang, Divya Mishra, Wei Zhang, and Katie Jordan. Jun Huang and Divya Mishra assembled and annotated the genomes of strains Guy11, O-135, O-137, and O-219, and J. Huang conducted the CHEF analysis. Wei Zhang constructed the ChIP-seq libraries. Katie Jordan assisted with the population level analysis of the strains.

Abstract

How functional elements of DNA are organized along physical chromosomes is not random, but our knowledge of what defines organization is incomplete. For plant pathogenic fungi, there are different descriptions of genome organization, and we lack a framework to understand the mechanisms and evolutionary consequences of these organizational patterns. To further understand genome organization and its potential drivers, we utilized highly contiguous genome assemblies and histone modification profiles to compare four strains of *Magnaporthe oryzae* representing different lineages in the population. We report that single nucleotide polymorphisms (SNPs) colocalize with transposable elements and effector genes in sub-telomeric regions, showing an organization of variation on the chromosome. Genome-wide repressive histone marks, tri-methylation of H3 Lysine 27 and H3 Lysine 9, H3K27me3 and H3K9me3 respectively, are associated with higher SNP frequencies and higher SNP density, linking these epigenetic marks to variable genomic regions. Densely grouped SNPs occur at even frequencies in genes and TEs, indicating it is not a technical artifact or simply the result of selection. We conclude that in *M. oryzae,* gene density does not play a crucial role in organizing the genome, but rather, epigenetic marks associated with euchromatin and heterochromatin

define genomic regions displaying unique forms of genetic diversity. Effector genes are specifically associated with H3K27me3 enrichment. Additionally, regions defined as euchromatic tend to be syntenic between strains, while heterochromatic regions are more often non-syntenic. These associations were consistent across the four analyzed assemblies, indicating generalized patterns of organization are more broadly conserved across the species.

Introduction

Genome variation plays an important role in shaping phenotypic differences within and between species (Kasetsomboon et al. 2013; Ali et al. 2014; Wang et al. 2020). There are many forms of genomic variation, two of the most well studied forms being single nucleotide polymorphisms (SNPs) and small insertions and deletions (INDELs), both of which change short DNA segments (Katsonis et al. 2014; Mills et al. 2006). In fungal plant pathogens, there are contrasting findings regarding the dependence between the distribution and organization of SNPs and INDELs with other genomic features, such as genes and transposable elements (TEs). Recent studies show that variation, in the form of SNPs and INDELs, is found distributed unequally in some fungal genomes while variation appears random in others (Ajawatanawong and Baldauf 2013; Torres et al. 2020), meaning different fungal genomes are accumulating these mutations in different patterns. In *Magnaporthe oryzae* the accumulation of new SNPs and INDELs in an experimental evolution study were more frequent in non-coding regions and many TEs preferentially inserted more often near secreted protein coding genes than other types of genes (Zhong et al. 2020), however, a more comprehensive understanding of what contributes to the organization of the *M. oryzae* genome is still needed.

One pattern of genome organization seen in a variety of plant pathogens is the two-speed genome. First described in the oomycete *Phytophthora infestans*, the two-speed genome describes genes that are more densely grouped in *P. infestans* have a lower rate of accumulated non-synonymous SNPs than sparsely grouped genes (Raffaele et al. 2010). Often, effector genes, which encode secreted proteins that aid in infection, are organized into similar compartments and regions of the genome (Sexton and Cavalli 2015; Möller and Stukenbrock 2017). In *Puccinia graminis*, a pathogen of wheat, gene families involved in pathogenesis were under strong

diversifying selection which resulted in highly polymorphic genes (Sperschneider et al. 2014; Upadhyaya et al. 2015). However, not all fungal pathogens demonstrate a connection between variation and genomic features. In the grass powdery mildew *Blumeria graminis* TE insertions have occurred fairly evenly throughout the genome and are found in close proximity to genes regardless of the function of their protein product (Frantzeskakis et al. 2018). Interestingly, there are even differences in genome organization within families of fungi. In the Magnaporthaceae family, *M. oryzae* demonstrated a connection between unique secreted proteins and DNA variation, but *Magnaporthe poae* and *Gaeumannomyces graminis* had similar associations between variation and genes of any function (Okagaki et al. 2016). Together, there appears to be diverse types of genome organization that have arisen in plant pathogenic fungi.

The genome is organized not only by DNA sequence information, but also by the epigenome that imparts additional information and regulation. Chemical modifications to histone proteins form an important component of the epigenome (Lai et al. 2022). Histones form a multi-protein complex, termed nucleosome, that interacts with DNA and provides the local and global structure termed chromatin (Bannister and Kouzarides 2011). These compartments of the genome, brought about by DNA and protein interactions, play an important role in many cellular functions including regulating transcription, cellular division, and organismal development (Fukagawa and Earnshaw 2014; Zhang et al. 2021). Histone proteins undergo a variety of post-translational chemical modifications, including methylation and acetylation, which alters their DNA interactions and ultimately their chromatin structure (Wolffe and Hayes 1999; Marmorstein and Zhou 2014). Trimethylation on histone H3 lysine 27 (H3K27me3) and lysine 9 (H3K9me3) are considered repressive marks associated with lower transcription and termed heterochromatin (Lai et al. 2022), while methylation of other residues, such as H3K4me3, are

associated with active gene transcription and open chromatin termed euchromatin (Miao and Natarajan 2005). In addition to methylating histones, acetylation of histones can impact DNA usage, such as H3K27ac that is enriched in the promoter regions of transcribed genes (Creyghton et al. 2010; Beacon et al. 2021). In addition to histone modifications controlling transcription, there is increasing interest in how they may also impact other aspects of genome biology, including evolution. For example, in the plant pathogen *Verticillium dahliae*, regions coding for genes involved in plant infection and adaptation are associated with high levels of variation in the population and specific epigenetic profiles (Cook et al. 2020). In the fungal wheat pathogen *Zymoseptoria tritici*, base substitutions accumulated at higher rates in regions marked with H3K27me3 (Möller et al. 2019; Habig et al. 2021). Additionally, those same regions accumulated base substitutions at lower rates when the gene encoding the histone lysine-methyltransferase were knocked out and the histone mark was absent (Habig et al. 2021). Interestingly, in *Z. tritici* deletions are more commonly found in H3K27me3 and H3K9me3 marked regions (Seidl et al. 2016), however, in human cell lines H3K27me3 was negatively associated with the occurrence of new INDELs at double strand breaks (Johnson et al. 2019; Schep et al. 2021). Together, these results indicate associations between histone marks and variation, but the mechanisms and occurrence remain unclear. It is therefore important to further research the connection between DNA variation and epigenetic histone marks, especially in plant pathogens, in order to understand their genome biology and evolution.

To further investigate genome organization in filamentous plant pathogens and its connection with genome variability, we conducted a comparative genomic and epigenomic study in *M. oryzae* a model plant pathogenic fungus and an important global pathogen, capable of infecting a wide range of poaceous hosts including important agricultural crops such as rice,

wheat, barley, and millet (Kang et al. 1995; Takan et al. 2012; Gladieux et al. 2018; Chung et al. 2020). Additionally, *M. oryzae* is emerging as a good model to study the epigenome of filamentous fungi (Pham et al. 2015; Wu et al. 2022).

Utilizing new highly-contiguous genome assemblies and epigenome maps for four closely related *M. oryzae* rice infecting strains, we detail at high-resolution the relationship between DNA variation and epigenome organization. We identified contrasting associations between SNPs and INDELs, and syntenic and non-syntenic genomic regions that are organized according to histone modification profiles. The results support the model that fungal plant pathogen genomes exist on a continuum from stable to hyper variable regions, that largely co-occur with variability in the epigenome, and evolutionary analysis indicates that the association may be causative.

Results

Complete assemblies of four *M. oryzae* rice-infecting strains have similar coding and repetitive sequence content

To better understand genomic variation between closely related strains of *M. oryzae*, we generated high quality genome assemblies of four rice-infecting strains: Guy11, O-135, O-137, and O-219. The Guy11 strain is a field isolate collected from French Guiana in 1979 (Leung et al. 1988), strains O-135 and O-137 are field isolates collected in 1985 from Hangzhou, Zhejiang, China (Valent et al. 1991), and strain O-219 is a field isolate collected in 1985 from the Ivory Coast (Valent et al. 1991; Orbach et al. 1996). Complete genome assemblies were generated using Oxford Nanopore long-read sequencing, ranging from 62.1x to 149.3x coverage, and supported with Illumina short-read sequencing (Supplemental Table A.1). The genomes were assembled into 21-27 contigs each, which were reduced to 8-12 contigs based on quality assessment (see methods for details). Total assembly size for the final versions ranged from 42.6 megabases (Mb) to 46.8 Mb with an N50 between 5.7 Mb and 6.5 Mb (Table 2.1). The contigs - were assigned chromosome numbers based on alignment and convention with the 70-15 (Dean et al. 2005) and B71 (Peng et al. 2019). Assembly completeness based on Benchmarking Universal Single-Copy Orthologs (BUSCO) (Manni et al. 2021), which assess the presence of conserved coding sequences, determined that the four strains had similarly high completeness, 97.0%-97.9% (Table 2.1). Short read genome sequencing data of 82 additional rice-infecting strains of *M. oryzae* were mapped against our Guy11 genome assembly to identify single nucleotide polymorphisms (SNPs). Whole genome SNP phylogeny of the 86 strains formed four distinct clades, consistent with previously published phylogenies of rice infecting *M. oryzae* strains (Thierry et al. 2022; Latorre et al. 2020; Zhong et al. 2018; Gladieux et al. 2018) (Figure 2.1A).

Guy11 and O-219 are both present in clade I, which has been described as a diverse group capable of sexual recombination, while O-135 and O-137 represent clades II and III, respectively, both of which are asexually reproducing lineages. Overall, these four strains represent a diverse group of rice-infecting strains.

Gene annotation of the assemblies indicate that the four strains code for a similar number of genes, ranging from 12,879 to 13,530 (Figure 2.1B). Plant pathogenic fungi, including *M. oryzae*, produce a number of secreted proteins that aid in host infection by subverting or modifying the host immune response (Seidl et al. 2016; Jeong et al. 2007; Mentlak et al. 2012). The presence or absence of effectors can strongly influence what plant hosts the fungal pathogen is able to infect and how successful infection will be. To identify potentially secreted proteins, SignalP (Almagro Armenteros et al. 2019) and EffectorP (Sperschneider and Dodds 2022) were used to process the annotated protein sequences. Between the four genomes, a range of 1,574 to 1,702 genes were found to have secretion signals (Figure 2.1D). Of these, 634 to 670 were predicted to be effectors that may aid infection (Figure 2.1D). We additionally identified carbohydrate-active enzymes (CAZymes), a group of proteins that catalytically alter or break glycosidic bonds that fungal plant pathogens use to degrade carbohydrates such as components of plant cell walls (Sista Kameshwar and Qin 2019). We identified 519 to 534 CAZymes (Figure 2.1D). Secondary metabolites (SM) are also commonly produced during host infection, and we identified between 57 and 61 SM gene clusters in the four analyzed genomes (Figure 2.1D). In addition to protein coding sequences, we annotated transposable elements (TEs) (see methods for details) and found that TEs accounted for 12.5% to 14.4% of the total genome size among the strains (Table 2.1). Long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs), both of which are retrotransposons, were the two most common classes of TEs, where

LINEs accounted for 3.27-4.98% of the genomes and LTR elements represented 6.40-8.31% (Figure 2.1C). A notable difference between the strains is that three of the four strains were also identified to contain supernumerary 'mini' chromosomes. Mini-chromosomes are present in many *M. oryzae* strains but are not necessary for growth (Orbach et al. 1996). O-135 and O-137 genomes each contained one mini-chromosome (Figure 2.1E). The O-219 genome has previously been reported to contain a large mini-chromosome of approximately ~2 Mb, and three additional mini-chromosomes on the order of ~ 1 Mb each (Orbach et al. 1996). Our assembly indicated three mini-chromosomes, ranging in size from 915,411 to 1,401,154 bp, which is consistent with the results from clamped homogenous electric field (CHEF) gel analysis (Figure 2.1E). Our CHEF analysis was not able to resolve the individual smaller mini-chromosomes, but their approximate size is consistent with the original report and our long-read assemblies. The four assemblies are of high quality and provide new reference genomes for the different clades of *M. oryzae* rice infecting strains.

The four strains contain similar numbers of single-nucleotide polymorphisms and short insertion and deletion DNA variants

We next investigated SNPs and short insertion and deletion (INDEL) variants with respect to each of the four new assemblies. For SNP analysis, the group of 86 rice-infecting strains were mapped to the four new assemblies and SNPs were called using the Genome Analysis Toolkit (Auwera and O'Connor 2020). The identified SNPs were filtered for biallelic SNPs present in at least 80% of strains for further analysis. The majority of SNPs were found in intergenic regions, averaging approximately 65% (Figure 2.2A). Of the SNPs occurring in exons of coding sequences, more than half were predicted to be nonsynonymous and cause a change in the resulting amino acid sequence (Figure 2.2B), which are similar results to previously

published data on related strains of *M. oryzae* (Gowda et al. 2015). In the four *M. oryzae* genomes, INDELs were identified using a combination of Assemblytics, an assembly based variant caller (Nattestad and Schatz 2016), and Sniffles, which utilizes unassembled long-read data (Sedlazeck et al. 2018). To generate a high confidence set of INDELs, consensus variants for each genome were identified between the two pipelines (Table 2.2). For this analysis, we defined INDELs as short insertion and deletions that affected DNA sequences between 4 and 30 bp long, while larger identified variants were considered structural variants (SV). The majority of the identified variants were classified as INDELS; Guy11 contained 3,834 (91.9% of total variants), O-135 contained 4,154 (92.5% of total variants), O-137 contained 4,343 (93.3% of total variants), and O-219 contained 4,210 (93.1% of total variants) INDELs. All four genomes had similar percentages of genes containing INDELs, 9.1-9.8%, and similar percentages of TEs containing INDELs, 1.4-2.5% (Figure 2.2C). The median INDEL size is 6 bp in all four genomes (Figure 2.2D). These results suggest that the number and type of DNA variants identified were similar between the four strains.

DNA variants are organized in the genome based on physical location and density

After determining that there were few differences in the quantity of genes, TEs, SNPs, and INDELs between the four strains, we addressed the genomic location of these elements. We investigated both the distribution and density of genomic features (genes, predicted effectors, and TEs) and DNA variants (SNPs and INDELs) (Figure 2.3). To assess distribution, we divided each of the core chromosomes into ten equally sized bins and summarized the genomic features and variants across all chromosomes and four strains (Figure 2.3A, C, E, G, I). Genes were evenly located in the genome across the chromosomes of the four strains (Figure 2.3A). Each bin representing one tenth of the chromosome had approximately 10% of the total genes of any

given chromosome (Figure 2.3A). To assess density, we measured the distance between genes in both the 3' and 5' directions. Gene density follows a normal distribution with a 1 kb mean distance between genes, consistent in all four strains (Figure 2.3B). We also examined the subset of predicted effector genes and found that this specific category of genes was not evenly distributed along chromosomes. Outer bins, representing sub-telomeric sections of the chromosomes, had higher percentages of effectors than inner bins representing the center of the chromosome (Figure 2.3C). Effector genes are not in gene sparse regions of the genome, with the mean distance between effectors and their nearest gene around 1 kb (Figure 2.3D). TEs were also unevenly located along the chromosomes. Outer bins had on average nearly 20% of the TEs, while central genomic bins contained only 1-5% of total TEs (Figure 2.3E). The distribution of TE density also differs from genes and displayed a bimodal distribution in all four genomes (Figure 2.3F). There is a peak of densely occurring TEs within 10 bp up and downstream of another TE, and a second broader peak, with the distance between TEs centered around 1 kb (Figure 2.3F). This means that some TEs are densely clustered together in the genome. Among eukaryotic genomes, TEs often are inserted near or inside of other TEs which forms nested groups of TEs in repeat-rich regions (Gao et al. 2012). It is likely that the densely clustered TEs observed in our four genomes are the result of similar nested TE insertions.

Analyzing DNA variants, we found that sub-telomeric regions of chromosomes generally had slightly higher average SNP counts compared to the center of chromosomes (Figure 2.3G). Interestingly, SNP density forms a multi-modal distribution with a peak around 1 bp, 10 bp, and 1000 bp (Figure 2.3H). These patterns could also be treated as bimodal, indicating a SNP dense region of the genome, where SNPs occur within less than 100 bp, and a SNP sparse region occurring at a distance of greater than 100 bp. The distribution of INDELs across chromosomes

is slightly more common in middle sections of chromosomes than telomeres, and their density forms a left-skewed distribution with the distance between most INDELs being around 10 kb (Figure 2.3I, J). INDELs occur sparsely throughout the genome with a small number of them occurring near one or two other INDELs (Figure 2.3J). These results show that variance is unequally distributed in the genome. Genes as a whole are equally distributed in the genome, however, certain functional groups of genes such as effectors are biased towards sub-telomeres.

Genome variation is associated with regions containing specific histone modifications

We sought to further understand the factors influencing the genome feature and variant distributions. Previous results indicate an association between genome organization, variation, and the epigenome (Cook et al. 2020; Habig et al. 2021), and we set out to determine whether specific histone modifications are associated with genome organization. Chromatin immunoprecipitation and sequencing (ChIP-seq) data were collected for each of the four *M. oryzae* strains for histone modifications canonically associated with repressed transcription and heterochromatin, H3K9me3 and H3K27me3, and two marks associated with active gene transcription, H3K27ac and H3K4me3 (Lai et al. 2022; Kim et al. 2021; Pasini et al. 2010; Atanasoff-Kardjalieff et al. 2021). We previously validated the antibodies and experimental protocol for H3K27me3 and H3K27ac in *M. oryzae* (Zhang et al. 2021), and correlation and principal component analysis suggests that the other two marks captured genomic regions as expected (Supplemental Figures A.1-4). Across the four genomes, H3K27ac marked 19.8% to 28.8% of the genomes, H3K27me3 marked 18.8% to 21.1% of the genomes, and H3K9me3 marked 12.4% to 20.7% of the genomes (Figure 2.4A). Peak calling of the ChIP data was performed using MACS2 (see methods) to determine regions of the genome enriched for

H3K27ac, H3K27me3, and H3K9me3 compared to input controls, and peak position was used to identify overlaps with genes and TEs. For genes, H3K27ac marked 13.2% to 22.3% across the genomes, and H3K27me3 marked 19.7% to 21.7% of genes (Figure 2.4B). The mark H3K9me3 covered a more variable number of genes between the four genomes, covering 14.0% and 18.9% of genes in Guy11 and O-135, respectively, but only 3.6% and 8.1% of genes in O-137 and O-219 (Figure 2.4B). As for TEs, H3K27ac covered between 1.2% and 7.3% of TEs, while H3K27me3 marked 36.4% to 54.5% of TEs and H3K9me3 marked 50.6% to 84.4% of TEs (Figure 2.4C). Given our observation that some SNPs are highly clustered in the genome (Figure 2.3F), we grouped SNPs by proximity, so that SNPs within 100 bps of another SNP in both the 5' and 3' directions were categorized as dense, and everything else was termed sparse. Dense SNPs account for nearly half of the total SNPs in the genomes (45.0% to 54.4%), and of those, 26.1% to 38.4% of dense SNPs occur in genes and 26.8% to 40.6% of dense SNPs occur in TEs. Dense SNPs were intersected with peaks for each of the histone marks, and 5.1% to 8.7% of dense SNPs occur in H3K27ac regions, which stands in contrast to the 39.4% to 53.3% of dense SNPs that occur in H3K27me3 regions and 46.8% to 65.8% of dense SNPs that occur in H3K9me3 regions (Figure 2.4D). Sparse SNPs show a different pattern where most genomes have the highest percentage of sparse SNPs in H3K27ac regions (19.8 to 30.9%) (Figure 2.4E).

To gain a broad understanding of the relationships between the analyzed variables, all of the data, including counts of genes, TEs, SNPs, INDELs, and the normalized ChIP signal for each histone modification were summarized in 2.5 kb windows and correlated within each strain (Figure 2.5A-D). Heterochromatic marks, H3K27me3 and H3K9me3 were highly correlated with one another and negatively correlated with H3K27ac. Repressive histone marks were associated with SNPs occurring in TEs while activating histone marks were associated with

SNPs occurring in genes. INDELs, interestingly, were positively correlated with H3K27ac. To further understand associations between DNA variation, genomic features and histone modifications, we tested for independence between histone modifications and variation by genomic feature (see methods for details). The resulting chi-squared values greater than the critical value of 10.8 (Bonferroni multiple testing correction) indicated dependence between DNA variants and the presence of histone modifications for some genomic features (Figure 2.5E, Supplemental Figures A.9-12). For genes, the chi-square test indicated that DNA variants occurred significantly differently than expected based on the presence of the histone marks in all four genomes (Figure 2.5E). Interestingly, variation in TEs mainly occurred independently of histone modifications for both SNPs and INDELs, except for variation in TEs in O-137 and O-219 associated with H3K9me3 (Figure 2.5E). For other intergenic regions of the genome, the results indicated that the occurrence of SNPs was dependent on H3K9me3 and to a lesser extent H3K27me3, and both histone marks influenced the occurrence of INDELs (Figure 2.5E). To understand the direction and nature of the dependence between histone modifications and variation as indicated by the chi-squared test, we calculated the frequency of SNPs and INDELs per kb for each genomic feature marked by either H3K27ac, H3K27me3, or H3K9me3 (Table 2.3 and Table 2.4). All four genomes showed higher SNPs/kb in genes and intergenic regions when the features were present in heterochromatin domains H3K27me3 or H3K9me3 (Figure 2.5F). We observed that the SNPs/kb were 1.5 to 3 times higher for genes in H3K27me3 or H3K9me3 domains versus the genes not in the domains (Figure 2.5F). In intergenic regions, SNPs/kb were 1.7 to 2.6 times higher in H3K27me3 domains, and those in H3K9me3 domains were 2.5 to 4 times higher (Figure 2.5F). Compared to the results for genes and intergenic regions, the SNPs/kb in TEs were contrasting between the heterochromatic histone

modifications. The TEs in H3K27me3 domains had on average 50% lower SNPs/kb frequency compared to those not in H3K27me3 domains, while TEs in H3K9me3 domains had on average 3.8 times higher SNPs/kb than TEs not in H3K9me3 domains (Table 2.3). Given that each comparison within a histone modification used the same set of TEs, the results indicate that TEs in H3K9me3 domains have higher SNP/kb frequency than those in H3K27me3 domains. There was not the same elevated INDELs/kb associated with the heterochromatin marks. Instead, there is an elevated INDELs/kb rate for TEs and genes in H3K27ac marked regions, but it is not significant based on the chi-squared results. The results from the chi-squared test indicated a dependence between INDELs and heterochromatic marks, which based on the rate of INDELS/kb indicates that INDELs occur less frequently than expected in these regions. These results indicate a clear association between SNP and INDEL frequency and the presence of histone modifications, but the directions and magnitude seem to differ between different genomic features.

Variation and synteny is organized by chromatin states

To move beyond pair-wise correlations and to understand the genome-wide associations, dimensional reduction through principal component analysis (PCA) was used. Also, we further divided the genome into syntenic and non-syntenic regions based on whole genome alignments, and classified regions as syntenic if they were shared between the four genomes using Syri (Goel et al. 2019). This classification was used to more broadly capture genome variation at a larger scale than the short DNA tracts captured by SNPs and INDELs. All genomic data were integrated in 2.5 kb windows, classified as syntenic if at least half the window was called syntenic from the alignment analysis, all other windows were classified non-syntenic (Supplemental Table A.2). The first principal component explained 29.54% of the variance,

largely separating euchromatin and heterochromatin and also syntenic and non-syntenic windows (Figure 2.6A). Gene number and H3K27me3 are the main drivers separating the regions on the second principal component, explaining 15.53% of the variance (Figure 2.6A). Genomic regions from the PCA grouping sharing similar profiles were identified using k-means clustering (Figure 2.6). In Guy11, k-means clustering identified three groups of windows (Figure 2.6A). The first cluster, containing 73.1% of the genome, can be characterized by the presence of INDELs and the enrichment of H3K27ac and H3K4me3. 69.4% of the windows that form this group also are from regions of the genome that are syntenic between the four genomes. The second group, 15.1% of the genome, is associated with genes and effectors. The third group, containing only 11.8% of the genome, is strongly associated with the presence of TEs, SNPs, and the enrichment of H3K9me3. Interestingly, H3K27me3 appears to be associated with both the second and third clusters. Related to DNA variation, a higher percentage of windows in the first and second clusters had low numbers of SNPs compared to the third cluster (Figure 2.6A). Also interesting, the third cluster had the highest percentage of windows without INDELs while the first cluster had the highest percentage of windows with INDELs (Figure 2.6A). For the other three genomes, we found that the first principal component separated the same features, namely chromatin features and genome synteny (Figure 2.6, B, C, D). For the second principal component, both O-135 and O-137 were more similar showing Gene and INDEL largely driving the separation, while O-219 was more similar to Guy11 having Gene and H3K27me3 being most important. O-135 and O-137 share similar patterns of clustering windows of the genome. In both genomes the first cluster is associated with INDELs and H3K27ac, the second cluster is associated with genes and H3K4me3, and the third cluster is associated with TEs, SNPs, H3K27me3, and H3K9me3 (Figure 2.6B, C, Table 2.5). Differences between euchromatin and heterochromatin and syntenic

and non-syntenic windows contribute to 31.3% of variance in the first principal component, while genes and H3K27me3 contribute the most to the second principal component which represents 16.6% of variance.

Discussion

In this work we tested the hypothesis that DNA variation is associated with physical organization of the genome and epigenome. Previous inquiries in the connections between DNA variation and genome organization have used the term 'two-speed genome' to describe observations of rapidly evolving regions of the genome and more stable regions seen in some plant pathogenic fungi (Raffaele et al. 2010; Dong et al. 2015; Faino et al. 2016). In the original publication of the two-speed genome, the hypothesis was defined in *P. infestans* by gene-dense regions and gene-sparse regions that contain an abundance of effector genes and have higher rates of non-synonymous substitutions (Raffaele et al. 2010). However, as seen in our data, *Magnaporthe oryzae* does not have gene-dense or sparse regions so it does not fit the original definition of a two-speed genome. Instead, the *M. oryzae* genome is organized by epigenetic histone marks into stable euchromatin regions and variable heterochromatin regions.

The four strains examined in this study represent three divergent lineages of rice infecting *M. oryzae*, but they have remarkable similar epigenetic histone mark associations. Guy11 and O-219 are members of a lineage that has evidence of past sexual recombination, whereas O-135 and O-137 represent two asexually reproducing lineages (Thierry et al. 2022; Gladieux et al. 2018). Chromatin and epigenetic histone marks are involved in DNA repair mechanisms (Papamichos-Chronakis and Peterson 2013) and it is proposed that genomic rearrangements provide important sources of variation in asexual organisms (Seidl and Thomma 2014). In our analysis of synteny between the four strains, we found that roughly 50% of the genomes were syntenic between all strains, yet the correlations between the epigenome and genomic features were largely consistent. SNPs correlated with H3K27me3, H3K9me3, and TEs while INDELs correlated with H3K27ac and H3K4me3. The four genomes ultimately share patterns of genomic and epigenomic

organization. Taken together, this may highlight an underlying role in the different associations between histone marks and variation in the genomes.

Our *M. oryzae* genome analysis shows evidence of an organizational pattern. Effector genes, TEs, and SNPs are unevenly distributed throughout the genome, with most localized to sub-telomeric regions of chromosomes. SNPs in particular show a bimodal distribution of dense and sparse SNPs. Dense SNPs occur much more frequently in repressive histone marks, H3K27me3 and H3K9me3, than they do in activating marks. An argument could be made that the high occurrence of dense SNPs is a byproduct of heterochromatin and associated TEs (Gao et al. 2012; Liu et al. 2020). However, this is not the case, as dense SNPs are found in similar proportions in genes and TEs and the result does not appear to be an artifact of transcriptionally repressed TEs accumulating mutations. The frequency of SNPs in H3K27me3 marked genes is higher than unmarked genes. SNPs are also more frequent in H3K9me3 marked features regardless of whether those features were genes, TEs, or non-coding intergenic regions. Interestingly, not all variation in the genome follows this pattern. INDELs occur more frequently in H3K27ac marked euchromatin which could mean that different types of DNA variation are somehow linked to different chromatin states. These findings are supported by previously published work in the Magnaporthaceae family which conclude that in *M. oryzae* there is an association between diversity and the proximity to TEs (Okagaki et al. 2016). Furthermore, even in more distantly related organisms such as *V. dahliae*, chromatin appears to be more important in defining genomic organization than gene density (Cook et al. 2020). Our understanding of chromatin has traditionally been in relation to the physical accessibility of DNA and the transcriptional regulation of genes (Huisinga et al. 2006). The associations and correlations between DNA variation and epigenetic histone marks add an additional layer to the role of

chromatin that incorporates genetic variation. DNA repair mechanisms are another important component in creating or avoiding variation. In eukaryotes, the types of mutations as well as kinetics of repair enzymes have been reported to change between euchromatin and heterochromatin (Han et al. 2016). One potential reason for the observed differences between the associations of SNPs and INDELs with histone marks would be that the different chromatin states influence the type of DNA repair in the region and thus alter the frequency of various mutations as a result of favoring different repair pathways (Huang and Cook 2022). If DNA damage is occurring unevenly throughout the genome this would help explain why certain regions accumulate more mutations as a consequence of that damage than other regions. The overlapping distributions of histone modifications and DNA variation in *M. oryzae* suggests that the epigenome may play a role in shaping the accumulation of DNA variation.

It is currently unclear whether epigenetic marks lead to specific types of variation in the genome or whether it is a response to variation occurring. Experimental evolution in *Z. tritici* revealed that TEs had significantly higher base substitution mutation rates when the sole methyltransferase responsible for H3K9me3 (*kmt1*) was deleted (Möller et al. 2019; Habig et al. 2021). This suggests that when H3K9me3 is present the mark plays a role in reducing the frequency of mutations. However, in other organisms, such as *Neurospora crassa*, H3K9me3 is enriched in response to mutation of the DNA sequence after repeat-induced point mutation (RIP) (Lewis et al. 2009). These findings are compounded by the tendency for TEs and effectors to be in sub-telomeric regions of the chromosomes which can be more fragile due to highly repetitive sequences and the challenges of high fidelity DNA replication and repair at the ends of chromosomes (Barnes et al. 2023). However, it is still unknown whether the presence of effectors helps drive variability in or around them by the nature of the selective pressures they

face or if effectors arise in regions that are already variable because the organization of the genome to contain variable regions benefits the fitness of the pathogen. Effector genes are frequently under diversifying selective pressures to evade host immunity (Sperschneider et al. 2014), meaning that the occurrence of mutations and variation to alter the protein product of effector genes can often be beneficial for the pathogen. Selection may contribute to genome organization because small changes to the structure of an effector protein may help the pathogen avoid detection by the host plant, while still retaining effector function to aid infection. In this way, the presence of effectors may influence the local accumulation of mutations due to the specific selective forces they experience. On the other hand, the physical location of effector genes may be due to effects of TEs, telomeres, and heterochromatin in creating that diversity. These are regions of the genome that are inherently less stable due to the repetitive sequences and physical structure of DNA at the telomeres (Barnes et al. 2023), so these regions may be predisposed to favor the evolution of effector genes because mutations in them are more tolerated than other essential genes. Under this hypothesis, selection is responding to the organization of the genome because repressive histone marks and TEs are destabilizing nearby DNA which is generally only favorable to effector genes.

Materials and Methods

Genome assemblies and repeat and gene annotation

Oxford Nanopore sequencing was performed using MinION sequencing platform. Raw MinION fast5 files were processed into fastq files by Guppy (version 3.4.4) (Wick et al. 2019) with the following parameters: --disable_pings --compress_fastq --flowcell FLO-MIN106 --kit SQK-LSK109. Adaptors were trimmed from reads by Porechop (version 0.2.4) (Wick et al. 2017). Canu (version 1.9 for O-135 and 2.2.1 for O-137, O-219 and Guy11) (Koren et al. 2017) was used for genome assemblies with the following parameters: genomeSize=45m, corMhapSensitivity=normal, corOutCoverage=80, minOverlapLength=2000, minReadLength=5000, correctedErrorRate=0.12, useGrid=1. The draft assemblies were further polished by Nanopolish (two times, version 0.13.2) (Loman et al. 2015) and Pilon (two times, version 1.22) (Walker et al. 2014), with Nanopore and Illumina reads. Each contigs were renamed based on the NUCmer (version 3.1) (Kurtz et al. 2004) alignment with the core chromosomes from B71ref2 and 70-15 genome assemblies. Unplaced contigs with (>50%) duplicated with the core chromosomes or bacterial DNA were manually removed from the assemblies. The size of mitochondrial DNA was trimmed according to B71ref2 assembly.

Predicted results for repeated sequences in the four isolates from LTRfinder (version 1.07) (Xu and Wang 2007), LTRharvest (version 1.5.10) (Ellinghaus et al. 2008), MGEscan (version 3.0.0) (Lee et al. 2016), LTR_retriever (version 2.9.0) (Ou and Jiang 2018), MITE-Tracker (version 2.7.1) (Crescente et al. 2018) and MITE-Hunter (Han and Wessler 2010) were combined with CD-HIT (version 4.8.1) (Li and Godzik 2006), parameters: -c 0.8 -G 0.8 -s 0.85 -T 15 -aL 0.85 -aS 0.85, to remove highly similar sequences. The non-redundant library for

repeated sequences were used as input for annotation with Repeatmasker (version 4.0.9) (Smit et al. 2015).

The soft-masked genome was used as the input for gene annotation following funannotate pipeline (version1.8.5) (Palmer and Stajich). The completeness of assembled genomes and predicted gene annotations were evaluated by BUSCO (version 5.1.2) (Manni et al. 2021) by using ascomycota_odb10 dataset. The presumed secreted proteins were predicted by SignalP (version 5.0ba) (Almagro Armenteros et al. 2019) and further annotated with EffectorP (version 3) (Sperschneider and Dodds 2022).

Chromatin immunoprecipitation sequencing and analysis

Chromatin Immunoprecipitation sequencing was performed following our previously published methodology (Zhang et al. 2021). Briefly, proteins were crosslinked to DNA in mycelia grown in liquid culture using formaldehyde and then flash frozen in liquid nitrogen. 200 mg of frozen tissue per sample was ground and DNA was sheared using sonication. Sheared DNA was treated with MNase and then precleared using an 80:20 ratio of A and G beads. Samples were then incubated for 14 hours with the following antibodies: H3K27ac (Diagenode, C15410196), H3K27me3 (abcam 6002), H3K36me3 (abcam 9050), H3K4me3 (Diagenode, C15410003), or H3K9me3 (Diagenode C15410056). Antibodies were then pulled down with the A/G beads and washed thoroughly. DNA was purified using the iPure kit from Diagenode. Libraries were prepared using NEBNext FFPE DNA repair mix and Ultra II End-prep mix. A universal adaptor was then ligated to DNA samples and IDT i5 and i7 primers were used in a PCR reaction to add unique barcodes to each of the samples. DNA fragments were double size-selected after PCR to enrich for approximately 300 bp size libraries. Sequencing was performed

using Illumina NextSeq high output, single end sequencing. Peak calling was performed using MACS2 (Zhang et al. 2008).

INDEL detection

Highly confident structural variations (SV) among Guy11, O-219, O-137 and O-135 were detected by using assembly-based caller Assemblytics (Nattestad and Schatz 2016) and long-read-based caller Sniffles (version 1.0.12) (Sedlazeck et al. 2018). Briefly, NUCmer (version 3.1) (Kurtz et al. 2004) was used for creating the alignment between one of the assemblies (REFERENCE) and the rest of three assemblies (ASSEMBLY) with the default setting (nucmer -maxmatch -l 100 -c 500 REFERENCE.fa ASSEMBLY.fa -prefix OUT). The delta file output from NUCmer was used as the input file for running Assemblytics with the below setting (Unique sequence length required=1,000, Maximum variant size=100,000 Minimum variant size=10). Raw long-reads were mapped to one of our assemblies by ngmlr (version 0.2.7, -x ont) (Sedlazeck et al. 2018) following indexing and sorting by samtools (version 1.9) (Li et al. 2009). Sorted BAM files served as input for Sniffles detection (version 1.0.12, -s 5 -l 10). SURVIVOR (version 1.0.7, SURVIVOR merge 1000 2 1 0 0 0) was used to identify the SVs identified by both Assmeblytics and Sniffles, which were considered as the highly confident SVs. The highly confident SVs identified using identical reference, but different queries were further merged with SURVIVOR (version 1.0.7, SURVIVOR vcftobed 9 100001) (Jeffares et al. 2017) to remove the redundant SVs among isolates. Insertion and deletion variants of size 4 to 30 bp were gathered as a subset of identified SVs and used for further analysis.

Single-nucleotide polymorphism detection

Single nucleotide polymorphisms (SNPs) were detected in the 86 rice-infecting strains using the Genome Analysis Toolkit (GATK v3.4.4) (Auwera and O'Connor 2020). Each of the

86 strains was mapped against each of the four strains (Guy11, O-135, O-137, and O-219) using gatk MarkDuplicates, then indexed and HaplotypeCaller was run using -ploidy 1 -ERC GVCF. Vcf files were combined using gatk CombineGVCFs and genotyped using gatk GenotypeGVCFs. Then variants with low data to support them were removed using gatk VariantFiltration –filterexpression "AN < 69" and vcftools –gzvcf –remove-filtered-all –recode (Danecek et al. 2011) to remove filtered variants. AN < 69 was chosen so that only variants that had sequencing data, either for the reference or the variant, in at least 80% of the 86 strains were used in further analysis. Finally, vcf entries with more than one variant at a position were removed so that all variants were biallelic SNPs.

SnpEff (Cingolani et al. 2012) was used to identify the impact of SNPs and determine whether they were synonymous or nonsynonymous. Custom databases were made using gff files from our genome assemblies. Biallelic high-confidence filtered vcf files were used as input files to generate the SNP calls.

Average pairwise distance between the strains

Filtered vcf files of *M. oryzae* isolates aligned to assemblies were split by chromosome. Single chromosome variant files were converted into LDHat (2.2) (McVean and Auton 2007) input files. Once the sites and loci input files were generated by chromosome, using a custom Perl script to split the chromosome into sliding windows of 5 kb total, stepping 2.5 kb each step, we calculated genetic diversity statistics by window. Average pairwise distance was normalized by window size to generate the nucleotide diversity stat on a per site basis. Window estimates for diversity statistics were plotted along the chromosome and outliers were generated based on while genome distribution.

Synteny Analysis

To analyze synteny between the four genomes, a pipeline was used to map the genome assemblies to each other and then extract syntenic regions. Assembled genomes were mapped in a one-to-one fashion using Minimap2 (Li 2018). With the following options '-x asm5' preset option for comparison between related genomes with less than 5% divergence. The output alignment files were parsed using the Syri program with default parameters (Goel et al. 2019). To identify syntenic regions between the four genomes, syntenic regions were intersected using bedtools intersect to keep only regions that are syntenic between all four strains (Quinlan and Hall 2010). Finally, the syntenic information was integrated into the 2.5kb windows for each genome. For this, a 2.5kb window was classified as syntenic if at least half of the window overlapped a syntenic region. All 2.5kb windows not classified as syntenic from this analysis were classified as non-syntenic.

Integrated genomic analysis

The expected frequency of SNPs or INDELs per kb was calculated by dividing the total number of variants, SNPs or INDELs, over the total size of a gene, TE, or intergenic region. Then the number of variants in features was split whether it was in a marked (H3K27ac, H3K27me3, H3K9me3) region, previously determined using MACS2, and divided by the size of those sub-groups. Deviation between expected and observed variants was determined using a chi-square analysis, chi2_contingency (Virtanen et al. 2020). Bonferroni multiple testing correction was performed to adjust the critical value to 10.8 at df = 1. Principal component analysis was performed on the data subdivided into 2.5 kb windows. Genomes features were counted within these windows and ChIP enrichment was mapped into 2.5 kb windows. K-means

clustering was performed on the first two components. The distribution of SNPs and INDELs in these clusters was measured.

Figures

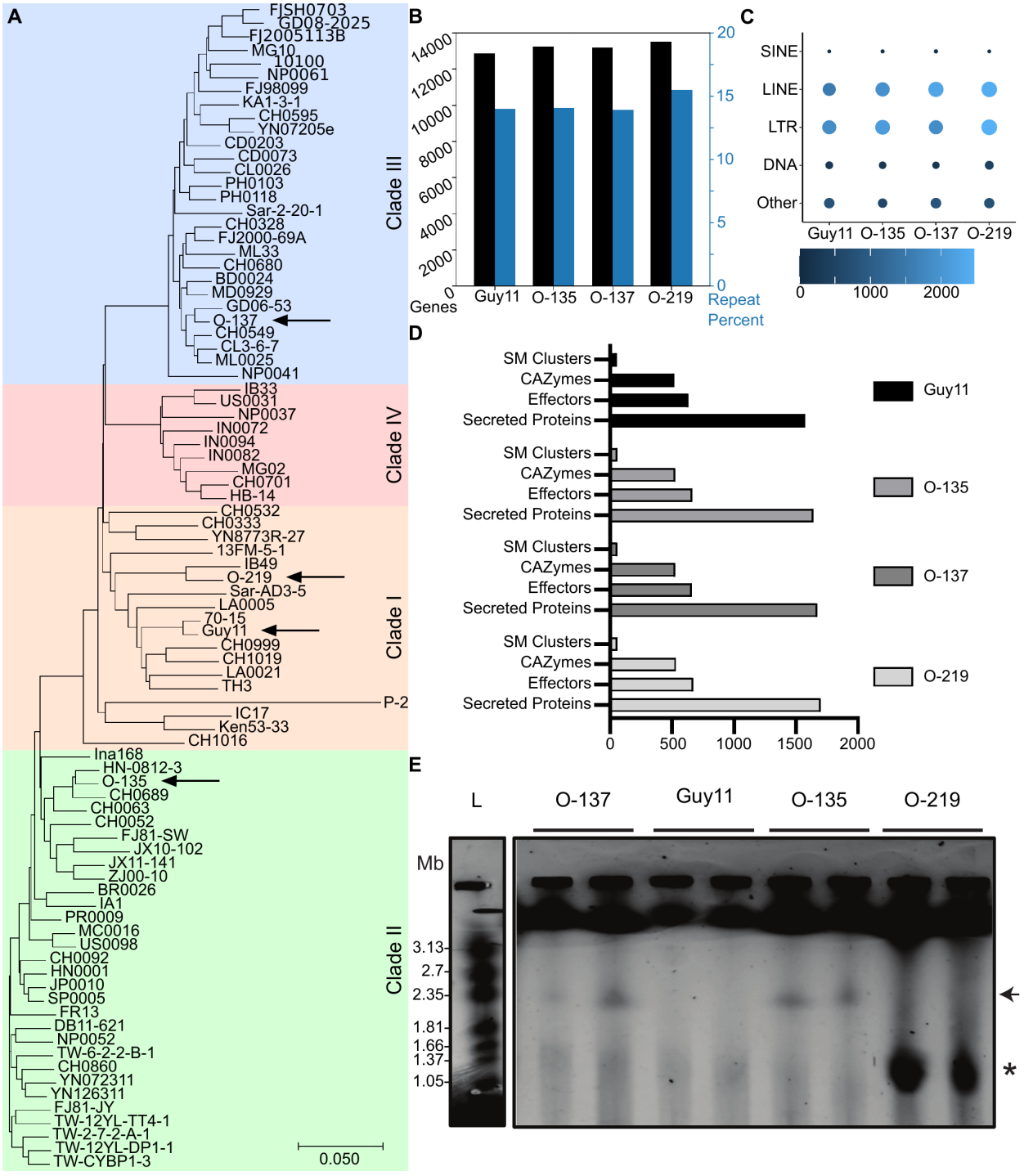

A
FJSH0703
GD08-2025
FJ2005113B
MG10
10100
NP0061
FJ98099
KA1-3-1
CH0595
YN07205e
CD0203
CD0073
CL0026
PH0103
PH0118
Sar-2-20-1
CH0328
FJ2000-69A
ML33
CH0680
BD0024
MD0929
GD06-53
Q-137
CH0549
CL3-6-7
ML0025
NP0041
IB33
US0031
NP0037
IN0072
IN0094
IN0082
MG02
CH0701
HB-14
CH0532
CH0333
YN8773R-27
13FM-5-1
IB49
O-219
Sar-AD3-5
LA0005
70-15
Guy11
CH0999
CH1019
LA0021
TH3
P-2
IC17
Ken53-33
CH1016
Ina168
HN-0812-3
O-135
CH0689
CH0063
CH0052
FJ81-SW
JX10-102
JX11-141
ZJ00-10
BR0026
IA1
PR0009
MC0016
US0098
CH0092
HN0001
JP0010
SP0005
FR13
DB11-621
NP0052
TW-6-2-2-B-1
CH0860
YN072311
YN126311
FJ81-JY
TW-12YL-TT4-1
TW-2-7-2-A-1
TW-12YL-DP1-1
TW-CYBP1-3
0.050
Clade III
Clade IV
Clade I
Clade II
B
14000
12000
10000
8000
6000
4000
2000
0
Genes
Guy11
O-135
O-137
O-219
20
15
10
5
0
Repeat Percent
C
SINE
LINE
LTR
DNA
Other
Guy11 O-135 O-137 O-219
0
1000
2000
D
SM Clusters
CAZymes
Effectors
Secreted Proteins
Guy11
O-135
O-137
O-219
0
500
1000
1500
2000
E
L
O-137
Guy11
O-135
O-219
Mb
3.13
2.7
2.35
1.81
1.66
1.37
1.05
*

Figure 2.1. Four rice infecting *M. oryzae* strains have similar genome compositions based on complete assemblies

(A) Single nucleotide polymorphism (SNP) phylogeny of 86 rice-infecting strains of M. oryzae. Clades are labeled to be consistent with previous M. oryzae population descriptions (Latorre et al. 2020). Arrows indicate the four newly characterized strains from this work. (B) Total gene count (black) and percent repeat content (blue) of the four strains. (C) Summary counts of TEs separated by class from the four strains. (D) Counts of secondary metabolite (SM) clusters, proteins with conventional secretion signals, predicted effectors, and CAZyme genes annotated in the four strains. (E) CHEF gel visualizing the presence of mini-chromosomes in strains O-135, O-137, and O-219. The arrow represents the expected size based on our assembly of the mini chromosome in O-135 and O-137. The star represents the possible three bands that are too close together on the gel to determine the number of mini chromosomes present in O-219.

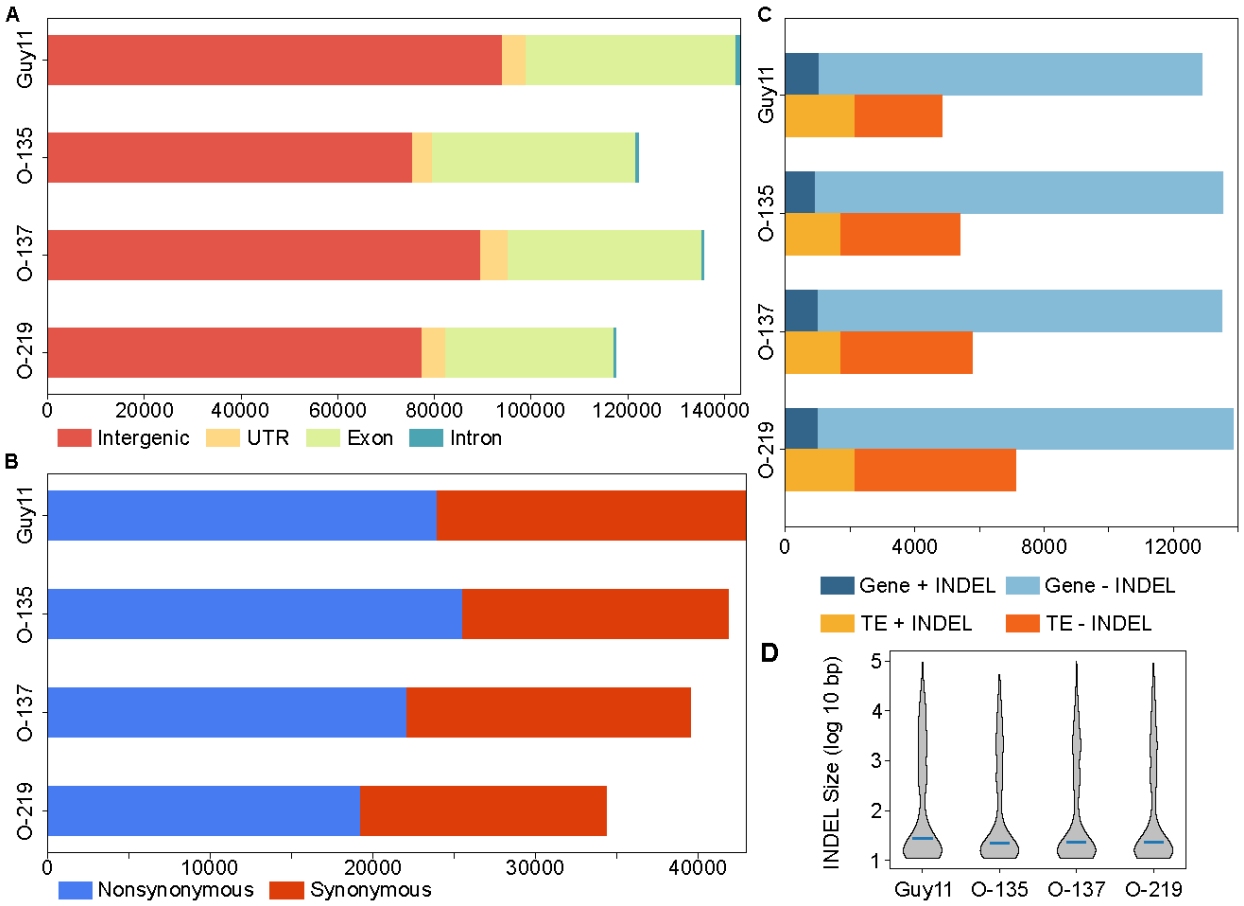


Figure 2.2. Genes and TEs are impacted by multiple types of sequence variation

(A) Gene features where SNPs occur in the four strains. (B) Impact of SNPs in protein coding sequences. (C) Number of genes and TEs that do or do not contain INDEL mutations. (D) Size distribution of INDELs found in the four strains.

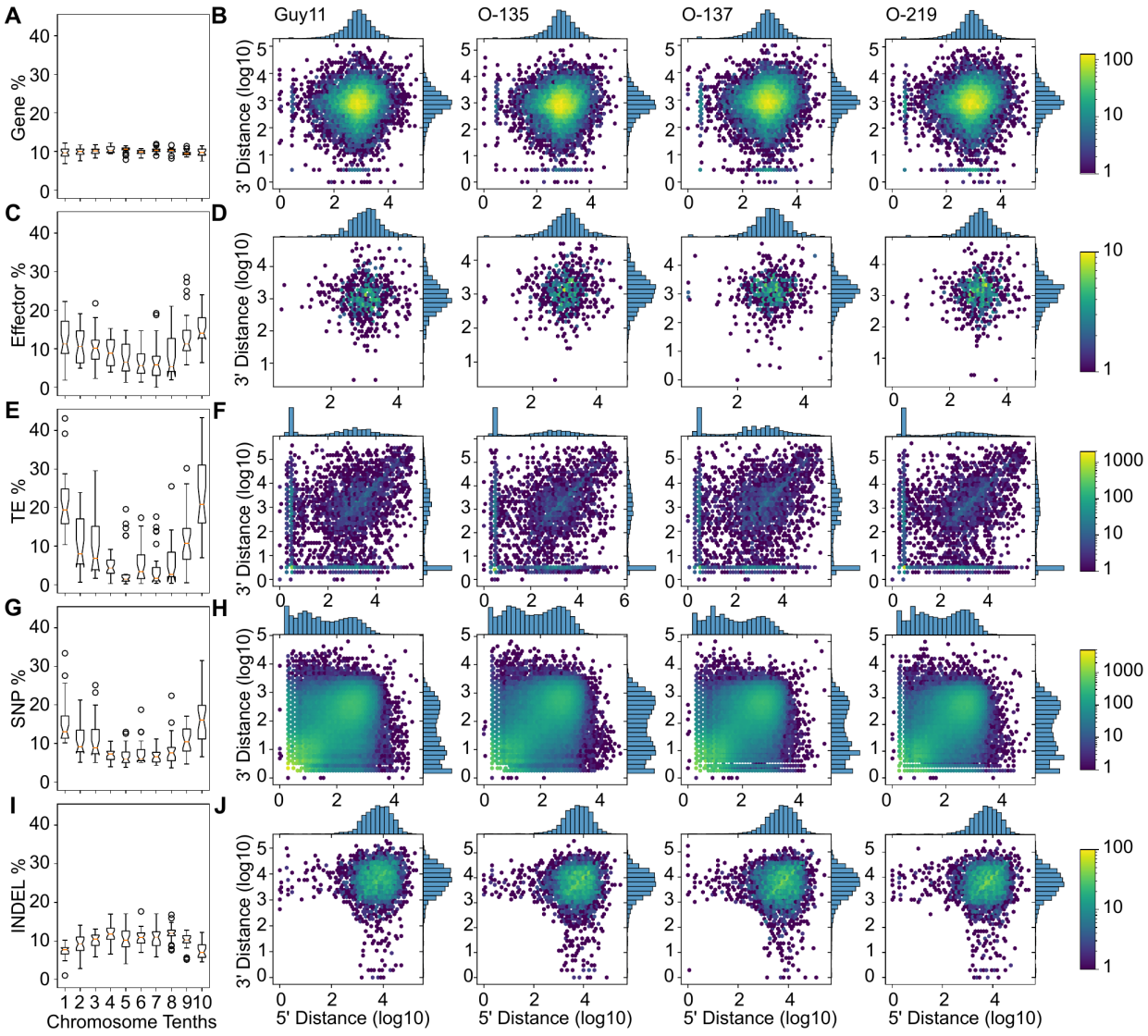


Figure 2.3. Genomic features display different distributions across the genome from the four analyzed genomes

(A) Positional grouping of genes into bins across the chromosomes. Each bin represents 10% of a chromosome. The percentages of genes from the same bin in all core chromosomes from all four strains is shown. (B) Log scaled density of gene distribution for each of the four genomes. Color bar represents number of genes per hex. (C) Positional grouping of predicted effector genes into bins across the. Each bin represents 10% of a chromosome. The percentages of effectors from the same bin in all core chromosomes from all four strains is shown. (D) Log scaled density of effector distribution for each of the four genomes. Color bar represents number

of effectors per hex. (E) Positional grouping of TEs into bins across the. Each bin represents 10% of a chromosome. The percentages of TEs from the same bin in all core chromosomes from all four strains is shown. (F) Log scaled density of TE distribution for each of the four genomes. Color bar represents number of TEs per hex. (G) Positional grouping of SNPs into bins across the. Each bin represents 10% of a chromosome. The percentages of SNPs from the same bin in all core chromosomes from all four strains is shown. (H) Log scaled density of SNP distribution for each of the four genomes. Color bar represents number of SNPs per hex. (I) Positional grouping of INDELs into bins across the. Each bin represents 10% of a chromosome. The percentages of INDELs from the same bin in all core chromosomes from all four strains is shown. (J) Log scaled density of INDEL distribution for each of the four genomes. Color bar represents number of INDELs per hex.

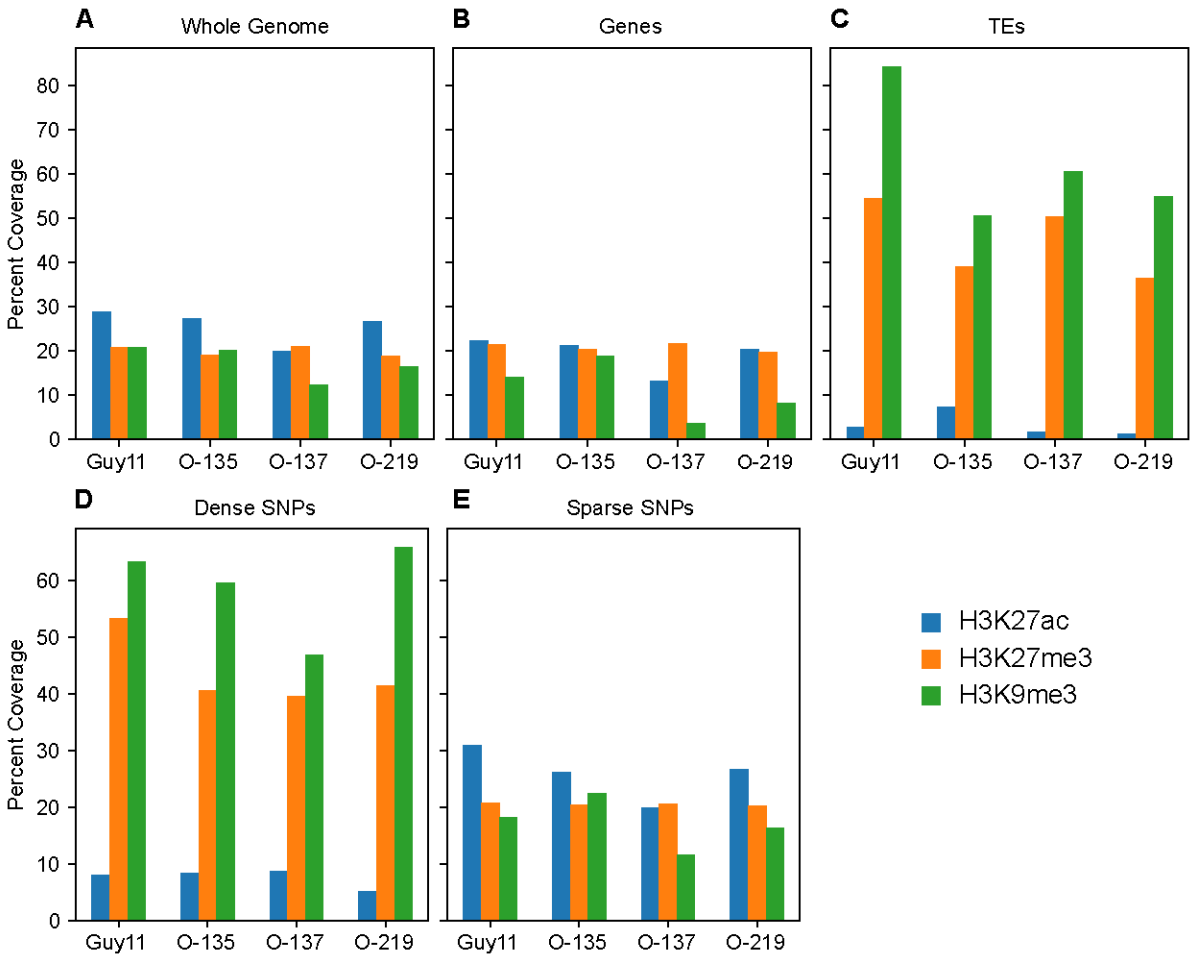


Figure 2.4. Densely grouped SNPs more frequently occur in repressive histone marks
(A) Percentage of the genomes covered by H3K27ac, blue, H3K27me3, orange, and H3K9me3, green. (B) Percentage of genes in regions enriched for H3K27ac, H3K27me3, and H3K9me3. (C) Percentage of TEs in regions enriched for H3K27ac, H3K27me3, and H3K9me3. (D) Percentage of dense SNPs, <100 bp distance between nearest SNP in 5' and 3' direction, in regions enriched for H3K27ac, H3K27me3, and H3K9me3. (E) Percentage of sparse SNPs, >100 bp distance between either nearest SNP in 5' or 3' direction, in regions enriched for H3K27ac, H3K27me3, and H3K9me3.

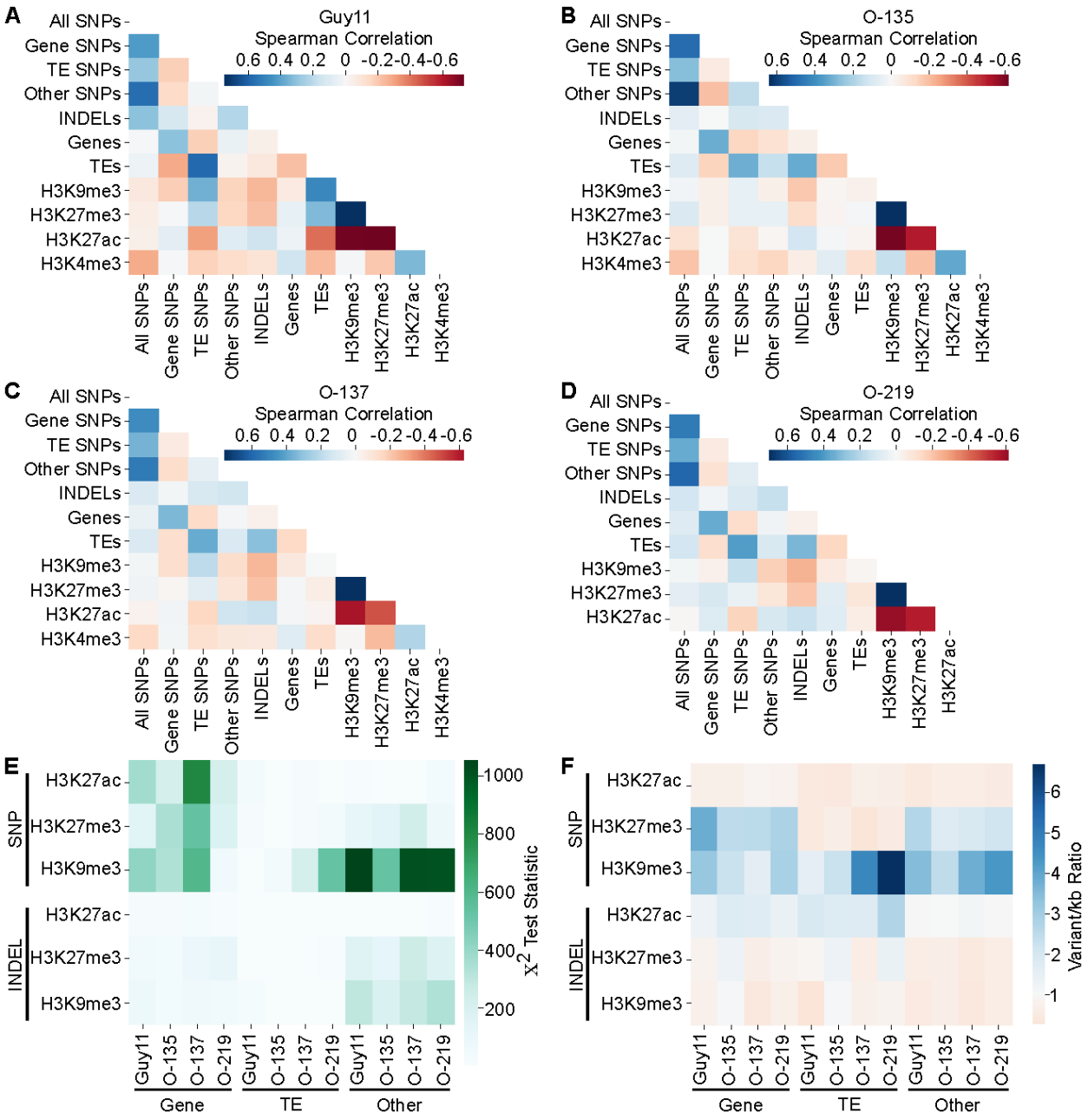


Figure 2.5. Variants in different genomic features are correlated with different epigenetic histone marks

(A) Heatmap of Spearman correlation between genes, TEs, SNPs, INDELs, epigenetic histone marks, and the average pairwise distance (PWD) of the 86 rice-infecting strain sequences in the Guy11 genome. (B) Heatmap of Spearman correlation between genes, TEs, SNPs, INDELs, epigenetic histone marks, and the PWD of the 86 rice-infecting strain sequences in the O-135

genome. (C) Heatmap of Spearman correlation between genes, TEs, SNPs, INDELs, epigenetic histone marks, and the PWD of the 86 rice-infecting strain sequences in the O-137 genome. (D) Heatmap of Spearman correlation between genes, TEs, SNPs, INDELs, epigenetic histone marks, and the PWD of the 86 rice-infecting strain sequences in the O-219 genome. O-219 H3K4me3 ChIP data was not high quality and has been omitted from our analysis. (E) Visualization of chi square analysis between the presence of SNPs or INDELs and the presence of H3K27ac, H3K27me3, or H3K9me3 in genes, TEs, and intergenic regions. Color represents the test statistic from chi square analysis of features that contain a given histone mark. The critical value was determined to be 10.8 at alpha = 0.001 based on Bonferroni multiple testing correction. (F) Heatmap of the ratio of variants per kb, SNPs or INDELs, between marked and unmarked regions of a given histone mark. Higher values indicate more variants present in marked regions than unmarked regions.

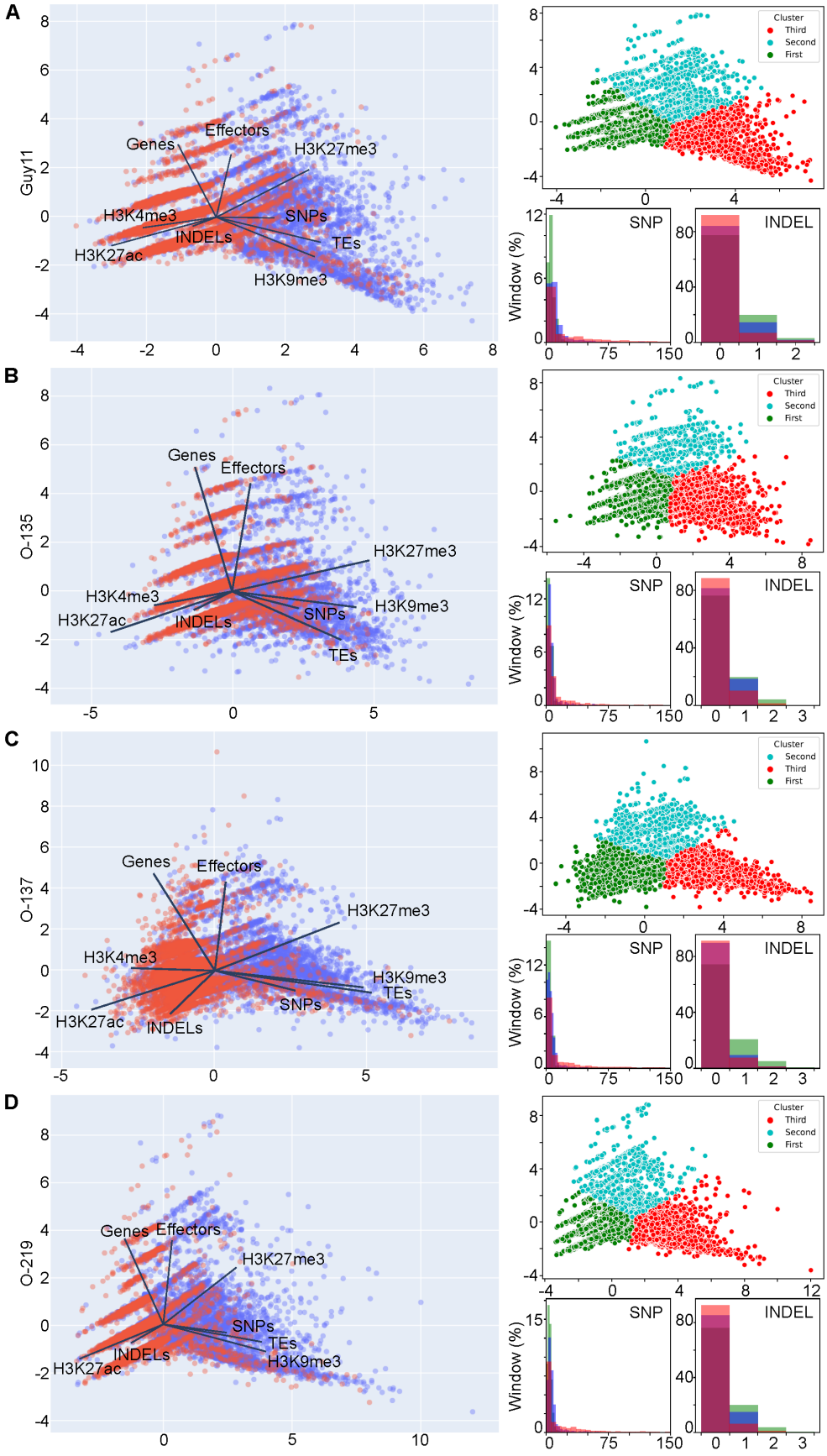

A
Guy11
Genes
Effectors
H3K27me3
H3K4me3
SNPs
INDELs
TEs
H3K27ac
H3K9me3
B
O-135
Genes
Effectors
H3K27me3
H3K4me3
H3K9me3
H3K27ac
INDELs
SNPs
TEs
C
O-137
Genes
Effectors
H3K27me3
H3K4me3
H3K9me3
TEs
SNPs
H3K27ac
INDELs
D
O-219
Genes
Effectors
H3K27me3
SNPs
TEs
INDELs
H3K9me3
H3K27ac
Cluster
Third
Second
First
SNP
INDEL
Window (%)

Figure 2.6. Chromatin clustering in PCA analysis

(A) Principal component analysis (PCA) of genes, effectors, TEs, SNPs, INDELs, H3K27ac, H3K27me3, H3K4me3, and H3K9me3 in Guy11; red points are syntenic windows and blue points are non-syntenic windows. K-means clustering of PCA variance shows statistically distinct groups of windows. Histograms show the distribution of the number of variants per window in each of the clusters. (B) PCA of genes, effectors, TEs, SNPs, INDELs, H3K27ac, H3K27me3, H3K4me3, and H3K9me3 in O-135; red points are syntenic windows and blue points are non-syntenic windows. K-means clustering of PCA variance shows statistically distinct groups of windows. Histograms show the distribution of the number of variants per window in each of the clusters. (C) PCA of genes, effectors, TEs, SNPs, INDELs, H3K27ac, H3K27me3, H3K4me3, and H3K9me3 in O-137; red points are syntenic windows and blue points are non-syntenic windows. K-means clustering of PCA variance shows statistically distinct groups of windows. Histograms show the distribution of the number of variants per window in each of the clusters. (D) PCA of genes, effectors, TEs, SNPs, INDELs, H3K27ac, H3K27me3, and H3K9me3 in O-219; red points are syntenic windows and blue points are non-syntenic windows. K-means clustering of PCA variance shows statistically distinct groups of windows. Histograms show the distribution of the number of variants per window in each of the clusters.

Table 2.1. Assembly metrics of the four genomes

Isolate	Guy11	O-135	O-137	O-219
Total long read sequence (Gb)	3.9	5.93	2.8	6.72
Coverage	86.6x	131.9x	62.1x	149.3x
Contigs	11	11	8	12
Scaffolds	2	0	1	1
Assembly size (bp)	42,642,675	44,503,382	44,234,998	46,802,257
N50 (bp)	6,590,161	5,950,546	5,740,312	6,382,644
Max contig size (bp)	8,304,599	8,309,243	8,325,531	8,280,420
Average contig size (bp)	3,280,206	4,045,762	4,914,999	3,600,174
Genome Completeness	97.5%	97.8%	97.6%	97.7%
tRNAs	241	285	303	330
Genes	12,879	13,253	13,191	13,530
Transcripts (mRNA)	13,050	13,341	13,344	13,462
Secretion Signals	1,574	1,641	1,670	1,702
Effectors	634	663	662	670
Protein Completeness	97.7%	97.7%	97.9%	97.0%
Repeat Content	14.0%	14.1%	13.9%	15.5%
GC Content	51.2%	51.0%	51.1%	51.0%
Mitochondrial Chromosome	1	1	1	1
Mating Type	Mat1-2	Mat1-1	Mat1-2	Mat1-2

Table 2.2. Types of structural variations in the four strains

Strains	Guy11	O-135	O-137	O-219
Insertions	992	589	659	579
Deletions	956	1561	1508	1592
Other	27	29	15	28

Table 2.3. Enrichment of SNPs/kb in repressive histone marks

		Expected	H3K37ac		H3K27me3		H3K9me3	
			No	Yes	No	Yes	No	Yes
		(SNPs/kb)	(SNPs/kb)		(SNPs/kb)		(SNPs/kb)	
Guy11	Gene	3.44	4.33	3.21	2.61	11.07	3.13	12.15
	TE	6.23	17.38	12.35	31.40	9.43	14.13	20.46
	Intergenic	5.79	8.73	4.04	4.81	14.12	4.52	17.03
O-135	Gene	2.45	3.33	2.56	2.25	7.15	2.37	6.97
	TE	4.03	10.28	3.32	10.56	7.15	6.29	13.75
	Intergenic	2.95	4.66	2.71	3.14	6.15	2.95	7.90
O-137	Gene	2.39	2.88	3.21	2.13	7.03	2.67	8.37
	TE	5.09	16.71	18.88	30.40	7.57	11.02	21.38
	Intergenic	3.49	5.20	3.27	3.46	7.74	3.39	15.68
O-219	Gene	2.04	2.72	2.28	1.87	6.38	2.18	8.11
	TE	3.79	17.49	11.40	28.70	8.65	13.70	21.29
	Intergenic	2.95	4.87	2.38	2.97	6.74	2.68	12.23

Table 2.4. Enrichment of INDELs/kb in genes, TEs, and intergenic regions

		Expected	H3K37ac		H3K27me3		H3K9me3	
			No	Yes	No	Yes	No	Yes
		(INDELs/kb)	(INDELs/kb)		(INDELs/kb)		(INDELs/kb)	
Guy11	Gene	0.07	0.56	0.69	0.37	0.44	0.38	0.80
	TE	0.03	0.59	3.45	0.77	0.53	1.96	0.64
	Intergenic	0.15	1.19	0.76	0.52	0.81	0.54	0.65
O-135	Gene	0.07	0.57	0.76	0.41	0.50	0.44	0.67
	TE	0.02	0.59	1.05	0.51	0.55	0.87	0.57
	Intergenic	0.16	0.82	0.76	0.53	0.75	0.56	0.76
O-137	Gene	0.07	0.53	1.02	0.40	0.57	0.40	1.17
	TE	0.02	0.62	11.55	0.73	0.74	1.33	0.62
	Intergenic	0.17	0.86	0.99	0.56	0.76	0.58	0.54
O-219	Gene	0.07	0.56	0.74	0.41	0.56	0.42	1.13
	TE	0.02	0.59	4.02	0.67	0.73	1.85	0.77
	Intergenic	0.17	0.90	0.67	0.52	0.80	0.55	0.85

Table 2.5. Percentage of genome represented by k-means clustering

	First Cluster	Second Cluster	Third Cluster
Guy11	73.1%	15.1%	11.8%
O-135	73.9%	5.8%	20.3%
O-137	72.3%	12.4%	15.3%
O-219	76.7%	12.1%	11.2%

Chapter 3 - Cross-talk between H3K36me3 and H3K27me3

Note: Jun Huang assisted with the first step of sample preparation for ChIP-seq.

Abstract

Epigenetic histone marks serve an important function in the genome to regulate gene transcriptional availability. The field of epigenetics is still uncovering ways in which histone marks interact with each other, called cross-talk, to further modulate their effects in the genome. In filamentous fungi there are contrasting results as to how H3K36me3 mediated by ASH1 and H3K27me3 mediated by polycomb repressive complex 2 (PRC2) engage in cross-talk. Using a Cas9 gene editing approach, we knocked out the two H3K36 methyltransferases in *Magnaporthe oryzae*, SET2 and ASH1, and used chromatin immunoprecipitation sequencing to determine their roles. We find that SET2 and ASH1 function in discrete regions of the genome and that transcriptionally repressed genes and transposable elements are primarily marked with H3K36me3 by ASH1. ASH1 also functions in many of the same regions that are marked with H3K27me3 by PRC2. Loss of ASH1-mediated H3K36me3 causes some regions of the genome that are also marked with H3K27me3 to partially lose H3K27me3. This research shows that ASH1 and PRC2 engage in positive cross-talk, where the loss of the ASH1 enzyme causes a reduction in the function of PRC2.

Introduction

Histone proteins have important roles in a variety of cellular functions, ranging from chromosomal organization during cell division to regulating gene expression (Lai et al. 2022; Roy and Sanyal 2011; Brosch et al. 2008). To fulfill these functions, nuclear DNA is wrapped around histone octamers to form nucleosomes (Luger et al. 2012). Histone proteins can undergo a large variety of post-translational modifications, called epigenetic marks, which modulate how histones function. The epigenetic mark the tri-methylation of H3 lysine 36 (H3K36me3) is generally understood to be an activating mark that is enriched at genes that are actively transcribed, seen across animal, plant, and fungal organisms (Kolasinska-Zwierz et al. 2009; Pajoro et al. 2017; Adhvaryu et al. 2005). However, in some fungi the role of this mark is not always consistent with that description, in *Fusarium graminearum* H3K36me3 is enriched in genes in general regardless of expression level (Connolly et al. 2013). Other epigenetic histone marks, such as H3K27me3, can function as repressive marks that are associated with low gene expression and repeat-rich regions of the genome (Zhang et al. 2021). These epigenetic marks play an important role in formatting the genome into euchromatin and heterochromatin (Freitag 2017). For example, euchromatin is typically defined by activating marks such as H3K4me3 and H3K36me3 and heterochromatin is typically defined by repressive marks such as H3K9me3 and H3K27me3 (Strauss and Reyes-Dominguez 2011). Furthermore, there is evidence to suggest that there is cross-talk between different histone marks that may form a more complex regulatory network in establishing chromatin and regulating gene expression (Zhang et al. 2015). In humans, mice, flies, and even Arabidopsis the presence of H3K4me3 and H3K36me3 could inhibit the production of H3K27me3 in the same local regions of the chromosome (Schmitges et

al. 2011). Less is known about the cross-talk of these histone marks in fungi, which may be more complicated due to the less conserved function of H3K36me3.

The mark H3K36me3 is catalyzed by two enzymes in filamentous fungi, SET2 and ASH1 (Bicocca et al. 2018; Janevska et al. 2018). Both enzymes contain a Su(var)3-9, Enhancer of zeste, Trithorax (SET) domain. Orthologs of these genes have been demonstrated to be specific to H3K36me3 in several organisms including yeast and humans (Strahl et al. 2002; Eram et al. 2015). Despite catalyzing the same epigenetic histone mark, there are reported differences in the function of SET2 and ASH1 in filamentous fungi. In *Neurospora crassa* SET2 was responsible for marking 80% of genes with H3K36me3 and the loss of SET2 resulted in both up- and down-regulation of marked genes (Bicocca et al. 2018). ASH1 was responsible for the remaining marked genes, most of which became more highly expressed after the loss of ASH1. This indicates that ASH1 and SET2 function independently to deposit H3K36me3, resulting in different gene expression outcomes. Even more interesting is the finding that ASH1 catalyzed H3K36me3 is co-localized with H3K27me3 and the loss of a functional ASH1 also results in the loss of H3K27me3 in regions where ASH1 normally functions. In *Fusarium fujikuroi* similar patterns of the division of SET2/ASH1 contributions to total H3K36me3 were found (Janevska et al. 2018). SET2 marked genes found in euchromatin, while ASH1 marked genes in heterochromatin, consistent with results in *N. crassa*. However, a major difference between *F. fujikuroi* and *N. crassa* is that the loss of ASH1 in *F. fujikuroi* resulted in more H3K27me3 at ASH1 marked regions. This stark contrast potentially points to different methods of cross-talk or regulation between H3K36me3 and H3K27me3 in different fungal organisms. Further complicating the picture, an investigation into the regulation of the *Ash1* homolog in *Zymoseptoria tritici* revealed that the epigenetic histone mark H4K20me3 regulated both ASH1-

mediated H3K36me3 and H3K27me3 (Möller et al. 2023). Deletion of *kmt5*, responsible for H4K20me3, resulted in loss of ASH1-mediated H3K36me3 and H3K27me3 and lowered levels of gene transcription, especially in heterochromatin regions. These findings show just how complicated epigenetic cross-talk and regulation may be in filamentous fungi.

Research into both Neurospora and Fusarium demonstrate ASH1-mediated H3K36me3 influencing H3K27me3 enrichment, but in our previous work to understand the dynamics between H3K27me3 and H3K27ac in *Magnaporthe oryzae* we saw evidence of H3K27me3 influencing H3K36me3 (Zhang et al. 2021). H3K27me3 marked domains lost H3K36me3 while unmarked regions gained H3K36me3. It appears that cross-talk between H3K36me3 and H3K27me3 exists in *M. oryzae*, however, it is unclear how ASH1 and SET2 participate in that cross-talk and how the loss of either SET2 or ASH1 will impact H3K36me3 or H3K27me3 in the genome. To address this gap in knowledge, we used a genetic approach to transform the Guy11 strain of *M. oryzae* to remove *Ash1* and *Set2* and used chromatin immunoprecipitation sequencing (ChIP-seq) to identify enzyme specific domains of ASH1 and SET2-mediated H3K36me3 and their impact on H3K27me3.

Results

Homologs of *Ash1* and *Set2* are required for full H3K36 methylation in *M. oryzae*

To study the function of ASH1 and SET2 in *M. oryzae*, we identified homologs of the *N. crassa* proteins in the strain Guy11 through a BLASTp search (Galagan et al. 2003; Dean et al. 2005; Altschul et al. 1990). The *Set2* homolog in *M. oryzae* had a 52% identity and 66% positive alignment against the *N. crassa Set2* and the SET domain had a 77% identity and 88% positive alignment (Figure 3.1A). The *Ash1* homolog had a 58% identity and 71% positive alignment, and the SET domain had an 85% identity and 90% positive alignment (Figure 3.1B). Gene deletion mutants for both coding sequences were generated in *M. oryzae* strain Guy11 (WT) using a Cas9 ribonucleoprotein editing approach (Huang and Cook 2021; Foster et al. 2018). A paired cut strategy, using two Cas9 single guide RNA (sgRNA) targeting near the start and stop sites, was used to fully remove the target gene (Figure 3.2A). Donor DNA coding for antibiotic resistance, such as against hygromycin (*Hyg*) or neomycin (*G418*), was included in protoplast transformation for genome integration at the target locus. Colonies growing on selection following transformation were PCR genotyped to confirm gene deletion using a series of PCR primers (Figure 3.2B, C). *Ash1* transformant #14 was PCR negative for the control and was PCR positive for integration of *Hyg* at the Ash1 locus. Knockouts of *Ash1* (Δ*ash1*) were identified to lack the ASH1 coding sequence and instead contain the hygromycin coding sequence at the locus (Figure 3.2B). *Set2* transformants were transformed using *G418* and genotyped with the same primer design methodology as *Ash1*. Primers upstream and downstream of *Set2* paired with primers in *G418* have predicted PCR products sizes of 2 kb and 2.1 kb respectively (Figure 3.2C). The positive control, using primers upstream and inside of *Set2*, produce PCR products 3.6 kb long. *Set2* transformant #20 was PCR negative for the control and showed positive PCR

results for *G418* integration in the *Set2* locus. Double knockout strains for both the ASH1 and SET2 coding sequence were also made. For these strains, both single knockout strains were used to make the other gene knockout in case there was a parental genotype impact on H3K36me3 deposition and inheritance. Double knockout transformants were collected, 1-8 were transformed from the Δ*ash1* strain and 9-22 were transformed from the Δ*set2* strain (Figure 3.2D). Double knockout strains 6, 8, 15, 17, and 19 were selected for continued investigation. Western blotting showed that both Δ*ash1* and Δ*set2* had less H3K36me3 compared to the WT (Figure 3.2E). The five tested Δ*ash1*Δ*set2* strains showed a complete loss of H3K36me3 (Figure 3.2E). We additionally checked the level of H3K27me3 in our transformants given our previous results in *M. oryzae* that showed loss of H3K27me3 altered H3K36me3 (Zhang et al. 2021). There were similar levels of H3K27me3 in the WT, Δ*ash1*, and Δ*set2* strains, however, all of the double knockouts showed reduced H3K27me3 (Figure 3.2F). Taken together, these results show that ASH1 and SET2 both catalyze the tri-methylation of H3K36. Previously, SET2 has been reported to contribute 80-95% of total H3K36me3 in other filamentous fungi (Bicocca et al. 2018; Janevska et al. 2018), however, we find in *M. oryzae* that SET2 and ASH1 appear to contribute more equally to the deposition of H3K36me3. These results are inconsistent with previously published findings in *M. oryzae*, deletion of *Set2* in one study did not result in a change in H3K36me3 abundance (Pham et al. 2015), and deletion of *Ash1* in another study also did not result in a change in H3K36me3 (Cao et al. 2016). No double knockout strains have previously been reported. In addition to the presumed direct impact on H3K36me3, our results suggests that the loss of H3K36me3 in Δ*ash1*Δ*set2* strains results in reduced H3K27me3. Interestingly, in Fusarium Δ*ash1* or Δ*set2* showed impacts on H3K36me3 and H3K27me3 (Janevska et al. 2018), while in Neurospora only Δ*ash1* was shown to also impact H3K27me3

(Bicocca et al. 2018). This suggests that the roles of ASH1 and SET2 may differ slightly in different filamentous fungi.

Loss of H3K36me3 results in slower *M. oryzae* growth and lower pathogenicity

To characterize the phenotypic impacts of altered H3K36me3 levels in *M. oryzae*, we examined growth rates and pathogenicity of the strains. Each of the strains were grown in triplicate over two experiments, and the colony diameter measured after 12 days of growth at 28°C (Figure 3.3A, B). WT colonies grew the most during the 12-day growth period, and both Δ*ash1* and Δ*set2* had diminished growth compared to the WT. All double knockouts grew at slower rates than either of the single knockouts, which suggests that there is a cumulative effect of removing both H3K36me3 lysine methyltransferases. These findings are consistent with data from other fungi (Cao et al. 2016; Bicocca et al. 2018; Janevska et al. 2018). To test pathogenicity, the susceptible rice cultivar YT-16 was infected with spores from each of the strains and visually rated (Valent et al. 1991) (Figure 3.3C, D). Infection by WT was the most successful, Δ*ash1* had reduced infectivity, and Δ*set2* and the double knockouts were compromised pathogens (Figure 3.3C). Infection by Δ*set2* primarily produced dark black spots on the leaves, and rarely formed larger lesions (Figure 3.3C). Infection by the double knockout strains produced a variety of infection results ranging from similar virulence of the single knockout strains to complete inability to cause infection (Δ*ash1*Δ*set2* #17), indicating severe defects on the normal infection process (Figure 3.3C, D). It is important to note that H3K36me3 has a general and global impact on gene transcription in many eukaryotes. Therefore, it is not surprising that the reported knockouts have altered growth and development, and these results alone do not suggest H3K36me3 is a regulator or specifically involved in host virulence. These

results show that H3K36me3 is important for normal growth and disruption impacts pathogenicity.

Ash1 and *Set2* deposit H3K36me3 in different regions of the genome

We next sought to understand the role of the two different enzymes depositing H3K36me3. Given that protein blotting indicated that both enzymes can deposit H3K36me3, we were interested to know if they catalyze H3K36me3 at the same or different genomic regions. Genome-wide histone modification maps were generated by chromatin immunoprecipitation and sequencing (ChIP-seq) for H3K36me3 in the WT and mutant strains (Supplemental Figures B.1-5). In WT, H3K36me3 covers approximately 44% of the genome. Of the parts of the genome that are marked by H3K36me3, ASH1 and SET2 independently contribute roughly 40% of the total H3K36me3 (Figure 3.4A). The remaining 20% of H3K36me3 appears to be overlapped by both ASH1 and SET2 which suggests that in those particular regions there is functional redundancy. SET2 has a larger role than ASH1 in marking genes with H3K36me3, however when specifically looking at effector genes ASH1 contributes the most (Figure 3.4A). This is a remarkable finding given that predicted effectors are a small subset of the total coding sequence (~5%), and it is not clear what would account for ASH1 catalyzing 80% of H3K36me3 at effectors. Transposable elements marked with H3K36me3 are also mostly marked by ASH1 or by both ASH1 and SET2. Results from H3K36me3 ChIP-seq reveal that the differences between the genes and TEs marked by ASH1 and SET2 form discrete and continuous regions of the chromosome (Figure 3.4B). These results confirm the previous results that both enzymes participate in H3K36me3, but they appear to function in different regions of the genome. In the double knockouts Δ*ash1*Δ*set2* #6 and #15, nearly all H3K36me3 is lost, consistent with results from the western blot, but there were regions that retained clear ChIP-signal, that seemed to

largely overlap with ASH1-mediated H3K36me3 (Figure 3.4B). Given the presumed function role of H3K36me3 in transcription, we sought to test if SET2 versus ASH1 had specific associations with gene transcription. Annotated genes were grouped into expression quartiles and the H3K36me3-ChIP signal was analyzed (Figure 3.4C). Counter to expectations, the lowest expressed genes had the highest H3K36me3 profile, while the other quartiles representing higher expressed genes had similar levels of lesser enrichment (Figure 3.4C). Splitting the results by mutant showed that SET2-mediated H3K36me3 specifically marked genes in the top three expression quartiles (Figure 3.4C). Conversely, the lowest expressed genes were the most enriched for ASH1-mediated H3K36me3 (Figure 3.4C). These findings confirm that both ASH1 and SET2 function to deposit H3K36me3, but that they function in different regions of the genome. It is interesting that SET2 marks highly expressed genes and essentially never marks TEs, while the function of ASH1 appears to be the opposite, marking lowly transcribed genes, many effectors and TEs. These results suggest that the genome is capable of differentially interpreting H3K36me3 deposited by the two lysine methyltransferases.

Loss of H3K36me3 alters the enrichment of H3K27me3

Our previous results indicated that PRC2 deposited H3K27me3 influenced a subset of H3K36me3 peaks (Zhang et al. 2021). To understand how H3K36me3 deposition might also influence H3K27me3, we compared ChIP-seq results between WT and gene deletion strains (Figure 3.5A). In the Δ*ash1* strain, H3K36me3 is reduced in regions that are also marked with H3K27me3 (Figure 3.5A). The Δ*set2* strain shows a nearly opposite result where much of the H3K36me3 that is lost is outside of H3K27me3 regions. This indicates that ASH1 and PRC2 are functioning in the same regions of the genome and that these regions are generally separate from regions marked by SET2. Peaks for each histone modification were called using MACS2 and

overlapped, which shows that SET2 deposits H3K36me3 in 16.0% of H3K27me3 regions and ASH1 deposits H3K36me3 in 87.7% of H3K27me3 regions (Supplemental Figure B.6). In the double knockouts the small amount of H3K36me3 left also occurs in H3K27me3 regions. To quantify the histone cross talk, the ratio of H3K36me3 enrichment in the various strains compared to the WT was calculated. Many windows in H3K27me3 peaks of Δ*ash1* had lower H3K36me3 enrichment compared to the WT, again indicating that ASH1 functions in H3K27me3 regions (Figure 3.5B). Conversely, windows in H3K27me3 peaks in Δ*set2* showed increased H3K36me3, possibly suggesting an inhibitory relationship between SET2 and ASH1 (Figure 3.5B). Interestingly, windows in Δ*ash1*Δ*set2* #6 and Δ*ash1*Δ*set2* #15 showed a diversity of responses, in some cases losing H3K36me3 and in other cases gaining H3K36me3. The cause of the H3K36me3 that does appear in these strains is unknown. Overall, these results show that H3K27me3 positively interacts with ASH1-mediated H3K36me3, but negatively interacts with SET2-mediated H3K36me3. H3K27me3 in SET2-catalyzed domains surprisingly showed a small increase in enrichment on average in Δ*ash1* and a small reduction in Δ*set2* (Figure 3.5C). H3K27me3 also slightly increased in SET2 domains in both Δ*ash1*Δ*set2* #6 and Δ*ash1*Δ*set2* #15. H3K27me3 behaved much differently in ASH1-catalyzed domains. In Δash1 some regions of ASH1 domains lost H3K27me3 while others maintained or even slightly gain H3K27me3 (Figure 3.5D). Similar patterns exist in Δ*set2*, Δ*ash1*Δ*set2* #6 and Δ*ash1*Δ*set2* #15, however, more regions lost H3K27me3 in those strains than in Δ*ash1* (Figure 3.5D). These results show that ASH1 functions in H3K27me3 regions and that the loss of H3K36me3 can change the enrichment of H3K27me3.

Discussion

The goal of this research was to characterize the role of ASH1 and SET2 in mediating the production of the epigenetic histone mark H3K36me3 in the filamentous fungus *Magnaporthe oryzae*. Genetic knockout analysis revealed that both ASH1 and SET2 contribute roughly equally towards the total genomic H3K36me3. We find that SET2-mediated H3K36me3 makes up a larger percentage of total genic H3K36me3, while ASH1-mediated H3K36me3 predominates transcriptionally silent genes. This also supports findings in *Neurospora crassa* where SET2-mediated H3K36me3 was enriched in actively transcribed genes but ASH1-mediated H3K36me3 was enriched in lowly expressed genes (Bicocca et al. 2018). It is interesting that effectors that contain H3K36me3 are nearly entirely mediated by ASH1. Related to this, we found that ASH1-mediated H3K36me3 was typically found in the same genomic regions as H3K27me3, which we previously documented as being enriched at effectors (Zhang et al. 2021). It is not clear why this organization exists in the genome, which epigenetic mark drives the relationship, or if they serve the same function. The results support the idea that ASH1 functions in heterochromatin-type regions of the genome, opposite of SET2. The SET2 protein contains a domain responsible for interacting with RNA polymerase II, called Set2 Rpb1 Interacting (SRI) domain (Krogan et al. 2003). This domain is required for normal transcription elongation in yeasts (Kizer et al. 2005). This domain is absent from ASH1, and likely explains at least why AHS1-mediated H3K36me3 is not associated with actively transcribed genes. Further research to understand the targeting mechanism of ASH1, and ultimately its role in genome regulation will provide new information on epigenome function for this conserved modification.

ASH1-mediated H3K36me3 and PRC2-mediated H3K27me3 exhibit cross-talk. In H3K27me3 domains, ASH1-mediated H3K36me3 increased in the Δ*set2* strain compared to the

WT, while SET2-mediated H3K36me3 decreased in the Δ*ash1* strain. This demonstrates a positive interaction between ASH1 and PRC2. The data support our previous observations that loss of H3K27me3 resulted in H3K27me3 rich regions to lose H3K36me3 but other regions to gain H3K36me3 (Zhang et al. 2021). The H3K36me3 in H3K27me3 rich regions is likely ASH1-mediated based on our findings. The cross-talk between ASH1 and PRC2 and their respective marks does not appear to be necessary for PRC2 to deposit H3K27me3. In all tested transformed strains, H3K27me3 marks were still present in the same regions as the WT, however, the relative abundance of H3K27me3 was decreased in Δ*ash1* and Δ*ash1*Δ*set2* strains. This indicates that ASH1 is not required to determine where PRC2 functions, but in some way promotes its function. This is in contrast to what is reported in *N. crassa*. ASH1 has been demonstrated to be required in *N. crassa* for H3K27me3 deposition, and a nonfunctional *Ash1* gene results in the complete loss of H3K27me3 (Bicocca et al. 2018). Research in *F. fujikuroi* shows Δ*ash1* strains gain H3K27me3 in gene bodies that are typically marked with H3K36me3 by ASH1 in heterochromatic regions (Janevska et al. 2018). It appears that in *F. fujikuroi* ASH1 is neither necessary to direct PRC2 to function nor positively interacting with PRC2 to facilitate H3K27me3 deposition. Interestingly, a modified H3 protein (H3K36A) which cannot be methylated showed a lower impact on H3K27me3 changes which indicates that the enzyme ASH1 is more important in the cross-talk with PRC2 than the H3K36me3 mark in Fusarium. In both Drosophila and human systems, PRC2 activity is reduced in the presence of H3K36me3 (Schmitges et al. 2011) indicating a different pattern of cross-talk in animals, although only SET2 tri-methylates H3K36 in animals which makes it difficult to separate the mark from the enzyme. Recent characterization of *Kmt5*, responsible for H4K20me3, in *Zymoseptoria tritici* has shown a more complicated regulatory network of histone marks. Deletion of *kmt5* results in the

loss of H4K20me3 and in some regions of the genome that is accompanied by the loss of ASH1-mediated H3K36me3 and PRC2-mediated H3K27me3, however, other regions of the genome only lose ASH1-mediated H3K36me3 but not H3K27me3 (Möller et al. 2023). This shows that *Kmt5* and H4K20me3 regulate ASH1 and H3K36me3, and sometimes both also regulate H3K27me3. Through these comparisons, it becomes apparent that there likely is not a conserved pattern of cross-talk between ASH1 and PRC2 in filamentous fungi.

Although H3K36me3 is frequently discussed in the context of euchromatin (Freitag 2017), it is possible that that association is not due to an activating function of the mark or reader protein that interprets the mark. In yeast, several deacetylase enzymes have been characterized and recognize SET2-mediated H3K36me3 (Carrozza et al. 2005). These deacetylases remove activating histone marks to temporarily lower the transcriptional availability. In this way, RNAPII transcribes the gene while an associated SET2 methylates H3K36, then deacetylases recognize H3K36me3 and remove acetyl groups to prevent spurious transcription and defects in transcription elongation (Carrozza et al. 2005). Some genes also contain cryptic promoters which are intragenic locations where transcription can begin in the middle of a gene (Smolle and Workman 2013). Transcripts initiated from these cryptic promoters was repressed by H3K36me3 (Carrozza et al. 2005). SET2-mediated H3K36me3 may function like a “gene speedbump” to regulate transcription. This framework alleviates the problem of having one histone mark somehow recognized for two separate and opposing functions of simultaneously being associated with actively transcribed genes but also found with repressed genes. It appears that SET2 functions to prevent over-transcription of active genes in euchromatin, while ASH1 and PRC2 work in concert to maintain longer-term repression in heterochromatin. Our results show that typically repressed genes, such as effectors, are more often marked by ASH1, however, future

work to demonstrate potential transcriptional changes as a result of disruption of either ASH1 or SET2 needs to be done to better support that hypothesis.

Materials and Methods

Fungal strain preparation

The Guy11 strain of *Magnaporthe oryzae* was used as the wild type (WT) strain through this research. Fungal strains were grown on complete media (10 g/L sucrose, 6 g/L casamino acids, 6 g/L yeast extract, 20 g/L agar; for solid media) for transformations and growth assays or on oatmeal media (72.5 g/L oatmeal agar) to promote sporulation. To test the growth of *M. oryzae*, strains were grown on plates of complete media at 28 °C in the dark for 12 days. The diameter of the fungal colonies was measured on the twelfth day.

Fungal transformation

Transformed strains for Δ*ash1* and Δ*set2* and double knockout strains Δ*ash1*/Δ*set2* were made using CRISPR/Cas9 gene editing based on a previously published protocol (Huang and Cook 2021). Briefly, *Hyg* and *G418* donor DNA were synthesized from plasmids pFGL821 and pFGL921, respectively. sgRNA was synthesized using EnGen sgRNA Synthesis Kit (NEB #E3322) (Supplemental Table B.1), then assembled with Cas9 to form the RNP. Guy11 was grown in liquid complete media at 28 °C in the dark on a shaker at 120 RPM for 3 to 4 days. Mycelia was lysed to release protoplast which were then incubated over night at 28 °C in the dark on a shaker at 90 RPM with donor DNA and a pair of RNPs to target DSBs up- and downstream of the target gene. The next day they were plated in solid complete media containing antibiotics appropriate for the donor DNA. After 7-10 days of growth on selective media, colonies were collected and PCR genotyped to determine if the donor DNA had integrated at the target site and deleted the target gene (Supplemental Table B.1).

Plant materials and infection assay

The susceptible rice cultivar YT-16 was used for leaf tissue infection assays. Six germinating seeds were sown in each pot with Baccto soil. Rice plants were grown for 21 days before their leaves were spray inoculated with $5x10^4$ spores/mL in 0.25% gelatin solution as described in (Valent et al. 1991). Spores were collected from strains grown on oatmeal agar in light conditions. Two pots were sprayed for each tested strain and the experiment was performed two times separately. Plants were incubated for 24 hours in dark conditions immediately following inoculation, then incubated for 6 days in a growth chamber. On the seventh day after inoculation, the infection was scored on the youngest leaf that was present during inoculation (Valent et al. 1991). Leaves were scored on the following scale: 0 – Fully green leaf, no evidence of infection, 1 – Black/brown lesions with no center and diameter < 0.5 mm, 2 – Small black/brown lesion with tan center and diameter < 1 mm, 3 – Lesion with diameter < 2 mm, 4 – Lesion with diameter < 4 mm, 5 – Large lesion 5 mm or larger diameter (Valent et al. 1991).

Western blotting assay

Strains were grown in liquid complete media for 3 to 4 days at 28 °C in the dark on a shaker at 120 RPM. 300 mg of mycelia was collected from each strain and homogenized using a bead mill. Ground mycelia was resuspended in protein lysis buffer (163 mL water, 10 mL 1M Tris-HCl pH 7.5, 5 mL 4M NaCl, 2 mL 0.5M EDTA, 20 mL 10% Triton X, 100 uL PMSF, 50 uL Leupeptin, 5 uL Pepstatin, 50 uL E64). 4x Laemmli buffer was added to samples and incubated at 95 °C for 5 minutes, then moved to ice. Proteins were separated on a precast 4-20% protein gel (RioRad, #4561093) at 130V for 40-50 minutes. Proteins were wet-transferred onto nitrocellulose membranes overnight at 20 V in 4 °C. Western blotting was done using iBind following manufacturer's protocol (Thermo Fisher). The primary antibodies used were anti-

H3K36me3 (Diagenode, C15210013) and anti-H3K27me3 (Diagenode, C15200181). The secondary antibody used was Goat Anti-mouse IgG H&L (HRP) (Abcam, #ab6789) and Goat Anti-rabbit (HRP) (Epigentek, #A12004-1). Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare, #45-002-401) was used to image the proteins on the membrane.

Chromatin immunoprecipitation sequencing

Chromatin Immunoprecipitation sequencing was performed following our previously published methodology (Zhang et al. 2021). Briefly, proteins were crosslinked to DNA in mycelia grown in liquid culture using formaldehyde and then flash frozen in liquid nitrogen. 200 mg of frozen tissue per sample was ground and DNA was sheared using sonication. Sheared DNA was treated with MNase and then precleared using an 80:20 ratio of A and G beads. Samples were then incubated for 14 hours with the following antibodies: H3K27me3 (abcam 6002) and H3K36me3 (abcam 9050). Antibodies were then pulled down with the A/G beads and washed thoroughly. DNA was purified using the iPure kit from Diagenode. Libraries were prepared using NEBNext FFPE DNA repair mix and Ultra II End-prep mix. A universal adaptor was then ligated to DNA samples and IDT i5 and i7 primers were used in a PCR reaction to add unique barcodes to each of the samples. DNA fragments were double size-selected after PCR to enrich for approximately 300 bp size libraries. Sequencing was performed using Illumina. Replicates were normalized using deepTools bamCoverage BPM (bins per million) (Ramírez et al. 2016). Then replicates were combined using deepTools multiBigwigSummary. Peak calling was performed using MACS2 (Zhang et al. 2008). H3K27me3 and H3K9me3 peaks that were less than 1 kb from each other were merged using BEDTools (Quinlan and Hall 2010).

Figures

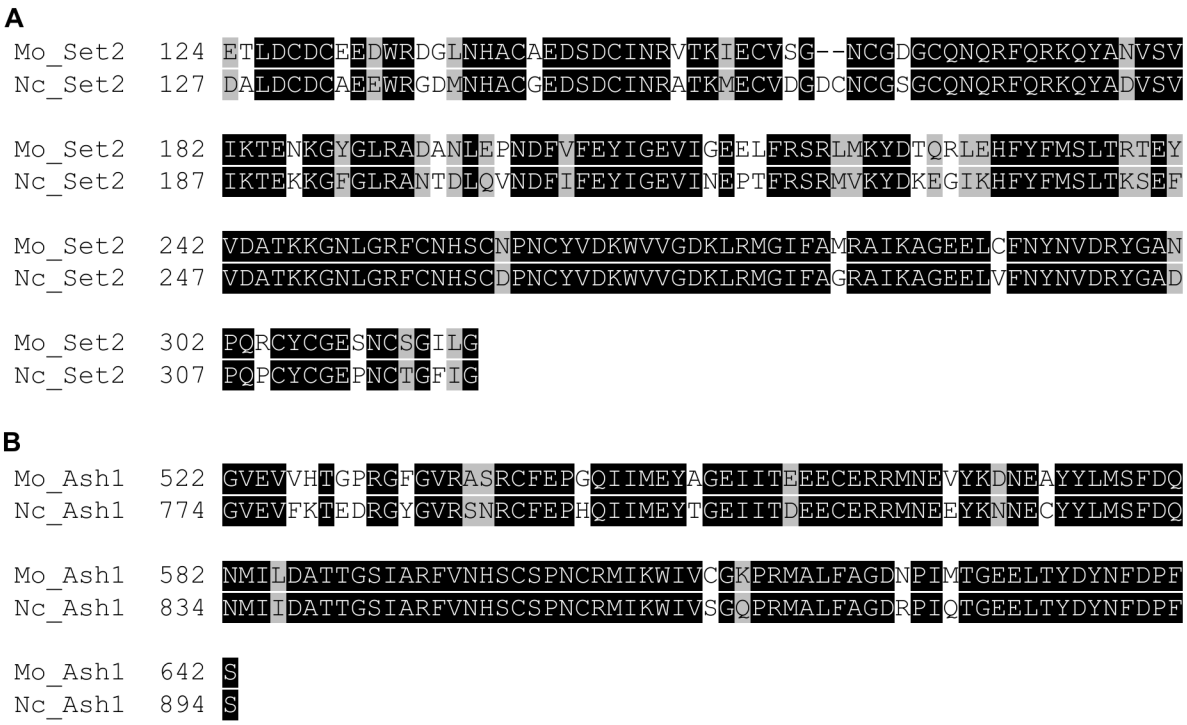


Figure 3.1. Alignment of *M. oryzae* SET2 and ASH1 against *N. crassa* homologs

(A) Alignment of SET2 SET domain between *M. oryzae* and *N. crassa*. (B) Alignment of ASH1 SET domain between *M. oryzae* and *N. crassa*.

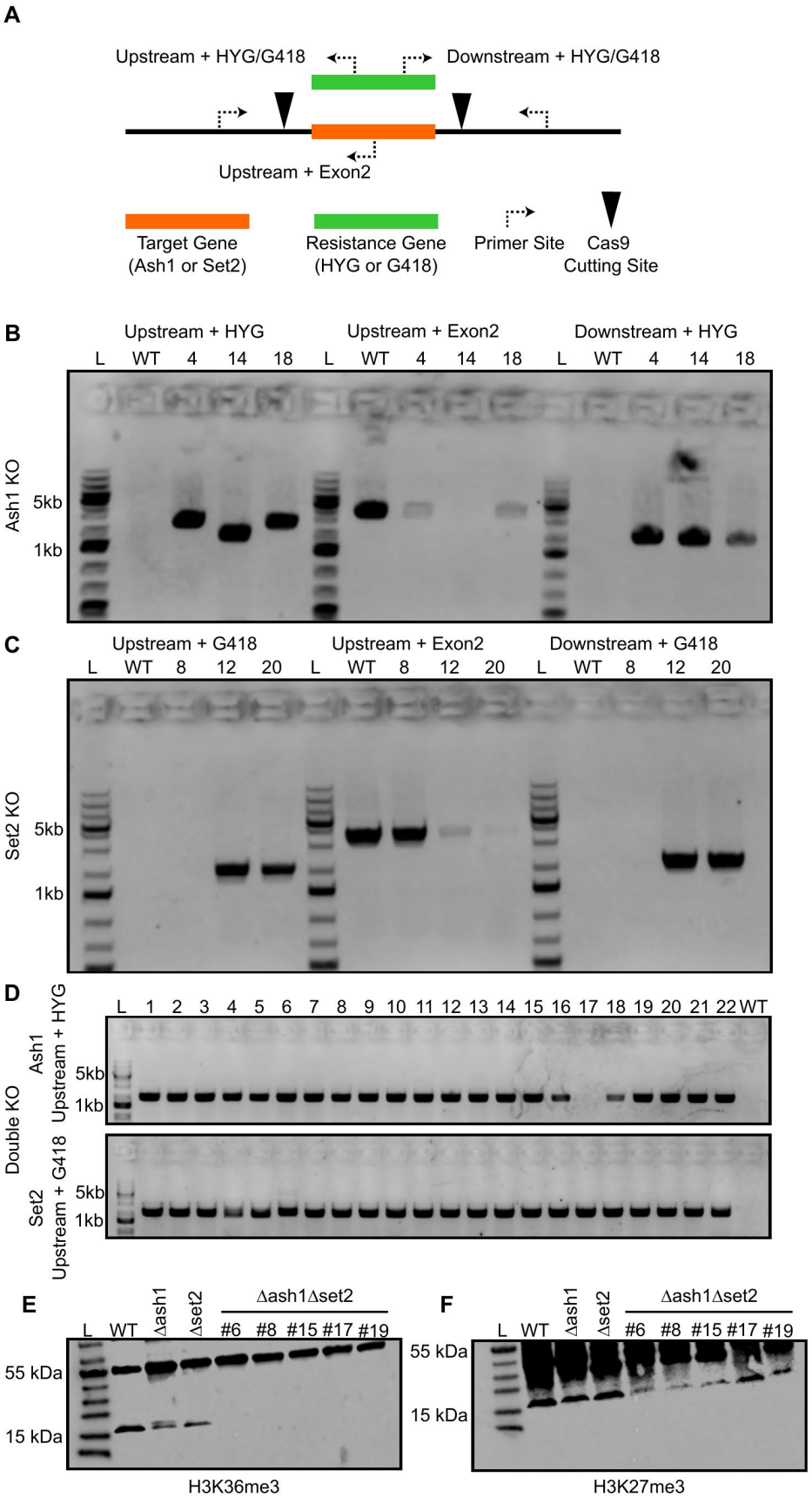

A
Upstream + HYG/G418
Downstream + HYG/G418
Upstream + Exon2
Target Gene
(Ash1 or Set2)
Resistance Gene
(HYG or G418)
Primer Site
Cas9
Cutting Site
B
Upstream + HYG
Upstream + Exon2
Downstream + HYG
L WT 4 14 18 L WT 4 14 18 L WT 4 14 18
Ash1 KO
5kb
1kb
C
Upstream + G418
Upstream + Exon2
Downstream + G418
L WT 8 12 20 L WT 8 12 20 L WT 8 12 20
Set2 KO
5kb
1kb
D
L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 WT
Double KO
Ash1
Upstream + HYG
Set2
Upstream + G418
5kb
1kb
E
L WT Δash1 Δset2 Δash1Δset2
#6 #8 #15 #17 #19
55 kDa
15 kDa
H3K36me3
F
L WT Δash1 Δset2 Δash1Δset2
#6 #8 #15 #17 #19
55 kDa
15 kDa
H3K27me3

Figure 3.2. Validation of Cas9 editing of *Ash1* and *Set2*

(A) Graphic outline of gene editing procedure. Double strand breaks (DSBs) were made at target genes <50 bp upstream from the start codon and <50 bp downstream from the stop codon using Cas9. An antibiotic resistance gene, *Hyg* or *G418*, was inserted between the DSBs. (B) Polymerase chain reaction (PCR) gel image of the wild type (WT) strain and three potential *Ash1* transformants. Primer pairs are upstream of the target gene and in the antibiotic resistance gene, upstream and inside an exon of the target gene, and downstream of the target gene and in the antibiotic resistance gene. (C) PCR gel image of the WT strain and three potential *Set2* transformants. Primer pairs are upstream of the target gene and in the antibiotic resistance gene, upstream and inside an exon of the target gene, and downstream of the target gene and in the antibiotic resistance gene. (D) PCR gel images of potential double KO transformants. (E) Western blot of H3K36me3 in the WT, Δ*ash1*, Δ*set2*, Δ*ash1*Δ*set2* #6, 8, 15, 17, and 19 strains. (F) Western blot of H3K27me3 in the WT, Δ*ash1*, Δ*set2*, Δ*ash1*Δ*set2* #6, 8, 15, 17, and 19 strains.

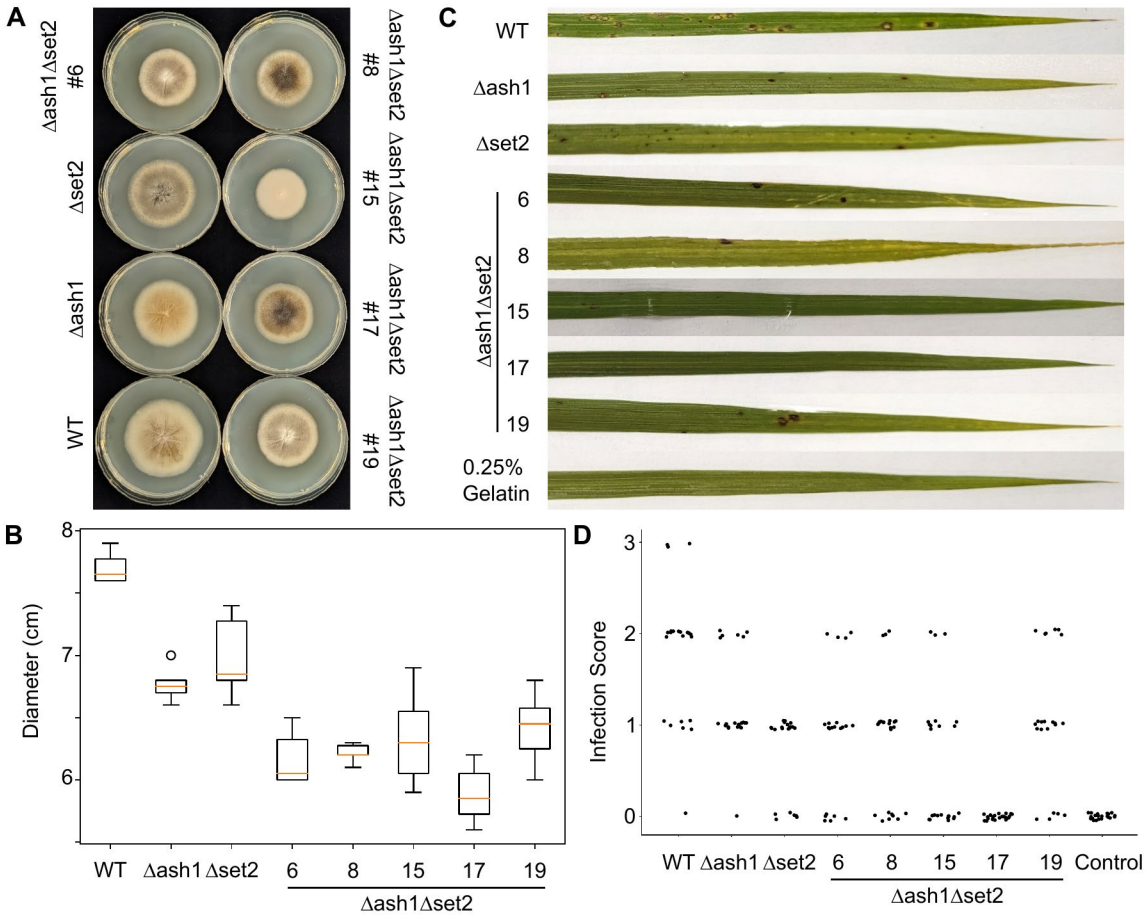


Figure 3.3. Growth and infection phenotypes change in transformed strains

(A) Image of different growth phenotypes in the WT, Δ*ash1*, Δ*set2*, Δ*ash1*Δ*set2* #6, 8, 15, 17, and 19 strains. (B) Diameter size of fungal colonies after 12 days of growth, n=6. (C) Images of leaf lesions on YT-16 rice plants. Rice was three weeks old at the time of inoculation, leaf tissue collected seven days after inoculation. (D) Qualitative infection scoring based on the size and appearance of lesions using the scale established in (Valent et al. 1991).

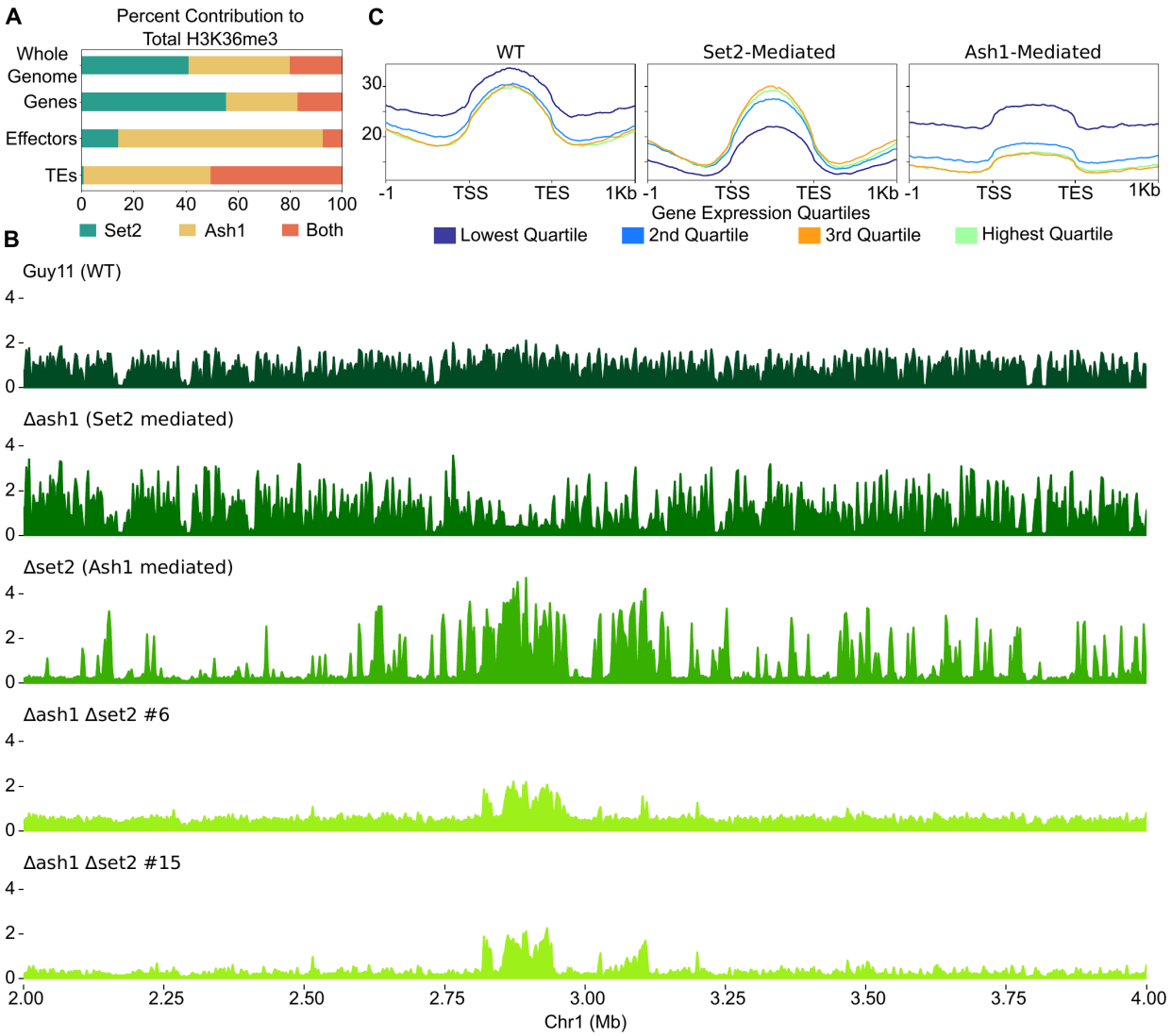


Figure 3.4. *Ash1* and *Set2* function in separate regions

(A) Bar graph showing the difference between ASH1 and SET2 function in genes, effectors, and transposable elements. (B) Normalized chromatin immunoprecipitation sequencing (ChIP-seq) data of H3K36me3 enrichment. (C) Enrichment of H3K36me3 across gene bodies in the WT, Δ*ash1*, and Δ*set2* strains. Genes were grouped into four quartiles based on expression data.

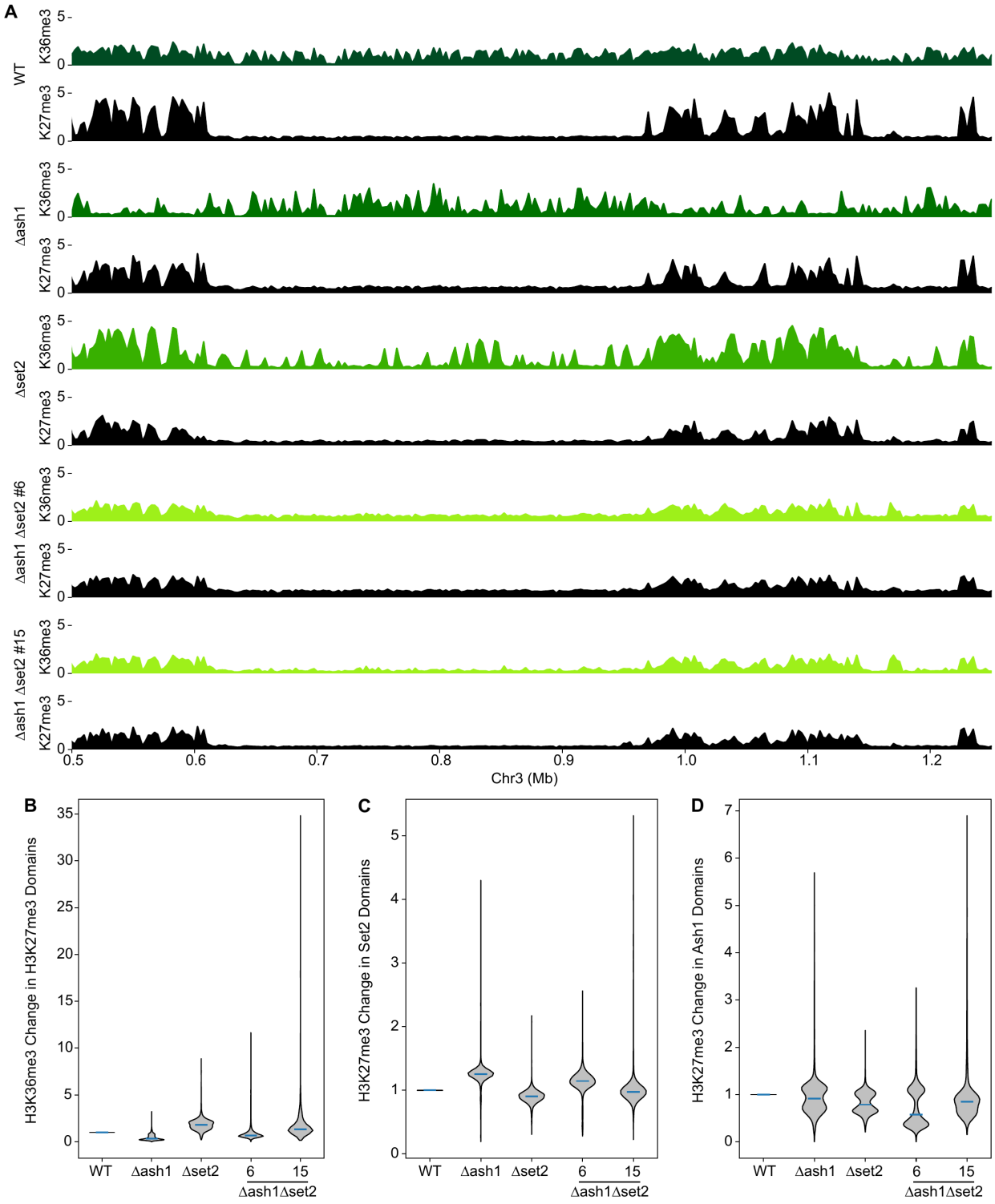


Figure 3.5. Alterations in H3K36me3 enrichment impact H3K27me3 enrichment

(A) Normalized ChIP-seq data of H3K36me3 and H3K27me3 in the WT, Δ*ash1*, Δ*set2*, Δ*ash1*Δ*set2* #6, 8, 15, 17, and 19 strains. (B) Violin plot of changes in the enrichment of

H3K36me3 in H3K37me3 domains with respect to the WT. Higher values indicate more H3K36me3 enrichment than in WT. (C) Violin plot of changes in the enrichment of H3K27me3 in Set2-mediated H3K36me3 domains with respect to the WT. (D) Violin plot of changes in the enrichment of H3K27me3 in Ash1-mediated H3K36me3 domains with respect to the WT.

Chapter 4 - General Discussion

Active and repressive histone modifications associate with different types of DNA variation

This research sought to better understand the multifaceted role of epigenetic histone modifications in genome organization in *Magnaporthe oryzae*. The genome organization of plant pathogenic fungi has been described in contrasting terms, from the two speed genome organized into gene dense and sparse regions (Raffaele et al. 2010; Rouxel and Balesdent 2017), to the one speed genome without such differences in gene density (Frantzeskakis et al. 2018). In genomes such as *Verticillium dahliae*, TEs and heterochromatin appear to be more relevant than gene density in terms of genome organization (Cook et al. 2020). In *M. oryzae*, we demonstrate that there are not gene-dense or gene-sparse regions. We show the *M. oryzae* genome has dense and sparse groups of single nucleotide polymorphisms (SNPs), and that dense SNPs are associated with the repressive histone marks H3K27me3 and H3K9me3. Under the hypothesis that the lack of selection against variation in TEs is responsible for dense SNPs occurring in heterochromatin, it would be expected to observe a majority of SNPs in those regions occur in TEs. However, dense SNPs appear in nearly equal frequencies in genes and TEs. This is important because it indicates that selection alone cannot explain the bimodal SNP distribution and association between dense SNPs and repressive histone marks. Instead, it appears that heterochromatin defines variable regions of the genome.

It is not known why the genome follows these patterns. One hypothesis that could create such a distinction between different types of DNA variation in separate regions of the genome, is that euchromatin and heterochromatin histone marks interact with DNA repair mechanisms which creates different mutation accumulation profiles. In human cell cultures, DNA damage

binding protein 2 (DDB2), a protein involved in nucleotide excision repair, was shown to facilitate the removal of UV-induced DNA dimers only in heterochromatin marked regions (Han et al. 2016). Another protein, p53 binding protein (53BP1) has been shown to interact with H4K20me3 to suppress homology directed repair pathways (Xie et al. 2007). In yeasts, the phosphorylation of C-terminal amino acids on H2A signal to efficiently recruit non-homologous end joining repair proteins at double strand breaks (Downs et al. 2000). This suggests that chromatin is important in determining how DNA damage is repaired, and more importantly that DNA repair mechanisms do not function equally in all parts of the genome. This difference in repair pathway choice may contribute to why heterochromatin and euchromatin in *M. oryzae* are associated with different types of DNA variation. Another hypothesis is that as the pathogen has evolved to infect a host, selection has benefited individuals with organized genomes that separate necessary and conserved genes that face purifying selection from genes that face diversifying selection. Effector genes, which are often under this diversifying selection (Sperschneider et al. 2014), are likely associated with H3K27me3 and TE rich sub-telomeric sequences because those regions are more likely to accumulate mutations due to their repetitive sequences which can be prone to errors in DNA replication (Barnes et al. 2023). This increase in replication errors could then contribute to variation seen in heterochromatic regions. Variation in effector genes such as *AvrL567* in *Melampsora lini*, has been demonstrated to be advantageous in overcoming plant host immunity (Dodds et al. 2006). In *M. oryzae*, variation in the form of gain and loss of effector genes has been important in its adaption to different host plants (Yoshida et al. 2009). It is still an open question as to whether, in the course of evolving to gain or lose effector genes to overcome host immunity, epigenetic marks are 1) creating variable regions where it is favorable for effectors to evolve; or 2) whether these marks are merely correlated with these regions that

are predisposed to variability due to the presence of TEs and genes that are under diversifying selection. Overall, the variation and genome organization in *M. oryzae* are shaped by chromatin states maintained by epigenetic histone marks.

ASH1 and PRC2 engage in cross-talk

In addition to the connections between H3K27me3 and variation, we also explored the relationship between H3K36me3 and H3K27me3 to understand how ASH1 and SET2-mediated H3K36me3 interact with polycomb repressive complex 2 (PRC2) mediated H3K27me3. We report that SET2 primarily functions in euchromatin regions while ASH1 functions in H3K36me3 marked regions. Furthermore, TEs and effector genes that are marked with H3K36me3 are more frequently marked by ASH1. In ASH1 marked regions of the genome, the loss of *Ash1* resulted in a reduction of H3K27me3. This indicates that ASH1 somehow promotes PRC2 to deposit H3K27me3, but it is not required for PRC2 function. These findings demonstrate a different pattern of cross-talk between ASH1 and PRC2 than reported in *Neurospora crassa*, where ASH1 was needed to direct PRC2 to certain regions (Bicocca et al. 2018). The histone mark H3K36me3 is associated with genes and frequently found in euchromatin (Freitag 2017), however, our work suggests that H3K36me3 may function as a repressive mark, given the association between ASH1 and PRC2. In yeast, SET2-mediated H3K36me3 has been demonstrated to recruit deacetylases to temporarily remove activating acetylation marks in a dynamic process after transcription (Carrozza et al. 2005). It is thought that this dynamic process of temporary repression of a gene immediately after transcription prevents transcriptional errors either from erroneous transcription or from collisions between transcriptional machinery (Carrozza et al. 2005). This may address the question as to why multiple H3K36me3-specific methyltransferases have evolved. SET2 likely functions to

transiently reduce the availability of a gene to be transcribed, whereas ASH1 likely functions with PRC2 to inactivate the transcription of genes or TEs more permanently than SET2. This idea is supported in our data where the lowest quarter of expressed genes are marked by ASH1 while other genes are marked by SET2. Future research into ASH1 and SET2 should investigate the changes in transcription of those groups of genes in response to the loss of ASH1 or SET2. It is expected that the loss of SET2 would result in the deregulation of highly expressed genes, while the loss of ASH1 would result in the upregulation of transcriptionally inactive genes. Research in this direction would further solidify our findings that ASH1 and SET2 mark genes with different transcriptional activity in discrete chromatin regions.

Implications and future directions

While filamentous fungi generally only have two H3K36me3-specific methyltransferases, other eukaryotes have many more homologous genes. For example in humans, there are three nuclear receptor-binding SET domain (NSD) proteins, a SET2 homolog (SETD2) and an ASH1 homolog (ASH1L) (Lam et al. 2022). A major difference between fungi and mammalian H3K36me3 methyltransferases is that in mammals only SETD2 can tri-methylate H3K36. ASH1L and NSD proteins can only mono- and di-methylate H3K36me3. An interesting area of research is how these proteins with similar functions evolved differently and why some organisms have more lysine methyltransferases with seemingly redundant functions. A potential explanation of the purpose of these additional methyltransferases is that the NSD family of proteins contain various chromatin methylation reader domains (PWWP) while ASH1 and its homologs contain a different chromatin methylation reader domain, the bromo domain (Marmorstein and Zhou 2014). It is possible that these domains provide an additional layer of

regulation and cross-talk between histone marks, but under that hypothesis it is unclear why filamentous fungi lack these additional epigenetic regulators.

Differences in cross-talk between histone marks may also have implications for genome organization in other eukaryotes. Human genomes, for example, do contain gene-dense and gene-sparse regions, and gene-dense regions are associated with euchromatin histone marks while gene-sparse regions associated with heterochromatin (Choi and Kim 2007). Interestingly, it was not highly expressed genes that clustered densely, but transcriptionally robust (i.e. consistently expressed across variable environments) genes were clustered (Choi and Kim 2007). It was also found that gene-sparse regions had the most variable chromatin states between related individuals. This suggests that gene-dense and sparse regions in eukaryotes may also be informed by or shape the epigenetic landscape of their chromosomes. Further research should be directed toward investigating the potential connections between epigenetic histone marks and variation in fungi that follow the two-speed genome hypothesis as well as other eukaryotic organisms to better understand conserved or ancestral patterns of genome organization. For plant pathogenic fungi specifically, identifying variable and stable regions of genomes has potential importance in focusing the breeding of resistant plant cultivars against conserved and less variable effector genes. This research has only focused on the seven core chromosomes of *M. oryzae*, however numerous strains have mini-chromosomes which may also contribute to genome organization and evolution (Orbach et al. 1996). There is strong evidence that mini-chromosomes exchange DNA with the core chromosomes and that these rearrangements can move effector genes in the genome (Peng et al. 2019; Langner et al. 2021). Some effector genes, such as *AvrPik* and *AvrK60*, are found on mini-chromosomes (Luo et al. 2007; Kusaba et al. 2008) and other highly mobile effector genes, such as *AvrPita*, are thought to move between

chromosomes using mini-chromosomes (Chuma et al. 2011). Future investigations should focus on understanding associations between larger scale structural variations, mini-chromosomes, and potential similarities in epigenetic histone mark profiles between DNA moved from a core chromosome to a mini-chromosome. This knowledge is likely an important piece in the evolution of *M. oryzae* as a plant pathogen.

References

Adhvaryu K, Morris S, Strahl B, Selker E. 2005. Methylation of Histone H3 Lysine 36 Is Required for Normal Development in Neurospora crassa. *Eukaryotic cell* **4**: 1455–64.

Ajawatanawong P, Baldauf SL. 2013. Evolution of protein indels in plants, animals and fungi. *BMC Evolutionary Biology* **13**: 140.

Ali S, Gladieux P, Leconte M, Gautier A, Justesen AF, Hovmøller MS, Enjalbert J, Vallavieille-Pope C de. 2014. Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. tritici. *PLOS Pathogens* **10**: e1003903.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**: 420–423.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Asibi AE, Chai Q, Coulter JA. 2019. Rice Blast: A Disease with Implications for Global Food Security. *Agronomy* **9**: 451.

Atanasoff-Kardjalieff AK, Lünne F, Kalinina S, Strauss J, Humpf H-U, Studt L. 2021. Biosynthesis of Fusapyrone Depends on the H3K9 Methyltransferase, FmKmt1, in Fusarium mangiferae. *Frontiers in Fungal Biology* **2**. https://www.frontiersin.org/articles/10.3389/ffunb.2021.671796 (Accessed February 23, 2023).

Auwera GAV der, O'Connor BD. 2020. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Inc.

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* **21**: 381–395.

Barnes RP, Thosar SA, Opresko PL. 2023. Telomere Fragility and MiDAS: Managing the Gaps at the End of the Road. *Genes* **14**: 348.

Basenko EY, Sasaki T, Ji L, Prybol CJ, Burckhardt RM, Schmitz RJ, Lewis ZA. 2015. Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. *Proceedings of the National Academy of Sciences* **112**: E6339–E6348.

Beacon TH, Delcuve GP, López C, Nardocci G, Kovalchuk I, van Wijnen AJ, Davie JR. 2021. The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. *Clinical Epigenetics* **13**: 138.

Bender J. 2004. Dna Methylation and Epigenetics. *Annual Review of Plant Biology* **55**: 41–68.

Bicocca VT, Ormsby T, Adhvaryu KK, Honda S, Selker EU. 2018. ASH1-catalyzed H3K36 methylation drives gene repression and marks H3K27me2/3-competent chromatin eds. J.L. Workman and K. Struhl. *eLife* **7**: e41497.

Brosch G, Loidl P, Graessle S. 2008. Histone modifications and chromatin dynamics: a focus on filamentous fungi. *FEMS Microbiology Reviews* **32**: 409–439.

Cao Z, Yin Y, Sun X, Han J, Sun Q peng, Lu M, Pan J, Wang W. 2016. An Ash1-Like Protein MoKMT2H Null Mutant Is Delayed for Conidium Germination and Pathogenesis in *Magnaporthe oryzae*. *BioMed Research International* **2016**: e1575430.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia W-J, Anderson S, Yates J, Washburn MP, et al. 2005. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell* **123**: 581–592.

Choi JK, Kim SC. 2007. Environmental Effects on Gene Expression Phenotype Have Regional Biases in the Human Genome. *Genetics* **175**: 1607–1613.

Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, Yoshida K, Terauchi R, Fujita Y, Nakayashiki H, et al. 2011. Multiple Translocation of the AVR-Pita Effector Gene among Chromosomes of the Rice Blast Fungus Magnaporthe oryzae and Related Species. *PLOS Pathogens* **7**: e1002147.

Chuma I, Tosa Y, Taga M, Nakayashiki H, Mayama S. 2003. Meiotic behavior of a supernumerary chromosome in Magnaporthe oryzae. *Curr Genet* **43**: 191–198.

Chung H, Goh J, Han S-S, Roh J-H, Kim Y, Heu S, Shim H-K, Jeong DG, Kang IJ, Yang J-W. 2020. Comparative Pathogenicity and Host Ranges of Magnaporthe oryzae and Related Species. *Plant Pathol J* **36**: 305–313.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**: 80–92.

Connolly LR, Smith KM, Freitag M. 2013. The Fusarium graminearum Histone H3 K27 Methyltransferase KMT6 Regulates Development and Expression of Secondary Metabolite Gene Clusters. *PLOS Genetics* **9**: e1003916.

Cook DE, Kramer HM, Torres DE, Seidl MF, Thomma BPHJ. 2020. A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen. *eLife* **9**: e62208.

Costa FF. 2008. Non-coding RNAs, epigenetics and complexity. *Gene* **410**: 9–17.

Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics* **19**: 348.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**: 21931–21936.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu J-R, Pan H, et al. 2005. The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature* **434**: 980–986.

Dodds PN, Lawrence GJ, Catanzariti A-M, Teh T, Wang C-IA, Ayliffe MA, Kobe B, Ellis JG. 2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences* **103**: 8888–8893.

Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development* **35**: 57–65.

Downs JA, Lowndes NF, Jackson SP. 2000. A role for Saccharomyces cerevisiae histone H2A in DNA repair. *Nature* **408**: 1001–1004.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**: 18.

Eram MS, Kuznetsova E, Li F, Lima-Fernandes E, Kennedy S, Chau I, Arrowsmith CH, Schapira M, Vedadi M. 2015. Kinetic characterization of human histone H3 lysine 36 methyltransferases, ASH1L and SETD2. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**: 1842–1848.

Faino L, Seidl MF, Shi-Kunne X, Pauper M, Berg GCM van den, Wittenberg AHJ, Thomma BPHJ. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res* **26**: 1091–1100.

Foster AJ, Martin-Urdiroz M, Yan X, Wright HS, Soanes DM, Talbot NJ. 2018. CRISPR-Cas9 ribonucleoprotein-mediated co-editing and counterselection in the rice blast fungus. *Scientific Reports* **8**. https://ueaeprints.uea.ac.uk/id/eprint/69934/ (Accessed March 8, 2023).

Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics* **19**: 381.

Freitag M. 2017. Histone Methylation by SET Domain Proteins in Fungi. *Annual Review of Microbiology* **71**: 413–439.

Fukagawa T, Earnshaw WC. 2014. The Centromere: Chromatin Foundation for the Kinetochore Machinery. *Developmental Cell* **30**: 496–508.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, et al. 2003. The genome sequence of the filamentous fungus Neurospora crassa. *Nature* **422**: 859–868.

Gao C, Xiao M, Ren X, Hayward A, Yin J, Wu L, Fu D, Li J. 2012. Characterization and functional annotation of nested transposable elements in eukaryotic genomes. *Genomics* **100**: 222–230.

Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, Chen L, Terauchi R, Lebrun M-H, Tharreau D, et al. 2018. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae. *mBio* **9**: e01219-17.

Goel M, Sun H, Jiao W-B, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology* **20**: 277.

Gowda M, Shirke MD, Mahesh HB, Chandarana P, Rajamani A, Chattoo BB. 2015. Genome analysis of rice-blast fungus Magnaporthe oryzae field isolates from southern India. *Genomics Data* **5**: 284–291.

Habig M, Lorrain C, Feurtey A, Komluski J, Stukenbrock EH. 2021. Epigenetic modifications affect the rate of spontaneous mutations in a pathogenic fungus. *Nat Commun* **12**: 5869.

Han C, Srivastava AK, Cui T, Wang Q-E, Wani AA. 2016. Differential DNA lesion formation and repair in heterochromatin and euchromatin. *Carcinogenesis* **37**: 129–138.

Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res* **38**: e199.

Harr JC, Luperchio TR, Wong X, Cohen E, Wheelan SJ, Reddy KL. 2015. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *Journal of Cell Biology* **208**: 33–52.

Huang J, Cook DE. 2021. CRISPR-Cas12a ribonucleoprotein-mediated gene editing in the plant pathogenic fungus *Magnaporthe oryzae*. https://star-protocols.cell.com/protocols/1324 (Accessed March 6, 2023).

Huang J, Cook DE. 2022. The contribution of DNA repair pathways to genome editing and evolution in filamentous pathogens. *FEMS Microbiology Reviews* **46**: fuac035.

Huisinga KL, Brower-Toland B, Elgin SCR. 2006. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**: 110–122.

Jamieson K, Rountree MR, Lewis ZA, Stajich JE, Selker EU. 2013. Regional control of histone H3 lysine 27 methylation in Neurospora. *Proceedings of the National Academy of Sciences* **110**: 6027–6032.

Janevska S, Baumann L, Sieber CMK, Münsterkötter M, Ulrich J, Kämper J, Güldener U, Tudzynski B. 2018. Elucidation of the Two H3K36me3 Histone Methyltransferases Set2 and Ash1 in Fusarium fujikuroi Unravels Their Different Chromosomal Targets and a Major Impact of Ash1 on Genome Stability. *Genetics* **208**: 153–171.

Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bähler J, Sedlazeck FJ. 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* **8**: 14061.

Jenuwein T, Allis CD. 2001. Translating the Histone Code. *Science* **293**: 1074–1080.

Jeong JS, Mitchell TK, Dean RA. 2007. The Magnaporthe grisea snodprot1 homolog, MSP1, is required for virulence. *FEMS Microbiology Letters* **273**: 157–165.

Johnson DP, Spitz-Becker GS, Chakraborti K, Bhaskara S. 2019. Assessment of epigenetic mechanisms and DNA double-strand break repair using laser micro-irradiation technique developed for hematological cells. *EBioMedicine* **43**: 138–149.

Jonge R de, Bolton MD, Kombrink A, Berg GCM van den, Yadeta KA, Thomma BPHJ. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res* **23**: 1271–1282.

Kang S, Lebrun MH, Farrall L, Valent B. 2001. Gain of Virulence Caused by Insertion of a Pot3 Transposon in a Magnaporthe grisea Avirulence Gene. *MPMI* **14**: 671–674.

Kang S, Sweigard JA, Valent B. 1995. The PWL host specificty gene family in the blast fungus Magnaporthe grisea. *Molecular plant-microbe interactions : MPMI (USA)*. https://scholar.google.com/scholar_lookup?title=The+PWL+host+specificty+gene+family+in+the+blast+fungus+Magnaporthe+grisea&author=Kang%2C+S.+%28University+of+New+Mexico%2C+Albuquerque%2C+NM.%29&publication_year=1995 (Accessed February 16, 2023).

Kasetsomboon T, Kate-Ngam S, Sriwongchai T, Zhou B, Jantasuriyarat C. 2013. Sequence variation of avirulence gene AVR-Pita1 in rice blast fungus, Magnaporthe oryzae. *Mycol Progress* **12**: 617–628.

Katsonis P, Koire A, Wilson SJ, Hsu T-K, Lua RC, Wilkins AD, Lichtarge O. 2014. Single nucleotide variations: Biological impact and theoretical interpretation. *Protein Science* **23**: 1650–1666.

Khang CH, Berruyer R, Giraldo MC, Kankanala P, Park S-Y, Czymmek K, Kang S, Valent B. 2010. Translocation of Magnaporthe oryzae Effectors into Rice Cells and Their Subsequent Cell-to-Cell Movement. *The Plant Cell* **22**: 1388–1403.

Kim J, Park S, Kwon S, Lee E-J, Lee J-S. 2021. Set1-mediated H3K4 methylation is required for Candida albicans virulence by regulating intracellular level of reactive oxygen species. *Virulence* **12**: 2648–2658.

Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. 2005. A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. *Molecular and Cellular Biology* **25**: 3305–3316.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41**: 376–381.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**: 722–736.

Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C, et al. 2003. Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. *Molecular and Cellular Biology* **23**: 4207–4218.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12.

Kusaba M, Luo C-X, Hanamura H, Misaka M, Mochida T, Fujita Y, Tosa Y. 2008. An avirulence gene to rice cultivar K60 is located on the 1.6-Mb chromosome in Magnaporthe oryzae isolate 84R-62B. *J Gen Plant Pathol* **74**: 250–253.

Lai Y, Wang L, Zheng W, Wang S. 2022. Regulatory Roles of Histone Modifications in Filamentous Fungal Pathogens. *Journal of Fungi* **8**: 565.

Lam UTF, Tan BKY, Poh JJX, Chen ES. 2022. Structural and functional specificity of H3K36 methylation. *Epigenetics & Chromatin* **15**: 17.

Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Malmgren A, Latorre SM, Burbano HA, Win J, Kamoun S. 2021. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLOS Genetics* **17**: e1009386.

Latorre SM, Reyes-Avila CS, Malmgren A, Win J, Kamoun S, Burbano HA. 2020. *Recently expanded clonal lineages of the rice blast fungus display distinct patterns of presence/absence of effector genes*. Evolutionary Biology http://biorxiv.org/lookup/doi/10.1101/2020.01.09.900308 (Accessed September 22, 2020).

Lee H, Lee M, Mohammed Ismail W, Rho M, Fox GC, Oh S, Tang H. 2016. MGEScan: a Galaxy-based system for identifying retrotransposons in genomes. *Bioinformatics* **32**: 2502–2504.

Lee Y, Park D, Iyer VR. 2017. The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes. *Nucleic Acids Research* **45**: 7180–7190.

Leung H, Borromeo E, Bernardo M, Notteghem J. 1988. Genetic analysis of virulence in the rice blast fungus Magnaporthe grisea. *Phytopathology* **78**: 1227–1233.

Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schübeler D, Selker EU. 2009. Relics of repeat-induced point mutation direct heterochromatin formation in Neurospora crassa. *Genome Res* **19**: 427–437.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Liu J, Ali M, Zhou Q. 2020. Establishment and evolution of heterochromatin. *Annals of the New York Academy of Sciences* **1476**: 59–77.

Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* **12**: 733–735.

Luger K, Dechassa ML, Tremethick DJ. 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* **13**: 436–447.

Luo C-X, Yin L-F, Ohtaka K, Kusaba M. 2007. The 1.6Mb chromosome carrying the avirulence gene AvrPik in Magnaporthe oryzae isolate 84R-62B is a chimera containing chromosome 1 sequences. *Mycological Research* **111**: 232–239.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**: 4647–4654.

Marmorstein R, Zhou M-M. 2014. Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition. *Cold Spring Harb Perspect Biol* **6**: a018762.

McVean G, Auton A. 2007. LDhat 2.2: A package for the population genetic analysis of recombination. https://www.semanticscholar.org/paper/LDhat-2.2%3A-A-package-for-the-population-genetic-of-McVean-Auton/37c071c2bcd115bcce4bb2e4eb89a7dbb861c136 (Accessed January 5, 2023).

Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BPHJ, et al. 2012. Effector-Mediated Suppression of Chitin-Triggered Immunity by Magnaporthe oryzae Is Necessary for Rice Blast Disease. *The Plant Cell* **24**: 322–335.

Miao F, Natarajan R. 2005. Mapping Global Histone Methylation Patterns in the Coding Regions of Human Genes. *Molecular and Cellular Biology* **25**: 4650–4661.

Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* **16**: 1182–1190.

Möller M, Ridenour JB, Wright DF, Freitag M. 2023. H4K20me3 controls Ash1-mediated H3K36me3 and transcriptional silencing in facultative heterochromatin. 2022.11.25.517763. https://www.biorxiv.org/content/10.1101/2022.11.25.517763v2 (Accessed March 3, 2023).

Möller M, Schotanus K, Soyer JL, Haueisen J, Happ K, Stralucke M, Happel P, Smith KM, Connolly LR, Freitag M, et al. 2019. Destabilization of chromosome structure by histone H3 lysine 27 methylation. *PLOS Genetics* **15**: e1008093.

Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol* **15**: 756–771.

Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B. 2009. Interaction Transcriptome Analysis Identifies Magnaporthe oryzae BAS1-4 as Biotrophy-Associated Secreted Proteins in Rice Blast Disease. *The Plant Cell* **21**: 1273–1290.

Nalley L, Tsiboe F, Durand-Morat A, Shew A, Thoma G. 2016. Economic and Environmental Impact of Rice Blast Pathogen (Magnaporthe oryzae) Alleviation in the United States. *PLOS ONE* **11**: e0167295.

Nattestad M, Schatz MC. 2016. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32**: 3021–3023.

Okagaki LH, Sailsbery JK, Eyre AW, Dean RA. 2016. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. *BMC Genomics* **17**: 135.

Orbach M, Chumley F, Valent B. 1996. Electrophoretic karyotypes of Magnaporthe grisea pathogens of diverse grasses. *MPMI* **9**: 261–271.

Osés-Ruiz M, Cruz-Mireles N, Martin-Urdiroz M, Soanes DM, Eseola AB, Tang B, Derbyshire P, Nielsen M, Cheema J, Were V, et al. 2021. Appressorium-mediated plant infection by Magnaporthe oryzae is regulated by a Pmk1-dependent hierarchical transcriptional network. *Nat Microbiol* **6**: 1383–1397.

Ou S, Jiang N. 2018. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiol* **176**: 1410–1422.

Pajoro A, Severing E, Angenent GC, Immink RGH. 2017. Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. *Genome Biol* **18**: 102.

Palmer J, Stajich J. nextgenusfs/funannotate: funannotate. https://doi.org/10.5281/zenodo.2604804.

Papamichos-Chronakis M, Peterson CL. 2013. Chromatin and the genome integrity network. *Nat Rev Genet* **14**: 62–75.

Pasini D, Malatesta M, Jung HR, Walfridsson J, Willer A, Olsson L, Skotte J, Wutz A, Porse B, Jensen ON, et al. 2010. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* **38**: 4958–4969.

Peng Z, Oliveira-Garcia E, Lin G, Hu Y, Dalby M, Migeon P, Tang H, Farman M, Cook D, White FF, et al. 2019. Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. *PLOS Genetics* **15**: e1008272.

Pham KTM, Inoue Y, Vu BV, Nguyen HH, Nakayashiki T, Ikeda K, Nakayashiki H. 2015. MoSET1 (Histone H3K4 Methyltransferase in Magnaporthe oryzae) Regulates Global Gene Expression during Infection-Related Morphogenesis. *PLOS Genetics* **11**: e1005385.

Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, Grant PA. 2005. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**: 434–438.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, Donofrio NM, et al. 2010. Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage. *Science* **330**: 1540–1543.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**: W160–W165.

Rouxel T, Balesdent M-H. 2017. Life, death and rebirth of avirulence effectors in a fungal pathogen of Brassica crops, Leptosphaeria maculans. *New Phytologist* **214**: 526–532.

Roy B, Sanyal K. 2011. Diversity in Requirement of Genetic and Epigenetic Factors for Centromere Function in Fungi. *Eukaryotic Cell* **10**: 1384–1395.

Schep R, Brinkman EK, Leemans C, Vergara X, van der Weide RH, Morris B, van Schaik T, Manzo SG, Peric-Hupkes D, van den Berg J, et al. 2021. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Molecular Cell* **81**: 2216-2230.e10.

Schmitges FW, Prusty AB, Faty M, Stützer A, Lingaraju GM, Aiwazian J, Sack R, Hess D, Li L, Zhou S, et al. 2011. Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell* **42**: 330–341.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15**: 461–468.

Seidl MF, Cook DE, Thomma BPHJ. 2016. Chromatin Biology Impacts Adaptive Evolution of Filamentous Plant Pathogens. *PLOS Pathogens* **12**: e1005920.

Seidl MF, Thomma BPHJ. 2014. Sex or no sex: Evolutionary adaptation occurs regardless. *BioEssays* **36**: 335–345.

Sexton T, Cavalli G. 2015. The Role of Chromosome Domains in Shaping the Functional Genome. *Cell* **160**: 1049–1059.

Sista Kameshwar AK, Qin W. 2019. Systematic review of publicly available non-Dikarya fungal proteomes for understanding their plant biomass-degrading and bioremediation potentials. *Bioresour Bioprocess* **6**: 30.

Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Smolle M, Workman JL. 2013. Transcription-associated histone modifications and cryptic transcription. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**: 84–97.

Sperschneider J, Dodds PN. 2022. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. *MPMI* **35**: 146–156.

Sperschneider J, Ying H, Dodds PN, Gardiner DM, Upadhyaya NM, Singh KB, Manners JM, Taylor JM. 2014. Diversifying selection in the wheat stem rust fungus acts predominantly on pathogen-associated gene families and reveals candidate effectors. *Frontiers in Plant Science* **5**. https://www.frontiersin.org/articles/10.3389/fpls.2014.00372 (Accessed February 20, 2023).

Strahl BD, Grant PA, Briggs SD, Sun Z-W, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF, et al. 2002. Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Molecular and Cellular Biology* **22**: 1298–1306.

Strauss J, Reyes-Dominguez Y. 2011. Regulation of secondary metabolism by chromatin structure and epigenetic codes. *Fungal Genetics and Biology* **48**: 62–69.

Takan JP, Chipili J, Muthumeenakshi S, Talbot NJ, Manyasa EO, Bandyopadhyay R, Sere Y, Nutsugah SK, Talhinhas P, Hossain M, et al. 2012. Magnaporthe oryzae Populations Adapted to Finger Millet and Rice Exhibit Distinctive Patterns of Genetic Diversity, Sexuality and Host Interaction. *Mol Biotechnol* **50**: 145–158.

Talbot NJ. 2003. On the Trail of a Cereal Killer: Exploring the Biology of Magnaporthe grisea. *Annual Review of Microbiology* **57**: 177–202.

Tanaka Y, Katagiri Z, Kawahashi K, Kioussis D, Kitajima S. 2007. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* **397**: 161–168.

Thierry M, Charriat F, Milazzo J, Adreit H, Ravel S, Cros-Arteil S, Borron S, Sella V, Kroj T, Ioos R, et al. 2022. Maintenance of divergent lineages of the Rice Blast Fungus Pyricularia oryzae through niche separation, loss of sex and post-mating genetic incompatibilities. *PLOS Pathogens* **18**: e1010687.

Torres DE, Oggenfuss U, Croll D, Seidl MF. 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews* **34**: 136–143.

Upadhyaya NM, Garnica DP, Karaoglu H, Sperschneider J, Nemri A, Xu B, Mago R, Cuomo CA, Rathjen JP, Park RF, et al. 2015. Comparative genomics of Australian isolates of the wheat stem rust pathogen Puccinia graminis f. sp. tritici reveals extensive polymorphism in candidate effector genes. *Frontiers in Plant Science* **5**. https://www.frontiersin.org/articles/10.3389/fpls.2014.00759 (Accessed February 20, 2023).

Valent B, Farrall L, Chumley F. 1991. Magnaporthe grisea genes for pathogenicity and virulence identified through a series of backgrosses. *Genetics* **127**: 87–101.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**: 261–272.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**: e112963.

Wang Q, Li J, Lu L, He C, Li C. 2020. Novel Variation and Evolution of AvrPiz-t of Magnaporthe oryzae in Field Isolates. *Frontiers in Genetics* **11**. https://www.frontiersin.org/articles/10.3389/fgene.2020.00746 (Accessed February 16, 2023).

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics* **3**: e000132.

Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology* **20**: 129.

Wilson RA, Talbot NJ. 2009. Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. *Nat Rev Microbiol* **7**: 185–195.

Wolffe AP, Hayes JJ. 1999. Chromatin disruption and modification. *Nucleic Acids Research* **27**: 711–720.

Wu Z, Qiu J, Shi H, Lin C, Yue J, Liu Z, Xie W, Naqvi NI, Kou Y, Tao Z. 2022. Polycomb Repressive Complex 2-Mediated H3K27 Trimethylation Is Required for Pathogenicity in Magnaporthe oryzae. *Rice Science* **29**: 363–374.

Xie A, Hartlerode A, Stucki M, Odate S, Puget N, Kwok A, Nagaraju G, Yan C, Alt FW, Chen J, et al. 2007. Distinct Roles of Chromatin-Associated Proteins MDC1 and 53BP1 in Mammalian Double-Strand Break Repair. *Molecular Cell* **28**: 1045–1057.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**: W265–W268.

Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano Y, Win J, et al. 2009. Association Genetics Reveals Three Novel Avirulence Genes from the Rice Blast Fungal Pathogen Magnaporthe oryzae. *The Plant Cell* **21**: 1573–1591.

Zhang S, Xu J-R. 2014. Effectors and Effector Delivery in Magnaporthe oryzae. *PLOS Pathogens* **10**: e1003826.

Zhang T, Cooper S, Brockdorff N. 2015. The interplay of histone modifications – writers that read. *EMBO reports* **16**: 1467–1481.

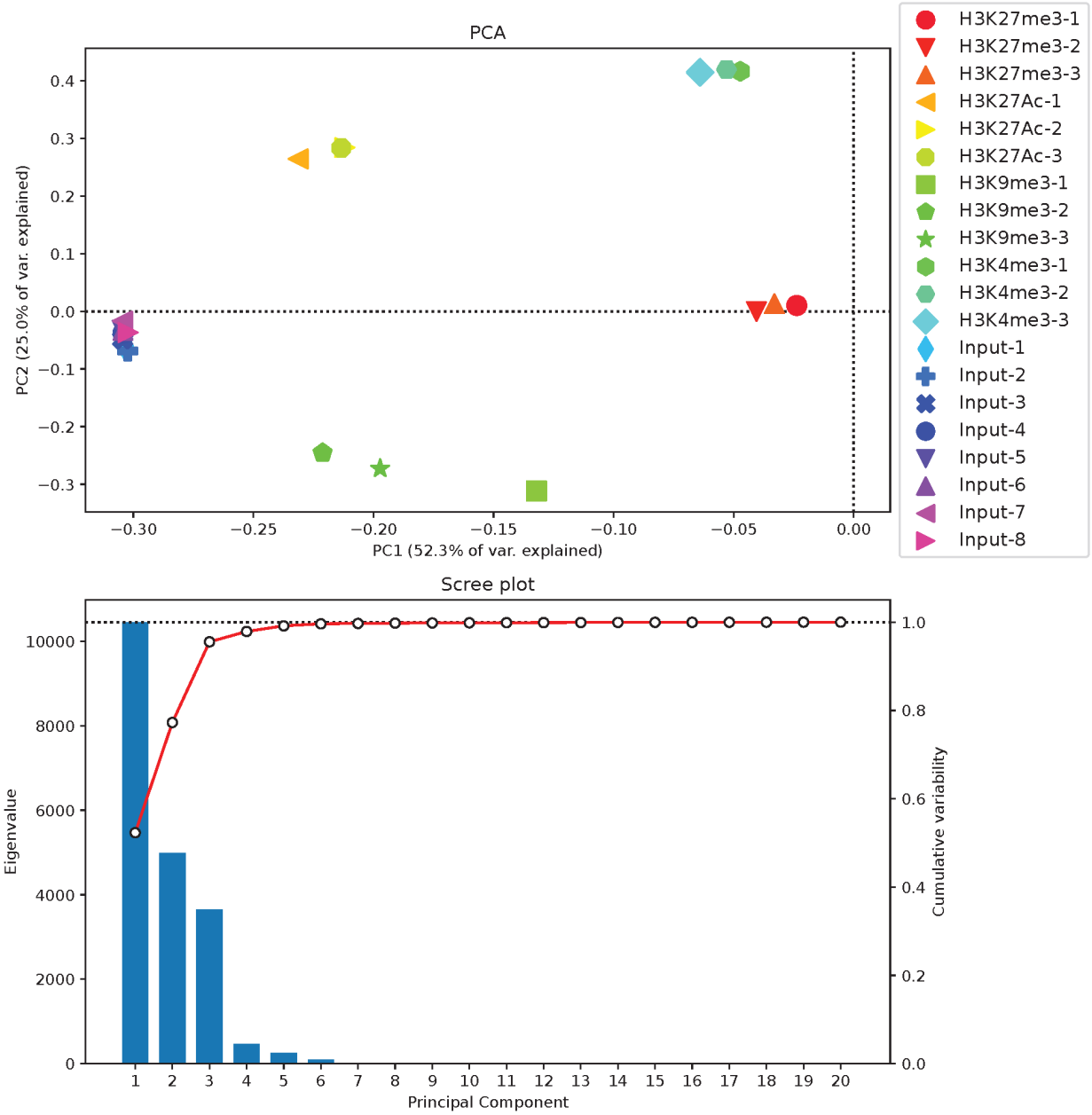
Zhang W, Huang J, Cook DE. 2021. Histone modification dynamics at H3K27 are associated with altered transcription of in planta induced genes in Magnaporthe oryzae. *PLOS Genetics* **17**: e1009376.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**: R137.

Zhong Z, Chen M, Lin L, Chen R, Liu D, Norvienyeku J, Zheng H, Wang Z. 2020. Genetic Variation Bias toward Noncoding Regions and Secreted Proteins in the Rice Blast Fungus Magnaporthe oryzae. *mSystems* **5**: e00346-20.

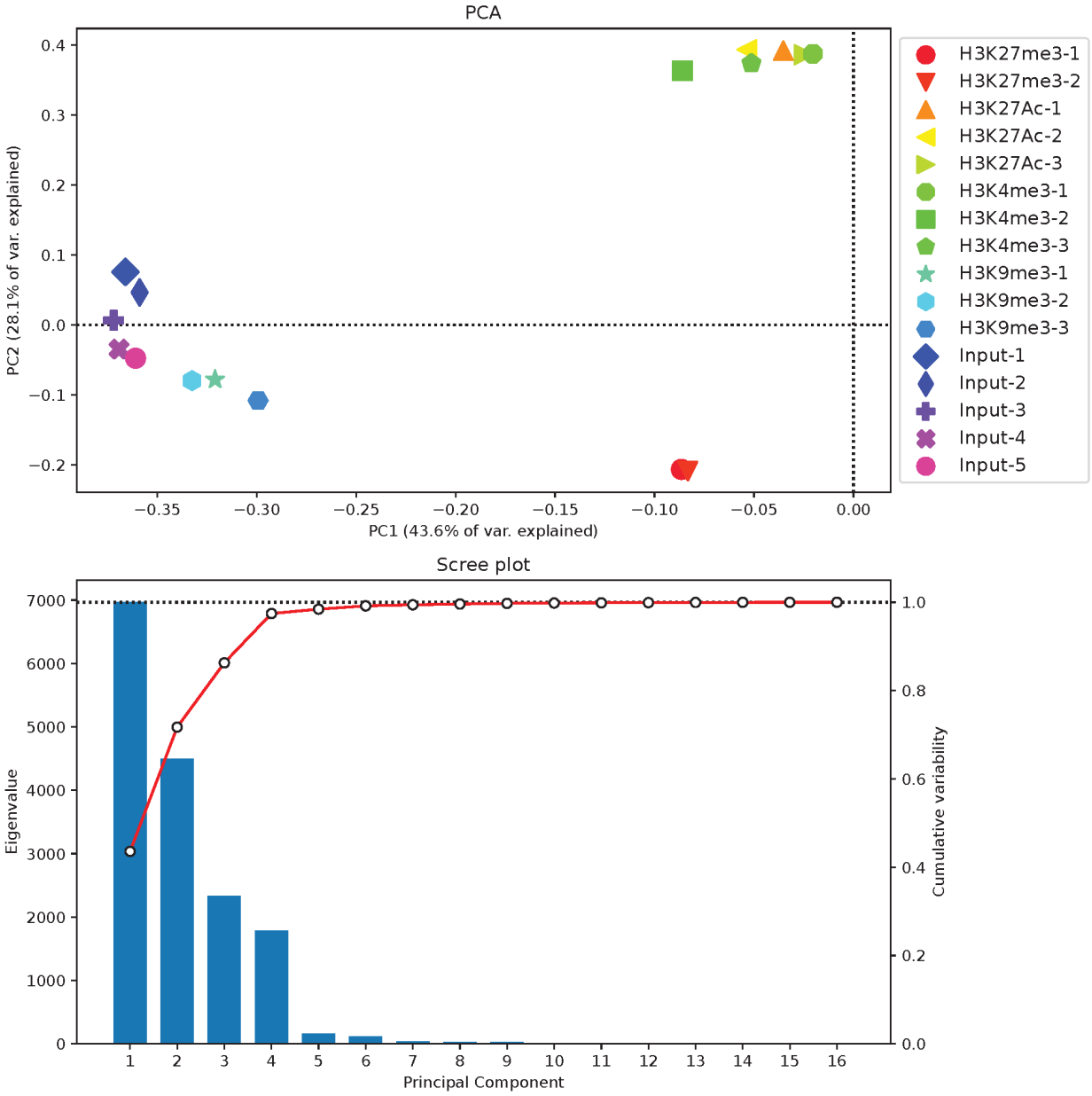
Zhong Z, Chen M, Lin L, Han Y, Bao J, Tang W, Lin L, Lin Y, Somai R, Lu L, et al. 2018. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *The ISME Journal* **12**: 1867–1878.

Appendix A - Supplemental Figures and Tables of Chapter 2



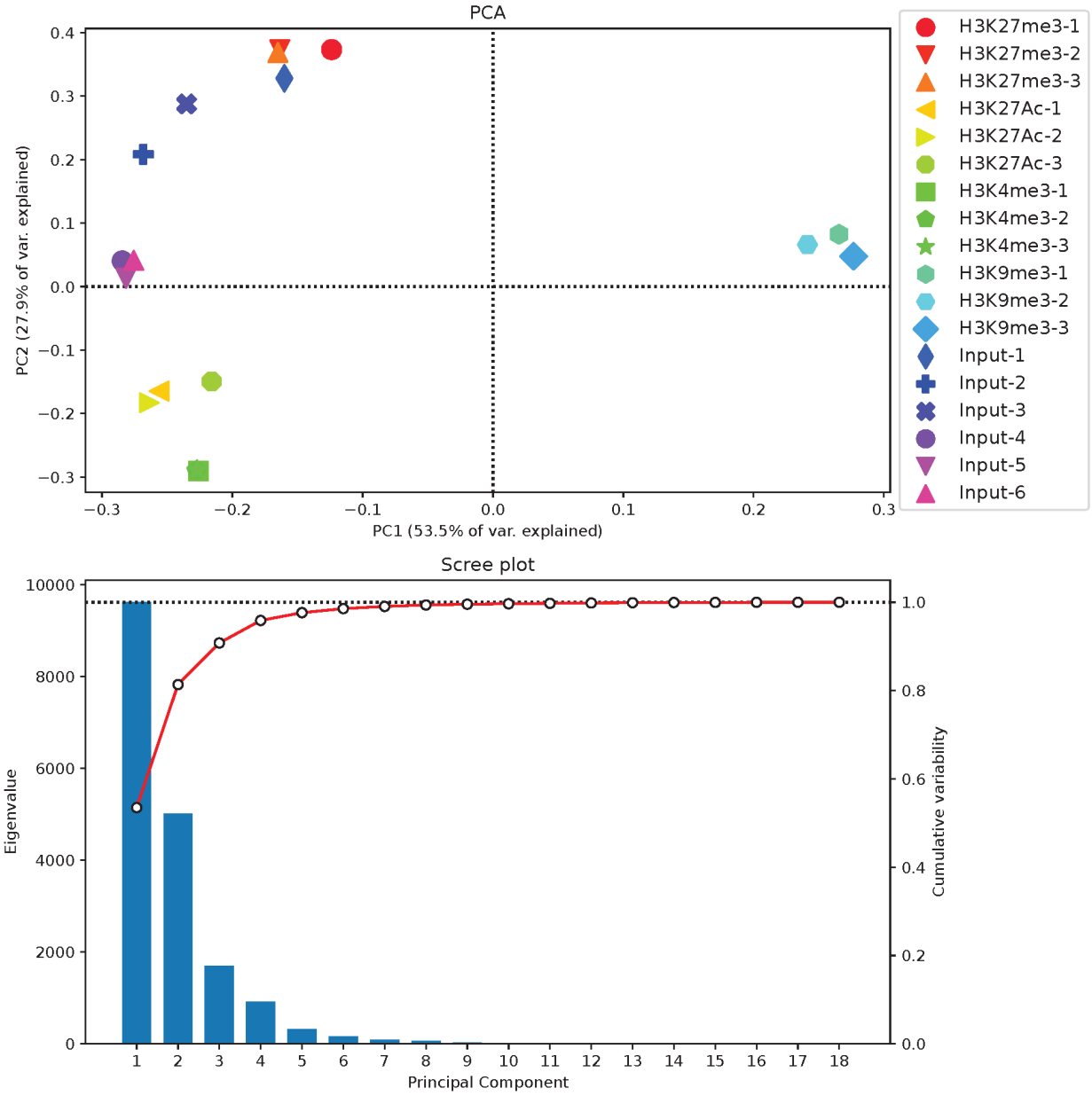
Appendix A Figure A.1. Guy11 ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in Guy11 for the histone marks H3K27me3, H3K27ac, H3K9me3, and H3K4me3. The highest contributing components explain 52.3% and 25.0% of variance.



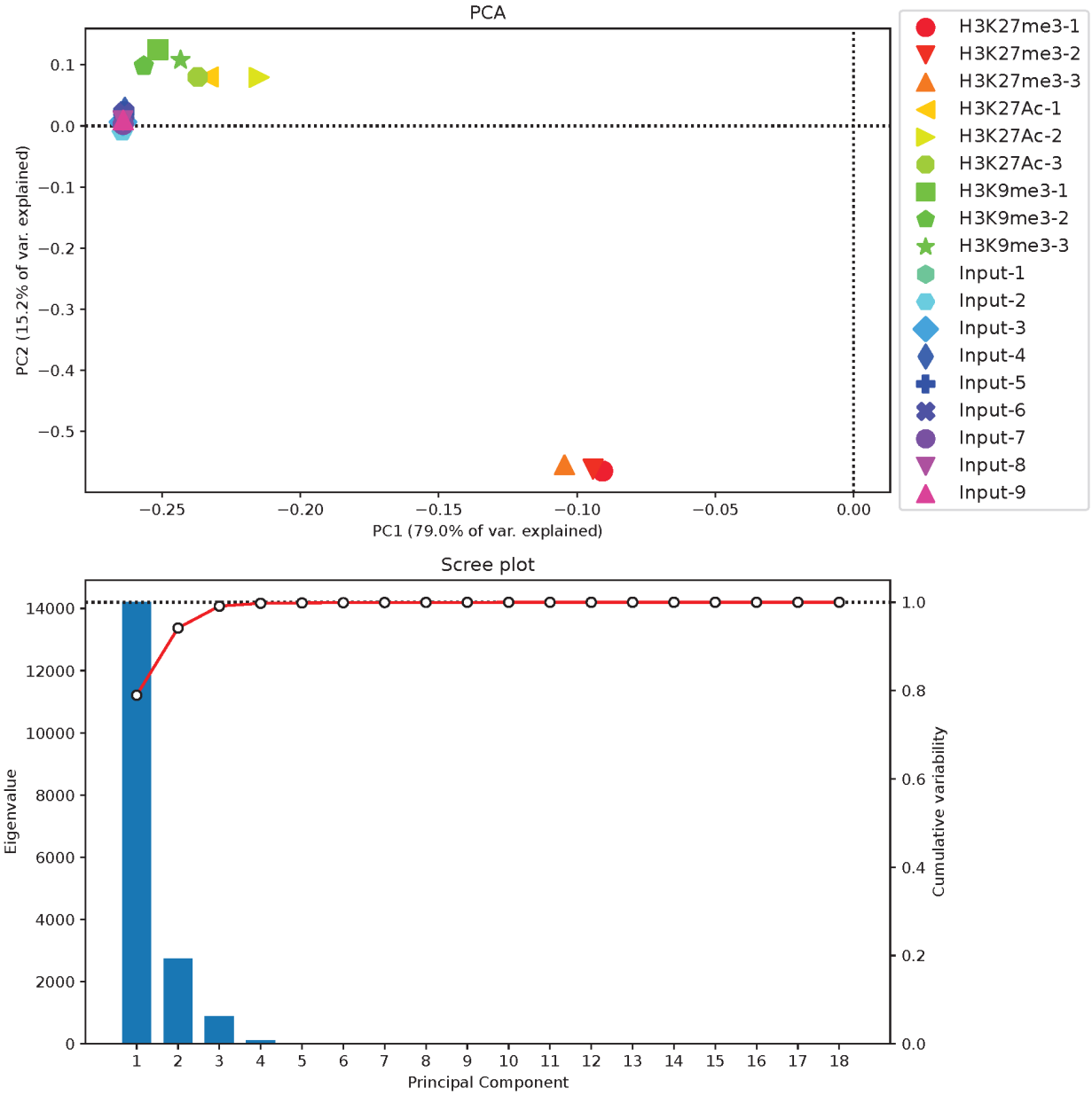
Appendix A Figure A.2. O-135 ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in O-135 for the histone marks H3K27me3, H3K27ac, H3K9me3, and H3K4me3. The highest contributing components explain 43.6% and 28.1% of variance.



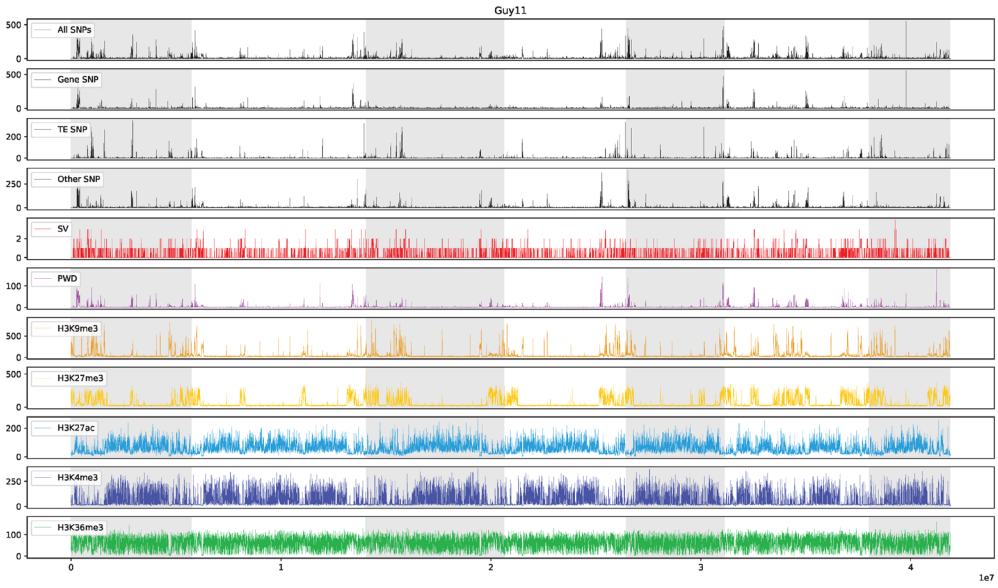
Appendix A Figure A.3. O-137 ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in O-137 for the histone marks H3K27me3, H3K27ac, H3K9me3, and H3K4me3. The highest contributing components explain 53.5% and 27.9% of variance.



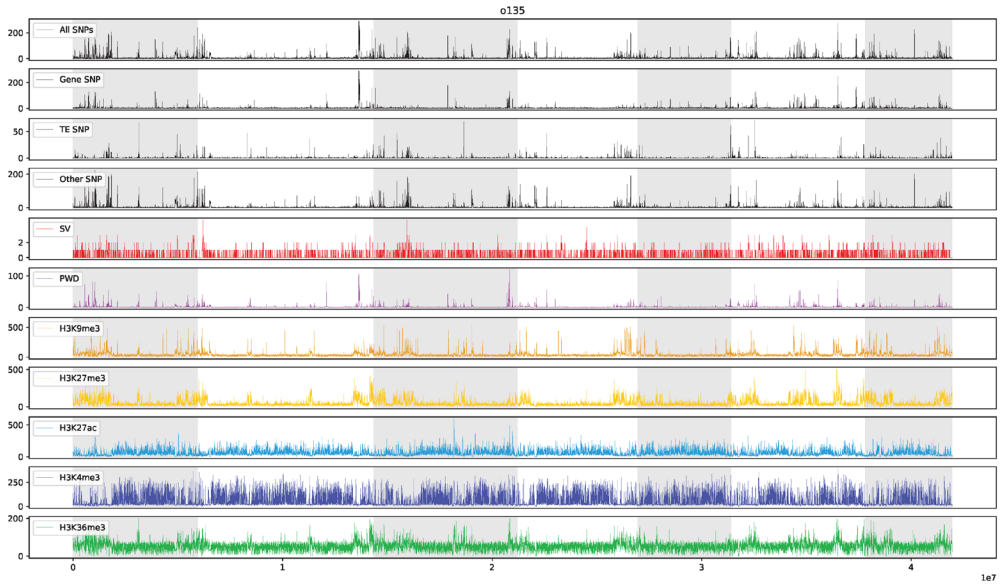
Appendix A Figure A.4. O-219 ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in O-219 for the histone marks H3K27me3, H3K27ac, and H3K9me3. The highest contributing components explain 79.0% and 15.2% of variance.



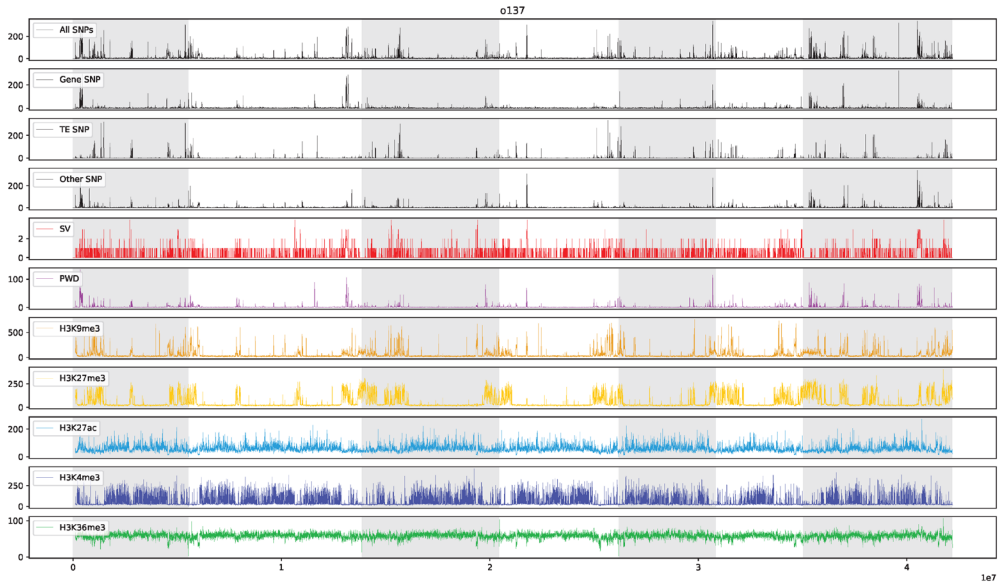
Appendix A Figure A.5. Guy11 whole genome association between variation and histone marks

Graph of SNPs, INDELs, average pair-wise distance (PWD) and ChIP enrichment of H3K9me3, H3K27me3, H3K27ac, H3K4me3, and H3K36me3. Alternating shaded regions indicate core chromosomes 1 through 7.



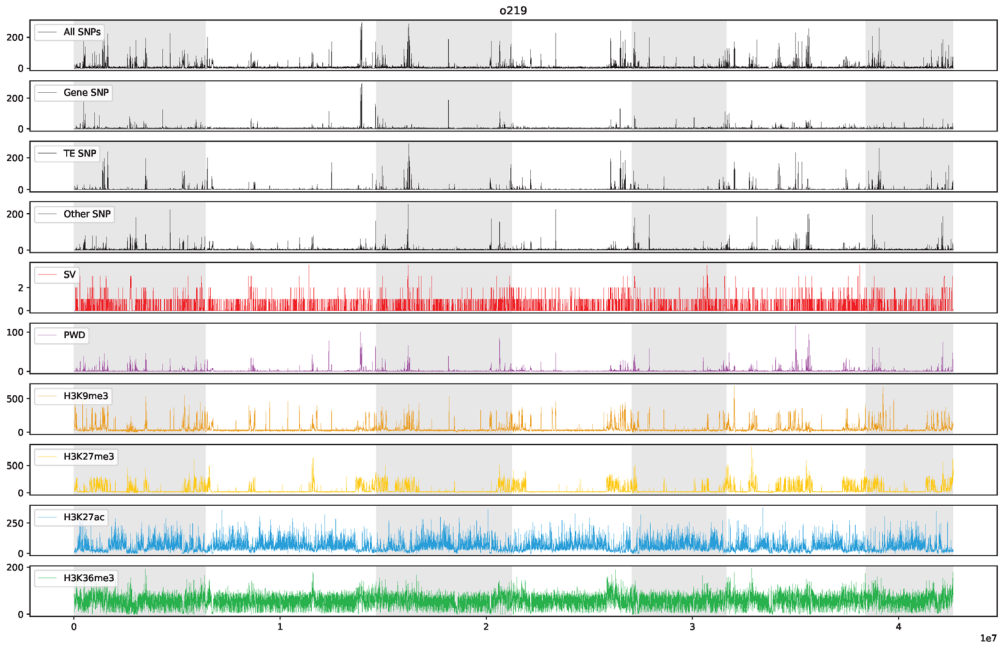
Appendix A Figure A.6. O-135 whole genome association between variation and histone marks

Graph of SNPs, INDELs, average pair-wise distance (PWD) and ChIP enrichment of H3K9me3, H3K27me3, H3K27ac, H3K4me3, and H3K36me3. Alternating shaded regions indicate core chromosomes 1 through 7.



Appendix A Figure A.7. O-137 whole genome association between variation and histone marks

Graph of SNPs, INDELs, average pair-wise distance (PWD) and ChIP enrichment of H3K9me3, H3K27me3, H3K27ac, H3K4me3, and H3K36me3. Alternating shaded regions indicate core chromosomes 1 through 7.



Appendix A Figure A.8. O-219 whole genome association between variation and histone marks

Graph of SNPs, INDELs, average pair-wise distance (PWD) and ChIP enrichment of H3K9me3, H3K27me3, H3K27ac, and H3K36me3. Alternating shaded regions indicate core chromosomes 1 through 7.

Appendix A Table A.1. Illumina short read assembly metrics

Strains	Guy11	O-135	O-137	O-219
Number of Reads	120828	164979	121132	161767
Coverage	82.8x	144.6x	60.2x	128.4x
Median Size	29065	35736	18662	32804

Appendix A Table A.2. Genomic synteny between the four strains

Strains	Guy11	O-135	O-137	O-219
Syntenic portion of the genome	55.0%	52.6%	52.1%	52.0%

Appendix A Table A.3. Guy11 Chi2 Analysis

		K27ac				K27me3				K9me3		
		With Mark	Without Mark	Chi2		With Mark	Without Mark	Chi2		With Mark	Without Mark	Chi2
Guy11	Gene + SNP	5189	9561	371.79	Gene + SNP	2117	8712	154.61	Gene + SNP	1384	9552	418.83
	Gene - SNP	3548	3738		Gene - SNP	833	1892		Gene - SNP	803	2016	
	Gene + INDEL	515	957	14.12	Gene + INDEL	181	1115	50.68	Gene + INDEL	100	1203	72.14
	Gene - INDEL	8222	12342		Gene - INDEL	2769	9489		Gene - INDEL	2087	10365	
	TE + SNP	78	1752	29.78	TE + SNP	904	937	18.2	TE + SNP	1477	350	0.65
	TE - SNP	290	3205		TE - SNP	1882	1521		TE - SNP	2736	609	
	TE + INDEL	6	117	0.52	TE + INDEL	66	57	0	TE + INDEL	80	43	21.38
	TE - INDEL	362	4840		TE - INDEL	2720	2401		TE - INDEL	4133	916	
	Inter + SNP	5218	9336	8.12	Inter + SNP	2479	8185	122.71	Inter + SNP	2060	8671	1056.42
	Inter - SNP	3070	5954		Inter - SNP	1649	3594		Inter - SNP	2302	2988	
	Inter + INDEL	729	1247	2.79	Inter + INDEL	239	1560	168.59	Inter + INDEL	182	1617	298.46
	Inter - INDEL	7559	14043		Inter - INDEL	3889	10219		Inter - INDEL	4180	10042	

Appendix A Table A.4. O-135 Chi2 Analysis

		K27ac				K27me3				K9me3		
		With Mark	Without Mark	Chi2		With Mark	Without Mark			With Mark	Without Mark	
O-135	Gene + SNP	4165	8528	218.99	Gene + SNP	1716	8258	346.79	Gene + SNP	1597	8460	339.35
	Gene - SNP	3903	5256		Gene - SNP	1277	2788		Gene - SNP	1233	2960	
	Gene + INDEL	427	921	16.73	Gene + INDEL	170	1074	47.17	Gene + INDEL	160	1087	41.94
	Gene - INDEL	7641	12863		Gene - INDEL	2823	9972		Gene - INDEL	2670	10333	
	TE + SNP	321	1695	3.72	TE + SNP	699	1146	1.9	TE + SNP	1039	815	41.19
	TE - SNP	621	3798		TE - SNP	1547	2337		TE - SNP	1847	2086	
	TE + INDEL	17	92	0.02	TE + INDEL	53	57	3.42	TE + INDEL	62	47	1.91
	TE - INDEL	925	5401		TE - INDEL	2193	3426		TE - INDEL	2824	2854	
	Inter + SNP	4002	7681	1.23	Inter + SNP	1814	7380	151.25	Inter + SNP	1764	7508	531.65
	Inter - SNP	3895	7712		Inter - SNP	1933	4965		Inter - SNP	2350	4319	
	Inter + INDEL	790	1464	1.39	Inter + INDEL	266	1755	131.95	Inter + INDEL	255	1768	197.09
	Inter - INDEL	7107	13929		Inter - INDEL	3481	10590		Inter - INDEL	3859	10359	

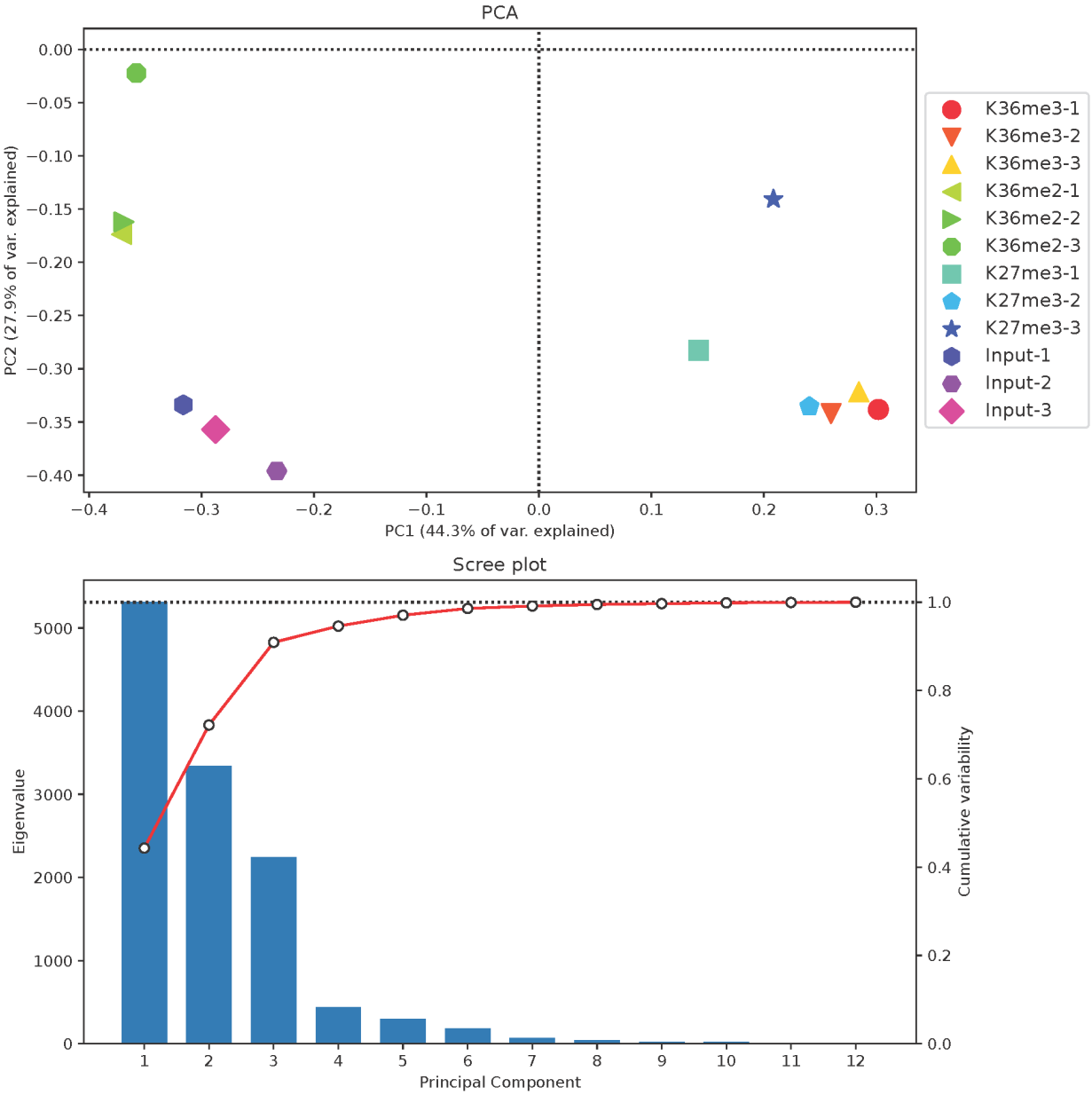
Appendix A Table A.5. O-137 Chi2 Analysis

		K27ac				K27me3				K9me3		
		With Mark	Without Mark	Chi2		With Mark	Without Mark			With Mark	Without Mark	
O-137	Gene + SNP	3166	9570	805.48	Gene + SNP	1755	8456	532.8	Gene + SNP	247	9866	600.21
	Gene - SNP	3535	4550		Gene - SNP	1304	2354		Gene - SNP	479	3212	
	Gene + INDEL	379	1051	22.42	Gene + INDEL	152	1168	93.62	Gene + INDEL	9	1305	59.98
	Gene - INDEL	6322	13069		Gene - INDEL	2907	9642		Gene - INDEL	717	11773	
	TE + SNP	61	1603	19.96	TE + SNP	746	912	12.62	TE + SNP	1235	436	222.27
	TE - SNP	312	4333		TE - SNP	2268	2256		TE - SNP	2426	2162	
	TE + INDEL	4	121	1.23	TE + INDEL	54	71	1.43	TE + INDEL	72	53	0.01
	TE - INDEL	369	5815		TE - INDEL	2960	3097		TE - INDEL	3589	2545	
	Inter + SNP	3712	8778	2.49	Inter + SNP	2048	7408	233.74	Inter + SNP	827	8649	1016.63
	Inter - SNP	3495	7902		Inter - SNP	2070	4289		Inter - SNP	1819	4734	
	Inter + INDEL	704	1600	0.16	Inter + INDEL	234	1818	261.37	Inter + INDEL	73	1976	284.56
	Inter - INDEL	6503	15080		Inter - INDEL	3884	9879		Inter - INDEL	2573	11407	

Appendix A Table A.6. O-219 Chi2 Analysis

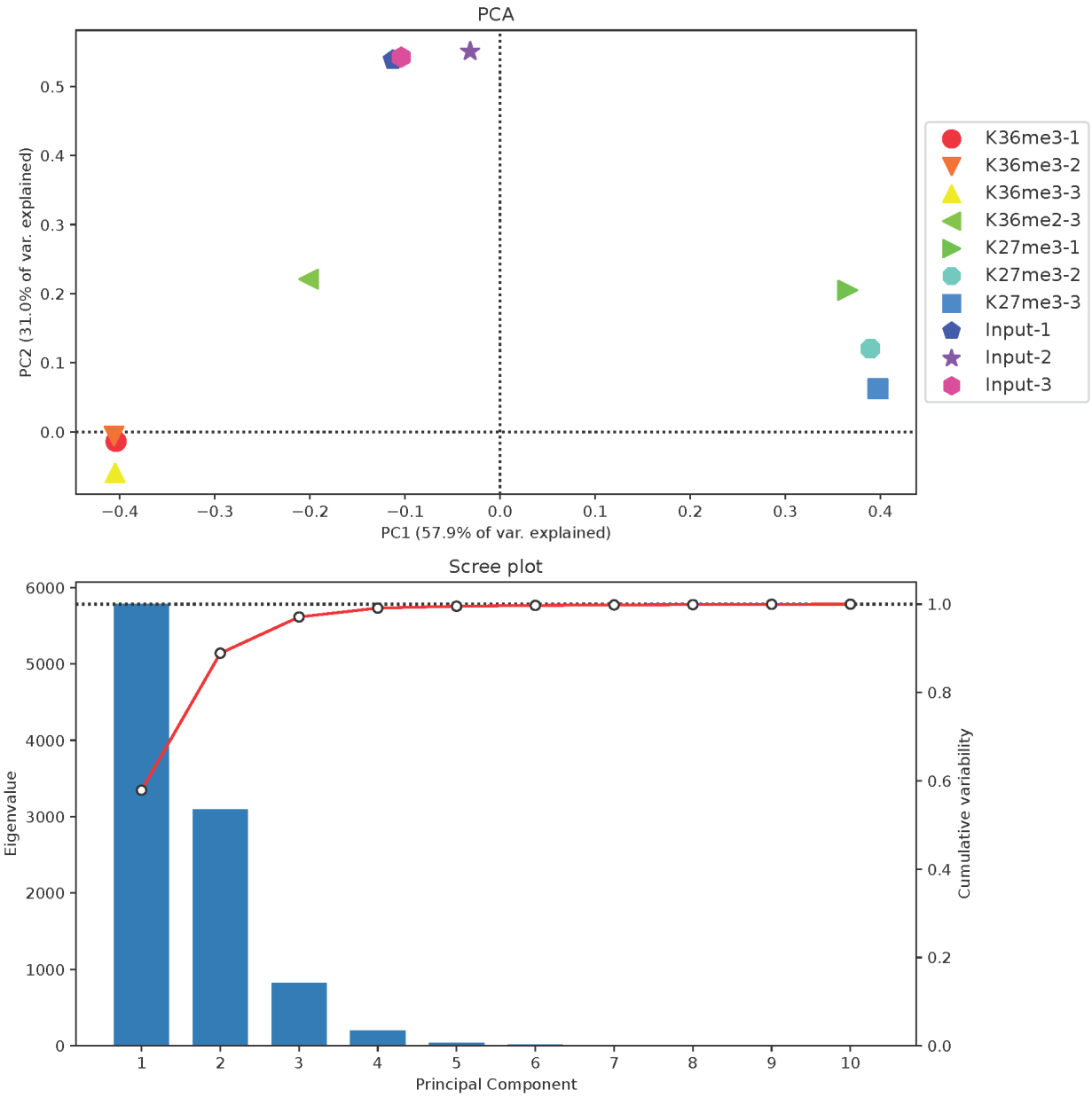
		K27ac				K27me3				K9me3		
		With Mark	Without Mark	Chi2		With Mark	Without Mark			With Mark	Without Mark	
O-219	Gene + SNP	4022	8598	215.47	Gene + SNP	1733	8319	185.66	Gene + SNP	805	9284	108.16
	Gene - SNP	3718	5216		Gene - SNP	1159	3091		Gene - SNP	584	3713	
	Gene + INDEL	437	934	10.17	Gene + INDEL	157	1106	51.59	Gene + INDEL	47	1218	55.35
	Gene - INDEL	7303	12880		Gene - INDEL	2735	10304		Gene - INDEL	1342	11779	
	TE + SNP	54	1516	0.26	TE + SNP	675	916	37.79	TE + SNP	1248	333	529.34
	TE - SNP	222	5697		TE - SNP	2049	3964		TE - SNP	2814	3241	
	TE + INDEL	5	96	0.17	TE + INDEL	48	54	5.19	TE + INDEL	62	41	1.78
	TE - INDEL	271	7117		TE - INDEL	2676	4826		TE - INDEL	4000	3533	
	Inter + SNP	4125	7922	49.73	Inter + SNP	1994	7388	71.18	Inter + SNP	1333	8113	1005.24
	Inter - SNP	3429	8026		Inter - SNP	1911	5191		Inter - SNP	2509	4642	
	Inter + INDEL	811	1496	10.49	Inter + INDEL	253	1827	174.2	Inter + INDEL	151	1925	335.12
	Inter - INDEL	6743	14452		Inter - INDEL	3652	10752		Inter - INDEL	3691	10830	

Appendix B - Supplemental Figures and Tables of Chapter 3



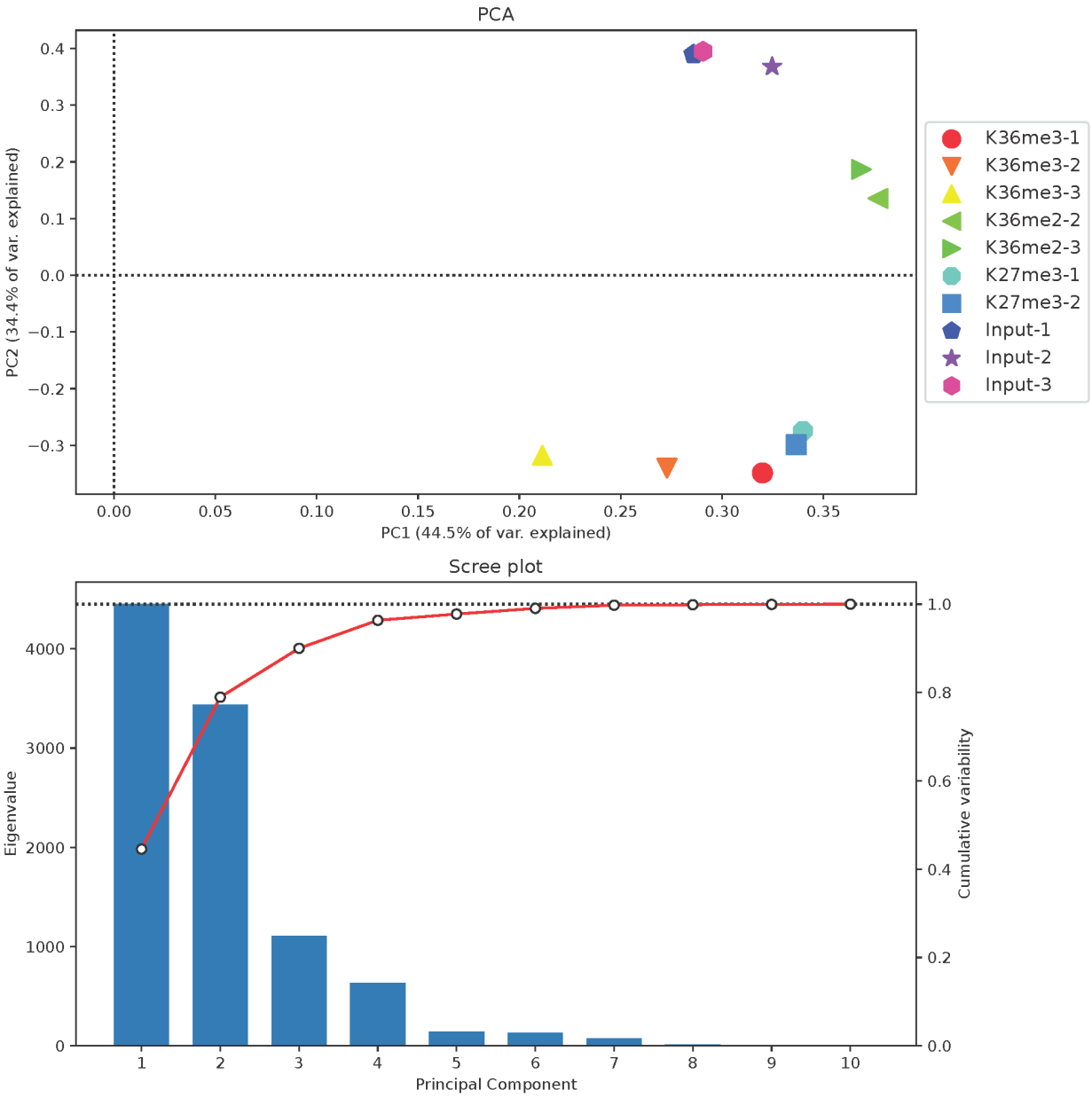
Appendix B Figure B.1. WT (Guy11) ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in WT for the histone marks H3K36me3, H3K36me2, and H3K27me3. The highest contributing components explain 44.3% and 27.9% of variance.



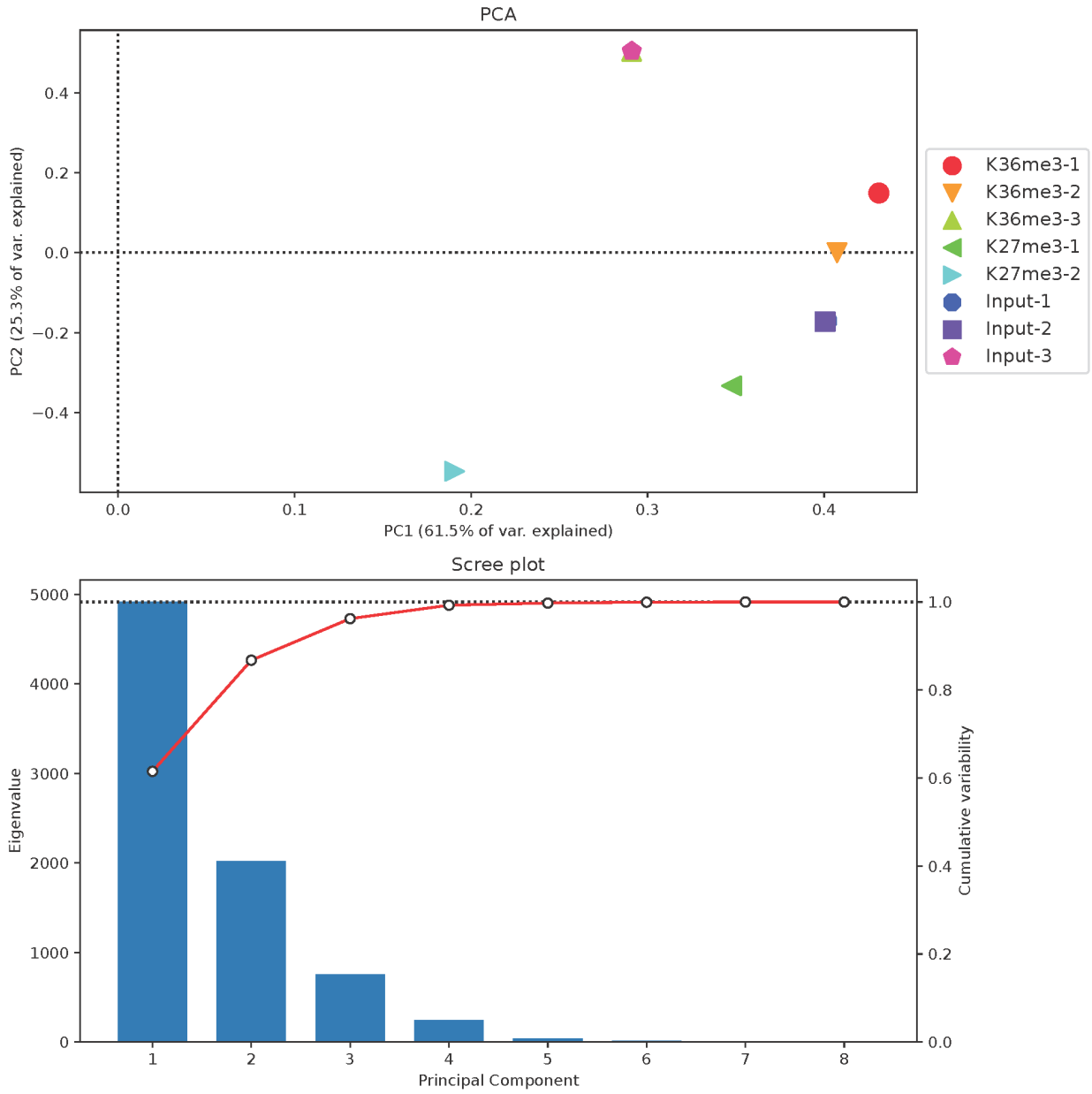
Appendix B Figure B.2. Δ*ash1* ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in Δ*ash1* for the histone marks H3K36me3, H3K36me2, and H3K27me3. The highest contributing components explain 57.9% and 31.0% of variance.



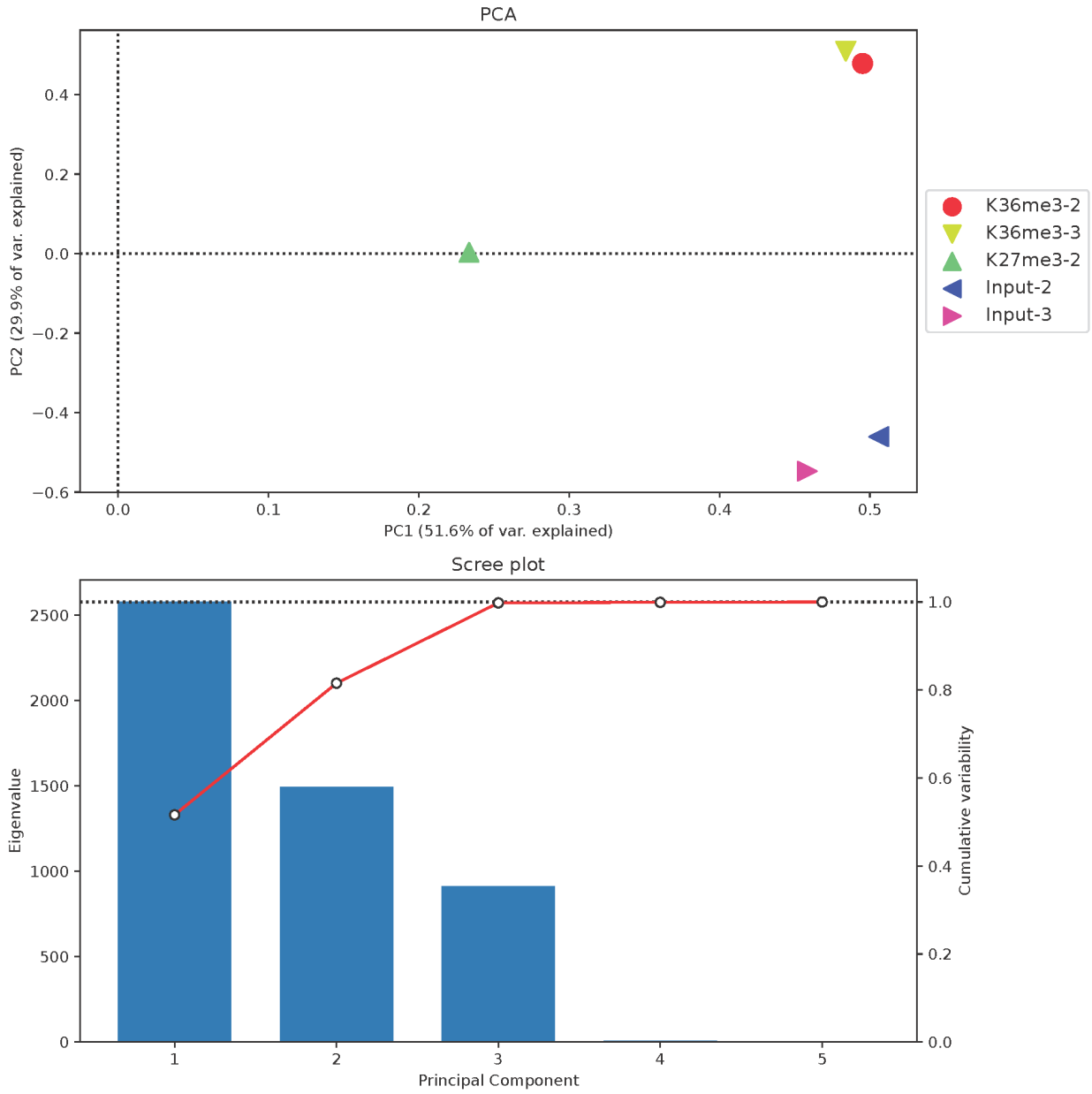
Appendix B Figure B.3. Δ*set2* ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in Δ*set2* for the histone marks H3K36me3, H3K36me2, and H3K27me3. The highest contributing components explain 44.5% and 34.4% of variance.



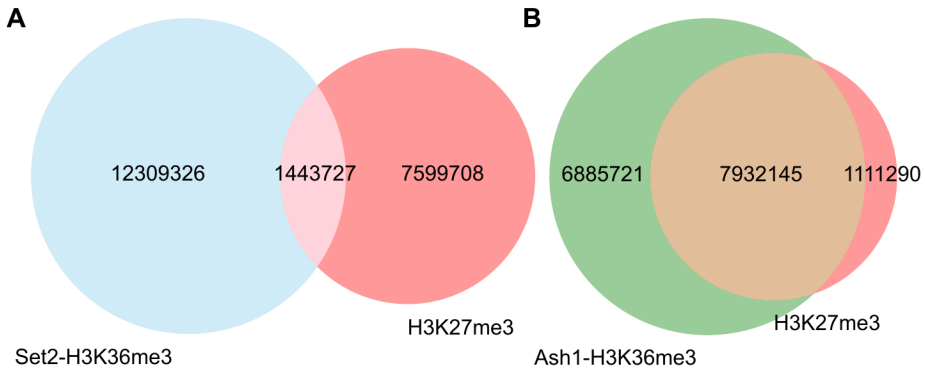
Appendix B Figure B.4. Δ*ash1*/Δ*set2* #6 ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in Δ*ash1*/Δ*set2* #6 for the histone marks H3K36me3, H3K36me2, and H3K27me3. The highest contributing components explain 61.5% and 25.3% of variance.



Appendix B Figure B.5. Δ*ash1*/Δ*set2* #15 ChIP PCA

Principal component analysis of individual replicates of ChIP-seq in Δ*ash1*/Δ*set2* #15 for the histone marks H3K36me3, H3K36me2, and H3K27me3. The highest contributing components explain 51.6% and 29.9% of variance.



Appendix B Figure B.6. Venn diagram of overlap between ASH1 and SET2 and H3K27me3

(A) Venn diagram of SET2-mediated H3K36me3 overlap with H3K27me3. (B) Venn diagram of ASH1-mediated H3K36me3 overlap with H3K37me3.

Appendix B Table B.1. sgRNA and primer oligo sequences

Name	Sequence
Cas9 Ash1 Left Guide	TTCTAATACGACTCACTATAGGAGAACTGAACCCTCCCACGGTTTTAGAGCTAGA
Cas9 Ash1 Right Guide	TTCTAATACGACTCACTATAGCACCACGACAGTTGAGACTCGTTTTAGAGCTAGA
Cas9 Set2 Left Guide	TTCTAATACGACTCACTATAGCAAGGAGGGCAGAGCTGAGCGTTTTAGAGCTAGA
Cas9 Set2 Right Guide	TTCTAATACGACTCACTATAGCGCCACCAGGCACCGATGCCGTTTTAGAGCTAGA
Ash1 Upstream	GGGGTTCGTCGCCATGATATG
Ash1 Exon2	AGGAGGGGTTGTGTCTTGTG
Ash1 Downstream	GCTCTGTCTTGTACCCCAATCTAT
Ash1 Hyg Upstream	TGAAAGCACGAGATTCTTCGC
Ash1 Hyg Downstream	CAAGGAATCGGTCAATACACTACA
Set2 Upstream	GTCGGCTACTTATTGTTTTTACGC
Set2 Exon2	GGGCACACTGATTGATGATG
Set2 Downstream	CTACTACTTTGGCAACCCTGAACT
Set2 G418 Upstream	AGTTGGATATGGAATGAATTGGAG
Set2 G418 Downstream	AGGATCTCCTGTCATCTCACCTT