

Histone modification dynamics and genome engineering in *Magnaporthe oryzae*

by

Jun Huang

B.S., Fujian Agriculture and Forestry University, 2013

M.S., Fujian Agriculture and Forestry University, 2016

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

Magnaporthe oryzae (synonym of *Pyricularia oryzae*) is a filamentous fungal plant pathogen, best known for causing rice (*Oryza sativa*) and wheat (*Triticum aestivum*) blast diseases that threaten world food security. The estimated annual loss of rice production due to rice blast is equivalent to the amount needed to feed 60 million people. *M. oryzae* is also a well-established model to study molecular pathogen-host interactions.

In order to suppress plant host immunity, numerous secreted proteins and chemicals, termed effectors, are secreted into host cells to promote disease. Many fungal pathogen effectors identified to-date, including those of *M. oryzae*, are highly expressed *in planta* but remain silent during *in vitro* growth conditions, and are located in highly variable regions of the genome. Little is known about the molecular mechanisms of transcriptional regulation and biased genome distribution and evolution of fungal effectors.

Previous reports indicate that trimethylation at H3-Lys27 (i.e., H3K27me3), a repressive histone modification, plays an important role in transcriptional regulation of secondary metabolite gene clusters. Like effectors, secondary metabolites are crucial for growth in changing environments and for specific niche adaptation, but they are dispensable for normal growth. Due to the transcriptional and genomic location similarities between effectors and secondary metabolites, we hypothesized that H3K27me3 plays an important role in silencing effector gene expression *in vitro*. Moreover, experimental evidence in animal systems suggested that histone modification dynamics involving other modifications, such as acetylation at the H3K27, could be part of a pathway for chromatin switching between inactive and activate states, similar to that observed during infection. Following chromatin immunoprecipitation and sequencing (ChIP-seq) and RNA-sequencing (RNA-seq), we found the enrichment of silent effectors in H3K27me3

domains. Genetic deletion of three core components of Polycomb Repressive Complex 2 (PRC2), resulted in a complete loss of H3K27me₃, and derepressed ~10% of transcripts during *in vitro* growth, including ~44% of high-confidence effectors. Concomitant with the loss of H3K27me₃, we found many regions gained the activation mark, H3K27ac, which appeared to play a role in effector activation, accounting for ~30% of increased transcription observed in the PRC2 mutant. ChIP-qPCR from infected materials indicated that some loci underwent H3K27 modification dynamics *in planta*, while other tested loci were not altered. Our result revealed an unexplored connection between histone modification dynamics and fungal transcriptional regulation, where H3K27me₃ is important for effector silencing *in vitro*, while *in planta* activated genes may instead be H3K27ac and contributes to effector gene activation.

Considering that roughly half of the presumed effectors showed no transcriptional changes in the PRC2 mutants, we additionally considered other genome functions that H3K27me₃ may play. One possibility is that H3K27me₃ contributes to genome instability, as has been reported in the plant pathogenic fungus *Zymoseptoria tritici*. Information connecting genome instability and histone modifications remains scarce. To further explore a potential mechanism, we addressed the contribution of DNA repair pathways to mutation outcomes and the potential for intragenomic variation. We first developed a genome editing tool based on CRISPR-Cas12a ribonucleoprotein (RNP) in *M. oryzae*. Surprisingly, we frequently observed Cas12a RNP edited mutants had abnormal genotyping outcomes. Using hundreds of edited mutants, significant variation in the mutational profile was observed, ranging from small INDELs to several kilobase deletions or insertions. Sequencing outcomes using a range of techniques suggested that multiple DNA double-strand break (DSB) repair pathways were active and responsible for the repair patterns. To exclude that the results were unique to the single tested locus, we expanded the editing to four additional

loci. Interestingly, our results suggest that the frequency of the different DSB repair outcomes varies across the tested loci. Taken together, these results suggest that DSBs are repaired in a locus-dependent manner by different DSB repair pathways. Based on these observations, we speculate that choice of DSB repair pathway may contribute to genome instability and biased genome variation. Further efforts to dissect the connection between genomic and epigenomic features that impact DSB repair choices and influence fungal genome evolution will greatly benefit our understanding about fungal genome variation in natural populations.

Histone modification dynamics and genome engineering in *Magnaporthe oryzae*

by

Jun Huang

B.S., Fujian Agriculture and Forestry University, 2013

M.S., Fujian Agriculture and Forestry University, 2016

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Dr. David E. Cook

Copyright

© Jun Huang 2022.

Abstract

Magnaporthe oryzae (synonym of *Pyricularia oryzae*) is a filamentous fungal plant pathogen, best known for causing rice (*Oryza sativa*) and wheat (*Triticum aestivum*) blast diseases that threaten world food security. The estimated annual loss of rice production due to rice blast is equivalent to the amount needed to feed 60 million people. *M. oryzae* is also a well-established model to study molecular pathogen-host interactions.

In order to suppress plant host immunity, numerous secreted proteins and chemicals, termed effectors, are secreted into host cells to promote disease. Many fungal pathogen effectors identified to-date, including those of *M. oryzae*, are highly expressed *in planta* but remain silent during *in vitro* growth conditions, and are located in highly variable regions of the genome. Little is known about the molecular mechanisms of transcriptional regulation and biased genome distribution and evolution of fungal effectors.

Previous reports indicate that trimethylation at H3-Lys27 (i.e., H3K27me3), a repressive histone modification, plays an important role in transcriptional regulation of secondary metabolite gene clusters. Like effectors, secondary metabolites are crucial for growth in changing environments and for specific niche adaptation, but they are dispensable for normal growth. Due to the transcriptional and genomic location similarities between effectors and secondary metabolites, we hypothesized that H3K27me3 plays an important role in silencing effector gene expression *in vitro*. Moreover, experimental evidence in animal systems suggested that histone modification dynamics involving other modifications, such as acetylation at the H3K27, could be part of a pathway for chromatin switching between inactive and activate states, similar to that observed during infection. Following chromatin immunoprecipitation and sequencing (ChIP-seq) and RNA-sequencing (RNA-seq), we found the enrichment of silent effectors in H3K27me3

domains. Genetic deletion of three core components of Polycomb Repressive Complex 2 (PRC2), resulted in a complete loss of H3K27me₃, and derepressed ~10% of transcripts during *in vitro* growth, including ~44% of high-confidence effectors. Concomitant with the loss of H3K27me₃, we found many regions gained the activation mark, H3K27ac, which appeared to play a role in effector activation, accounting for ~30% of increased transcription observed in the PRC2 mutant. ChIP-qPCR from infected materials indicated that some loci underwent H3K27 modification dynamics *in planta*, while other tested loci were not altered. Our result revealed an unexplored connection between histone modification dynamics and fungal transcriptional regulation, where H3K27me₃ is important for effector silencing *in vitro*, while *in planta* activated genes may instead be H3K27ac and contributes to effector gene activation.

Considering that roughly half of the presumed effectors showed no transcriptional changes in the PRC2 mutants, we additionally considered other genome functions that H3K27me₃ may play. One possibility is that H3K27me₃ contributes to genome instability, as has been reported in the plant pathogenic fungus *Zymoseptoria tritici*. Information connecting genome instability and histone modifications remains scarce. To further explore a potential mechanism, we addressed the contribution of DNA repair pathways to mutation outcomes and the potential for intragenomic variation. We first developed a genome editing tool based on CRISPR-Cas12a ribonucleoprotein (RNP) in *M. oryzae*. Surprisingly, we frequently observed Cas12a RNP edited mutants had abnormal genotyping outcomes. Using hundreds of edited mutants, significant variation in the mutational profile was observed, ranging from small INDELs to several kilobase deletions or insertions. Sequencing outcomes using a range of techniques suggested that multiple DNA double-strand break (DSB) repair pathways were active and responsible for the repair patterns. To exclude that the results were unique to the single tested locus, we expanded the editing to four additional

loci. Interestingly, our results suggest that the frequency of the different DSB repair outcomes varies across the tested loci. Taken together, these results suggest that DSBs are repaired in a locus-dependent manner by different DSB repair pathways. Based on these observations, we speculate that choice of DSB repair pathway may contribute to genome instability and biased genome variation. Further efforts to dissect the connection between genomic and epigenomic features that impact DSB repair choices and influence fungal genome evolution will greatly benefit our understanding about fungal genome variation in natural populations.

Table of Contents

List of Figures	xvi
Acknowledgements.....	xviii
Dedication	xx
Chapter 1 - General literature reviews.....	1
The contribution of DNA repair pathways to genome editing and evolution in filamentous pathogens	1
One-sentence summary	1
Abstract	1
Keywords	2
Introduction.....	2
1. Overview of eukaryotic DNA double-strand break repair pathways in model systems, and knowledge gaps in filamentous pathogens	4
Repairing DNA through canonical non-homologous end joining and homologous recombination	5
Repairing DNA through microhomology-mediated end joining and single-strand annealing	7
Interactions between DNA DSB repair pathways and knowledge gaps in filamentous pathogens	13
2. Factors affecting the usage of DNA repair pathways and their conservation in filamentous pathogens	15
Multiple repair pathways for a single DNA DSB-who repairs what?	15
Cell cycle	15
Chromatin status and nuclear compartmentalization.....	18
Sequence content and physical chromosome location.....	22
3. The influence of DNA repair pathway usage on the full spectrum of CRISPR genome editing mutations.....	24
Genome editing in animals	25
Genome editing in microbes	27
4. The contribution of DNA repair pathways to genome evolution in filamentous pathogens- the two-speed genome observation.....	31

Refining the description of the two-speed genome and associated characteristics	32
A unifying cause for the two-speed genome.....	35
Biased Variation model: synthesis of the epigenome, chromatin, and DNA repair as a mechanism for biased genome evolution.....	38
Concluding remarks and future perspectives.....	42
Acknowledgements.....	43
Funding	43
Conflicts of Interest	43
Figures	44
Chapter 2 - Histone modification dynamics at H3K27 are associated with altered transcription of in planta induced genes in <i>Magnaporthe oryzae</i>	51
Abstract	51
Introduction.....	52
Results.....	55
Trimethylation and acetylation at H3K27 occupy distinct genomic regions in <i>M. oryzae</i> ...	55
PRC2-mediated H3K27me3 contributes to normal <i>in vitro</i> growth	56
Differential gene expression upon loss of H3K27me3 and epigenetic state change	58
H3K27me3 suppresses <i>in planta</i> induced genes during <i>in vitro</i> growth	61
Effectors regulated by H3K27 modification dynamics	62
Discussion	67
Methods	70
Fungal strains and incubation condition	70
Plasmid construction and fungal transformation	71
Plant materials, leaf infection and leaf sheath assay	72
Western blotting assay	73
RNA extraction, cDNA synthesis and RT-PCR	74
ChIP sequencing	74
<i>In planta</i> ChIP-qPCR.....	76
Stranded RNA sequencing.....	76
ChIP-seq and RNA-seq data analysis	77
Phylogenetic and domain structure analysis	78

Statistical analysis	79
Availability of data and material	79
Funding	79
Figures	80
Chapter 3 - Repair of CRISPR-Cas12a induced DNA double-strand breaks in <i>Magnaporthe</i>	
<i>oryzae</i> generates locus-dependent mutation profiles	93
Abstract	93
Introduction.....	95
Results.....	98
Cas12a ribonucleoprotein editing causes unexpected DNA mutations	98
Microhomology can mediate donor DNA integration	100
Long-read sequencing and <i>de novo</i> assembly resolve genotypes following Cas12a-mediated	
DSB repair	103
DNA repair outcomes following Cas12a editing differ between multiple loci	106
Locus-dependent DNA mutation frequency still occurs under a different editing scheme	109
Discussion.....	111
Materials and Methods.....	116
Fungal strains and incubation condition	116
<i>In vitro</i> crRNA synthesis and LbCas12a RNP assembly	116
DNA donor preparation	117
Protoplast preparation and polyethylene glycol (PEG) mediated transformation	117
PCR genotyping	118
High-molecular-weight DNA extraction and library preparation for Nanopore sequencing	
.....	119
Genome assembly and long-read mapping	119
Synteny analysis.....	120
Statistical analysis.....	120
Phylogenetic and domain analysis	120
Data availability	121
Funding	121
Acknowledgments	121

Author contributions	122
Competing interests	122
Figures	123
Chapter 4 - CRISPR-Cas12a ribonucleoprotein-mediated gene editing in the plant pathogenic fungus <i>Magnaporthe oryzae</i>	131
Summary	131
Before You Begin	132
Fungal culture	132
Key Resources Table	132
Materials and Equipment	135
Prepare medium and stock solution	135
Step-by-Step Method Details.....	139
Design oligos for LbCas12a guide RNA	139
<i>In vitro</i> guide RNA synthesis and RNP assembly	141
Donor DNA preparation	145
<i>M. oryzae</i> protoplast preparation	147
<i>M. oryzae</i> protoplast transformation	148
Quick fungal DNA extraction for genotyping	151
PCR genotyping.....	152
Expected Outcomes	154
Limitations	155
Troubleshooting.....	155
Problem 1:.....	155
Potential Solution:.....	155
Problem 2:.....	156
Potential Solution:.....	156
Problem 3:.....	156
Potential Solution:.....	156
Problem 4:.....	157
Potential Solution:.....	157
Problem 5:.....	157

Potential Solution:.....	157
Acknowledgments	158
Author Contributions	158
Declaration of Interests	159
Figures	159
Chapter 5 - Re-establishing avirulence in a <i>Magnaporthe oryzae</i> -rice interaction using CRISPR-Cas12a based transposable element editing.....	164
Abstract.....	164
Introduction.....	165
Results.....	168
Discussions	170
Methods and materials	171
Fungal strains	171
<i>In vitro</i> crRNA synthesis and LbCas12a RNP assembly.....	171
Polyethylene glycol (PEG) mediated transformation	172
Pathogenicity assay	172
Southern blotting.....	173
Figures	174
Chapter 6 - Development of FuNCaP, a method for isolating fungal nuclei from <i>in-planta</i> infected sample	177
Abstract.....	177
Introduction.....	178
Results and Discussions.....	180
Experimental Procedures	183
Fungal strains	183
DNA constructions and transformation	183
Microscopy	183
Pull down by streptavidin beads	183
Figures	185
References.....	188
Appendix A - Supplementary Figures and Tables of Chapter 2.....	207

Appendix B - Supplementary Figures and Tables of Chapter 3	227
Appendix C - Supplementary Table of Chapter 5	264
Appendix D - Supplementary Table of Chapter 6	265

List of Figures

Figure 1.1. Schematic overview of the 4 DNA double-strand break repair pathways and diagram on the terminology for a-EJ, MMEJ and TMEJ.....	45
Figure 1.2. Overview of how physical, chemical, and biological conditions influence the usage of different DNA repair mechanisms.....	46
Figure 1.3. Aberrant DNA editing outcomes in filamentous pathogens.....	48
Figure 1.4. The association among DSB repair, genome compartments and biased genome evolution in filamentous pathogens	49
Figure 2.1. Methylation and acetylation at H3K27 occupy repressive and active genomic regions with distinct distribution profiles	80
Figure 2.2. PRC2 mediated H3K27me3 is required for normal fungal <i>in vitro</i> growth.....	82
Figure 2.3. Loss of H3K27me3 leads to redistribution of H3K27ac and H3K36me3.....	84
Figure 2.4. Loss of trimethylation and gain of acetylation at H3K27 contribute to H3K27me3-marked gene expression	86
Figure 2.5. H3K27me3 marked genes are enriched for <i>in planta</i> induced genes.....	88
Figure 2.6. Fungal effector genes are regulated in H3K27me3-dependent and -independent manners	90
Figure 2.7. H3K27 modifications undergo dynamics at specific genomic locations <i>in planta</i>	92
Figure 3.1. Cas12a RNP mediated editing results in unexpected genotyping at <i>BUF1</i> locus....	123
Figure 3.2. Microhomologous sequences are associated with donor DNA integrates at Cas12a target site	125
Figure 3.3. Long-read assemblies reveal three non-canonical error-prone DNA repair outcomes	127
Figure 3.4. DNA repair outcomes following Cas12a editing differ among multiple loci	128
Figure 3.5. Dramatic DNA deletions at subtelomeric <i>BAS4</i> locus	129
Figure 3.6. Distinct DSB repair patterns following donor insertion at a second integration site	130
Figure 4.1. Schematic diagram of Cas12a cutting	159
Figure 4.2. Schematic example of oligo design.....	160
Figure 4.3. Different protoplast status after lysing enzyme digestion	161
Figure 4.4. Schematic illustration of genotyping PCR primer pairs designing	162

Figure 4.5. Representative phenotyping and genotyping outcomes at <i>BUF1</i> locus with <i>BUF1</i> - guide1 RNP and no-homology <i>HYG</i> DNA donor.....	163
Figure 5.1. Deletion Pot3 in the promoter of <i>avrpiz-t</i> restores the avirulence to <i>Piz-t</i> rice.....	174
Figure 5.2. Illustration diagram of 3 components used in fungal transformation.....	176
Figure 5.3. Strategy of Southern blotting confirmation.....	176
Figure 6.1. Workflow of FuNCaP.....	185
Figure 6.2. Localization and biotinylating status for the transformants expressing nuclear envelope proteins	186

Acknowledgements

I would like to thank my major professor, Dr. David E. Cook, who gave me this precious chance to be involved in these cool projects and most importantly learn how to think with him. Before I joined the Cook lab, I was not sure whether it is the best place for me. But now, I think I made an absolutely right decision at that time. No matter how bad my results were, Dr. Cook always went through all the details with me and helped me to find a better way to move forward. He convinced me the philosophy “less is more”, and always inspired me to do the “right” experiment rather than “more and more” experiments, which has greatly benefited me. My achievements are a reflection of his patience and continuous support during this wonderful Ph.D. journey.

I would also like to thank my committee members, Dr. Barbara Valent, Dr. Richard Todd, and Dr. Stephanie Shames for sharing a wealth of knowledge on blast fungus, fungal genetics, and effector biology with me.

I would like to extend my thanks to the current and former lab members from the Cook lab. It was always fun to chat with Drs. Wenguang Zheng, Divya Mishra, Veerendra Sharma, Sandeep Marla, Wei Zhang about life and science. I really appreciate all the great help I got from Dr. Wei Zhang, David Rowe, Pratima Subedi, Melinda Dalby, Nathan Ryan, Dr. Velazhahan Rethinasamy, and Joel Oberkrom for different projects. Thanks to my colleagues and friends especially Yuzhou Xu, Dr. Huijuan Yan, Carla Redila, Dr. Guifang Lin, Dr. Liangliang Gao, Dr. Wei Wang, Javier Kiyuna, Haolang Jiang, Kai Wang, Dr. Zhenhui Zhong, Dr. Xiaofeng Chen and Dr. Sanzhen Liu and Dr. Zonghua Wang for their support and encouragement for the past five years. Thanks to all the help and support I received from the Tillman family, the Warren family,

the Sarachek family, the great Plant Pathology Department and the great KSU. I am so proud to be a wildcat.

Finally, I would like to thank my parents Yuangeng Huang and Mingfang Zhang and my girlfriend Ruiyun Zeng for their love and support.

Dedication

To my parents, Yuangeng Huang and Mingfang Zhang. Nothing can be done without their constant encouragement and support.

Chapter 1 - General literature reviews

The contribution of DNA repair pathways to genome editing and evolution in filamentous pathogens

Note: This chapter has been published in FEMS Microbiology Reviews (10.1093/femsre/fuac035). Jun Huang and David E. Cook contributed to the writing.

One-sentence summary

This review summarizes and compares the molecular mechanism, hierarchy, and regulation of four DNA double-strand break repair pathways in animal and fungal models, with the aim to connect these DNA repair pathways to genome engineering outcomes and biased genome evolution in filamentous pathogens.

Abstract

DNA double-strand breaks require repair or risk corrupting the language of life. To ensure genome integrity and viability, multiple DNA double-strand break repair pathways function in eukaryotes. Two such repair pathways, canonical non-homologous end joining and homologous recombination, have been extensively studied, while other pathways such as microhomology-mediated end joint and single-strand annealing, once thought to serve as back-ups, now appear to play a fundamental role in DNA repair. Here, we review the molecular details and hierarchy of these four DNA repair pathways, and where possible, a comparison for what is known between animal and fungal models. We address the factors contributing to break repair pathway choice and

aim to explore our understanding and knowledge gaps regarding mechanisms and regulation in filamentous pathogens. We additionally discuss how DNA double-strand break repair pathways influence genome engineering results, including unexpected mutation outcomes. Finally, we review the concept of biased genome evolution in filamentous pathogens, and provide a model, termed Biased Variation, that links DNA double-strand break repair pathways with properties of genome evolution. Despite our extensive knowledge for this universal process, there remain many unanswered questions, for which the answers may improve genome engineering and our understanding of genome evolution.

Keywords

DNA double-strand break repair; CRISPR-Cas; microhomology-mediated end joining; two-speed genome; filamentous pathogens; Chromatin

Introduction

The last 75 years have produced the structure of DNA, the birth of biotechnology, the development of high-throughput sequencing and advanced computational power, fueling discovery and understanding of the genome. This has led to an understanding of DNA synthesis, usage, and repair, and the ability to precisely manipulate DNA to alter life. This has revolutionized our understanding of biology.

One area of DNA biology that has received significant attention is how cells repair DNA that has undergone damage. There are a number of ways that DNA can be damaged (Chatterjee and Walker 2017) and specific repair mechanisms have evolved to fix or respond to the different types of physiochemical DNA damage, generally referred to as the DNA-damage response

(Jackson and Bartek 2009). This review focuses on one specific type of DNA damage, termed DNA double-strand breaks (DSBs), in which the covalent phosphodiester bond between adjacent sugar atoms of DNA nucleotides are broken on both DNA strands. Repairing DNA DSBs is critical for cell viability and genome integrity, and cells cannot undergo replication and division with such lesions (Ceccaldi *et al.* 2016). While repair is a critical process, it is also dangerous to the cell, in that improper DNA rejoining can result in mutations with potentially negative effects. As such, cells have evolved extensive mechanisms to detect DNA DSBs, to stabilize the broken DNA sites, and to manipulate the nucleic acid in an effort to repair the DNA while ensuring normal function. Given this critical and complex task, it is not surprising that eukaryotes have evolved multiple DNA DSB repair pathways that are generally conserved across eukaryotic life (Ceccaldi *et al.* 2016; Friedberg 2003; Mehta and Haber 2014). At least two of the DNA DSB repair pathways in eukaryotes are also active in prokaryotes, reviewed previously (Aravind *et al.* 1999; Cromie *et al.* 2001; Wright *et al.* 2018), highlighting the critical need to resolve DNA DSBs to maintain normal genome function. As detailed in this review, DNA repair is not a one-size fits all mechanism, and variation exists for the preference, protein components, and hierarchy of DNA DSB repair pathways (Bertrand *et al.* 2019; Lieber 2010). This review does not cover programmed DNA DSBs induction (Borde and de Massy 2013) needed for genome function, such as class-switch recombination important for antibody diversification (Stavnezer *et al.* 2008), gene-conversion such as mating type switching (Haber 2012), or the regulation of DSB repair pathways required for telomere maintenance (Doksani and de Lange 2014).

The aim of this review is to detail DNA DSB repair, and to explore how these pathways influence modern genome engineering efforts and genome evolution in natural populations. We focus on experimental knowledge from animal, plant, and yeast models, and compare this with

experimental results from filamentous pathogens. We specifically identify the study species with respect to experimental results where possible, but generally, our references to yeasts refer to conventional species, such as Baker's yeast, *Saccharomyces cerevisiae*, which has been a critical model for studying eukaryotic DNA repair mechanisms. In addition to filamentous fungi, dimorphic fungi such as the model plant pathogens *Ustilago maydis* and *Zymoseptoria tritici*, along with oomycetes are discussed as well. The review is organized in four sections: 1) The basic molecular mechanisms of the four major DNA DSB repair pathways, including canonical non-homologous end joining (C-NHEJ), homologous recombination (HR), microhomology-mediated end joining (MMEJ) and single-strand annealing (SSA). We pay special attention to the MMEJ pathway, and the variety of other alternative end-joining pathways that have been described, as many details, function, and impact of these pathways remain unclear. 2) Knowledge about the hierarchy of DNA repair pathways, and the factors that influence the interplay between pathway activity. 3) How DNA repair influences genome engineering outcomes. 4) The impact of DNA repair on genome evolution. We pay special attention to filamentous pathogens in section four, as previous observations on genome variation and evolution have detailed biased genome evolution, and we suggest this may be driven in part by variation in DNA repair, which could influence the emergence of pathogen genome variation.

1. Overview of eukaryotic DNA double-strand break repair pathways in model systems, and knowledge gaps in filamentous pathogens

Repairing DNA through canonical non-homologous end joining and homologous recombination

Proper repair of DNA double-strand breaks (DSBs), caused by inherent mechanisms or external factors, are vital to the maintenance and function of the genome (Mehta and Haber 2014; Srinivas *et al.* 2019; Vitor *et al.* 2020). The association between aberrant DNA DSB repair and genetic disease has been reported in many systems (Aparicio *et al.* 2014; Helleday *et al.* 2014). The two most well characterized DNA DSB repair pathways are canonical non-homologous end joining (C-NHEJ) and homologous recombination (HR), also called homology direct repair (HDR). The mechanisms of repair for these two pathways are quite different, and largely determined by the initial fate of the DNA ends at the break site. During C-NHEJ, the Ku70-Ku80 heterodimer binds to the DNA DSB ends, protecting them from end resection, and recruits other enzymes central to C-NHEJ repair (Mimitou and Symington 2010). This includes a DNA-dependent protein kinase catalytic subunit (DNA-PKcs), endonuclease Artemis, and DNA polymerase μ and λ to promote DNA end processing suitable for subsequent ligation. The DNA ligase IV (Lig4)-XRCC4 complex is required for this final ligation to seal the DNA DSB (Chang *et al.* 2017) (Figure 1.1A). As the name suggests, the C-NHEJ process does not rely on homologous DNA templates for repair, however, short microhomology (up to 4 base pair) has been found at repair junctions during C-NHEJ (Daley and Wilson 2005; Pannunzio *et al.* 2018). DNA DSB repair from the C-NHEJ pathway often results in small sequences changes at the repair site. This is due to the action of nucleases such as Artemis and the associated DNA polymerases, which process and prepare the broken DNA ends for ligation. This error-prone repair is characterized by small insertions and deletions (INDEL), and tandem duplications (Her and Bunting 2018; Schimmel *et al.* 2017). It is important to note that while C-NHEJ is often described as inducing

INDELS, and therefore as error-prone, precise C-NHEJ ligation has also been reported (Betermier *et al.* 2014; He *et al.* 2016). Such high-fidelity DNA repair is often unreported or intractable with common selection and sequencing-based DNA DSB repair assays, and may occur more frequently than currently appreciated.

Unlike C-NHEJ, which suppresses end resection, the HR pathways utilizes end resection to generate single-stranded DNA ends compatible for homologous recombination. During the initial steps of HR, the Mre11 endonuclease cleaves 5'-terminated DNA as part of the Mre11-Rad50-Xrs2 complex that is termed MRX in budding yeast *Saccharomyces cerevisiae*, while the homologous complex of Mre11-Rad50-Nbs1 is termed MRN in mammalian systems (Wright *et al.* 2018). The endonuclease activity of the MRX/MRN complex is further promoted by the nuclease termed Sae2 in *S. cerevisiae*, or CtIP in animals (Anand *et al.* 2016; Cannavo and Cejka 2014). An interesting feature of the MRX complex is that it can process DNA bidirectionally. The nick caused by MRX and Sae2 serves as the entry point for 3'-5' exonuclease activity by Mre11, and 5'-3' exonuclease activity by Exo1 and/or helicase Sgs1 with endonuclease Dna2 (Garcia *et al.* 2011; Symington 2014). The result of DNA end resection by MRX and accompanying nucleases are long 3' ssDNA overhangs that are bound by replication protein A (RPA), and subsequently replaced by DNA recombinase protein Rad51 with the help from accessory protein (i.e., Rad52 in budding yeast) (Chen *et al.* 2013a; Daley *et al.* 2014). The association of Rad51 on ssDNA creates nucleoprotein filaments that promote base-pairing between ssDNA and homologous dsDNA in the genome. This process is referred to as strand invasion, and forms a DNA structure called a displacement loop (D-loop) (Wright *et al.* 2018) (Figure 1.1A). The complementary base pairing by the 3' DNA sequence from the DSB site can serve as a DNA polymerase initiation site, allowing for extension of the invading strand with the homologous locus

serving as a template (Wright *et al.* 2018). There are different mechanisms to resolve the DNA DSB depending on the D-loop disassociation, engagement of a second resected end, and the method resolving the double Holliday junction (dHJ) (Daley *et al.* 2014). The different HR outcomes can be achieved by resolving the extended D-loop by synthesis-dependent strand annealing (SDSA) pathway (the invaded strand with extended sequences gets dismantled for the D-loop and anneals with ssDNA from the other side of DSB); double-strand break repair (DSBR) pathway (second strand capture with dHJ resolution); double Holliday Junction dissolution pathway (second strand capture with dHJ dissolution) and break-induced replication (BIR) pathway (only one side of DSB gets repaired) (Daley *et al.* 2014; Kramara *et al.* 2018). In general, DNA DSBs repaired with HR result in a more precise sequence repair compared to C-NHEJ, owing to the use of homologous DNA template from either sister chromatids or homologous chromosomes (Wright *et al.* 2018). This fact serves as the basis for knock-in or gene insertion genome engineering projects that utilize extensive homologous sequence to stimulate site-specific insertion of exogenous DNA.

Repairing DNA through microhomology-mediated end joining and single-strand annealing

There are additional DNA DSB repair pathways active in eukaryotes in addition to C-NHEJ and HR. These pathways are termed microhomology-mediated end joining (MMEJ) and single-strand annealing (SSA). Both pathways involve homologous DNA sequence to resolve the DNA DSB, but the pathways are different from each other, and from HR directed repair, based on the proteins involved, characteristics of the homologous DNA used for repair, and their resulting sequence profiles.

The generalized model for MMEJ across kingdoms of life involves end resection, annealing of microhomologous sequences from the exposed ssDNA ends, nuclease activity to remove DNA-flaps if necessary, DNA polymerase activity to fill gaps around the microhomology annealed sequences, and finally DNA ligation to seal the DNA DSB (Figure 1.1A). The role of microhomologous sequence pairing during DSB end joining was first demonstrated in monkey cell lines (Roth and Wilson 1986). Genetic evidence of a Ku-independent repair pathway was first clearly demonstrated in *S. cerevisiae* using a $\Delta ku70$ strain that could *in vivo* repair a plasmid based DNA DSB that resulted in sequence deletion at the repair site (Boulton and Jackson 1996b). The size of the observed deletions ranged from 6 to 811 bp and were flanked by 3 to 16 bp microhomology. This work demonstrated the activity of two HR-independent repair pathways, and found Ku-dependent repair was preferentially active and less error-prone than Ku-independent repair (Boulton and Jackson 1996b). Another early study in *S. cerevisiae* using repair of non-complementary end sequences, also revealed that Ku-independent end joining resulted in sequence deletions and involved annealing microhomology sequences near the repair junction (Ma *et al.* 2003). This pathway was termed microhomology-mediated end joining (MMEJ) owing to the usage of microhomology to direct DNA repair (Ma *et al.* 2003). The terms alternative-NHEJ (a-NHEJ) and alternative end joining (a-EJ) have also been used to describe these DNA DSB repair outcomes that do not require known components of C-NHEJ (Fattah *et al.* 2010; Sallmyr and Tomkinson 2018). The occurrence of a-EJ repair mediated by microhomologous sequence has been readily observed in animal, plant, and fungal systems, suggesting an evolutionarily conserved outcome, however, the molecular mechanisms and complexes that mediate a-EJ across eukaryotic domains of life appear different (Figure 1.1B). Work in multiple metazoan models, including mouse, *Drosophila melanogaster* and *Caenorhabditis elegans*, discovered that DNA polymerase

theta (Pol Θ) plays a key role in mediating a-EJ, characterized as C-NHEJ independent repair that utilizes short homologous sequence and often creates templated insertions (Chan *et al.* 2010; Roerink *et al.* 2014; Schimmel *et al.* 2019; Shima *et al.* 2003). The term theta-mediated end joining (TMEJ) has been proposed as a more precise definition for this repair, which falls under the broader umbrella term a-EJ that might include additional unknown repair mechanisms (Chan *et al.* 2010; Roerink *et al.* 2014) (Figure 1.1B). In plants, there are clear homologs to animal Pol Θ , and experimental evidence in the moss *Physcomitrella patens* suggests that the majority of DNA mutations following repair of Cas9-induced DSB are the result of Pol Θ -mediated repair, independent of C-NHEJ (Mara *et al.* 2019). Many details regarding TMEJ repair appear similar between plants and animals (e.g., observed microhomology in the repair junctions) (Mara *et al.* 2019; van Kregten *et al.* 2016). Interestingly, the two most well studied phyla of fungi, namely *Basidiomycota* and *Ascomycota*, do not possess a clear homolog containing the canonical Pol Θ domains (Huang *et al.* 2021; Meyer *et al.* 2015). Therefore, we will refer to DNA DSB repair results in fungi that involved microhomology and were C-NHEJ independent as MMEJ, while we will use the term TMEJ for similar DNA repair outcomes that involve Pol Θ from animal systems.

Similar to the HR pathway, end resection is a key initial step for TMEJ to generate short ssDNA ends that may contain microhomology to direct annealing in animals. This initial end resection is carried out by the MRN complex, together with CtIP, as described for HR (Ahrabi *et al.* 2016; Truong *et al.* 2013; Zhang and Jasin 2011). This initial end resection by MRN results in relatively short DNA sequence removal prior to TMEJ. For HR, further end resection is carried out to expose longer ssDNA mediated by Exo1 or BLM-Dna2, which are generally dispensable for TMEJ (Truong *et al.* 2013). Another early molecular event implicated in TMEJ is PARylation (i.e., post-translational deposition of ADP-ribose molecules) of proteins associated with the DSB

site by poly (ADP-ribose) polymerase 1 (PARP1) (Ray Chaudhuri and Nussenzweig 2017). The model is that PARylation serves as a signal to recruit additional proteins (e.g., Mre11) to the DSB site to facilitate TMEJ-mediated repair, and genetic and chemical inhibition of PARP1 activity reduces TMEJ repair (Mansour *et al.* 2010; Ray Chaudhuri and Nussenzweig 2017; Wray *et al.* 2013). Following unpaired DNA end trimming by Xpf-Ercc1 (Bai *et al.* 2021), the 3'OH group is eligible for polymerase-mediated DNA synthesis (Sijbers *et al.* 1996). The major factor defining TMEJ in animals and plants is the function of a low-fidelity DNA polymerase theta (Pol Θ) encoded by *POLQ* (Kent *et al.* 2015; Mara *et al.* 2019). Pol Θ homolog (mus308) was first identified from a screening assay with hypersensitivity to DNA crossing-linking agents in *Drosophila* (Boyd *et al.* 1981; Boyd *et al.* 1990). Further molecular data suggested that Pol Θ is able to extend the single-strand DNA with either template-independent or -dependent activities (Hogg *et al.* 2012; Kent *et al.* 2015; Kent *et al.* 2016). The final step for DSB repair is to seal the nick between the two ends through a phosphodiester bond, catalyzed by two DNA ligases in animals, termed ligase I (Lig1) and ligase III (Lig3) (Liang *et al.* 2008) (Figure 1.1A).

There is comparatively less mechanistic evidence for Ku70-80 independent DSB repair pathway in fungi, referred to here as MMEJ. The majority of work published to date has been conducted in yeast, an incredibly productive system, but hardly representative of all fungi. The initial end resection during MMEJ is carried out by the same complex as in HR, MRX with Sae2 (Sfeir and Symington 2015). Conflicting results have been reported in budding yeast regarding the role of the MRX complex and MMEJ. Using a chromosomal end joint assay, one group reported that *Asae2* and *Amre11* mutants without nuclease activity appeared to have normal MMEJ repair (Deng *et al.* 2014), while earlier observations using a dual endonuclease HO-mediated cleavage assay reported that *Asae2* and *Amre11* caused MMEJ defects (Lee and Lee 2007). The difference

may be due to the distinct experimental conditions or point to unknown differences in end resection requirements in this yeast. In animals, experimental evidence links PARylation by PARP1 to TMEJ repair, but the role of PARylation during fungal DNA repair remains unclear (Audebert *et al.* 2004; Citarelli *et al.* 2010; Ray Chaudhuri and Nussenzweig 2017; Tao *et al.* 2009). In *Aspergillus nidulans*, a putative PARP homolog, termed PrpA, was identified that contained high identity to the catalytic domain of PARP1, but lacked the N-terminal DNA binding domain (Semighini *et al.* 2006). Assays to test for altered sensitivity to DNA damaging agents, protein localization and transcriptional responses indicated that PrpA may play a role in DNA damage, but direct evidence was not provided (Semighini *et al.* 2006). The PrpA homolog in *Neurospora crassa*, termed NPO, was shown to be a *bona fide* PAR-polymerase, and the gene was transcriptionally induced by the DNA damage agent methyl methanesulfonate (MMS), but unlike in *A. nidulans*, the *Δnpo* strain in *N. crassa* was not sensitive to DNA damaging agents (Kothe *et al.* 2010). The observations that PrpA was required for normal growth and development in *A. nidulans*, but was dispensable in *N. crassa* indicates divergent function, and further research is needed regarding the role of PARP homologs for fungal DNA damage response. Following end resection, ssDNA with homologous sequence anneals, and 3' flap DNA is excised by the complex of Rad1 and Rad10, endonucleases discovered in *S. cerevisiae* (Ma *et al.* 2003). The absence of a clear Pol Θ homolog, containing the characteristic domains, is the major difference suggesting altered mechanisms for DNA DSB repair utilizing microhomology in animals and fungi (Huang *et al.* 2021; Wood and Doublet 2016). In *S. cerevisiae*, there is evidence that DNA polymerases 3 (Pol δ) and DNA polymerase 4 (Pol λ) function in gap-filling (i.e., DNA synthesis) following microhomology annealing (Meyer *et al.* 2015). Why there are two polymerases, how they interact, and their overall function in MMEJ in filamentous fungi remains to be determined (Meyer *et al.*

2015). For the final DNA ligation step, the main ligase from animals, Lig3, is absent in *S. cerevisiae*, and while a Lig1 homolog exists, its role in MMEJ has not been tested (Audebert *et al.* 2004; Liang *et al.* 2008; Sfeir and Symington 2015). In addition to the molecular differences described between TMEJ in animals and MMEJ in fungi, the sequence requirements and outcomes also appear to vary. In *C. elegans*, only ~1 bp of microhomology appears to be required to initiate TMEJ, and 2-4 bp of microhomology is frequently observed in *Xenopus* (Chandramouly *et al.* 2021; Koole *et al.* 2014; van Schendel *et al.* 2015). However, in two separate studies in *S. cerevisiae*, longer stretches of microhomology between 8-20 bp were observed (Lee *et al.* 2019; Villarreal *et al.* 2012). Following TMEJ repair in metazoans, small insertions/deletions (INDELS) and templated insertions are predominant at the DSB repair sites, while insertions are rare in yeast (Ceccaldi *et al.* 2016; Schimmel *et al.* 2019; Sfeir and Symington 2015) (Figure 1.1A).

Single-strand annealing (SSA) initiates through the same end resection step mediated by the MRN complex to generate free 3'-ssDNA for homology search, a shared requirement for SSA and TMEJ/MMEJ (Bhargava *et al.* 2016). However, one of the main differences between SSA and TMEJ/MMEJ is the length requirement between annealed homologous sequence (Sallmyr and Tomkinson 2018). Estimates vary, but SSA is generally characterized to require homology > 25 bp, which is longer than what is generally thought to be used for TMEJ/MMEJ (Sallmyr and Tomkinson 2018). The use of longer homologous sequence is likely the results of a second round of end resection mediated by Exo1 or Sgs1-Dna2 (Sturzenegger *et al.* 2014). This more extensive end resection generates longer ssDNA filaments that are bound by RPA and can undergo homology directed annealing, stimulated by RAD52 and small acidic protein DSS1 (Grimme *et al.* 2010; Stefanovie *et al.* 2020). The 3' unpaired tail DNA is removed by Xpf1-Ercc1 during both SSA and TMEJ/MMEJ repair (Al-Minawi *et al.* 2008). The final steps include gap-filling by DNA

polymerase and ligation by DNA ligase, but the exact protein requirements for these steps are poorly defined (Bhargava *et al.* 2016). The result of SSA repair is often large deletions flanked by homologous sequence, which are frequently genomic repeats, that result from homologous pairing of distantly located loci followed by non-homologous 3' flap trimming before ligation (Bhargava *et al.* 2016) (Figure 1.1A).

Interactions between DNA DSB repair pathways and knowledge gaps in filamentous pathogens

Given the function of multiple DNA repair pathways with a similar purpose, it is no surprise that multiple studies have detailed molecular interactions between the pathways. Evidence suggests the interactions depend on multiple cellular, genomic, and physiological factors that are discussed in section two of this review. Here we discuss general observations and molecular interactions between the repair pathways that influence the hierarchy of DNA repair. Generally, C-NHEJ is the predominant pathway to resolve DNA DSBs in multiple eukaryotic organisms, including mammalian cells, somatic plant cells, and filamentous fungi (Chang *et al.* 2017; Knoll *et al.* 2014; Ninomiya *et al.* 2004). The binding of the Ku70-Ku80 heterodimer to DSB ends represses end resection that is required for HR, TMEJ/MMEJ and SSA (Shao *et al.* 2012). The general evidence for this is that other repair pathways are generally difficult to detect, but are easily observed when core C-NHEJ components (e.g., Ku70-Ku80) are deleted (Boulton and Jackson 1996b; Chu *et al.* 2015; Ninomiya *et al.* 2004). Following MRX and Sae2-mediated short end resection, Ku70-Ku80 DNA binding is inhibited and TMEJ/MMEJ repair are more prominent (Ceccaldi *et al.* 2016; Lee and Lee 2007). This is seen in *S. cerevisiae*, where removal of Sae2 or expression of a nuclease defective Mre11 variant results in higher repair activity of C-NHEJ (Lee and Lee 2007). Similarly, the amount of Ku protein recruited to DSB sites was increased in a

Δmre11 mutant (Zhang *et al.* 2007b). For other pathway interactions, removal of Pol Θ in mammalian cells leads to increased HR repair (Mateos-Gomez *et al.* 2015). If extensive end resection is initiated, several hundred or thousands of DNA bases can be removed, and the SSA and HR pathways become predominant (Ceccaldi *et al.* 2016). Following long DNA end resection, the TMEJ/MMEJ pathway is inhibited by RPA binding ssDNA ends (Deng *et al.* 2014). Disrupting the HR pathway, such as through *Δrad51* deletion, promotes the frequency of MMEJ or SSA (Deng *et al.* 2014; Mansour *et al.* 2008). Due to the overall similarity between TMEJ/MMEJ and SSA, the molecular interactions and hierarchy between them is less clear. Collectively, evidence from genetic and biochemical experiments show that there is competition and inhibition between repair pathways for a given DNA DSB.

The majority of data on DNA repair in fungi come from studies on a few species. Early work in fungi on DNA repair and recombination were critical to the development of modern theory, relying heavily on experimental results from *S. cerevisiae*, *N. crassa*, and the *Basidiomycete Ustilago maydis*, the causal agent of corn smut (Boulton and Jackson 1996a; de Sena-Tomas *et al.* 2015; Inoue and Schroeder 1988; Kojic *et al.* 2003; Ninomiya *et al.* 2004; Stahl 1994; Steinberg and Perez-Martin 2008). However, this represents a small sampling of fungal diversity, and the early research does not incorporate current knowledge from metazoan models to understand if the rules and hierarchy discovered in those systems pertain to filamentous microbes. For instance, it is not clear how exciting new research defining the role of Pol Θ influencing genome stability and cancer development, pertain to fungi given that the major phyla lack a clear homolog of Pol Θ (Brambati *et al.* 2020; Huang *et al.* 2021). As the molecular mechanisms, outcomes, and importance of MMEJ are not well defined across diverse fungi, it is difficult to compare and contrast results between metazoans and fungi. It is also important to note that some

DNA repair details in *S. cerevisiae*, such as the predominant action of HR, do not appear to reflect the majority of filamentous fungi that exhibit dominant repair by C-NHEJ (Haber 1999; Ninomiya *et al.* 2004). Thus, caution is warranted when extrapolating details across fungal taxa. More research is needed across filamentous fungi and oomycetes to molecularly characterize and document that action of DNA DSB repair mechanisms, interactions, and how this critical cellular process influences stability and evolution of fungal genomes.

2. Factors affecting the usage of DNA repair pathways and their conservation in filamentous pathogens

Multiple repair pathways for a single DNA DSB-who repairs what?

Given the potential lethality of unresolved DNA DSBs, it is not a surprise evolution would give rise to multiple DNA repair pathways in the cell. However, the occurrence of partially overlapping and redundant pathways to achieve the same function raises the question of which pathway should repair any given DNA DSB? In the following section, we summarize current knowledge for how cell physiology and genomic features influence DNA DSB repair pathway choice.

Cell cycle

DNA DSB repair is tightly controlled during the cell cycle (Ceccaldi *et al.* 2016; Chang *et al.* 2017) (Figure 1.2A). C-NHEJ is activated throughout all phases (i.e., G1, S, G2, M) and dominates in G1. The mammalian protein, 53BP1 together with its downstream effector RIF1, inhibit end resection at DSB sites, limiting the recruitment of homology-dependent repair proteins to DSB sites in a cell cycle dependent manner. Inactivation of 53BP1 or RIF1 causes increased

HR, MMEJ and SSA repair frequencies (Escribano-Diaz *et al.* 2013). Experiments in human cell lines show that during the S and G2 phases, C-NHEJ is suppressed by cell cycle regulator of NHEJ (CYREN, also known as MRI) via interacting with Ku proteins (Arnoult *et al.* 2017). There is also a report that CYREN can promote C-NHEJ in the absence of the XRCC4-like factor, termed XLF in the G1 stage (Hung *et al.* 2018). The XLF protein is crucial during the final ligation reaction of C-NHEJ by interacting with Lig4-XRCC4 (Ahnesorg *et al.* 2006), and these observations indicate diverse control of DNA repair pathways in a cell-cycle dependent manner.

The HR pathway is more active during the S and G2 cell phases, when DNA replication has taken place and nucleic acids to serve as repair templates are available. The activity of cyclins and cyclin-dependent kinases (CDKs) promote DNA end resection and HR activity in a cell-growth dependent manner (Aylon *et al.* 2004; Mathiasen and Lisby 2014). In *S. cerevisiae*, Cdk1 (also known as Cdc28) phosphorylates Sae2 (Ser267) to facilitate short end resection in the S/G2 phase. A serine to alanine (S267A) substitution, creating a Sae2 phosphorylation deficient mutant, mimics DNA repair defects observed in a Δ sae2 null mutant, while serine to glutamic acid (S267E) substitution (i.e., phosphomimic) phenocopies wild-type Sae2 in the absence of Cdk1 (Huertas *et al.* 2008). Further studies in *S. cerevisiae* suggest that phosphorylation causes tetramerization of Sae2, followed by activation of Mre11 and interaction with Rad50, which promotes MRX-mediated end resection (Cannavo *et al.* 2018). The resection nuclease, Dna2, involved in extensive end resection, is also regulated by Cdk1 phosphorylation (Chen *et al.* 2011). Additionally, multiple HR related proteins involved in end resection, DNA damage checkpoint or recombination have been also identified as Cdk substrates from yeast and mammalian systems (Zhao *et al.* 2017). A recent study in *S. cerevisiae* found that Rad51 and Rad52 are phosphorylated by Cdk1 during G2/M phase with related cyclins (e.g., Clb2 and Clb3), and phosphorylated Rad51 and Rad52

promote DNA binding and strand annealing, respectively (Lim *et al.* 2020). Given the role of Rad52 in mediating ssDNA annealing during SSA repair, it is possible that SSA activity is cell cycle dependent. It would be interesting to further test how cell cycle regulated Rad52 impacts DNA repair pathway activity between HR and SSA (Bhargava *et al.* 2016; Grimme *et al.* 2010; Lim *et al.* 2020). An important future question is the role, hierarchy, and interaction of various Cdk phosphorylation targets. Understanding this network will provide a clearer understanding of DNA repair pathway activity in the context of the cell cycle.

It is less clear how TMEJ/MMEJ and SSA repair pathways are impacted by the cell-cycle. There are conflicting reports on the role of 53BP1, which was shown to facilitate TMEJ during G1 in human cell lines (Xiong *et al.* 2015), but it was shown to suppress TMEJ, HR, and SSA repair under similar human cell line conditions (Munoz *et al.* 2012). The TMEJ/MMEJ pathway is presumably favored over SSA in G1 in the absence of C-NHEJ due to the shorter homologous sequence requirement (Chang *et al.* 2017). Previous models proposed that end resection dependent repair pathway preference is governed by the availability of repair templates. This model suggests that TMEJ/MMEJ and SSA are preferred to HR when sister chromatids are not available in early S phase, and then later following DNA synthesis, HR is the predominant DNA DSB repair pathway (Bhargava *et al.* 2016). To our knowledge, there is little experimental evidence regarding DNA repair choice and cell cycle dependence in filamentous pathogens.

Knowledge generated in animal and yeast systems may transfer to filamentous fungi, but it is reasonable to believe that significant differences may exist because of inherent biological differences between classes of eukaryotes. We ask the reader to keep in mind that while DNA repair is critical, and the process is generally conserved across prokaryotic and eukaryotic life, it should not be assumed that the mechanisms or the impact on the genome are the same (Steenwyk

et al. 2019). Factors such as ploidy, where animal models are commonly diploid and many filamentous fungi are haploid, likely impacts DNA repair. Many animal systems, and some fungal systems (e.g., yeasts) frequently use the sexual cycle, but it is absent or not demonstrated in many filamentous pathogens. Also, some filamentous fungi are multi-nucleate, which can impact genome evolution (Roper *et al.* 2011), but details regarding how multiple nuclei in a cell could impact the DNA repair process remains unknown.

Chromatin status and nuclear compartmentalization

Local chromatin (i.e., DNA-RNA-protein interactions) can influence DNA DSB repair outcomes (Clouaire and Legube 2019; Scully *et al.* 2019) (Figure 1.2B). A majority of studies to-date have focused on how chromatin impacts C-NHEJ versus HR, but this could be caused more by experimental bias than the importance of chromatin influencing all types of DNA DSB repair pathway preference. Methylation at histone 3 lysine 36 (H3K36me) has been extensively studied in animals and yeast for its role in DNA repair (Sun *et al.* 2020). For example, di-methylation at histone 3 lysine 36 (H3K36me₂) catalyzed by a SET-domain containing methyltransferase (i.e., Metnase), localizes to induced DSB sites and promotes C-NHEJ repair through recruitment of Ku70 in human cell line experiments (Fnu *et al.* 2011). Post-translational modification of H3K36 appears to also influence DNA repair in fission yeast *Schizosaccharomyces pombe*, where data show antagonism between Set2-mediated H3K36 methylation promoting C-NHEJ, and Gcn5-mediated H3K36 acetylation promoting HR (Pai *et al.* 2014). However, in human cell lines, Setd2-dependent H3K36me₃ can recruit the nuclease CtIP, through the H3K36me₃ reader protein LEDGF, to promote end resection and HR (Pfister *et al.* 2014). Biological variation, such as differences in histone modifications or their reader proteins (i.e., epigenome), may influence organism specific DSB preference, such as seen between humans and yeast (Pai *et al.* 2014; Pfister

et al. 2014). For instance, the difference for H3K36me3 being associated with C-NHEJ in yeast, but with HR in humans, may be attributed to yeast lacking a LEDGF homolog. In addition to H3K36 modifications, other histone modifications, such as H3K4me2, H3K9me3, H3K79me2, H4K16ac and H4K20me0/2, can affect DSB repair choice (Alagoz *et al.* 2015; Clouaire and Legube 2019; Clouaire *et al.* 2018; Horikoshi *et al.* 2019; Pellegrino *et al.* 2017; Saredi *et al.* 2016). For example, structural and biochemical studies found that new histones deposited during DNA replication have uniquely unmethylated lysine residues (H4K20me0), while old histones are mainly methylated (H4K20me1/2/3) (Saredi *et al.* 2016). The post-replicative chromatin accompanying H4K20me0 can promote HR repair through interaction with the HR related protein complex TONSL-MMS22L. However, mono- or di- methylation at H4K20 are not compatible with TONSL binding (Saredi *et al.* 2016). Additionally, dilution of H4K20me2 after replication affects 53BP1 binding, which helps limit end resection, and thus promotes HR repair (Pellegrino *et al.* 2017). Moreover, genome-wide mapping of histone modifications via ChIP-seq in human cell lines revealed the association between the enhanced levels of multiple histone modifications (e.g., H3K36me3, H3K79me2 and H3K4me2) and potential HR repaired regions. These transcription activation markers might provide a preferred chromatin status for HR machinery (Clouaire *et al.* 2018; Her and Bunting 2018). This is consistent with other mechanistic data, such as the bromodomain containing protein, ZMYND8 identified in human cell lines, that interacts with acetylated histones at transcriptionally active chromatin, and recruits the nucleosome remodeling and histone deacetylase (NuRD) complex to promote HR repair (Gong *et al.* 2015). In *Arabidopsis*, a cohesion accessory protein, PDS5C (also known as RDM15), interacts with H3K4me1 at actively transcribed chromatin, and the complex is important for HR DSB repair (Niu *et al.* 2021; Pradillo *et al.* 2015; Quiroz *et al.* 2022). In addition to euchromatin, transient formation

of H3K9me3-mediated heterochromatin can accumulate at DNA DSB sites, which is thought to stabilize the damaged chromatin and activate further repair via regulating the level of acetyltransferase Tip60 and ATM kinase (Ayrapetov *et al.* 2014). Collectively, these observations show that multiple chromatin factors related to accessibility and transcriptional status influence the choice between C-NHEJ and HR repair (Aymard *et al.* 2014; Jha and Strahl 2014; Pai *et al.* 2014).

Compared to chromatin effects on C-NHEJ and HR, much less is known regarding the influence on TMEJ/MMEJ and SSA. A short reporter sequence inserted at over 1,000 genomic loci in human cell was used to track DNA repair following Cas9 induced DSBs, and found that C-NHEJ is generally biased toward euchromatin, while TMEJ is more probable in H3K27me3 associated heterochromatin. Inhibition of H3K27me3 methyltransferase EZH2, a component of the PRC2 complex, shifts the repair preference toward C-NHEJ for previously TMEJ preferred sites (Schep *et al.* 2021). A separate study showed that chromatin decompaction followed by hypotonic stress leads to enhanced SSA and repressed HR in DSB repair (Krieger *et al.* 2021).

In addition to chemical modifications at histone residues, the physical nuclear location of DNA (i.e., nuclear compartments) appears to affect DNA DSB repair choice (Figure 1.2C). Nuclear compartments are membraneless boundaries that have been documented based on morphological and functional studies, and include multiple regions occurring in the interior, such as the nucleolus, nuclear speckles, Cajal bodies, PML bodies, and regions on the nuclear periphery, such as nuclear lamina, nuclear envelope, and nuclear pores (Belmont 2021). In *S. cerevisiae*, DSB sites occurring on DNA located in internal nuclear regions can migrate to nuclear pore complexes for repair (Nagai *et al.* 2008). Dual mutations of Nup84, a nuclear pore complex, and Rad27, a HR related protein, results in a synthetic lethal phenotype (Loeillet *et al.* 2005). These observations

suggest that nuclear compartments and nuclear positioning contribute to DNA repair (Lamm *et al.* 2021; Loeillet *et al.* 2005; Nagai *et al.* 2008). Concordantly, a pioneering study in *Drosophila* found that DSBs in heterochromatic DNA relocate to the nuclear periphery, mediated by SUMOylation and other protein components. This relocation avoids ectopic recombination at heterochromatic DSBs composed largely of repeated sequences, and promotes higher fidelity HR repair (Ryu *et al.* 2015). Interestingly, taking advantage of human cell lines, Lemaître and colleagues found that DSBs occurring in DNA at the nuclear lamina did not relocate and remained positionally stable. This result reported that repair of DSBs at nuclear lamina with compact heterochromatin were impaired in HR while C-NHEJ and TMEJ were activated (Lemaitre *et al.* 2014). A distinct DNA repair pattern has also been reported to occur for centromeric and pericentric heterochromatin in mouse cells, indicating that repair of pericentric heterochromatin favors C-NHEJ repair within the nuclear core in the G1 phase, and relocates to the nuclear periphery for HR repair in S/G2 phase, while DSBs arising at centromeric heterochromatin are located in the nuclear periphery and recruit C-NHEJ or HR factors in a cell cycle independent manner (Tsouroula *et al.* 2016). These data provide a clue for an additional layer of information and regulation of DNA repair in eukaryotes.

ChIP-seq and ATAC-seq enable genome-wide mapping of histone modifications and chromatin accessibility, and have been applied in multiple filamentous pathogens (Connolly *et al.* 2013; Cook *et al.* 2020; Fang *et al.* 2020; Moller *et al.* 2019; Zhang *et al.* 2021). Most work in filamentous pathogens has focused on gene regulation or genome structure, not DNA repair. Interestingly, evidence from short-term lab evolution assays in the wheat pathogen *Z. tritici*, found that H3K27me3 is associated with certain unstable genomic regions and increased mutation rates, while H3K9me3 stabilizes the genome (Habig *et al.* 2021; Moller *et al.* 2019). Alterations in DNA

repair mediated by a specific histone code is one hypothesis to explain these observations (Habig *et al.* 2021). Interestingly, many gene deletion assays using either traditional HR strategies or CRISPR-mediated editing have reported variation in gene deletion efficiency among the different loci in filamentous pathogens (Ah-Fong *et al.* 2021; Villalba *et al.* 2008). The influence of chromatin in determining these results offers an attractive hypothesis, but future studies are needed to understand the role of chromatin features in DSB repair choice and editing efficiency, especially in filamentous microbes.

Sequence content and physical chromosome location

The underlying DNA sequence and physical location along the chromosome can influence the outcomes of DSB repair (Figure 1.2D and E). For example, the resulting mutations from Cas9 editing and repair, namely insertions and deletions, can be precisely predicted based on the fourth nucleotide upstream of the PAM sequence based on experiments in human cells (Chakrabarti *et al.* 2019). In support of this, target flanking sequence dependent repair was found in a large-scale investigation of Cas9 induced DNA DSB, using more than 40,000 guide RNAs tested in multiple genetics backgrounds (Allen *et al.* 2018). The authors found that editing outcomes can be accurately predicted through local DNA sequence alone (Allen *et al.* 2018). Additionally, mismatch, GC-content and length within microhomologous sequence influences the efficiency of TMEJ/MMEJ, probably by affecting the stability of annealed intermediate (Daley and Wilson 2005; Kent *et al.* 2015; Lee *et al.* 2019; Meyer *et al.* 2015; Villarreal *et al.* 2012). It is reasonable to speculate that genomic regions rich in repetitive sequences, that can serve as homologous templates during HR, TMEJ/MMEJ and SSA, might be involved in these types of repair more frequently than other regions lacking high proximal-homology. Consistently, HR repair at repetitive sequences has been identified as the source of genomic variation (Argueso *et al.* 2008).

Similarly, we have observed frequent DNA deletions and insertions following CRISPR induced DSBs at a repeat rich locus in *Magnaporthe oryzae* (Huang *et al.* 2021).

Genome-wide targeting in mouse cells of ~1,000 integrated reporter genes in different locations revealed that genomic location contributed to variation in mutation frequency owing to unknown factors (Gisler *et al.* 2019). A series of studies in mammalian cells reveal that telomeric DSB results in more frequent large deletion and chromosome rearrangements owing to Mre11-mediated 3' to 5' exonuclease processing during HR and TMEJ accompanying with reduced C-NHEJ in telomere, compared to interstitial DSB. Interestingly, the reduced rate of C-NHEJ was also observed in interstitial DSBs with adjacent telomeric repeat sequences (Miller *et al.* 2011; Muraki *et al.* 2015; Zschenker *et al.* 2009). Investigation on yeast chromosome XI showed that the rate of C-NHEJ decreased from the internal chromosome to the telomere, while other repair pathways (e.g., microhomology mediated BIR) were involved in sub-telomeric repair (Ricchetti *et al.* 2003). Difference in DNA sequences might also contribute to DNA repair variation between centromeric and pericentric heterochromatin (Tsouroula *et al.* 2016). In filamentous pathogens, such repair preference may explain observations such as the variable stability of Avr proteins. For example, the avirulence gene *AvrPi9* and corresponding resistance gene *Pi9* from the rice blast pathosystem, has provided durable resistance in commercial rice production that may have to do with *AvrPi9* stability (Qu *et al.* 2006; Wu *et al.* 2015). Interestingly, this effector is located near the centromere in *M. oryzae*, and it is interesting to speculate that this chromosomal location impacts its variability in the population (Wu *et al.* 2015). While in the human pathogen *Cryptococcus neoformans*, C-NHEJ and HR both appear to be involved in repairing DSBs at repeated centromere sequences (Yadav *et al.* 2020). Additional efforts are needed to further

understand the role of chromosome location in affecting the choice and outcome of DNA DSB repair in filamentous pathogens.

3. The influence of DNA repair pathway usage on the full spectrum of CRISPR genome editing mutations

Following the first *in vivo* genome editing using CRISPR-Cas9 in human cell lines, CRISPR-Cas DNA editing systems have revolutionized genetics and functional genomics (Cong *et al.* 2013; Hsu *et al.* 2014; Mali *et al.* 2013). The majority of CRISPR-Cas systems used for genome engineering rely on single effector Type-II Cas proteins with intrinsic nuclease activity. Therefore, following the generation of Cas-mediated DNA DSBs, the endogenous DNA repair machinery is activated for repair. Recent reviews have detailed how DNA repair pathways affect genome editing outcomes in model systems (Nambiar *et al.* 2022; Xue and Greene 2021), but our knowledge remains limited for filamentous pathogens. More recent efforts for genome engineering have developed modified CRISPR platforms that do not create DSBs, such as PRIME-editing, CAST and Drag-and-drop approaches (Anzalone *et al.* 2019; Ioannidi *et al.* 2021; Strecker *et al.* 2019). The development of these modified systems is in part a response to observations that CRISPR-Cas genome engineering can create substantial unintended DNA modifications, likely owing to the creation of DNA DSBs. Such observations have focused on so-called off-target mutations, where a secondary locus to the CRISPR targeted locus has an induced mutation. However, CRISPR-Cas systems that create DSBs also frequently cause unexpected and aberrant outcomes at the primary, CRISPR targeted locus (Kosicki *et al.* 2018; Lee and Kim 2018). These on-target but unintended mutations are likely the results of the interplay of endogenous DNA repair pathways, unknown genomic rules, and physiological conditions during editing. In this section,

we focus on reported low-fidelity and aberrant CRISPR on-target editing outcomes in model systems and filamentous pathogens and the potential link to DNA DSB repair pathway function.

Genome editing in animals

In the absence of homologous DNA template for DSB repair, the majority of CRISPR-Cas endonuclease editing (e.g., Cas9 and Cas12a) results in small insertions or deletions (INDELs) at the target locus, thought to be mediated by C-NHEJ (Figure 1.3A). The goal of most projects employing this approach is the creation of null-mutants, identified using amplification and sequencing based techniques to assess the target locus. Projects relying on amplicon sequencing have a limited ability to capture only certain DNA mutations at the target locus, as any mutation that alters primer binding or creates a sequence that is unamenable to amplification (e.g., large DNA insertion) will not be identified. This creates ascertainment bias and such analysis underestimates large-scale, complicated, error-prone DNA repair outcomes. Research specifically interested in identifying the broader class of DNA mutations following Cas-based editing have employed third generation long fragment DNA sequencing and more precise analysis, allowing for the characterization of previously hidden DNA mutations (Adikusuma *et al.* 2018; Alanis-Lobato *et al.* 2021; Cullot *et al.* 2019; Kosicki *et al.* 2018; Skryabin *et al.* 2020). For example, using PacBio long-fragment DNA sequencing, several classes of complex genome mutations were characterized at the primary on-target Cas9 editing locus in a mouse cell line. This systematic analysis of individual cell clones revealed that large deletions (up to 9.5 kb), large insertions (up to 2.5 kb), deletions plus large insertions, inversions, and INDEL mutations occurred at the on-target locus (Kosicki *et al.* 2018). Such mutations would have been missed with simple amplicon-based analysis (Kosicki *et al.* 2018). In another study using microinjected mouse embryos, frequent large DNA deletions, ranging from 100 bp to 3.2 kb were found at multiple Cas9 targeted

loci. In addition, approximately 23% of sequencing reads suggested that exogenously supplied Cas9 and gRNA plasmid DNA were inserted at the target locus (Adikusuma *et al.* 2018). Similarly, unexpected segmental gains and losses near the Cas9 editing locus, and even the loss of an entire chromosome, were observed in experiments editing human embryos (Alanis-Lobato *et al.* 2021; Zuccaro *et al.* 2020). A key observation resulting from Cas9-mediated large deletions in mouse cell lines is the significant overrepresentation of microhomology (2 to 5 bp) observed at the break junction sites (Owens *et al.* 2019). This led to the hypothesis that TMEJ is responsible for creating the observed large deletions, but direct experimental evidence needed to rule out the involvement of other pathways was not provided (Owens *et al.* 2019). A recent approach more directly determines the involvement of specific repair pathways mediating Cas9 induced large deletions in mice cell lines by conducting editing in lines lacking 32 DNA repair genes (Kosicki *et al.* 2022). The authors reported an increased frequency of Cas9 mediated large deletions when editing in mice knock-out lines lacking major C-NHEJ components, including Ku70-Ku80, Lig4, and Xrcc4, while there was not an increase in large deletion outcomes for other minor C-NHEJ factors (e.g., Artemis). In addition, a reduced frequency of large deletions was detected in knock-out lines for genes involved in end resection and TMEJ, such as Nbs1 (yeast Xrs2 homolog) and the core TMEJ polymerase Pol Θ (Kosicki *et al.* 2022). A new analysis pipeline, termed PEM-Q, found both an increased occurrence of large deletions and large DNA insertions (i.e., >20 bp) following Cas9-mediated editing in mouse cell lines with mutated Ku80 or Lig4 (Liu *et al.* 2021). Collectively, these results demonstrate frequent on-target mutations during genome editing in animal systems that are more complex than small INDELs and cause substantial DNA sequence changes. There is ample evidence of the underlying competition between DNA DSB repair pathways, especially between C-NHEJ and TMEJ, where the profile of the observed mutation outcomes changes when

one pathway is inhibited. Interestingly, in mouse embryonic stem cells, Cas9 mediated INDELs created through C-NHEJ repair caused tandem duplications (e.g., short insertions creating repeat ‘ATT/ATT’), while TMEJ mediated repair mainly contributed to templated insertions (e.g., inserted sequence that is not contiguous, but in proximity ‘ATT(X)ATT’) (Schimmel *et al.* 2017). These results suggests that in addition to molecular competition between the pathways, there is functional variation for the resulting mutation profile dependent on the DNA repair pathways employed by the cell.

Genome editing in microbes

While the CRISPR-Cas system has been applied to more than 40 species of ecologically diverse filamentous fungi and oomycetes, there has been little systematic analysis of the full spectrum of on-target DNA mutations (Liu *et al.* 2015; Nodvig *et al.* 2015; Schuster and Kahmann 2019). This is not surprising given that most of these studies used CRISPR-Cas to generate gene mutations for further functional characterization, and therefore did not catalog or explore specific DNA mutations or the involvement of different DNA repair pathways (Schuster and Kahmann 2019; Wang and Coleman 2019). Despite the lack of systematic analysis, unexpected on-target DNA mutations following CRISPR-Cas editing in filamentous fungi have been reported. In the human pathogen *A. fumigatus*, frequent insertions of linearized plasmid DNA (up to 6.5 kb) were found at the Cas9 targeting region (Fuller *et al.* 2015). In the plant pathogenic fungus, *Sclerotinia sclerotiorum*, amplification-based genotyping unexpectedly found that 100% of mutant phenotype transformants were PCR-negative using a primer pair amplifying across the Cas9 targeting site. Further TAIL-PCR and short-read sequencing revealed all of the transformants contained large plasmid DNA insertions, that importantly, lacked long stretches of homologous sequence to the genome (Li *et al.* 2018). Similar observations, reporting unexpected plasmid or foreigner DNA

integration at Cas9 or Cas12a targeting sites, have been reported in other filamentous microbes, including *T. reesei*, *Phytophthora infestans*, *Nodulisporium* sp and encapsulated yeast *C. neoformans* (Ah-Fong *et al.* 2021; Hao and Su 2019; Yadav *et al.* 2020; Zheng *et al.* 2017) (Figure 1.3B). In fission yeast, insertion of carrier DNA from chum salmon (*Oncorhynchus keta*) has been found to integrate at the Cas9 cleavage site (Longmuir *et al.* 2019). There has been postulation that the observed large DNA integrations resulted from C-NHEJ, but experimental evidence was not provided (Fuller *et al.* 2015; Li *et al.* 2018; Zheng *et al.* 2017). In addition to foreign DNA integration, Cas9 gene editing in *T. reesei* resulted in tandem duplications at the targeted locus, which is a type of DNA repair profile that has been linked to C-NHEJ (Hao and Su 2019; Schimmel *et al.* 2017). Our lab recently reported detailed characterization of DNA repair events following Cas12a ribonucleoprotein (RNP) editing in the fungal plant pathogen, *M. oryzae*, that causes blast disease on rice and other monocots (Peng *et al.* 2019; Valent 2021; Zhong *et al.* 2016). This analysis revealed large donor DNA insertions as concatemer fragments of up to ~20 kb at the Cas12a targeted site (Huang and Cook 2022b; Huang *et al.* 2021). In addition to large insertions, we identified four other DNA repair outcomes in the presence of DNA donor- (i) simple insertion; (ii) large deletions; (iii) deletion plus insertions; and (iv) INDELs (Figure 1.3B, 3C and 3D). A substantial number of independent transformants were analyzed using Sanger sequencing and long-read sequencing and assembly, and the spectrum of observed DNA mutations suggested that multiple DNA repair pathways, including C-NHEJ, MMEJ and SSA were involved (Huang *et al.* 2021). For example, multiple large deletion mutants at one edited locus were found to results from the combination of repetitive flanking DNA into a single repeat, which is thought to be a common outcome of SSA (Bhargava *et al.* 2016; Huang *et al.* 2021). For other transformants, the integration junctions of simple and large insertion mutants often contained microhomology between the

insertion and genomic sequence, indicating the involvement of microhomology during DNA repair. The use of microhomology (~ 50 bp) for precise integration of exogenous templates during Cas-induced DSB repair has been reported in other fungal species (Al Abdallah *et al.* 2017; Foster *et al.* 2018; Lax *et al.* 2021; Leisen *et al.* 2020; Zhang *et al.* 2016a). The importance of homologous sequence for DSB repair also extends to a protozoan parasite that frequently underwent large deletions between two long-homology sequences during DSB repair (Zhang and Matlashewski 2019). As *Leishmania* lacks the known ligase involved in C-NHEJ (i.e., Lig4), the authors indicate that SSA is that most likely pathway responsible for generating these mutations (Zhang and Matlashewski 2019).

Prior to the CRISPR era, induced DNA repair was shown to be connected to specific repair pathways. A strategy termed Restriction Enzyme Mediated Integration (REMI) created DNA DSBs using restriction enzymes (Kahmann and Basse 1999). Using this approach, co-transferred DNA donor could be integrated at induced DSB sites, increasing transformation efficiency and the ratio of single-copy events (Kahmann and Basse 1999). Further genetic study revealed that Rad50, the core component of the MRX complex, was required for REMI in budding yeast, but Rad51 and Rad52 were dispensable (Schiestl *et al.* 1994). These results indicate that REMI might require the normal functioning MRX complex, which we can now infer is presumably dependent on MMEJ repair. It is also clear that the rate of HR mediated DNA DSB repair is not the same across species of filamentous pathogens. For instance, HR frequently results in highly efficient gene deletions in *Fusarium graminearum*, the causal agent of Fusarium head blight, while HR is infrequently observed in the maize pathogen *F. verticillioides*. (Choi and Shim 2008; Wang *et al.* 2011). To artificially increase the chances of HR as a means to improve functional genomics, mutant strains have been created that lack C-NHEJ (e.g., Ku70 or Ku80), with the rationale that if

C-NHEJ is blocked then HR frequency will increase (Ninomiya *et al.* 2004). This method was first applied in *N. crassa*, and subsequently used in other filamentous pathogens, such as plant pathogens, *M. oryzae*, *Botrytis cinerea* and animal pathogens, *A. fumigatus*, and *Trichophyton mentagrophytes* (Choquer *et al.* 2008; da Silva Ferreira *et al.* 2006; Ninomiya *et al.* 2004; Villalba *et al.* 2008; Yamada *et al.* 2009). This approach has proven successful at increasing the rate of donor integrations, with one report showing inhibition of C-NHEJ repair led to 100% of fungal transformants containing the desired HR integrations (Ninomiya *et al.* 2004).

The described results establish that DNA editing in fungi and oomycetes, similar to animals, is tightly regulated by DSB repair pathways that interact, and the pathways have different propensities for on-site DNA mutations following induced DSBs. Future research is needed to determine the molecular details of the DNA DSB repair pathways in filamentous fungi, including their hierarchy and regulation, and ultimately the functional DNA mutation outcomes resulting from the different pathways. This can help resolve inconsistencies in the literature, such as Cas-based DNA integration using short homologous sequence being referred to as both MMEJ repair (Al Abdallah *et al.* 2017; Foster *et al.* 2018; Zhang *et al.* 2016a) and HR (Lax *et al.* 2021). Further efforts are needed to test if knowledge from yeast are relevant in filamentous microbes, such as the involvement of two DNA polymerases, Pol δ and Pol λ , in MMEJ (Meyer *et al.* 2015). Additional areas of interest include understanding why there are two polymerases involved in MMEJ in yeast, versus one for TMEJ in animals and plants and the functional impact. Also, given the important role of Lig3 and PARP1 for TMEJ, do fungi contain analogous components during MMEJ given these two proteins are absent in fungi. Answering these questions will provide a deeper evolutionary and mechanistic understanding of DNA repair across eukaryotes.

4. The contribution of DNA repair pathways to genome evolution in filamentous pathogens-the two-speed genome observation

Filamentous plant pathogens must overcome host immunity to cause disease and reproduce. We note that other plant-associated microbes, such as mutualists and endophytes, also counter the plant immune response in a similar manner to pathogens, but our review is pathogen-centric (Brader *et al.* 2017; Snelders *et al.* 2022). The plant innate immune system is largely characterized by receptor proteins that detect molecular signatures of the invading pathogen or the infection process through varied mechanisms (Cook *et al.* 2015). Plant immune receptor proteins either reside in the plasma membrane, typified by receptor-like kinases and receptor-like proteins (DeFalco and Zipfel 2021), or they are located in the cytoplasm, characterized as nucleotide-binding leucine-rich repeat/NOD-like receptor (NLR) proteins (Saur *et al.* 2021). While mechanistic details vary across pathosystems, a key paradigm is that successful pathogens employ a range of secreted proteins and small molecules, termed effectors, that suppress the plant immune system, modulate plant physiology, and protect the invading microbe (Oliveira-Garcia and Valent 2015; Snelders *et al.* 2022). Thus, the evolutionary stage is set for plant-microbe co-evolution, where selection for immune receptor recognition of pathogen effectors and signatures of infection imposes selection for receptor evasion or effector based immune suppression. Plant-microbe interactions can promote adaptive genome evolution, whereby the continued engagement of specific plant-microbe species produces refined mechanisms for both resistance and pathogenicity (Moller and Stukenbrock 2017; Sanchez-Vallet *et al.* 2018).

To better understand pathobiology and genome evolution, significant effort has been made to sequence and assemble filamentous pathogen genomes, including whole-genome sequencing of the rice blast causing fungus *M. oryzae* in 2005, and a number of other important model pathogens

obtained through increasingly improved technology (de Jonge *et al.* 2013; Dean *et al.* 2005; Haas *et al.* 2009; Kamper *et al.* 2006; Ma *et al.* 2010; Peng *et al.* 2019; Tyler *et al.* 2006). Early analysis of the *P. infestans* genome led to a novel and insightful observation, whereby the authors described the occurrence of gene-sparse regions that contain protein coding sequences with elevated levels of presence/absence variation, directional selection, and other features of adaptation to host infection (Raffaele *et al.* 2010). The authors named this phenomenon the ‘two-speed genome’ configuration, and numerous subsequent genomic studies of filamentous pathogens have described similar observations (Dong *et al.* 2015a; Faino *et al.* 2016). The exact definition of the two-speed genome is vague, but is often characterized as the occurrence of (i) slow-evolving genomic regions coding for proteins with essential functions and sparse transposable element density, and (ii) fast-evolving regions coding for proteins associated with environmental or adaptive functions (e.g., host infection) and dense transposable element sequences. It is important to acknowledge that while the two-speed genome concept is an attractive generalization, it appears that some fungal phytopathogens do not display strong signals of genome compartmentalization (Frantzeskakis *et al.* 2018; Stam *et al.* 2018). As with many black and white dichotomies of biological systems, fungal pathogen genomes likely display a continuum of organizations with respect to gene and TE densities and the presence of variation (Frantzeskakis *et al.* 2019; Torres *et al.* 2020).

Refining the description of the two-speed genome and associated characteristics

One aspect that needs further consideration is what exactly constitutes the two-speed genome. The two-speed genome classification has involved different metrics across filamentous pathogens. One common description has been the equation of two-speeds with the occurrence of lineage-specific (LS) and core genomic regions. Here, LS regions are defined by the

presence/absence of DNA sequences between different strains of often the same species. LS regions identified from the blast fungus *M. oryzae*, and vascular wilt pathogen *Verticillium dahliae* are enriched in repetitive elements and *in planta* expressed genes (Bao *et al.* 2017; de Jonge *et al.* 2013). Subsequent study in *V. dahliae* refined the description of LS regions to include features of chromatin and transcription, and found these updated genomic regions better correlated with characteristics of host infection (Cook *et al.* 2020). The genomic regions appear enriched for coding sequences that are not essential but involved in specific life-history events, and therefore were renamed as Adaptive Genomic Regions (Cook *et al.* 2020). Another description of the two-speed genome is based on the occurrence of AT-rich blocks, called AT-rich isochores, defined by base pair composition divergence at distinct regions compared to genome or species wide averages. In the phytopathogenic fungus, *Leptosphaeria maculans*, AT-rich isochores contain ~33% GC content and are over-represented for sequences encoding effectors (Rouxel *et al.* 2011). Another important feature of filamentous pathogen genomes that can contribute to a two-speed genome is the occurrence of dispensable chromosomes, also called mini-chromosomes, supernumerary chromosomes, accessory chromosomes or A/B chromosomes (Coleman *et al.* 2009; Galazka and Freitag 2014; Peng *et al.* 2019). The occurrence of dispensable chromosomes has been documented across diverse plant, animal, and microbial systems, sometimes without a clear phenotypic or evolutionary impact (D'Ambrosio *et al.* 2017; Peng *et al.* 2019; Soyer *et al.* 2018). In fungal pathogens, seminal work and subsequent experimentation has clearly demonstrated the importance of dispensable chromosomes in the evolution of virulence of *Fusarium* species (Li *et al.* 2020; Ma *et al.* 2010). There is also strong evidence that dispensable chromosomes contribute to virulence and evolution in the blast fungus *M. oryzae* (referred to at present as mini-chromosomes in *M. oryzae*) (Chuma *et al.* 2011; Langner *et al.* 2021; Liu *et al.* 2022; Peng *et al.* 2019). Indeed,

previous consideration of dispensable chromosomes detailed how these regions offer fungal pathogens a genomic ‘cradle’ possibly accelerating pathogen evolution (Croll and McDonald 2012). Quantifying rates of adaptive evolution is challenging, but it is clear that dispensable chromosomes can be quite dynamic, even under axenic lab growth (Moller *et al.* 2018).

These various descriptions of the two-speed genome have not used a common vernacular or criteria, instead the characterization has been based on sequence conservation, content, or physical arrangement, while the original categorization in *P. infestans* was based on gene density (Raffaele *et al.* 2010). What appears common between the descriptions is that many filamentous pathogen genomes have a non-uniform distribution of coding sequences related to host infection, and these coding sequences show elevated rates of sequence variation (e.g., Single-Nucleotide Polymorphisms (SNPs), copy number variation (CNV), presence/absence variation (PAV)) (de Jonge *et al.* 2013; Raffaele *et al.* 2010; Rouxel *et al.* 2011). More simply, many fungal pathogen genomes have segments enriched for non-essential coding sequences that are highly variable. We suggest that while individual experimental details or analysis may have emphasized the occurrence of gene-sparse, AT-rich, LS regions or dispensable-chromosomes, the unifying theme is segments of heightened variation. This may lead to accessory chromosomes or accessory regions of pathogen genomes in a population (Moller and Stukenbrock 2017). There are likely biological differences between filamentous fungal pathogens, such as reproductive mode, ploidy, genome size, and repeat content that influence the exact manifestation of a compartmentalized genome. Further research and theory are needed to understand if differences in genome compartmentalization, such as gene-sparse, isochores, LS regions or dispensable-chromosomes, differentially impact fungal fitness and evolutionary potential. Alternatively, the different genome configurations may reflect that there are multiple routes to answering evolution's challenge of

maintaining fitness in a changing environment. Under the idea of the latter, dispensable chromosomes are a more extreme example of compartmentalized (i.e., non-uniform) genome organization compared to LS regions for example, but they both ultimately represent regions of the genome with a different evolutionary trajectory compared to regions outside the dispensable or LS regions. Interestingly, these observations are not restricted to plant filamentous pathogens, as there are recent reports of the similar bipartite genome composition in the animal pathogens *Batrachochytrium dendrobatidis* (Bd) and *B. salamandrivorans* (Wacker *et al.* 2021) and the model arbuscular mycorrhizal fungus *Rhizophagus irregularis* (Yildirim *et al.* 2022).

A unifying cause for the two-speed genome

This leads to the bigger question of what accounts for the formation and maintenance of gene-sparse, LS or dispensable chromosomes? Are these configurations driven only by selection, or are there additional intrinsic genomic mechanisms that contribute to this configuration? This is an important question for the two-speed genome observation, which currently lacks systematic understanding and clear testable hypotheses ultimately needed to advance our understanding of pathogen genome evolution (Torres *et al.* 2020). We propose two competing hypotheses for the creation and maintenance of the two-speed genome, a passive selection alone model, and an active mechanistic model termed Biased Variation.

Under the selection alone model, DNA variation (e.g., SNPs, INDELs, duplications, inversions) arises at a nearly uniform rate across the genome. Most sequence variation would be evolutionarily neutral, but variation in essential coding sequences with lethal or fitness reducing effects would be selected against and removed from the population (i.e., purifying selection). Variation that arises in regions with less selective constraints, such as regions contributing to specific environmental conditions or non-essential functions, would remain in the population for

a longer period of time, and could eventually increase fitness under specific environmental conditions. This description of conventional evolutionary theory can explain present-day genome configuration whereby essential coding sequences have relatively less sequence variation in a population than regions contributing to non-essential or adaptive functions. This model does not evoke any special status or rules for the two-speed genome, and any apparent correlations seen today are non-causative associations.

A competing hypothesis to the selection alone model attempts to incorporate numerous described associations between genomic features and the two-speed genome. There have been proposed links between the two-speed genome and the epigenome, owing to the unique chromatin features and composition of repetitive elements in dynamics compartments of the genome (Chen *et al.* 2022; Cook *et al.* 2020; Faino *et al.* 2016; Seidl *et al.* 2016; Torres *et al.* 2020), but a single, clear mechanism has remained elusive. We summarize a few associations here and propose the ‘Biased Variation’ model as a means that could create a two-speed genome. An important association is the co-occurrence of TEs and structural variation at distinct regions of the two-speed genome (de Jonge *et al.* 2013; Faino *et al.* 2016; Moller and Stukenbrock 2017). The first sequence and assembly of a *M. oryzae* mini-chromosome identified sequence similarity between the ends of some core-chromosomes and the mini-chromosome (Peng *et al.* 2019). This sequence similarity and enrichment for TEs and copy-number variation may mediate effector re-shuffling, providing an adaptive benefit (Figure 1.4A). This was speculated to occur in *M. oryzae* based on observations of the first cloned avirulence effector from *M. oryzae*, *AVR-Pita* (Orbach *et al.* 2000). Two functional members in *AVR-Pita*, *AVR-Pita1* and *AVR-Pita2*, were found located on different chromosomes across multiple isolates, while the non-functional member *AVR-Pita3* was fixed at one location. Sequence of a retrotransposon element, termed Inago1, was found flanking all tested

AVR-Pita1/2 loci, and thus ectopic recombination mediated by Inago1 was considered a potential mechanism for the frequent translocation of *AVR-Pita1/2* (Chuma *et al.* 2011) (Figure 1.4A). Error-prone DNA DSB repair is also proposed to be involved in the formation of LS regions through transposon mediated genome recombination in *V. dahliae* (Faino *et al.* 2016) (Figure 1.4A). In addition to translocations, transposable elements can influence genome variation through SSA mediated deletion, such as the spontaneous deletion of 44 kb between flanking copies of the transposon *Occan*, which resulted in deletion of three copies of the avirulence effector *AVR-Pia* (Sone *et al.* 2013) (Figure 1.4A). In addition to plant pathogens, copy number variation and loss of heterozygosity were proposed to result from SSA repair using long repeat sequences in the human pathogen *Candida albicans* (Todd *et al.* 2019). As such, there are clear occurrences of DNA repair mediated mechanisms at repetitive sequences resulting in significant genome variation for important loci in filamentous pathogens (Seidl and Thomma 2017). In addition to these observations, a clear association between the epigenome and dispensable chromosome has been found in the wheat pathogen *Z. tritici*. Enriched histone modification of H3K27me3 on dispensable chromosomes is a distinguishing feature when compared to core chromosomes (Schotanus *et al.* 2015). Connecting to the previous sections of this review, the MMEJ repair pathway, which creates more mutagenic outcomes than C-NHEJ and HR, has been reported to be more activated at H3K27me3 marked heterochromatin (Her and Bunting 2018; Schep *et al.* 2021). This would be consistent with the observed increase in genomic instability and increased mutation rate at H3K27me3 regions in *Z. tritici* (Habig *et al.* 2021; Moller *et al.* 2019). Histone methylation H3K36me3 by Ash1 has also been proposed to effect DNA repair and genome stability in the rice pathogen *F. fujikuroi* (Janevska *et al.* 2018). Therefore, it is reasonable to speculate that DNA repair, the epigenome, and chromatin status are actively involved in the formation and maintenance

of the two-speed genome. We propose the ‘Biased Variation’ model, whereby the creation of DNA variation is suppressed or enhanced at specific genomic regions, which is subsequently subject to natural selection.

Biased Variation model: synthesis of the epigenome, chromatin, and DNA repair as a mechanism for biased genome evolution

The ‘Biased Variation’ model proposes that variation for genomic factors influence DNA repair preferences and leads to mutational bias in the genome. That is, genome variation arises at unequal rates prior to selection. DNA mutation does not arise uniformly in the genome, but rather follows local probability functions for the creation of DNA mutation. It is important to note, this model does not suggest that the rise of DNA mutation is influenced by the environment or that it co-occurs with selection. The famous Luria-Delbruck fluctuation test demonstrated that DNA mutation arises pre-selection (Luria and Delbruck 1943). As discussed in this review, features such as base pair sequence content and repetitive DNA density, chemical modifications to DNA and histones, and nuclear DNA arrangement have all been implicated to influence DNA repair. As there are multiple DNA repair pathways that show variation in their mutation profiles, and specific genomic features influence the preference for specific DNA repair pathways, the interplay between DNA repair, the epigenome and chromatin could explain the occurrence of the two-speed genome (Figure 1.4B). As a simple illustration of this model, we focus on only two DNA DSB pathways, C-NHEJ and MMEJ. The C-NHEJ DSB pathway can result in either perfect repair, resulting in no detectable DNA mutation (i.e., error-free), or imperfect repair (i.e., mutation), resulting in altered sequence (i.e., INDEL) (Figure 1.4B). Experimental evidence supports that high-fidelity C-NHEJ is possible, but to our knowledge there is not a clear estimate of this rate, certainly not in filamentous fungi (Betermier *et al.* 2014). We therefore use 50% chance for either perfect or

imperfect C-NHEJ repair as an approximation. For DNA DSBs repaired by MMEJ, all resulting loci will have a DNA mutation due to end resection and flap trimming. There is clear evidence for DNA DSB pathway preference, but the exact quantification is currently not clear. We therefore assume a 5% deviation from equal repair pathway preference at certain regions of the genome for MMEJ over C-NHEJ, 55% versus 45% respectively (Figure 1.4B). This simple probability difference in DNA repair pathway preference, coupled with the probability to create a DNA mutation, results in an estimate that genome regions preferentially repaired by MMEJ will experience 5% more DNA mutations (Figure 1.4B). While a 5% increase in DNA mutation may seem modest, at least two factors in filamentous pathogen biology make this increase substantial. One fact is that fungi do not have a described separation between germline and somatic tissue. Theoretically, any somatic tissue can give rise to asexually produced spores, and therefore any somatic mutations could be heritable. This significantly increases the number of cells that could undergo a heritable mutation and therefore increases their frequency. Additionally, experimental evidence in animal models show that germline cells have lower rates of mutation compared to somatic tissue (Milholland *et al.* 2017). It is interesting to speculate that fungi may have comparatively elevated genomic instability compared to metazoans for instance because they lack germline separation. The second related piece is the incredible volume of progeny that microbes, including filamentous pathogens, release with respect to both absolute time and growth normalized time. The number of progeny, and therefore the number of individuals effected by Biased Variation, are orders of magnitude greater in microbes than animals or plants. As such, the biology of filamentous pathogens could result in Biased Variation having an increased impact on their genome evolution compared to organisms in other domains of life.

Our simple example of Biased Variation using differences for C-NHEJ and MMEJ repair preference and mutation rate highlight the potential impact on genome evolution. This example ignores many other variables, such as the impact of other DNA repair pathways and genomic variables that influence their activity and mutational profiles. Future research is needed to unpack these associations and generate estimates for their interactions. Mutation accumulation experiments in both a fungal plant pathogen and the model dicot *Arabidopsis*, showed a clear association between the observation of accumulated DNA variation and components of the epigenome (Habig *et al.* 2021; Monroe *et al.* 2022). Re-analysis of fast neutron induced mutations in a population of rice (Li *et al.* 2017) identified a significant reduction in mutation probability for genic and non-genic regions associated in or around H3K4me1 domains (Quiroz *et al.* 2022). Such findings are consistent with the Biased Variation model, whereby the histone modification H3K4me1 could influence the propensity of HR DSB repair in marked regions, providing higher fidelity repair. It is interesting to note that research on the topic of biased genome variation (i.e., intragenomic mutation variability) in filamentous pathogens has tended to focus on regions and causes of hypervariable genomic regions (Coleman *et al.* 2009; Cook *et al.* 2020; de Jonge *et al.* 2013; Dutheil *et al.* 2016; Raffaele *et al.* 2010), while recent research in plants has described the occurrence of hypovariable genomic regions (Monroe *et al.* 2022; Quiroz *et al.* 2022). It will be interesting to bridge these results, and investigate the occurrence of such mechanisms across domains of life, to understand how DNA repair hierarchy and regulation contribute to hyper- and hypovariable genomic compartments, and ultimately how such processes impact genome evolution.

In addition to genomic factors that influence DNA DSB repair preference, the occurrence of DNA DSBs themselves are likely not uniform across the genome. Genomic factors such as

repetitive sequence density, transcriptional activity, and DNA replication timing interact and influence the creation of DNA DSBs (Gadgil *et al.* 2017; Hamperl and Cimprich 2016; Sebastian and Oberdoerffer 2017). This would lead to regions of the genome being more prone to DNA DSBs, and therefore more impacted by Biased Variation for DNA repair. This would be consistent with the fact that multiple characterizations of faster evolving sections of filamentous pathogen genomes involve regions with higher TE densities (Faino *et al.* 2016; Moller and Stukenbrock 2017). In addition to TEs serving as homologous regions to anchor DSB repair, TEs can also impact the creation of DNA variation through active transposition that may further contribute to Biased Variation (Fouche *et al.* 2022).

Another factor we speculate will have importance in Biased Variation is the creation and function of extrachromosomal circular DNAs (eccDNAs). Previous studies in eukaryotes have detailed adaptation to stress through amplifying gene content on eccDNAs. For example, the herbicide target gene coding 5-enolpyruvylshikimate-3-phosphate synthase can be amplified in the form of eccDNAs and heritably transmitted, thus causing herbicide resistance in the agronomically important weed *Amaranthus palmeri* (Koo *et al.* 2018). Interestingly, a recent pre-print using whole-genome sequencing in *M. oryzae* revealed that eccDNA-associated genes have significant longer distance to the nearest gene but shorter distance to the nearest repeats than other genes, which suggests these eccDNA-producing regions are located in fast-evolving genomic compartments (Joubert and Krasileva 2021). Indeed, effector coding sequences are enriched in the production of eccDNAs (Joubert and Krasileva 2021) (Figure 1.4A). Interestingly, a recent study in human cell lines found that the formation of eccDNAs is facilitated by DNA DSB and dependent on the balance between C-NHEJ and the TMEJ pathway (Paulsen *et al.* 2021). As an important downstream component of TMEJ pathway, Lig3 is also found to be required for the generation of

eccDNAs (Wang *et al.* 2021b). Future research determining how mechanisms of DNA repair pathway choice and their resulting mutational profiles are influenced by the epigenome and chromatin will provide a comprehensive understanding of first principles affecting genome function and genome evolution.

Concluding remarks and future perspectives

The repair of DNA DSBs is a well-studied and critical function of eukaryotic genomes. Multiple repair pathways have evolved to fix DNA DSBs, likely highlighting their importance and the need for redundant mechanisms, but the use of multiple systems may also highlight levels of specialization and unique outcomes of the different pathways. Here we have summarized mechanistic details of four DNA DSB repair pathways with an emphasis of what is known or remains unclear in fungal and oomycete pathogens. It is clear that a number of physiological, molecular, and biochemical factors influence DNA DSB repair choice for a given break. The use of CRISPR-Cas, or other genome editing strategies that rely on DNA DSB creation and repair are subject to these rules, and influence resulting outcomes. Naturally occurring DNA DSBs are equally influenced by the interconnection and hierarchy of DNA DSB repair pathways and epigenetics and chromatin states. We describe here a model, termed Biased Variation, and provide a simple probabilistic example demonstrating how genomic factors could influence DNA repair choice and impact the creation of DNA variation. The model is an attempt to understand mechanisms that create unequal DNA variation in the genome. Future research is needed to expand and test the Biased Variation model, or generate other mechanistic models of observed compartmentalized genome evolution. Further developing the understanding of how DNA repair influences genome stability and evolution will impact our ability to engineer genomes and predict evolution related to industrial, clinical, and agricultural aspects of oomycetes and fungi.

Acknowledgements

We would like to thank Dr. Geoffrey P. Morris (Colorado State University) for engaged and lively discussions regarding two-speed genome evolution and associated hypotheses. We thank anonymous reviewers for their helpful comments and suggestions for improvement. We also thank Cook lab members for suggestions and feedback. We apologize to colleagues whose work is not discussed in this review.

Funding

Research to DEC is funded by the United State Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA) awards numbers 2018-67013-28492 and 2021-67013-35724, and the National Science Foundation (NSF) Biological Sciences award numbers 1936800 and 2011500. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Figures

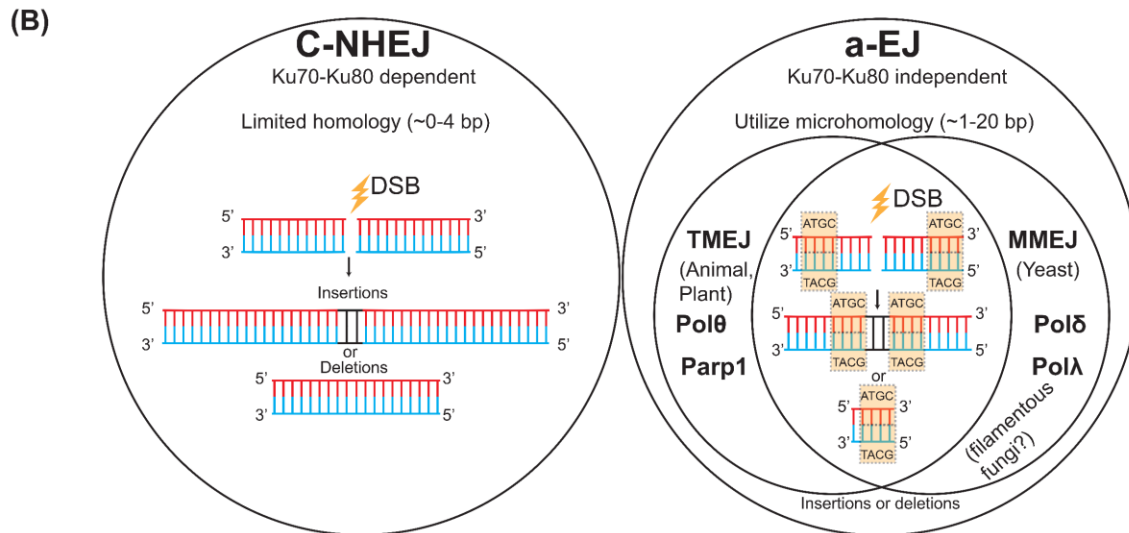
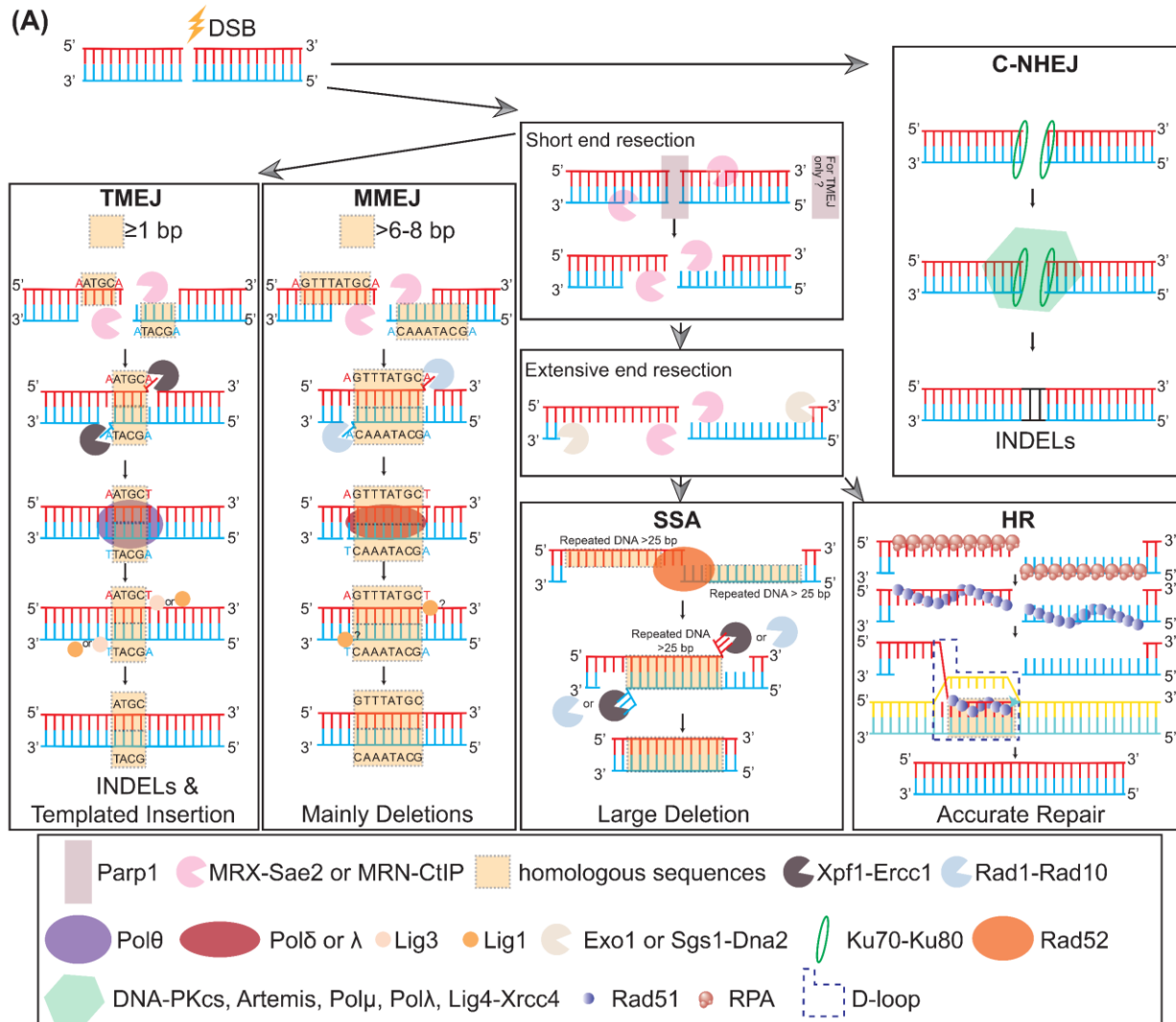


Figure 1.1. Schematic overview of the 4 DNA double-strand break repair pathways and diagram on the terminology for a-EJ, MMEJ and TMEJ

(A) Following the formation of a DNA double strand break (DSB), two routes of repair can occur. If end resection proceeds, one of TMEJ, MMEJ, SSA or HR will predominately repair the DSB, while C-NHEJ is dominant in the absence of end resection. TMEJ or MMEJ proceed after short end resection mediated by MRX-Sae2 (budding yeast) or MRN-CtIP (mammals) that expose external microhomologous (MH) sequences at the DSB ends. Complementary base pairing between MH sequence directs repair. Further 3' flap removal, gap-filling and ligation steps are required for proper TMEJ or MMEJ, while the protein components involved in these steps vary between TMEJ and MMEJ (e.g., Pol Θ vs Pol δ - λ in gap filling and Lig3 vs potentially Lig1 in ligation). The mutational outcome of TMEJ is INDELS and templated insertions, while deletions are more common following MMEJ. For SSA, extensive end resection mediated by Exo1 or Sgs1-Dna2 results in large homologous sequence sites (>25 bp) at the DSB ends, and annealing between homologous sequence is stimulated by Rad52. The 3' flap trimming by Xpf1-Ercc1(mammals) or Rad1-Rad10 (budding yeast) causes large deletions between the homologous sequences, which is a commonly found following SSA repair. For HR, ssDNA overhangs are bound by RPA then replaced with Rad51. Rad51 promotes strand invasion and forms D-loop between homologous sequences and typically results in accurate repair. For C-NHEJ, the Ku70-Ku80 dimer protects the DSB ends from end resection and recruit other key components (including DNA-PKcs, Artemis, Pol μ , Pol λ , Lig4-Xrcc4) for repair. Repair following C-NHEJ often results in INDELS. **(B)** C-NHEJ is defined as Ku70-Ku80 dependent DSB repair pathway, which utilizes minimal homologous sequences for repair and commonly results in INDEL mutations. The term a-EJ is defined as an umbrella term for Ku70-Ku80 independent repair that often utilizes microhomologous sequence for DSB repair. The a-EJ term can be further defined into TMEJ in mammals and plants, involving Pol Θ and PARP1, or MMEJ in yeast and presumably filamentous fungi, which might involve Pol δ and Pol λ . DSB repair by a-EJ should result in a DNA mutation owing to the initial end resection step and later 3' flap removal.

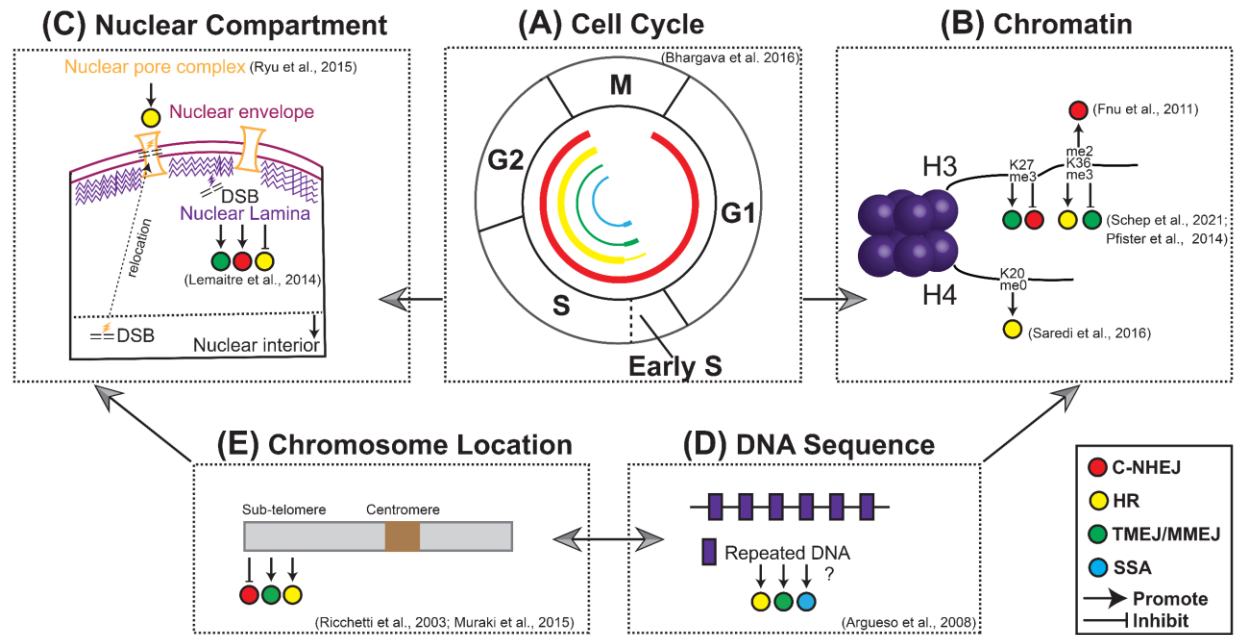


Figure 1.2. Overview of how physical, chemical, and biological conditions influence the usage of different DNA repair mechanisms

(A) The choice of DNA repair is tightly controlled by cell cycle. C-NHEJ is activated throughout cell cycles but dominates in G1 phase. TMEJ/MMEJ, SSA and HR are mainly functional from S to G2 phase, while HR becomes more dominant after early S phase when sister chromatids are available as repair templates (Bhargava et al., 2016). The thickness of the lines indicates the activity of different DNA repair pathway in the cell cycle. **(B)** Local chromatin status affects the choice of DNA repair pathway. For example, TMEJ has been found toward H3K27me3 involved heterochromatin from Schep et al., 2021. The positive or negative roles of H3K36me2/me3 in regulating C-NHEJ, HR or TMEJ has also been reported in multiple studies (Fnu et al., 2011 and Pfister et al., 2014). Besides, uniquely unmethylated lysine residues H4K20me0 on new histones deposited during DNA replication promotes HR repair (Saredi et al., 2016). **(C)** Nuclear compartments and nuclear positioning are associated with DNA repair patterns. For example, to avoid ectopic recombination, DSB inside heterochromatin relocates to nuclear periphery to induce the error-free HR repair (Ryu et al., 2015), while heterochromatic DSB arising from nuclear lamina remains positionally stable, promotes TMEJ and C-NHEJ and impairs HR (Lemaitre et al., 2014). **(D and E)** Sequence content and physical chromosome location have influence on DSB repairs. The influence of repeated DNA on HR-mediated genome variation has also been reported

(Argueso et al.,2008). We propose that repeated DNA might promote to HR, TMEJ/MMEJ and SSA due to the usage of homologous DNA in these repairs. Additionally, it has been found that sub-telomere DSB suffers more from TMEJ/MMEJ and HR than C-NHEJ (Ricchetti et al., 2003; Muraki et al., 2015). These factors do not function alone, but interact to influence DNA repair pathway choice, indicated by arrows between the panels.

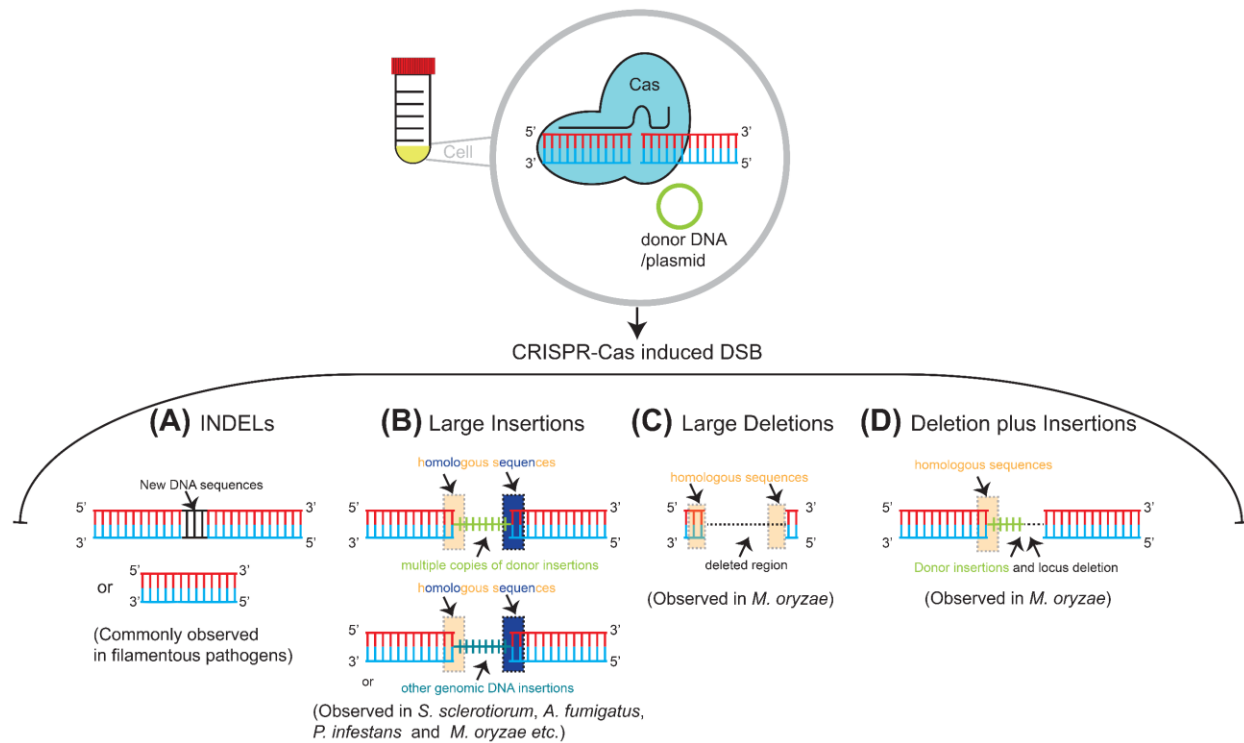


Figure 1.3. Aberrant DNA editing outcomes in filamentous pathogens

(A) INDELs are commonly observed Cas editing results in multiple organisms, including filamentous pathogens. Other types of DNA mutations following Cas-induced DNA repair, such as **(B)** large insertions, **(C)** large deletions, and **(D)** deletion plus insertions have been observed in multiple filamentous microbes.

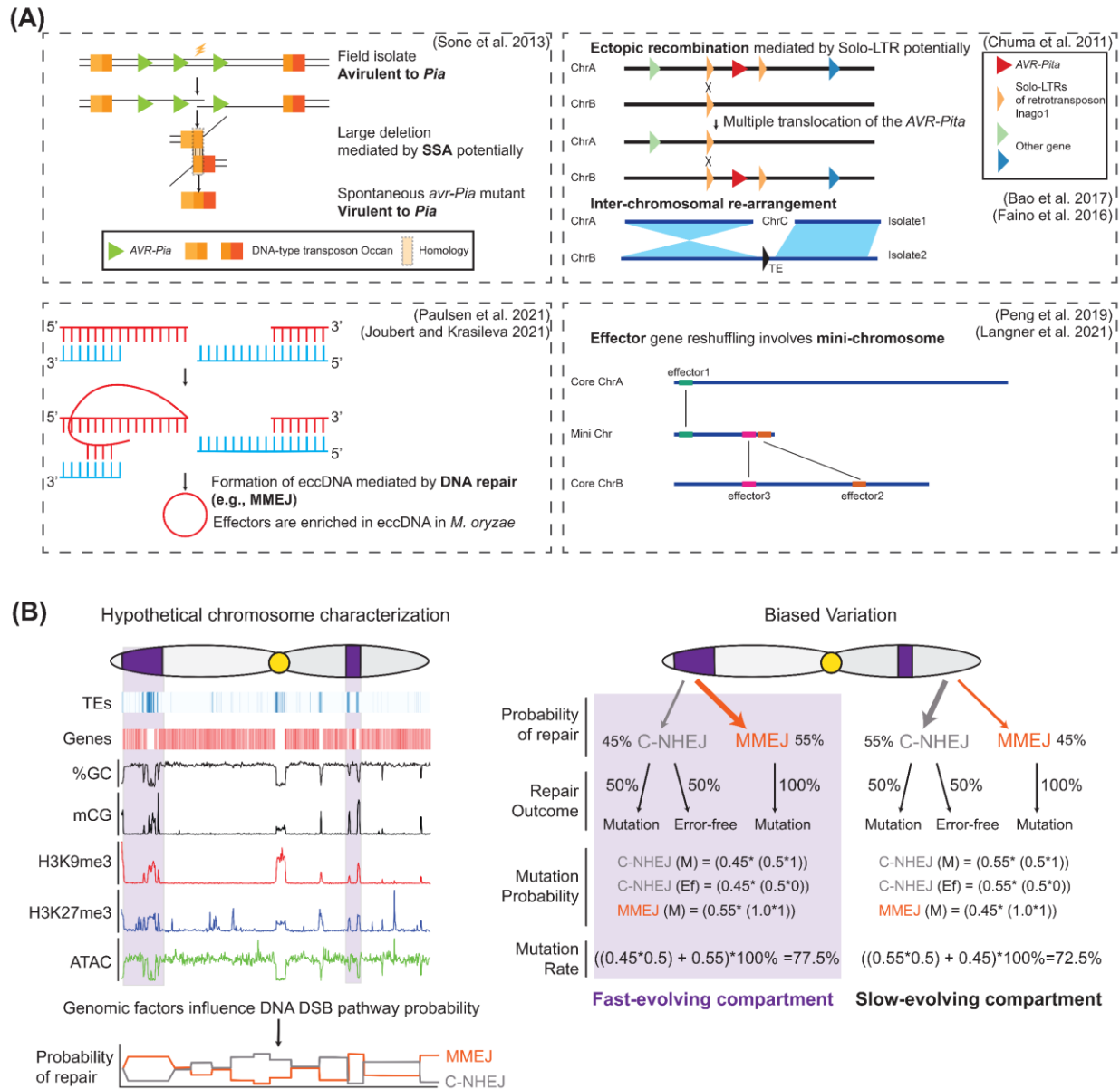


Figure 1.4. The association among DSB repair, genome compartments and biased genome evolution in filamentous pathogens

(A) Examples of observed genome variation influenced by DNA repair. Transposable element (TE) or other repeated DNA can mediate ectopic recombination or re-arrangements (top). DNA contained outside of the core genome, such as eccDNA and mini- or dispensable-chromosomes, may arise due to DNA repair mechanisms, or shuffle DNA with core chromosomes, contributing to biased genome variation in filamentous pathogens (bottom). **(B)** Illustration of a hypothetical chromosome characterized for multiple genomic, epigenomic and chromatin variables. These factors may collectively influence the DNA repair hierarchy at a given genomic locus. We provide

a specific example of how DNA repair choice between C-NHEJ and MMEJ could lead to different rates of DNA mutation in the genome, a model we refer to as Biased Variation. Note, half of the DNA DSBs repaired by C-NHEJ may not cause any DNA mutation, which we represent as 0 in the mutation probability calculation, while half and all DNA DSBs repaired by C-NHEJ and MMEJ will results in a DNA mutation, represented by a 1 in the mutation probability calculation. The rates and frequencies used here are approximations and remain to be experimentally determined. (M), mutation; (Ef), error-free.

Chapter 2 - Histone modification dynamics at H3K27 are associated with altered transcription of in planta induced genes in

Magnaporthe oryzae

Note: This chapter has been published in PLoS Genetics (10.1371/journal.pgen.1009376).

Wei Zhang performed the ChIP-seq and RNA-seq library preparations with the help from Jun Huang and analyzed the ChIP-seq and RNA-seq data. Jun Huang generated the gene knockouts, genotyped and phenotyped all the mutants. Wei Zhang and Jun Huang contributed equally to this work.

Abstract

Transcriptional dynamic in response to environmental and developmental cues are fundamental to biology, yet many mechanistic aspects are poorly understood. One such example is fungal plant pathogens, which use secreted proteins and small molecules, termed effectors, to suppress host immunity and promote colonization. Effectors are highly expressed *in planta* but remain transcriptionally repressed *ex planta*, but our mechanistic understanding of these transcriptional dynamics remains limited. We tested the hypothesis that repressive histone modification at H3-Lys27 underlies transcriptional silencing *ex planta*, and that exchange for an active chemical modification contributes to transcription of *in planta* induced genes. Using genetics, chromatin immunoprecipitation and sequencing and RNA-sequencing, we determined that H3K27me3 provides significant local transcriptional repression. We detail how regions that lose H3K27me3 gain H3K27ac, and these changes are associated with increased transcription. Importantly, we observed that many *in planta* induced genes were marked by H3K27me3 during axenic growth, and detail how altered H3K27 modification influences transcription. ChIP-qPCR

during *in planta* growth suggests that H3K27 modifications are generally stable, but can undergo dynamics at specific genomic locations. Our results support the hypothesis that dynamic histone modifications at H3K27 contributes to fungal genome regulation and specifically contributes to regulation of genes important during host infection.

Introduction

Epigenetic regulation of gene expression is common across many organisms including animals, plants and fungi, regulating diverse processes such as development, disease, and environmental response (Esteller 2007; Freitag 2017; Friedman and Rando 2015; Pfluger and Wagner 2007). In eukaryotes, nuclear DNA is packaged around histone proteins to form nucleosomes (Zentner and Henikoff 2013). Post-translational chemical modifications to histone tails provide epigenetic information contributing to genome regulation (Zhang et al. 2015b). These modifications contribute to active, poised, or silenced transcription by affecting DNA-protein interactions, referred to as chromatin, which contributes to DNA accessibility and positioning in the nucleus (Klemm et al. 2019; Rando and Ahmad 2007). Acetylation of H3 lys-27 (H3K27ac) is associated with transcriptional activation (Hilton et al. 2015), whereas trimethylation at H3 lys-27 (H3K27me3) is linked with transcriptional silencing (Connolly et al. 2013; Seidl et al. 2016). Histone modifications may function alone, but they can also influence other histone modifications, termed cross-talk, to produce different chromatin states giving rise to a histone code regulating DNA functions (Suganuma and Workman 2008; Zhang et al. 2015b). For example, in *Drosophila* and mouse H3K27me3 and H3K27ac are mutually exclusive, and H3K27ac can antagonize H3K27me3 mediated gene silencing (Pasini et al. 2010; Tie et al. 2009)

The H3K27me3 mark is catalyzed and maintained by Polycomb group proteins, termed Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2). While PRC1 appears to be absent in

fungi, PRC2, first identified in *Drosophila* as an important developmental regulator, is present in many eukaryotes (Freitag 2017; Lewis 1978; Ridenour *et al.* 2020; Schuettengruber *et al.* 2017). In the model filamentous fungus *Neurospora crassa*, three core components of PRC2 (*NcSet7*, *NcSu(z)12* and *NcEed*) are required for normal H3K27me3 deposition, and 130 genes of total 774 H3K27me3-marked genes are upregulated ~7.5-fold in the absence of H3K27me3 (Jamieson *et al.* 2013). The genome of *Fusarium graminearum*, the causal agent of wheat head scab, codes numerous secondary metabolite gene clusters that are enriched for H3K27me3 modification, and deletion of the methyltransferase responsible for H3K27me3, *FgKmt6*, resulted in transcriptional activation of many clusters (Connolly *et al.* 2013). Similarly, reduction of H3K27me3 in the rice pathogen *F. fujikuroi*, led to the induction of several secondary metabolite gene clusters and formation of novel chemical compounds (Niehaus *et al.* 2016; Studt *et al.* 2016). In addition to pathogenic interactions, H3K27me3 helps regulate gene expression in the fungus *Epichloë festucae* during endophytic growth in its host (Chujo and Scott 2014). Given these observations and the described function of H3K27me3 in facultative heterochromatin, genomic regions that switch between repressed and active transcription (Connolly *et al.* 2013; Lau-Corona *et al.* 2020; Pasini *et al.* 2010; Tie *et al.* 2009), dynamic regulation of H3K27 modifications provides an attractive regulatory model.

Magnaporthe oryzae (synonym of *Pyricularia oryzae*) is a filamentous fungal plant pathogen, best known for causing rice (*Oryza sativa*) and wheat (*Triticum aestivum*) blast diseases that threaten world food security (Inoue *et al.* 2017; Peng *et al.* 2019; Pennisi 2010; Zhang *et al.* 2016b; Zhong *et al.* 2016). The initial stages of *M. oryzae* infection, including spore germination, appressorium development and leaf penetration are well understood, but many details of growth and development inside the plant remain unknown (Khang *et al.* 2010; Li *et al.* 2012; Ryder *et al.*

2019; Xu and Hamer 1996). A central paradigm of plant-microbe interaction is the use of secreted proteins and chemicals by an invading organism to suppress host immunity, blunt the immune response and change host physiology to facilitate infection (Oliveira-Garcia and Valent 2015; Yi and Valent 2013). Numerous effectors identified in *M. oryzae* are highly expressed *in planta* but remain silent or lowly expressed *in vitro* (i.e. *ex planta*), a common phenomenon across filamentous pathogens (de Jonge *et al.* 2012; Li *et al.* 2009b; Mosquera *et al.* 2009; Tanaka *et al.* 2019; Wu *et al.* 2015). Little is known about the molecular mechanisms controlling repression of fungal effectors when growing outside of the host, and their subsequent induction during host infection. A previous study in a wheat-infecting strain of *M. oryzae* showed H3K4me2/me3 was positively correlated with active gene expression *in vitro* (Pham *et al.* 2015). Given that H3K4me2/me3 is a conserved modification marking active transcription, it provides little information addressing the mechanism of gene *in vitro* silencing and *in planta* induction. Whereas H3K27me3 contributes to secondary metabolite gene repression in fungi, how histone modification dynamics at H3K27 contributes to transcriptional regulation of effectors is largely unknown.

To address this knowledge gap, we tested two related hypotheses: that H3K27me3 is involved in suppressing effectors and other *in planta* induced genes during *in vitro* growth, and dynamic change between trimethylation and acetylation at H3K27 underlies the switch between a repressed and activated state (Figure 2.1A). Using a combination of fungal molecular genetics, epigenomics, and transcriptomics analysis, we demonstrate that histone modification dynamics at H3K27 can regulate specific genomic loci, often important during host infection, supporting the hypothesis.

Results

Trimethylation and acetylation at H3K27 occupy distinct genomic regions in *M. oryzae*

To understand the regulatory role of H3K27 modification, we performed ChIP-seq for H3K27me3 and H3K27ac to establish the genomic distribution of these marks (Supplementary Table A.1). The genome-wide domains for H3K27me3 and H3K27ac occupied discrete genomic regions in *M. oryzae*, with H3K27me3 forming longer discrete blocks, and H3K27ac having a more continuous distribution but shorter domains (Figure 2.1B and 2.1C). A total of 1,381 H3K27me3 domains were identified, ranging in size from 0.3 to 115.6 kb (average 6.4 kb). Collectively they occupy 8.8 Mb (20.6%) of the 42.8 Mb *M. oryzae* Guy11 wild type genome (Figure 2.1C, 2.1D, Supplementary Figure A.1A, and Supplementary Table A.2). This H3K27me3 genome distribution is greater than that reported in *Drosophila* (~3-7.5%), *Arabidopsis* (~5.7%), and the saprophytic fungus *Neurospora* (~6.8%) (Bonnet *et al.* 2019; Jamieson *et al.* 2013; Zhang *et al.* 2007a), but lower than the plant pathogenic fungus *F. graminearum* (~33%) (Connolly *et al.* 2013). The H3K27ac domains, total of 9,512, were significantly shorter than those of H3K27me3 and cover 10.9 Mb (25.6%) of the *M. oryzae* genome (Figure 2.1C, 2.1D, Supplementary Figure A.1A, and Supplementary Table A.3). In addition, the average H3K27me3 ChIP signal was significantly stronger than that for H3K27ac (Supplementary Figure A.1B).

In relation to the functional genome, H3K27me3 domains predominantly marked transposable elements (TEs) with 51.5% of the annotated TEs being occupied by a H3K27me3 domain, contrasted by only 1.8% being marked by H3K27ac (Figure 2.1D). Within the TEs, the H3K27me3-ChIP signals were significantly stronger for LINE and LTR retrotransposons compared to DNA Class II elements, while H3K27ac signals showed no significant difference between the classes (Supplementary Figure A.2A and B). We found the H3K27ac domains almost

exclusively associated with coding regions, marking promoters, terminators and 21.1% of annotated genes, while H3K27me3 was present at only 16.6% of genes (Figure 2.1D). There was no Gene Ontology (GO) enrichment for the H3K27me3 marked genes, but Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis suggested that many H3K27me3 marked genes were involved with secondary metabolism (Supplementary Table A.4 and Supplementary Table A.5). Profile plots of ensembles genes and TEs revealed that H3K27me3-marked genes and TEs had lower H3K27ac signals, and those lacking H3K27me3 displayed higher H3K27ac signal (Figure 2.1E and 2.1F). In addition, no genes or TEs were identified with both marks at H3K27 based on identified peaks, further showing the consistent and exclusive nature of the two marks. These results support the hypothesis that H3K27me3 is a repressive histone mark in *M. oryzae* and provides a distinct regulatory role from H3K27ac (Figure 2.1A).

PRC2-mediated H3K27me3 contributes to normal *in vitro* growth

The core components of the PRC2 complex were identified in *M. oryzae* based on homology, termed *MoKmt6*, *MoSuz12*, and *MoEed*, and deletion mutants were generated for each (Supplementary Figure A.3 and 4, and Supplementary Table A.6). Consistent with canonical PRC2 function, each of the three proteins were required for H3K27me3 (Figure 2.2A). Furthermore, the methyltransferase activity of PRC2 was restored when the individual gene deletions were complemented (*ΔMokmt6-com*, *ΔMosuz12-com* and *ΔMoeed-com*) (Figure 2.2B and Supplementary Figure A.5). Among the three mutants, only *ΔMoeed* showed significantly reduced growth on complete medium agar, while all three deletion mutants showed significant reduced growth on rice polish agar compared to the wild type (Figure 2.2C and 2.2D). The different growth rates observed on complete medium agar versus rice polish agar shows an interplay between PRC2

function and growth environment. Production of asexual conidia were not statistically significantly different for any mutant compared to wild type (Figure 2.2E).

Loss of H3K27me3 causes redistribution of H3K27ac and H3K36me3 at specific genomic regions

In other eukaryotes, H3K27me3 and H3K27ac have antagonistic effects on transcription, and their dynamics provide an important transcriptional regulatory layer (Pasini *et al.* 2010; Tie *et al.* 2009). We observed an increase in H3K27ac signal in the three *M. oryzae* PRC2 mutants, which indicated a direct antagonism between the two modifications (Figure 2.3A). To determine if the altered H3K27ac levels were the result of global or local H3K27ac redistribution, ChIP-seq was performed on Δ *Mokmt6*. Genome-wide Pearson correlation analysis showed H3K27me3 was negatively correlated with H3K27ac ($r = -0.61$, $p < 2.2e-16$, Supplementary Figure A.6) in wild type, and the loss of H3K27me3 in Δ *Mokmt6* caused a significant change in H3K27ac distribution ($r = 0.73$, $p < 2.2e-16$, Supplementary Figure A.6). This further supports the antagonistic nature of the two modifications. We observed that the largest change in H3K27ac in Δ *Mokmt6* occurred at genes and TEs that had previously been marked by H3K27me3 in wild type, in which these elements gained H3K27ac (Figure 2.3B). To monitor histone modification crosstalk, we checked the levels of another histone modification, H3K36me3, as it has been shown that H3K36me3 rarely co-exists with H3K27me3 (Schwammle *et al.* 2014; Yang *et al.* 2014; Yuan *et al.* 2011). We observed that H3K36me3 was positively correlated with H3K27me3 ($r = 0.61$, $p < 2.2e-16$, Supplementary Figure A.6). The largest change for H3K36me3 in Δ *Mokmt6* also occurred at genes and TEs that had been marked by H3K27me3 in wild type, but in this case these genes and TEs lost H3K36me3 (Figure 2.3B). Unsupervised K-mean clustering of genes based on H3K27me3 ChIP-seq from wild type resulted in cluster 1 and cluster 2, defined by the presence and absence

of H3K27me3 in wild type respectively (Figure 2.3C). The wild type cluster 1 genes, containing the highest levels of H3K27me3, were devoid of H3K27ac, while genes in cluster 2 showed increasing H3K27ac as H3K27me3 decreased (Figure 2.3C and 2.3D). Interestingly, the genes marked by H3K27me3 in cluster 1 showed a significant gain of H3K27ac in *ΔMokmt6*, which showed the change from trimethylated to acetylated H3K27 occurred at specific loci (Figure 2.3C and 2.3D). The opposite was seen for H3K36me3 levels, which were highest in cluster 1, but decreased in the *ΔMokmt6* strain lacking H3K27me3 (Figure 2.3C and 2.3E). Similar overall patterns were observed for TEs, where there was a gain of H3K27ac in *ΔMokmt6* for TEs that had been marked by H3K27me3 in the wild type (Supplementary Figure A.7). For example, TE rich regions on contig 5 contain H3K27me3 domains lacking H3K27ac in the wild type strain. However, these regions clearly gain H3K27ac and show a modest reduction for H3K36me3 when ChIP is conducted in *ΔMokmt6* (Figure 2.3F). Collectively, these results demonstrate specific regulatory machinery and rules are present governing H3K27 trimethylation and acetylation and crosstalk existing with H3K36me3. We did not observe a global redistribution of H3K27ac in the absence of H3K27me3, but rather local dynamics where previously trimethylated residues became acetylated.

Differential gene expression upon loss of H3K27me3 and epigenetic state change

Given our empirical evidence that H3K27me3 and H3K27ac are exclusive, and that previously H3K27me3 marked genes can become H3K27ac in *ΔMokmt6*, we next assessed how this change impacted transcription using RNA-seq. Gene expression analysis between wild type and *ΔMokmt6* grown *in vitro* identified that 1,115 genes were upregulated in the mutant (9.8% of transcriptome). Of these 1,115 up regulated genes, 980 (87.9%) of them were marked by H3K27me3 domains in the wild type, whereas only 9 in 269 downregulated genes in *ΔMokmt6* are

marked by H3K27me3 (Figure 2.4A, Supplementary Figure A.8, and Supplementary Table A.4). This shows significant local transcriptional repression by H3K27me3 in *M. oryzae*. We identified 229 genes that gained H3K27ac in the Δ *Mokmt6*. This set of genes had significantly higher expression in Δ *Mokmt6* compared to wild type (Supplementary Figure A.9A and Supplementary Figure A.9B). Conversely, we identified 857 genes that lost H3K27ac in Δ *Mokmt6*, which this set of genes was not significantly differently expressed between Δ *Mokmt6* and wild type (Supplementary Figure A.9C and Supplementary Figure A.9D). This suggested that gaining H3K27ac was associated with increased transcription, but H3K27ac is not required *per-se* for transcription as sites that lost the mark were not expressed at a lower level. We also found that 363 genes lost H3K36me3 in Δ *Mokmt6*, and this group of genes was significantly more highly expressed (Supplementary Figure A.10). If we further characterize genes as having either only gained H3K27ac, only lost H3K36me3, or both, we see the same impacts on transcription (Supplementary Figure A.11). That is, genes that gained H3K27ac or lost H3K36me3 were more highly expressed in Δ *Mokmt6* compared to wild type under the same axenic growth conditions. These associations do not prove causative, but they are consistent with the hypothesis that histone modification dynamics contribute the transcriptional regulation.

To further characterize how the redistribution of H3K27ac to previously H3K27me3 sites impacted transcription, we identified and deleted the coding sequence for *MoGcn5*, the protein responsible for acetylating histone lysines, including H3K27ac (Supplementary Figure A.12A and Supplementary Figure A.12B) (Zhang *et al.* 2017a). The double deletion mutant, Δ *Mokmt6* Δ *Mogcn5*, lacks H3K27me3 and the majority of H3K27ac (Figure 2.4B). The double mutant also had an altered morphology and significantly reduced growth on complete medium compared to the wild type (Supplementary Figure A.13A and Supplementary Figure A.13B). We

also noted the double mutant has an elongated and flattened conidial morphology, while both the single mutants, *ΔMokmt6* and *ΔMogcn5*, have conidia that resemble the wild type (Supplementary Figure A.13C). Rice infection showed that both single mutants, *ΔMokmt6* and *ΔMogcn5*, produced smaller lesions on rice, while *ΔMokmt6ΔMogcn5* completely lost the ability to infect (Supplementary Figure A.14). The double mutant's abnormal conidia and reduced growth likely resulted in the inability to infect rice. RNA-seq analysis of *ΔMokmt6ΔMogcn5* showed 29.5% of the H3K27me3 marked genes that were up regulated in *ΔMokmt6* were down regulated in *ΔMokmt6ΔMogcn5* (Figure 2.4A). These results were confirmed globally, where the average expression of H3K27me3 marked genes is significantly lower in wild type compared to *ΔMokmt6*, while the non-marked genes have a similar expression profile (Figure 2.4C). In *ΔMokmt6ΔMogcn5*, the average expression of H3K27me3 marked genes is significantly lower than in *ΔMokmt6*, showing the requirement for Gcn5 acetylation for part of the gene activation seen in *ΔMokmt6*. There is also a statistically significantly lower average expression for non-H3K27me3 marked genes in *ΔMokmt6ΔMogcn5* compared to wild type or *ΔMokmt6*, likely because Gcn5 acetylates additional lysine residues besides H3K27 that contribute to transcription (Liang *et al.* 2018; Rosler *et al.* 2016).

To further investigate how H3K27 modification status links with individual gene expression, we integrated RNA-seq from the wild type, *ΔMokmt6*, and *ΔMokmt6ΔMogcn5* with the ChIP-seq (Figure 2.4D). Genes were sorted in descending order based on wild type RNA-seq levels, combined into groups composed of 100 genes each, and the values for RNA-seq and ChIP-seq were summarized. A clear positive association between H3K27ac and gene expression can be observed, with genes in the highest expression groupings having the highest H3K27ac signal and decreasing as transcription decreases ($r_{\text{Pearson}} = 0.67$, $p < 2.2\text{e-}16$, Figure 2.4D). For H3K27me3,

there is a non-linear relationship, with the line connecting each group average H3K27me3 signal remaining relatively flat until after approximately group 80, after which there is a clear positive association between increased H3K27me3 and decreased transcription (Figure 2.4D). This result is consistent with the observation that H3K27me3 is not present uniformly across the genome, but that where present it is associated with repressed transcription. This also shows that the genes with the greatest transcriptional increase in $\Delta Mokmt6$ are those after group 80, where there is clear divergence between the average group RNA-seq levels (Figure 2.4D, blue versus red circles). These same gene groups are also where the largest bulk redistribution of H3K27ac occur in $\Delta Mokmt6$ compared to wild type (Figure 2.4D, blue versus red diamonds). Indeed, plotting differential gene expression and fold-change for ChIP signal between wild type and $\Delta Mokmt6$ showed that gene groups having the largest transcriptional activation in $\Delta Mokmt6$ correspond to a concomitant decrease in H3K27me3 and increase in H3K27ac in $\Delta Mokmt6$ (Figure 2.4E). Collectively, these analyses integrate the transcriptional changes associated with dynamic modifications at H3K27 to detail this control mechanism in *M. oryzae*. The data show that histone modification dynamics at H3K27 provide relatively local (*cis*) transcriptional control.

H3K27me3 suppresses *in planta* induced genes during *in vitro* growth

Genes required for host infection by *M. oryzae* are generally not expressed during *in vitro* growth (Dong *et al.* 2015b; Mentlak *et al.* 2012; Shimizu *et al.* 2019). As such, another prediction of our hypothesis is that H3K27me3 is present at host induced genes when the fungus is grown *in vitro*, functioning to repress transcription. To assess this, we compared *in vitro* RNA-seq and ChIP-seq results with RNA-seq collected from *M. oryzae* during rice infection. Interestingly, the expression profiles for H3K27me3-marked genes showed a statistically significantly higher *in*

planta expression compared to *in vitro* growth, while non-marked genes have a similar expression profile between the two conditions ($p_{\text{marked}} = 5.7\text{e-}14$, $p_{\text{non}} = 0.15$, Wilcoxon test, Figure 2.5A). Among the 1,154 *in planta* up regulated genes, 44.4% were marked by H3K27me3, which is significantly higher than expected given that H3K27me3 marks 16% of genes genome-wide ($\chi^2 = 939.09$, $p < 2.2\text{e-}16$, Figure 2.5B and 2.5C). Of the *in planta* up regulated genes marked by H3K27me3, two thirds (339/512) were also up regulated in ΔMokmt6 (Figure 2.5B). The *in vitro* ChIP-seq profiles were separated based on *in planta* expression profiles. A clear pattern is evident where *in planta* induced genes have the highest average H3K27me3 and H3K36me3 signal, but the lowest for H3K27ac during *in vitro* growth (Figure 2.5C). Interestingly, the *in planta* downregulated genes showed the highest average H3K27ac signal, but lowest H3K27me3 signal in wild type when grown *in vitro*. The ChIP-seq from ΔMokmt6 shows that H3K27ac is specifically re-distributed to sites that were previously H3K27me3 and devoid of H3K27ac, while the same loci lose H3K36me3 (Figure 2.5C). These results collectively show that *in planta* induced genes are significantly enriched for H3K27me3 when grown *in vitro* under transcriptionally repressive conditions. The significant association between *in* ΔMokmt6 upregulated genes and those induced *in planta* indicates that the H3K27me3 mark is not just present, but required for transcriptional repression.

Effectors regulated by H3K27 modification dynamics

To date, numerous *M. oryzae* effectors or presumed effectors have been identified, but little is known regarding their transcriptional regulation (Fernandez and Orth 2018). To further test the hypothesis that H3K27 modification dynamics contributes to *in planta* gene induction, we analyzed specific sets of putative and characterized genes contributing to *M. oryzae* host

colonization. The EffectorP software predicted 409 putative *M. oryzae* effector genes from our annotated proteins with a secretion signal (Sperschneider *et al.* 2018). Of these, only 349 gene models had read support from our RNA-seq data. We found that 92 of these putative effector genes (i.e. 26.4% of the total predicted) are marked by an H3K27me3 domain, which is a higher association than expected by chance ($\chi^2 = 10.178$, $p = 0.001422$, Supplementary Table A.7). Further restricting the set of 349 putative effector genes to those that are more highly expressed *in planta* versus media, identified 140 genes, of which 50 (35.7%) fall inside an H3K27me3 domain. This represents a highly significant association between *in planta* induced putative effector genes and H3K27me3 domains ($\chi^2 = 45.217$, $p < 0.00001$, Supplementary Table A.8).

As another approach, we curated a list of 31 published effectors and avirulence genes that have direct experimental evidence. Of these well studied effectors and avirulence genes, 27 are present in *M. oryzae* strain Guy11 (Supplementary Table A.9). We found that over half (14 of 27) of these high-confidence effectors were marked by H3K27me3 during *in vitro* growth, which is significantly higher than expected ($\chi^2 = 24.285$, $p = 0.0005$, Supplementary Table A.10). The H3K27me3 marked effector genes include five cloned *Avirulence* (AVR) genes, four Biotrophy-associated secreted (BAS) protein encoding genes, and five genes encoding proteins that can induce or repress programmed cell death (Supplementary Table A.9) (Bohnert *et al.* 2004; Chen *et al.* 2013b; Guo *et al.* 2019; Irieda *et al.* 2019; Mentlak *et al.* 2012; Mosquera *et al.* 2009; Ray *et al.* 2016; Saitoh *et al.* 2012; Shimizu *et al.* 2019; Wang *et al.* 2016; Wu *et al.* 2015; Zhong *et al.* 2018). Hierarchical clustering based on transcriptional profiles from four conditions identified four groups with similar expression profiles (Figure 2.6A). Genes in group 1 and 2 are *in planta* induced and transcriptionally responsive to genetic perturbation of H3K27. Specifically, the loss of H3K27me3 significantly increased expression for group 1 and 2 genes, which were dependent

on H3K27ac in group 1 and to a lesser extent for group 2 (Figure 2.6B, 6C, 6F, and 6G). Genes in group 3 were also *in planta* induced, but most of them did not show transcriptional regulation by H3K27 modification except for *BAS4* and *MoCDIP9*, while those in group 4 were not shown to be *in planta* induced in our dataset (Figure 2.6D, 6E, 6H and 6I). Locus specific examples for each group show the dynamics of histone modifications and transcriptional output (Figure 2.6F-6I, and Supplementary Figure A.15). For example, *SLP1* is present in an H3K27me3 domain, lacking H3K27ac and is not expressed during wild type *in vitro* growth (Figure 2.6F and Supplementary Figure A.15A). When the Δ *Mokmt6* strain is grown *in vitro*, peaks of H3K27ac can be seen at the presumed promoter and potential 3' regulatory region of the *SLP1* coding sequence along with induced transcription, which is partially lost in Δ *Mokmt6* Δ *Mogcn5*. The *MoCDIP5* locus showed a similar gene expression and histone modification pattern, by which loss of H3K27me3 results in increased transcription and a concomitant increase in H3K27ac (Figure 2.6G and Supplementary Figure A.15B). Although H3K27me3 is present at *BAS1*, the loss of the mark did not impact expression in Δ *Mokmt6*, suggesting induction of *BAS1* is likely dependent on other unknown transcriptional activators mediated by plant specific or induced signals (Figure 2.6H and Supplementary Figure A.15C). Of the 27 high-confidence effector genes, 12 (44.4%) were induced in Δ *Mokmt6* compared to wild type, which is far greater than the genome-wide 9.8% of genes upregulated by Δ *Mokmt6*, supporting an association between H3K27me3 and effector regulation ($\chi^2 = 36.327$, $p < 0.00001$, Supplementary Table A.11). These results support our hypothesis that H3K27me3 is responsible for the repression of many presumed effectors during growth *ex planta*, while H3K27ac can play an important role in activating these genes in the absence of H3K27me3.

To date, there are no published data sets of genome-wide *in planta* ChIP-seq for a filamentous pathogen. This is due to technical limitations of having heterogenous samples where

the pathogen (i.e. target genome of interest) is less abundant in the mixed sample by orders of magnitude. Despite this, we conducted *in planta* ChIP using a leaf-sheath assay in which *M. oryzae* spores infect rice leaf sheaths (Khang *et al.* 2010; Mosquera *et al.* 2009). We conducted H3K27me3, H3K27ac, and IgG ChIP from three independent infection samples, where an input fraction was collected after grinding, and then each sample was split into separate fractions for ChIP with one of the three antibodies. Subsequently, qPCR was performed (i.e. ChIP-qPCR) using two primer pairs that were either in the gene body or within the putative promoter (i.e. 200-300 bp upstream of the transcription start site) for two characterized *in planta* induced genes, one control gene (*Actin*) that lacked H3K27me3 during axenic growth, and another control (*MGG_09416*) which contained H3K27me3 during axenic growth. Results from the *in planta* ChIP-qPCR were compared to ChIP samples from axenic growth. Our results demonstrate that H3K27me3 was highly abundant and at comparable levels between *in plant* and *axenic* growth for the gene body location tested for *BAS4*, *SLP1* and control gene (*MGG_09416*) (Figure 2.7A, 7B, and 7D). Interestingly, the H3K27me3 levels dropped 6-fold at the *BAS4* promoter during *in planta* infection compared to axenic growth, while the levels remained unchanged for *SLP1* and *MGG_09416* (Figure 2.7A, 7B, and 7D). We did observe small changes for H3K27me3 levels between the growth conditions at the gene body and promoter for the *Actin* locus, but the amount of H3K27me3 was low, consistent with ChIP-seq data (Figure 2.7C). We did observe a large increase in H3K27ac levels during *in planta* growth at the gene body of *BAS4*, but we also observed this at the *Actin* and *MGG_09416* loci, which are not differentially expressed between the growth conditions. These results are consistent with our genome-wide analyses, in which H3K27ac was associated with increased expression, but it was not absolutely sufficient or required (e.g. Supplementary Figure A.9). The control ChIP-qPCR, where IgG was used for the pull-down, did

not produce significant qPCR signal, indicating the histone-modification antibody pull-down was specific (Figure 2.7). Overall, these results demonstrate dynamics for H3K27 modification at specific well characterized genes using both genetic perturbation and *in planta* infection experiments. These data provide important mechanistic information regarding the repression and activation of *in planta* induced genes.

Discussion

The aim of this research was to further understand the transcriptional regulation of *in planta* induced genes in *M. oryzae* by investigating the role of histone modification dynamics at H3K27. Our integrated multi-omics and molecular genetic investigation provides strong evidence that H3K27me3 is an important and specific negative regulator of *in planta* induced genes when *M. oryzae* is grown *ex planta*. We found that only 16.6% of genes were marked by H3K27me3, yet 44.4% of *in planta* induced genes were H3K27me3 marked, a striking enrichment. Not only is the mark present at many *in planta* induced loci, it appears to play a functional role based on the significant overlap between *in planta* induced genes and those upregulated in Δ *Mokmt6*, including many known or presumed effectors (i.e. *BAS4*, *BAS2*, *AVR-Pi9*, *SLP1*). This is consistent with reports that H3K27me3 represses transcription of specific secondary metabolite clusters in *Fusarium* species, and also the endophyte *E. festucae*, where knockout or knockdown of H3K27me3 results in transcriptional activation (Chujo and Scott 2014; Connolly *et al.* 2013; Niehaus *et al.* 2016; Studt *et al.* 2016). Secondary metabolite clusters have similar biological function to effectors in that they are transcriptionally repressed outside of specific growth conditions, they function outside of the producing organism, and they serve adaptive functions (Collemare and Seidl 2019; Woloshuk and Shim 2013). As such, there is a clear role of H3K27me3 playing a conserved role in silencing adaptive genes (i.e. effectors and secondary metabolite clusters) in pathogenic fungi, but further mechanistic understanding of the dynamics, signals and enzymes in this pathway are required.

The clear boundaries between H3K27me3 and H3K27ac domains in *M. oryzae*, and their association with TEs versus genes, respectively, highlight their distinct roles in transcriptional regulation. In *F. fujikuroi*, a global increase in H3K27ac was reported when H3K27me3 was

reduced by RNAi, but the alterations were not analyzed at specific loci (Niehaus *et al.* 2016). Here we present detailed analysis showing the specific re-distribution of H3K27ac in the absence of H3K27me3 is at local sites that were previously trimethylated, and not a general genome-wide increase of H3K27ac. Importantly, a significant amount of the upregulated genes in Δ *Mokmt6* showed lower transcription in Δ *Mokmt6* Δ *Mogcn5* (Zhang *et al.* 2017a). There is a confounding effect of GCN5 because it can acetylate additional lysine residues besides H3K27. Our results demonstrate a GCN5 dependence for some transcriptional activation in the absence of H3K27me3 ($\sim 1/3$), but this could be caused by additional means beyond H3K27ac. The specific accumulation of H3K27ac at previously H3K27me3 marked regions in Δ *Mokmt6* suggests a regulatory role, but further evidence is required. What protein(s) or signal recruits GCN5 to sites of previous H3K27me3 marked regions is not clear. Nor is it known how H3K27me3 is established at repressed sites, prior to and excluding H3K27ac. A number of *in planta* induced genes were not responsive to genetic perturbation of H3K27 modification, and further research is needed to test whether H3K27 modification dynamics cooperating with transcription factors contribute to *in planta* induced gene regulation in *M. oryzae*.

Similar to H3K27ac, H3K36me3 modification is thought to antagonize H3K27me3 in many systems and is associated with active transcription (Yang *et al.* 2014; Yuan *et al.* 2011). Our result show H3K36me3 is nearly ubiquitous for genes in *M. oryzae*, regardless of transcription, which was also found in *F. graminearum* (Connolly *et al.* 2013). However, in *F. graminearum* there was no interaction observed for H3K36me3 when H3K27me3 was genetically perturbed. In *M. oryzae*, we found H3K36me3 levels were significant reduced in Δ *Mokmt6* at loci that had H3K27me3 in the wild type, indicating a positive association between the presence of the marks. In *N. crassa*, the histone methyltransferase NcSet2 catalyzes the majority of H3K36me3, while,

another methyltransferase NcAsh1 deposits only 5% of H3K36me3 (Bicocca *et al.* 2018). Interestingly, NcAsh1 deposited H3K36me3 is associated with lowly transcribed genes and domains of H3K27me2/3. Unlike our findings, the level of NcAsh1 catalyzed H3K36me3 did not change in the absence of H3K27me2/3 (Bicocca *et al.* 2018). Additionally, in *F. fujikuroi*, ASH1 was responsible for a minor portion of genome-wide H3K36me3, but ASH1-mediated H3K36me3 was associated with regions enriched for H3K27me3 (Janevska *et al.* 2018). Interestingly, ASH1 has a reported role in genome stability, possibly through DNA repair, in *F. fujikuroi* (Janevska *et al.* 2018). The role of SET2 and ASH1 in catalyzing H3K36me3 in *M. oryzae* is not clear (Cao *et al.* 2016; Pham *et al.* 2015), but our results are consistent with *M. oryzae* also having two distinct chromatin states associated with H3K36me3. As suggested, it is an interesting hypothesis that distinct complexes may deposit H3K36me3 and H3K27me3 at specific facultative heterochromatin sites for a specific regulatory mechanism (Bicocca *et al.* 2018; Janevska *et al.* 2018). It will be interesting to determine if altered H3K36me3 distribution in the absence of H3K27me3 is SET2- or ASH1-dependent.

Our research specifically addressed the hypotheses that H3K27me3 represses transcription during *ex planta* growth, and that switching from a trimethylated to acetylated state at H3K27 underlies transcriptional activation at specific loci. While it is still not technically feasible to address histone modification dynamics at a global scale during plant infection, we sought to address predictions of this hypothesis using genetics and *in vitro* growth. For this model, we show that PRC2-mediated H3K27me3 represses transcription of many *in planta* induced genes, H3K27 modification dynamics from trimethylation to acetylation contribute to this transcriptional activation, and finally we report an association between H3K27 modification-mediated transcriptional changes and genes highly transcribed *in planta*. In addition, we report high-quality

data from successful *in planta* ChIP-qPCR. These results showed that H3K27me3 levels were generally stable between *in planta* and axenic growth at many tested genomic sites. We also detected a significant reduction at an effector promoter, where the level of H3K27me3 was reduced 6-fold compared to axenic growth. Our results show that histone modification status can be significantly altered *in planta*, but our analysis is limited to a few loci. Questions remain regarding how H3K27 modification dynamics, their crosstalk with additional histone modifications (e.g. H3K36me3) and their impact on DNA accessibility change across growth conditions. Technological advances are required to conduct genome-wide ChIP and chromatin analysis during *in planta* infection to more completely understand transcriptional regulation during host infection. Our observations support the hypothesis that modification dynamics at H3K27 is an important regulatory mechanism for host-adapted genes in *M. oryzae*.

Methods

Fungal strains and incubation condition

M. oryzae wild type strain Guy11 (French Guiana) was used throughout this study. Complete medium agar/broth was used for testing vegetative growth and preparing materials for DNA, RNA and protein extraction (Chen *et al.* 2008). Rice polish agar (rice polish 25 g/L, 16 g/L agar, autoclaved for 46 mins) was used for testing vegetative growth and conidiation. TB3 medium (Sucrose 200 g/L, Casamino acid 6 g/L, yeast extract 6 g/L, agar 15 g/L) was used for screening fungal transformants. For test vegetative growth, *M. oryzae* strains were grown on complete medium agar plates at 28 °C in the dark, and colony diameters were measured after 12 days of growth. *M. oryzae* strains were grown on rice polish agar plates at 28 °C for the first 7 days then cultured under light for the remaining 5 days. For conidiation, conidia were collected from rice polish agar plates under constant light for 10 days by using 5 mL sterilized water. The conidial

morphology was checked with Zeiss Axioplan 2 IE MOT microscope. For preparing materials for DNA, RNA and protein extraction, 5 to 7 pieces of fungal plugs from rice polish agar were cultured in the liquid complete medium for 3 to 4 days at 28 °C under dark conditions.

Plasmid construction and fungal transformation

The single deletion mutants for *MoKmt6*, *MoSuz12*, *MoEed*, *MoGcn5* and double deletion mutants for *MoKmt6/MoGcn5* were acquired by a homologous recombination strategy. Briefly, approximately 1.0 kb upstream and 1.0 kb downstream sequences of each gene were amplified from wild type Guy11 genomic DNA and fused with the hygromycin resistance cassette or G418 resistance cassette in binary vectors pFGL821 or pFGL921. pFGL821 was a gift from Dr. Naweed Naqvi (Addgene plasmid # 58223), and pFGL921 was modified from pFGL821 by replacing the hygromycin resistance cassette with a G418 resistance cassette. Drug resistant transformants were screened by PCR. The primers used in this study are listed in Supplementary Table A.6.

For gene complementation assays, entire *MoKmt6*, *MoSuz12* and *MoEed* genes with upstream about 1.5 kb potential promoter and downstream about 0.5 kb potential terminator were amplified and cloned into binary vector pFGL921 with the G418 resistance cassette. After Sanger sequencing, confirmed corresponding plasmids were transferred into each individual deletion mutant. G418 resistant transformants were screened by PCR.

Agrobacterium tumefaciens-mediated transformation was performed as previously described with minor modifications (Khang *et al.* 2006). Briefly, *M. oryzae* strains were cultured in growth chamber at 25 °C under light for 2 to 3 weeks. Spores were collected at the final concentration of 2×10^6 spores/mL. A single colony of *Agrobacterium* strain AGL1 with appropriate binary vector was mixed with around 100 μ L of above spore solution and co-cultivated on a mixed cellulose ester membrane (GE Healthcare, # 09-529-711) for 48 hours. After co-

cultivation, the membrane was transferred to TB3 with 400 mM cefotaxime and appropriate fungal selections (200 µg/mL hygromycin (Corning, # MT30240CR) or 800 µg/mL G418 (VWR, #97064-358)). After 8 to 12 days' culture, the fungal antibiotic resistant colonies were transferred to individual complete medium plates with appropriate antibiotic (200 µg/mL hygromycin or 500 µg/mL G418) for further screening. The fungal colonies with hygromycin or G418 resistance were screened by PCR analysis to validate the successful transformants.

Plant materials, leaf infection and leaf sheath assay

Susceptible rice cultivar YT16 was used in this study for infection assay. Six to seven seeds were sown in one pot filled with Baccto soil. Three-week old rice was used for inoculation. For rice leaf infection assay, 5×10^4 spores/mL was collected from related strains through two-layers of miracloth with 0.25% gelatin. The spore spray inoculation, further incubation and disease severity were performed as previously described (Valent *et al.* 1991). Two biological replications were performed for each strain and control (0.25% gelatin). The rice leaf sheath assay was conducted as previously described to collect samples for *in planta* RNA sequencing analysis (Mosquera *et al.* 2009). In brief, the rice leaf sheath was cut into 7 cm long sections with forceps and inoculated with around 80 µL of fungal spore solutions at the concentration of 1×10^5 spores/mL. Wild type Guy11 spores were harvested from about 3-week old plates with 0.25% gelatin solution for preparing spore suspensions. The inoculated rice leaf sheaths were supported horizontally in tray containing 1.5% agar such that the spores settled on the mid vein regions. The tray was placed at 28 °C in a growth chamber in a dark. At 36 hours post inoculation, the sheaths were hand-trimmed to remove the sides around the mid vein to enrich the fungal biomass. Three replicates of infected

rice leaf sheath were collected at 36 hours post inoculation for RNA sequencing analysis or ChIP-qPCR assay.

Western blotting assay

All strains were cultured in liquid complete medium with 150 rpm shaking at 28 °C for 2-3 days. About 200 mg of fungal mycelia were harvested and homogenized with beads mill homogenizer (OMNI international). The homogenized tissues were resuspended in 600 µL protein lysis buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM EDTA, 1% Triton X-100, 2 mM Phenylmethanesulfonyl fluoride, 100 µM Leupeptin, 1 µg/mL Pepstatin, 10 µM E-64). The lysates were centrifuged at 4°C at 12,000 rpm for 20 min. The supernatant containing proteins were collected in a fresh tube, mixed with 4x protein loading buffer (200 mM Tris-HCl (pH 6.8) 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol) and denatured at 95 °C for 5 min. The proteins were separated on 4-20% precast protein gel (Bio-Rad) and transferred onto PVDF membranes. Western blotting was performed by iBind western system (Thermo Fisher) followed manufacturer's protocol. Primary antibodies including anti-H3K27me3 (Diagenode, #C15200181, 1:1000 or 1:500) anti-H3K27ac (Abcam, #ab4729, 1:20000), and anti-H3 (Sigma, #06-755, 1:500) with secondary antibodies including Goat Anti-Mouse IgG H&L (HRP) (Abcam, #ab6789, 1:2000) or Goat Anti-Rabbit (HRP) (Epigentek, #A12004-1, 1:1000) were used for immunoblotting. Amersham™ ECL™ Prime Western Blotting Detection Reagent (GE Healthcare, #45-002-401) was used for signal detection. All Western blot assays were performed with at least two biological replicates and quantified by ImageJ (<https://imagej.nih.gov/ij/>).

RNA extraction, cDNA synthesis and RT-PCR

RNA extraction was performed using TRIzol (Invitrogen) following manufacturer's protocol. RNA quality was determined by NanoDrop spectrophotometer (Thermo Fisher). 1 µg of total RNA was used for cDNA synthesis by M-MuLv reverse transcriptase (Thermo Fisher) following the manufacturer's protocol.

ChIP sequencing

For ChIP-seq, samples from *M. oryzae* wild type and mutant *ΔMokmt6* grown in liquid complete medium were collected. Three independent biological replicates were performed for each sample in the same experiment. The ChIP procedure was performed as previously described with a modification (Sasaki *et al.* 2014). In brief, ChIP was performed 3-day-old fungal mycelia growing in liquid culture of CM. Fungal mycelia were washed twice in phosphate-buffered saline (PBS, Sigma) and dried by sterilized paper. For cross-linking, dried fungal mycelia were then transferred into a 50 mL fresh tube and incubated with 30 mL of 1% formaldehyde at room temperature on a shaker for 30 min. The reaction was quenched by adding 10% of 2.5 M glycine and incubating for 5 min. The cross-linked mycelia were washed with cold PBS for three times, dried, and flash frozen in liquid culture and stored at -80°C for further ChIP assays.

Total of 0.1 g of cross-linked fungal mycelia was used for each ChIP assay reaction. The desired amounts of fixed mycelia were transferred into 1.5 mL sonication-specific tube (Diagenode) and resuspended in lysis buffer with phosphatase inhibitor cocktail (Sigma). Cell lysis and chromatin shearing was conducted by two rounds of sonication (10 cycles, 30'' on/30'' off) on a Biorupter (Diagenode) at 4 °C. Lysates were centrifuged at 13,000 rpm for 10 min at 4 °C. The

MNase treatment was followed by adding final concentration of 2 unit of MNase and incubating at 37 °C for 15 min. The stop buffer of 0.5 M EDTA and 4M NaCl was added to quench the MNase digestion. The supernatant was collected by centrifuge and 20 µL of cell lysate was saved as input sample. A pre-clear procedure was conducted by incubating each sample with 20 µL of SureBeads protein A/G magnetic beads (1:4, v/v) (Bio-rad) for one hour to remove non-specific binding proteins. For detection of H3K27me3, H3K27ac, and H3K36me3, 2% (v/v) of antibodies of anti-H3K27me3 (Abcam, #ab6002), anti-H3K27ac (Epigentek, #A-4708-100), and anti-H3K36me3 (Abcam, #ab9050) was added to individual reactions, respectively. Following overnight incubation, protein A/G magnetic beads of 40 µL were added and incubated for three hours to bind the antibody. The beads with bound chromatin were then pelleted by magnetic rack, and washed twice by lysis buffer, once by high salt buffer, and finally by TE buffer.

Chromatin de-crosslinking and DNA purification were conducted for both input and antibody pulled down DNA by iPure Kit v2 (Diagenode). To obtain the insert DNA at a size of 150 to 450 bp, the purified DNA was sized selected by AmPure beads (Beckman Coulter) according to the dual size selection protocol. At least 4 ng of DNA from each reaction was used as input for library preparation. In brief, the DNA was first end repaired and A-tailed using NEBNext End Repair Module kit (NEB), then followed by universal dual adapter ligation (IDT). Adapter-ligated DNA fragments were enriched by PCR using Phusion Master Mix (NEB) and primers unique containing barcoding information (IDT). The PCR program was set for 8 cycles to reduce the bias caused during amplification. The cleaning up procedure was performed after each enzyme treatment by 0.8x AmPure beads in high-salt purification buffer or purification buffer. Successfully amplified libraries were size-selected and cleaned up by 0.8X AmPure beads to get rid of primer-dimer contaminants. There are total of 24 ChIP-seq libraries with three independent

biological replicates for each sample in this study. All libraries were pooled for Illumina sequencing. Single-end 75-bp sequencing was conducted using an Illumina NextSeq500 Instrument at the Integrated Genomics Facility of Kansas State University.

***In planta* ChIP-qPCR**

The infected rice leaf sheath samples for ChIP-qPCR assay were first performed the ChIP procedure as described in ChIP sequencing with a slight modification. After cross-linking, we combined five pieces of infected rice leaf sheathes as one biological replicate and fine ground the tissue in liquid nitrogen before adding the ChIP buffer. The purified input and antibody pulled down DNA were used for ChIP-qPCR assay. ChIP-qPCR was performed in CFX 96 Real-Time PCR system (BioRad), three biological replicates with two technical replicates for each primer pair were performed with SYBR Select Master Mix for CFX (Applied Biosystems). The length of all designed primer pairs was ranged from 80-120 bp, as listed in Supplementary Table A.6. ChIP-qPCR data was normalized with ‘Percent Input Method’ as described in (<https://www.thermofisher.com/us/en/home/life-science/epigenetics-noncoding-rna-research/chromatin-remodeling/chromatin-immunoprecipitation-chip/chip-analysis.html>)

Stranded RNA sequencing

The infected rice leaf sheath samples for RNA sequencing analysis were flash frozen in liquid nitrogen and stored at -80 °C for further total RNA extraction and library preparation. Stranded RNA-seq library preparation was mainly based on previous methods and TruSeq stranded mRNA sample prep guide from Illumina (Kumar *et al.* 2012; Zhang *et al.* 2017b; Zhang *et al.* 2019). After total RNA extraction, mRNA was isolated from total RNA using the Dynabeads Oligo(dT)₂₅ mRNA Purification Kit (Thermo Fisher). First-strand DNA/RNA hybrid was created by the

SuperScript III kit (Thermo Fisher) with random primers and RNaseOut (Thermo Fisher). Second-strand cDNA was synthesized by DNA Polymerase I (Thermo Fisher) and RNase H (NEB) in the second-strand buffer using dNTPs with dTTP replaced by dUTP. The double-strand cDNA was then fragmented using the Fragmentase enzyme (NEB) and used as input for library preparation as described above except 14 PCR cycles for amplification. There were total of twelve RNA sequencing libraries accounting for four samples with three independent biological replicates, including one Guy11 wild type strain, one single mutant $\Delta Mokmt6\#77$ strain, and one double mutant $\Delta Mokmt6\Delta Mogcn5\#18$ strain under *in vitro* liquid complete medium culture condition and one Guy11 wild type strain under *in planta* condition. Single-end 75-bp sequencing was conducted using an Illumina NextSeq500 Instrument at the Integrated Genomics Facility of Kansas State University.

ChIP-seq and RNA-seq data analysis

The quality of individual Fastq files derived from Illumina sequencing was checked by FASTQC (version 0.11.3; www.bioinformatics.babraham.ac.uk/projects/). Short reads were trimmed by TrimGalore (<https://github.com/FelixKrueger/TrimGalore>) to drop off the low-quality reads. Short reads for ChIP-seq and RNA-seq were mapped using Burrows-Wheeler Aligner (BWA) or STAR for ChIP-seq or RNA-seq, respectively, to the latest *M. oryzae* Guy11 genome annotated by our lab (Bao *et al.* 2017; Dobin *et al.* 2013; Li and Durbin 2009). The BWA-mem algorithm used for ChIP-seq alignments with default parameters. For STAR RNA-seq mapping, we used the output from `—quantMode GeneCounts` to summarize counts over gene features which only reports uniquely mapped reads. Peak calling was conducted using “callpeak” function in MACS2 (Zhang *et al.* 2008). ChIP-seq reads coverage was averaged, normalized and analyzed using deepTools (Ramirez *et al.* 2014). ChIP-seq summary was conducted by bedtools. The called

peaks were used to identify association to features, such as genes and TEs, by identifying where the peaks and the features overlap. Genes and TEs that were overlapped by a ChIP-peak for at least 50% of the gene or TE length were counted as ‘marked’ by that histone modification. Pearson correlation analysis was conducted to test the ChIP-seq data reproducibility among three biological replicates (Supplementary Figure A.16). Principal component analysis was conducted to investigate gene expression profiles at the whole genome level among different genotypes growing under different conditions with three biological replicates (Supplementary Figure A.17). Gene differential expression was identified using DESeq2 (Love *et al.* 2014). Genes were determined as differentially expressed if their expression had a minimum 2-fold changes with a statistical significance $p < 0.05$. The comparison was conducted on wild type under complete medium and *in planta* growth conditions or between wild type and mutants under complete medium growth conditions. We conducted genome-wide effector gene prediction by EffectorP 2.0 (Sperschneider *et al.* 2018). For functional analysis, we conducted gene ontology analysis by Blast2GO (Gotz *et al.* 2008) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for interested gene sets (Kanehisa and Goto 2000).

Phylogenetic and domain structure analysis

The protein sequences of potential PRC2 components in filamentous fungi were retrieved from FungiDB data base (<https://fungidb.org/fungidb/>). The MEGA X was used for sequence alignment and neighbor-joining tree generation with 1000 bootstrap replications (Kumar *et al.* 2018). The domain structure was analyzed by pfam (<https://pfam.xfam.org/>) and visualized by TBtools (<https://github.com/CJ-Chen/TBtools/releases>) (Chen *et al.* 2020).

Statistical analysis

All statistical analysis was conducted using the R version 3.5.3 statistical environment (R Core Team, 2019). Chi-square test was conducted using `chisq.test()` function from MASS package. Multiple comparison was conducted by `HSD.test()` function from agricolae package. Significant enrichment of overrepresented functional categories was determined using Fisher's exact test ($p < 0.05$) with adjusted p values by Benjamini–Hochberg procedure option. The ggplot2 package was used for data visualization. The Gviz package was used to visualize genome data (Hahne and Ivanek 2016). Heatmap plots were generated by heatmap3 package (Zhao *et al.* 2014).

Availability of data and material

Illumina sequence reads have been deposited at the National Center for Biotechnology Information, Short Reads Archive (BioProject accession No. PRJNA646251).

Funding

This work is supported by USDA's National Institute of Food and Agriculture (award no. 2018-67013-28492) through the Plant Biotic Interactions Program, which is jointly offered by the National Science Foundation (NSF) and the National Institute of Food and Agriculture (NIFA), and by the National Science Foundation (award no. 1936800) through Models for Uncovering Rules and Unexpected Phenomena in Biological Systems (MODULUS) program of the Division of Molecular and Cellular Biosciences – Systems and Synthetic Biology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figures

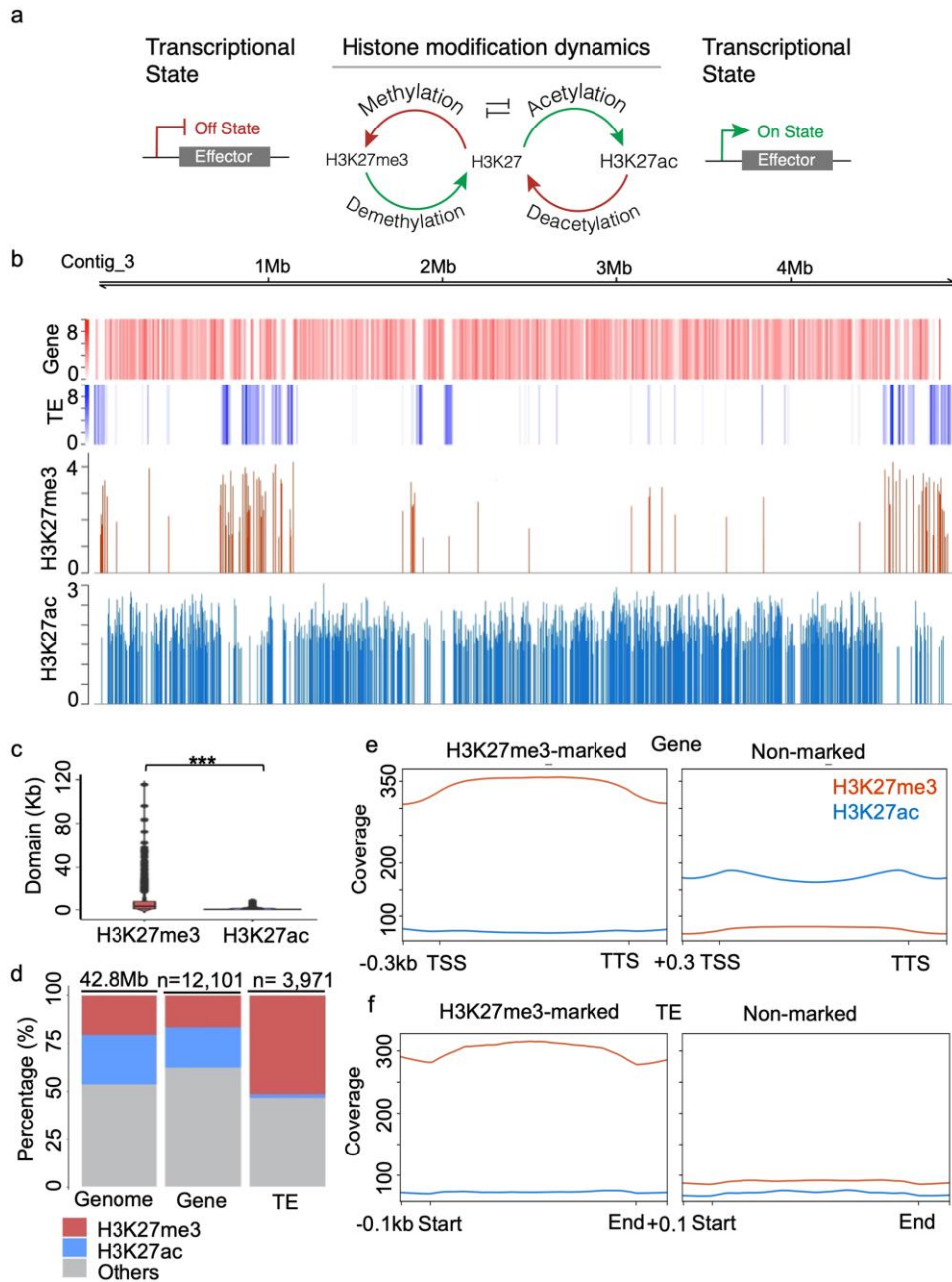


Figure 2.1. Methylation and acetylation at H3K27 occupy repressive and active genomic regions with distinct distribution profiles

(A) A working model for the dynamics of histone modification at H3K27 regulating fungal effector gene expression.

(B) ChIP signals for H3K27me3 and H3K27ac occupy distinct genomic regions in *M. oryzae* Guy11 wild type growing in complete medium. (C) Violin plots illustrating genome-wide distributions of domains of H3K27me3 and H3K27ac. (D) Bar chart illustrating the percentage of genome (Geno), gene and transposable element (TE) marked by H3K27me3 and H3K27ac in wild type grown in complete medium. (E) and (F) ChIP signal profiles of H3K27me3 and H3K27ac on genes (E) and TEs (F) that marked (left) and non-marked (right) by H3K27me3.

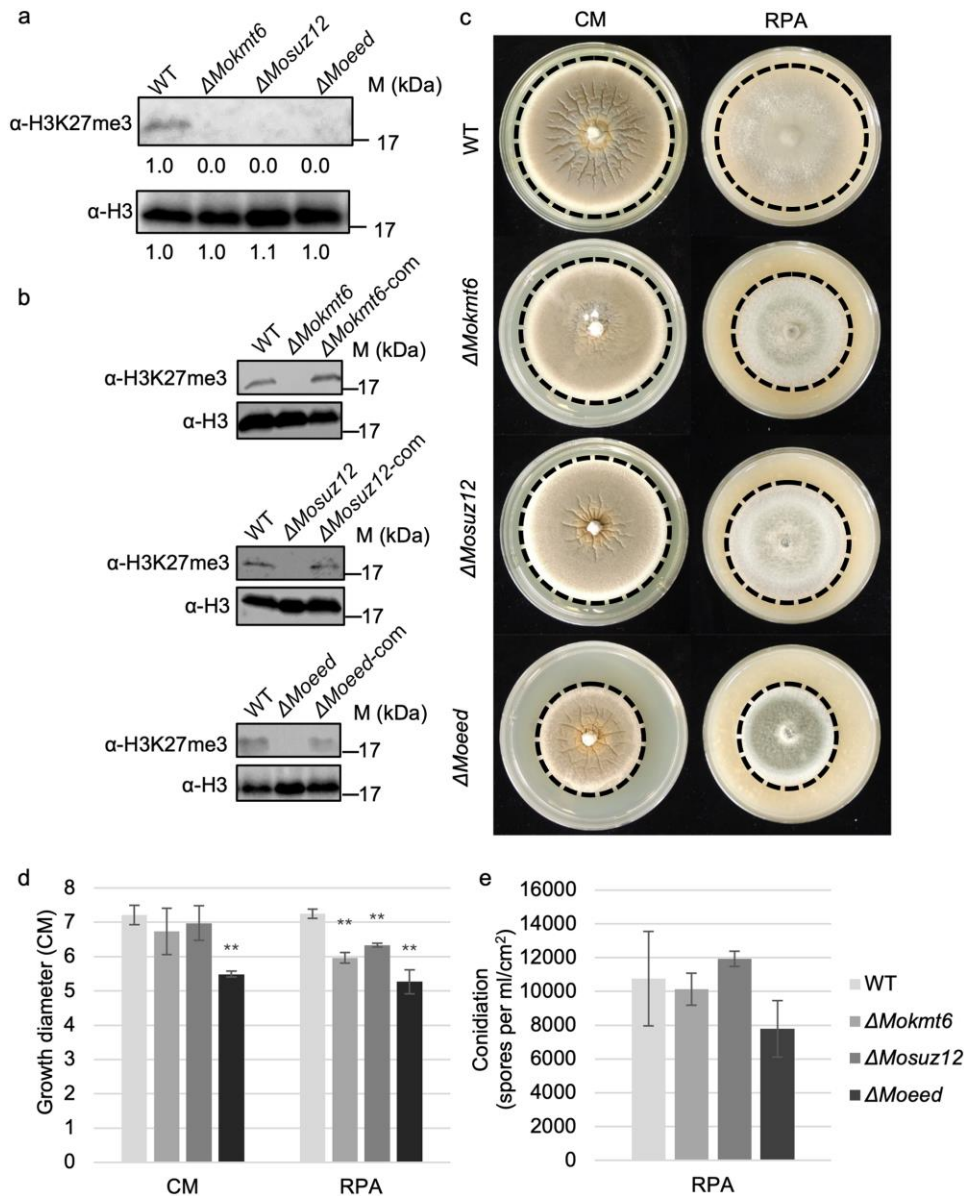


Figure 2.2. PRC2 mediated H3K27me3 is required for normal fungal *in vitro* growth

(A) Total proteins extracted from wild type (WT), Δ Mokmt6, Δ Mosuz12 and Δ Moeed were used for western blotting. The anti-H3K27me3 was used for detecting the global level of H3K27me3 in the samples and anti-H3 was used as loading control. Intensity of signals were analyzed by ImageJ. (B) Total proteins extracted from wild type, Δ Mokmt6, Δ Mosuz12, Δ Moeed, Δ Mokmt6-com, Δ Mosuz12-com and Δ Moeed-com were subjected to western blotting. (C) Colony morphology of wild type, Δ Mokmt6, Δ Mosuz12 and Δ Moeed on the complete medium agar (CM) and rice polish agar (RPA) at 12 days post inoculation. Dotted lines highlighted the edge of colonies. (D) Colony diameters were measured at 12 days post inoculation and subjected to statistical analysis.

Significance is present by double asterisks ($p < 0.01$) or single asterisk ($p < 0.05$) to indicate significant differences compared with wild type by student's t-test. (E) Production of asexual conidia in different strains were measured after culturing on rice polish agar at 10 days post inoculation under constant light.

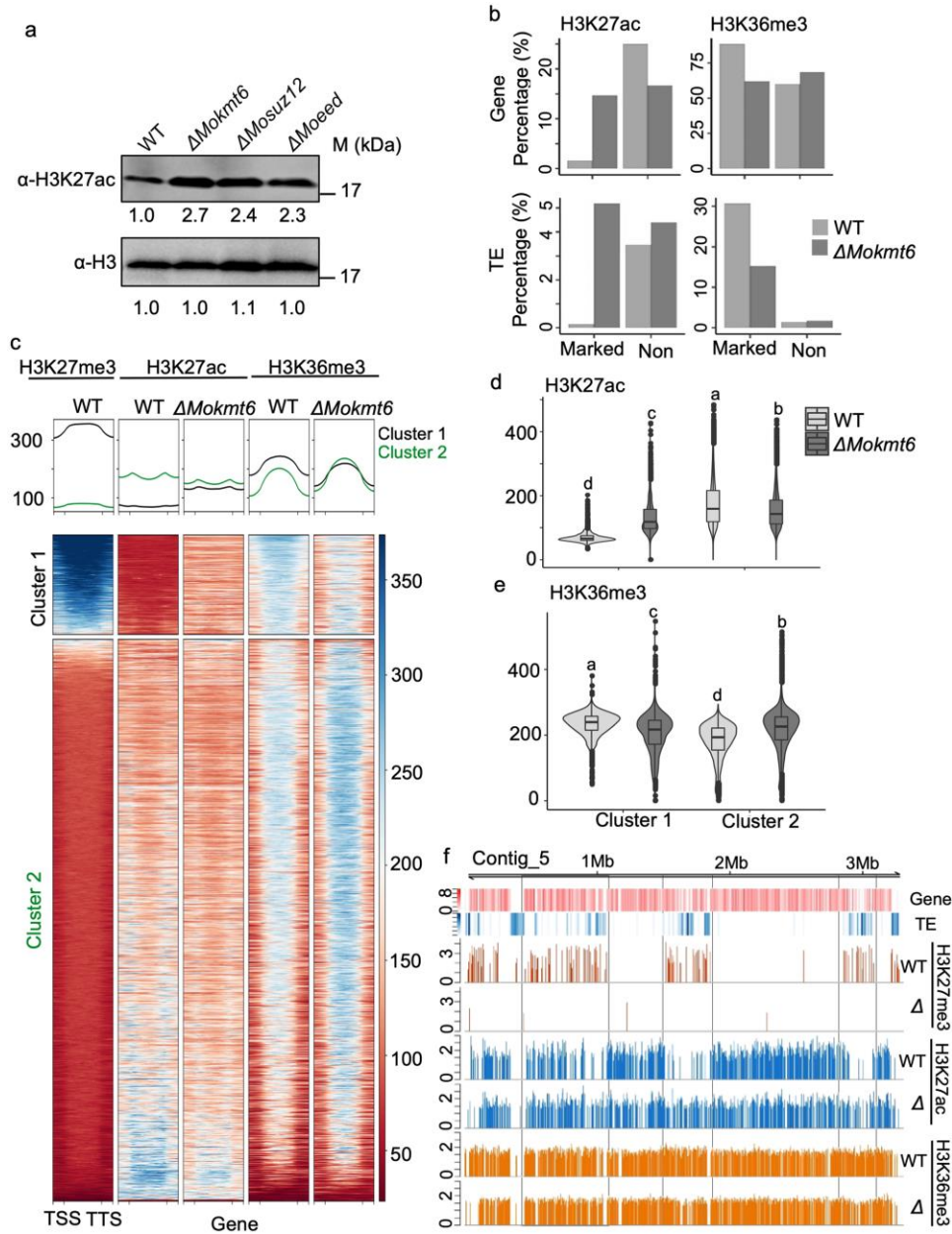


Figure 2.3. Loss of H3K27me3 leads to redistribution of H3K27ac and H3K36me3

(A) Western blotting analysis showed the global increased levels of H3K27ac in Δ Mokmt6, Δ Mosuz12 and Δ Moeed with absence of H3K27me3 compared with Guy11 wild type (WT). (B) Bar chart illustrating dynamics of H3K27ac (left) and H3K36me3 (right) marked genes and transposable elements (TEs) compared between wild type (blue) and Δ Mokmt6 mutant (red). (C) Heat maps visualizing profiles of ChIP signals for H3K27me3, H3K27ac and H3K36me3 across 12,101 genes in Guy11 wild type and mutant Δ Mokmt6. Two clusters of genes are generated by unsupervised k-means analysis and ranked according to their H3K27me3 enrichment from wild

type growing in liquid complete medium. (D) and (E) Violin plots illustrating redistribution of ChIP signals for H3K27ac and H3K36me3 in two clusters from the heatmap (C). Letters at the top indicating the significance. (F) ChIP signal showing the loss of H3K27me3 in mutant *ΔMokmt6* (*Δ*) resulted in the redistribution of H3K27ac and H3K36me3 at specific loci. ChIP signals are compared between Guy11 wild type and mutant *ΔMokmt6* growing in liquid complete medium.

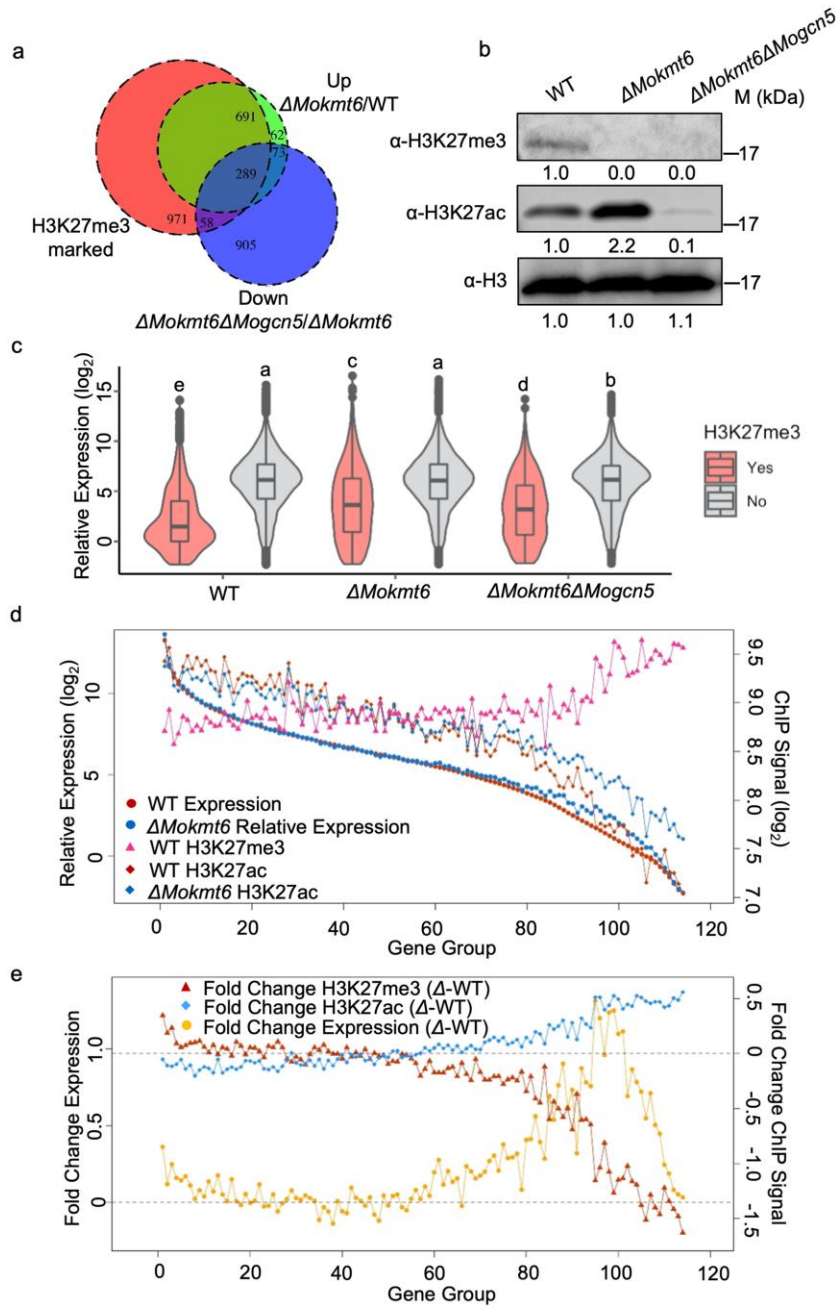


Figure 2.4. Loss of trimethylation and gain of acetylation at H3K27 contribute to H3K27me3-marked gene expression

(A) Venn diagram illustrating a subset of genes marked by H3K27me3 are up regulated in the mutant $\Delta Mokmt6$ compared with Guy11 wild type (WT) and some of them are downregulated by further loss of H3K27ac in the double mutant $\Delta Mokmt6 \Delta Mogcn5$. All genotypes are grown in liquid complete medium (CM). (B) Western blotting analysis of protein extracted from wild type (WT), the deletion mutant $\Delta Mokmt6$ with absence of H3K27me3, and the double deletion mutant $\Delta Mokmt6 \Delta Mogcn5$ with absence of H3K27me3 and reduced of H3K27ac. Blots were probed with

the indicated antibodies. (C) Violin plots showing H3K27me3-marked genes are significantly upregulated in mutants *ΔMokmt6* and *ΔMokmt6ΔMogcn5*. (D) Genome-wide visualization of relationships between ChIP signals for H3K27me3/H3K27ac and gene expression across wild type and mutant *ΔMokmt6* abolishing H3K27me3. Total of 11,327 expressed genes out of 12,101 gene models were divided into 114 groups (100 genes in one group) according to their expression levels from high to low (x axis). The dots, triangles, and diamonds represent the average levels of gene expression, ChIP signal for H3K27me3, and H3K27ac in each group, respectively. (E) Genome-wide visualization of relationships between fold change of ChIP signals for H3K27me3/H3K27ac and fold change of gene expression. The fold changes of ChIP signals or gene expression were calculated by subtracting the average levels of ChIP signals or gene expression of individual gene group in wild type (WT) from the mutant *ΔMokmt6* (Δ).

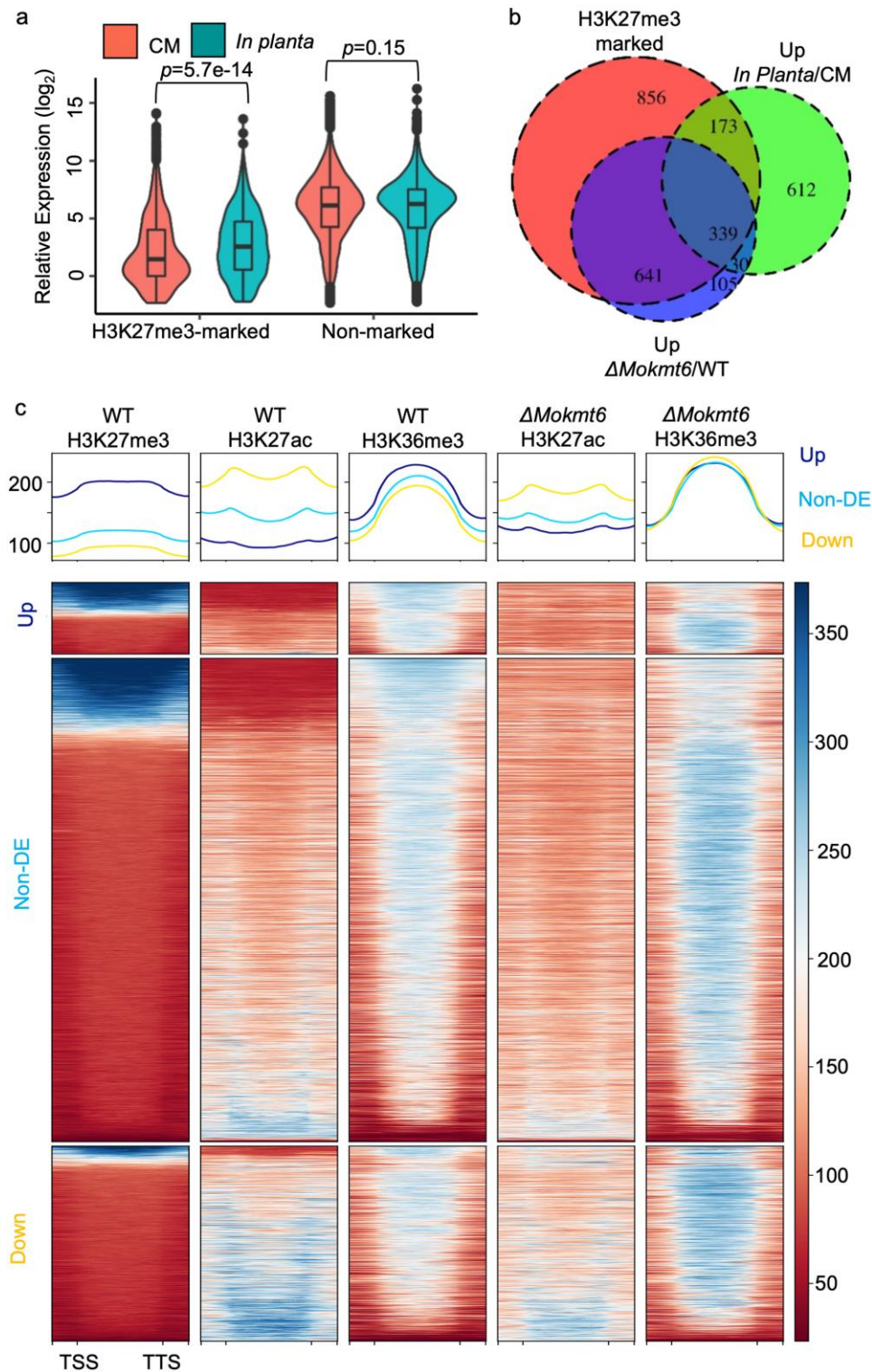


Figure 2.5. H3K27me3 marked genes are enriched for *in planta* induced genes

(A) Violin plots showed H3K27me3 marked genes identified in Guy11 wild type (WT) grown in complete medium (CM) are significantly upregulated *in planta*. (B) Venn diagram illustrating the

enrichment of H3K27me3 marked genes in upregulated genes *in planta* and in mutant $\Delta Mokmt6$ abolishing H3K27me3. (C) Heatmaps visualizing ChIP signals for H3K27me3, H3K27ac and H3K36me3 across differential expressed gene sets identified by comparison of relative gene expression from wild type growing *in planta* and liquid complete medium.

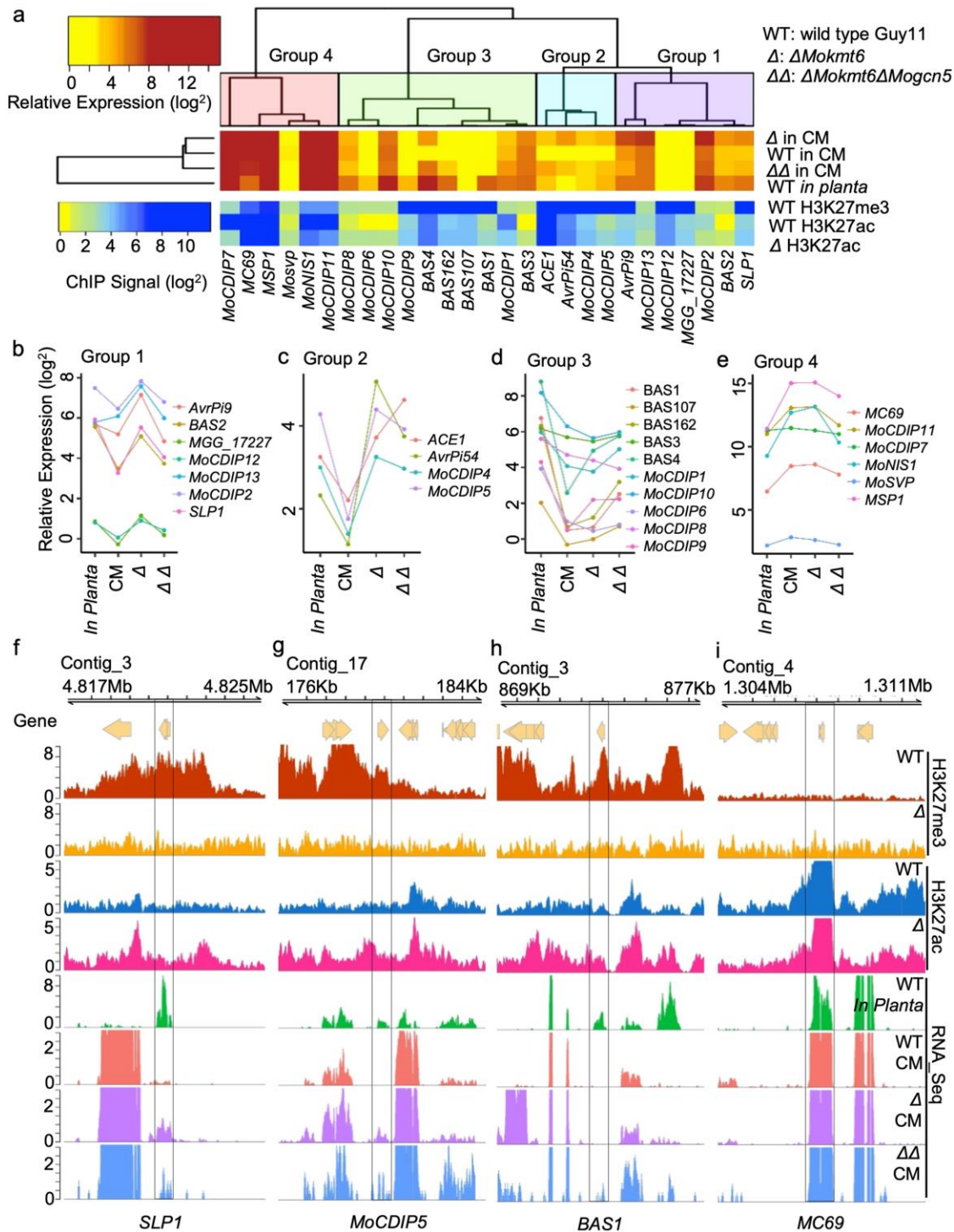


Figure 2.6. Fungal effector genes are regulated in H3K27me3-dependent and -independent manners

(A) The heatmap (top) identifying four groups of 27 characterized effector genes based on their gene expression pattern across three genotypes growing either *in vitro* complete medium (CM) or *in planta*. The heatmap (bottom) showing the status of H3K27me3 and H3K27ac for all effector

genes under the same experimental condition as their gene expression. The effector gene expression data were collected by RNA-seq from Guy11 wild type (WT) growing either *in vitro* complete medium or *in planta*, and single deletion mutant Δ *Mokmt6* (Δ) and double deletion mutant Δ *Mokmt6* Δ *Mogcn5* ($\Delta\Delta$) growing in complete medium. The Δ *Mokmt6* lacks H3K27me3 and Δ *Mokmt6* Δ *Mogcn5* lacks H3K27me3 and majority of H3K27ac. (B) to (E) showing the expression pattern of effector genes in each group. (F) to (I) showing the profiles of ChIP-seq reads (bins per million mapped reads, BPM) for H3K27me3 and H3K27ac, and RNA-seq reads (BPM) at specific effector gene loci from each of the four groups.

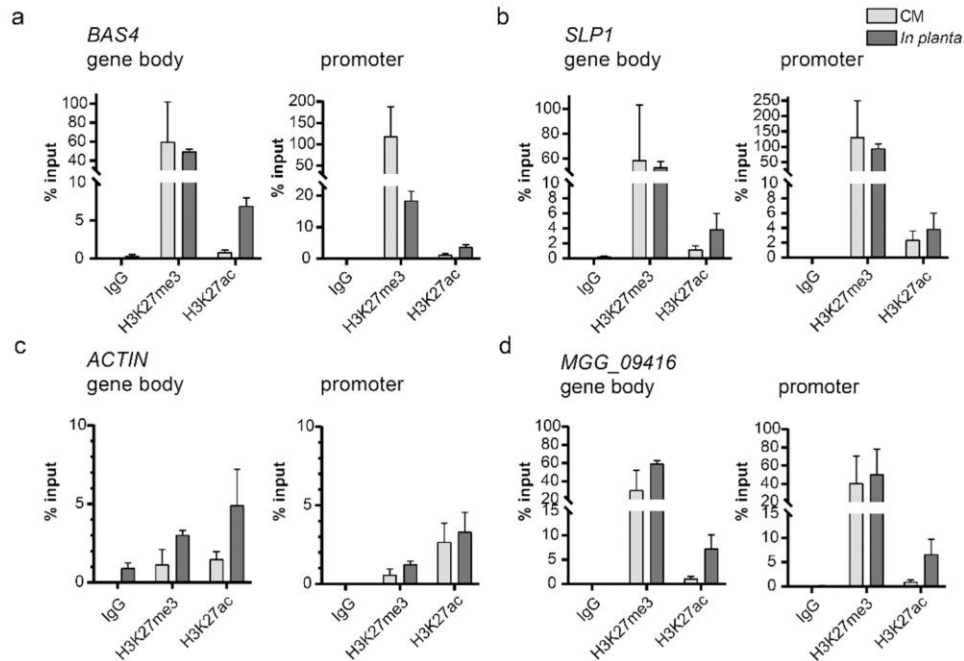


Figure 2.7. H3K27 modifications undergo dynamics at specific genomic locations *in planta*.

(A-D) The amount of DNA associated with H3K27me3 and H3K27ac was determined relative to the amount of input DNA (% Input) from *M. oryzae* during axenic growth (CM) and during plant infection (*in planta*). A control ChIP sample was conducted where the IgG antibody was used for the pull-down (IgG sample) and did not produce a significant signal for any primer pairs tested. For each gene locus, one primer pair was designed in the gene body while the was designed in the promoter (i.e. 200-300 bp in front of the transcription start site). (A) *BAS4* locus, (B) *SLP1* locus, (C) *ACTIN* locus, and (D) *MGG_09416* locus.

Chapter 3 - Repair of CRISPR-Cas12a induced DNA double-strand breaks in *Magnaporthe oryzae* generates locus-dependent mutation profiles

Note: This chapter has been posted in bioRxiv (<https://doi.org/10.1101/2021.09.08.459484>) and under the major revision in Nature Communications.

Jun Huang performed most of the experiments with the help from David Rowe, Pratima Subedi and Tyler Suelter. Jun Huang and Wei Zhang performed the Nanopore sequencing. Jun Huang performed the genome assembly and annotation in conjunction with David E. Cook. Jun Huang, Barbara Valent, and David E. Cook analyzed the experiments. All the authors contributed to writing.

Abstract

CRISPR-Cas mediated genome engineering has revolutionized functional genomics. However, basic questions remain regarding the mechanisms of DNA repair following Cas-mediated DNA cleavage. We developed CRISPR-Cas12a ribonucleoprotein genome editing in the fungal plant pathogen, *Magnaporthe oryzae*, and found frequent donor DNA integration despite the absence of long sequence homology. Interestingly, genotyping from hundreds of transformants showed that frequent non-canonical DNA repair outcomes predominated the recovered genome edited strains. Detailed analysis using Sanger and Nanopore long-read sequencing revealed five classes of DNA repair mutations, including simple donor DNA insertions, concatemer donor DNA insertions, large DNA deletions, deletions plus donor DNA insertions, and infrequently we observed INDELs. Our results show that different error-prone DNA repair pathways resolved the Cas12a-mediated double-strand breaks (DSBs) based on the DNA sequence of edited strains. Furthermore, we found that the frequency of the different DNA repair outcomes varied across the

genome, with some tested loci resulting in more frequent large-scale mutations. Our results suggest that DNA repair pathways provide preferential repair across the genome that could create biased genome variation, which has significant implications for genome engineering and the genome evolution in *M. oryzae* populations.

Introduction

The CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR associated protein 9) genome editing platform has been widely used in multiple organisms including animals, plants, and fungi for functional genomics studies (Cong *et al.* 2013; Hsu *et al.* 2014; Hwang *et al.* 2013; Liu *et al.* 2015; Mali *et al.* 2013; Shan *et al.* 2013). The basic requirement for CRISPR-Cas genome engineering is a Cas endonuclease protein complexed with a single-guide RNA targeting a genomic region following a protospacer adjacent motif (PAM), such as NGG in the case of the commonly used SpCas9 protein from *Streptococcus pyogenes* (Cong *et al.* 2013; Mali *et al.* 2013; Mojica *et al.* 2009). Another Cas effector, termed Cas12a (formerly named as Cpf1), is an alternative genome editing tool that has several unique features compared to Cas9 based effectors (Ah-Fong *et al.* 2021; Fagerlund *et al.* 2015; Ferenczi *et al.* 2017; Port *et al.* 2020; Tang *et al.* 2017; Zetsche *et al.* 2015; Zetsche *et al.* 2017). For instance, Cas12a recognizes a T-rich PAM, which can be better suited for editing some genomic regions (Zetsche *et al.* 2015). Also, the RNase activity of Cas12a can process an array (single RNA molecule with multiple guide sequences) into multiple RNA molecules of single sequences, which allows more convenient multiplex genome engineering (Zetsche *et al.* 2017). The nuclease activity of Cas12a generates staggered DNA breaks, resulting in double-stranded DNA with 5' overhangs, compared to blunt-end double-stranded DNA following Cas9 nuclease activity (Zetsche *et al.* 2015).

A critical component determining the outcome of genome engineering is DNA double-strand break (DSB) repair, mediated by endogenous DNA repair machinery (Xue and Greene 2021). Proper repair of DNA DSBs, whether induced by Cas effectors or under natural conditions, is critical to maintain genomic stability, where repair failure can result in altered genome function and be potentially lethal (Covo 2020; Her and Bunting 2018; Jackson 2002). DNA DSB repair is

mediated by two major pathways, canonical non-homologous end joining (C-NHEJ) and homology direct repair (HDR) (Ceccaldi *et al.* 2016; Chang *et al.* 2017; Pannunzio *et al.* 2018). One of the major differences between C-NHEJ and HDR is the initial processing of DNA ends at a break site, where HDR requires extensive DNA end resection (i.e., enzymatic nucleotide removal from DSB site), which is inhibited by NHEJ (Pannunzio *et al.* 2018; Truong *et al.* 2013). In the initial steps of C-NHEJ, the Ku70-Ku80 heterodimer interacts with broken DNA ends to inhibit resection (Walker *et al.* 2001), and recruits additional proteins to the site eventually repairing the DSB via DNA ligase IV (Ceccaldi *et al.* 2016; Nick McElhinny *et al.* 2000; Waters *et al.* 2014). The C-NHEJ pathway does not rely on a homologous DNA template for repair, and commonly results in small insertions and deletions (INDELs) (Ceccaldi *et al.* 2016; Pickar-Oliver and Gersbach 2019), but there are also examples of accurate NHEJ repair with and without DNA templates (He *et al.* 2016; Suzuki *et al.* 2016). For DNA DSB repair via the HDR pathway, template DNA with extended homologous sequences (typically >100 bp) are used for what is generally considered accurate repair (Moynahan and Jasin 2010). Two additional DNA DSB repair pathways, which also require end resection at DSB sites, are termed alternative end-joining (a-EJ), and single strand annealing (SSA) (Bhargava *et al.* 2016; Ceccaldi *et al.* 2016; Chang *et al.* 2017). The a-EJ pathway is also referred to as microhomology-mediated end-joining (MMEJ), theta-mediated end-joining (TMEJ), and has been called alternative NHEJ (A-NHEJ) depending on the system and report (Rodgers and McVey 2016; Schrempf *et al.* 2021). While the three pathways involving end resection rely on homologous sequence for DSB repair, the length of homologous sequence used by a-EJ, SSA and HDR is different. The a-EJ repair involves annealing microhomologous sequences (typically 2-20 bp) and gap filling by DNA polymerase theta (Pol θ) near the DSB (Sallmyr and Tomkinson 2018), resulting in small insertions, deletions and templated

insertions in mammalian and plant systems (Ceccaldi *et al.* 2016; Mara *et al.* 2019; Schimmel *et al.* 2019). The SSA pathway involves annealing with longer homologous sequences (>25 bp), often described to reside at longer distances from the DSB site and result in larger deletions as the result of removing 3' non-homologous ssDNA via Rad1-Rad10 endonuclease (Bhargava *et al.* 2016; Ceccaldi *et al.* 2016; Chang *et al.* 2017; Sallmyr and Tomkinson 2018).

Many questions remain for how the individual DNA repair pathways interact, such as their individual contributions to genome stability, their hierarchy for DSB repair, and variation in DSB repair pathways in microbial eukaryotes. There have been conflicting reports on the importance and role of a-EJ for repairing DSBs (Ceccaldi *et al.* 2016; Sfeir and Symington 2015), however, clear evidence shows that a-EJ substantially contributes to DNA repair in zebrafish embryos (Thyme and Schier 2016), mouse cell lines (Zelensky *et al.* 2017), *Caenorhabditis elegans* (van Schendel *et al.* 2015), human cells lines (Schimmel *et al.* 2017) and the model plant *Arabidopsis thaliana* (van Kregten *et al.* 2016). Interestingly, while the identification of C-NHEJ independent DSB repair was first described in *Saccharomyces cerevisiae* (Boulton and Jackson 1996b), there are no reports of TMEJ in yeast or filamentous fungi. There are also clear differences for DSB repair across fungi, such as HDR being highly active in yeast *S. cerevisiae*, while C-NHEJ predominates DNA DSB repair in most filamentous fungi (Haber 1999; Ninomiya *et al.* 2004; Villalba *et al.* 2008).

In this research, we developed efficient genome editing using Cas12a-based ribonucleoprotein (RNP) in *Magnaporthe oryzae* (synonym of *Pyricularia oryzae*), a filamentous fungal pathogen of monocots threatening world food security (Bao *et al.* 2017; Peng *et al.* 2019; Zhong *et al.* 2016). The use of CRISPR editing in fungi can increase the speed and efficiency of traditional gene replacement strategies, and the use of RNPs in *M. oryzae* can alleviate problems

related to cytotoxicity and off-target mutations (Foster *et al.* 2018; Leisen *et al.* 2020). Surprisingly, we found that Cas12a editing in *M. oryzae* resulted in numerous mutants that contained severe DNA alterations at the targeted locus. Using long-read DNA sequencing and *de-novo* assembly, we confirmed at nucleotide resolution, multiple classes of DNA mutations, suggesting the involvement of different DNA repair mechanisms. The frequency of DNA repair outcomes after Cas12a editing were found to be locus-dependent across five tested loci. These results provide a detailed report of variable DNA repair outcomes after Cas12a-RNP editing, which have significant implications for natural and induced DNA repair in fungal genomes.

Results

Cas12a ribonucleoprotein editing causes unexpected DNA mutations

To test Cas12a RNP gene editing in *M. oryzae*, we designed two gRNA targeting the *BUF1* locus that codes for a trihydroxynaphthalene reductase required for fungal melanin biosynthesis (Chumley and Valent 1990). A DNA nuclease competent RNP comprised of purified LbCas12a protein (*Lachnospiraceae* bacterium ND2006) and *BUF1*-gRNA1 or -gRNA2 were transferred with donor DNA coding for the hygromycin resistance gene (*HYG*) into *M. oryzae* field isolate O-137 using protoplast transformation (Figure 3.1a) (Giraldo *et al.* 2013). The donor DNA contained short (30 and 35 bp) flanking sequences at the ends, homologous to the *BUF1* locus, to direct microhomology-mediated end joining (MMEJ donor DNA integration following Cas12a DNA DSB) (Figure 3.1b). Transformed protoplasts were recovered on hygromycin selection, and hygromycin resistant (*HYG^R*) colonies were subsequently transferred to non-selective OTA plates to test for altered mycelial pigmentation. Across the experiments using both gRNA, the *BUF1* locus was edited at a rate of ~77% (81 buff mutants/105 total *HYG^R* transformants). Among the

48 HYG^R transformants from *BUF1*-gRNA1, 43 of them displayed buff phenotype, and 38 of the 57 HYG^R transformants from *BUF1*-gRNA2 showed the buff phenotype (Figure 3.1c, Supplementary Figure B.1 and Supplementary Table B.1). All 32 control transformants transformed with donor DNA alone exhibited wild-type hyphal pigmentation (Supplementary Figure B.2 and Supplementary Table B.1). To genotype and confirm Cas12a-mediated editing of *BUF1*, PCR was used to discriminate wild-type (~1.5 kb product) versus one-copy *HYG* donor insertion (~3.1 kb product) or other larger PCR product corresponding to the integration of *HYG* DNA donor, which we refer to as a ‘simple insertion’ (Figure 3.1b, d). The Cas12a-mediated DSB could also be repaired to create an INDEL resulting in a PCR product indistinguishable from the recipient wild-type, while HYG^R transformants with wild-type pigmentation were presumed to have integrated the selectable marker at a secondary locus (Figure 3.1d). Interestingly, ~93% of transformants (40/43) that displayed the mutant buff phenotype generated with *BUF1*-gRNA1, and almost 87% generated with the *BUF1*-gRNA2 (33/38), failed to produce a *BUF1* PCR product (Figure 3.1e, Supplementary Figure B.1 and Supplementary Table B.1). The other 2/43 buff mutants generated with *BUF1*-gRNA1 and 4/38 buff mutants with *BUF1*-gRNA2 produced a clear PCR band consistent with simple insertion of the *HYG* coding sequence (Figure 3.1e, Supplementary Figure B.1 and Supplementary Table B.1). Additionally, INDEL was detected in 1/43 buff mutant from *BUF1*-gRNA1 and 1/38 buff mutant from *BUF1*-gRNA2, which displayed the wild-type sized PCR amplifications (Figure 3.1e, f, Supplementary Figure B.1 and Supplementary Table B.1). All the recovered transformants displaying a wild-type hyphal color produced the anticipated wild-type sized PCR product (Figure 3.1e, Supplementary Figure B.1). The *BUF1*-gRNA1 targeted an intron and could therefore have generated mutants that failed to produce a visible phenotype. To assess this, transformants generated with *BUF1*-gRNA1 that had

a wild-type hyphal color and wild-type PCR amplicon were Sanger sequenced, which showed that all strains with normal pigmentation except for one transformant from replication2 had wild-type *BUF1* sequence (Figure 3.1e, f, Supplementary Figure B.1 and Supplementary Table B.1). In order to test if our observed unexpected negative PCR from buff mutants was isolate-specific, we repeated the experiments in a different rice blast field isolate, termed Guy11 (Bao *et al.* 2017). Using *BUF1*-gRNA1, we found 30 HYG^R transformants all showed the buff phenotype. For *BUF1*-gRNA2, we recovered 32 HYG^R transformants, of which 30 had a buff phenotype. Further PCR genotyping these transformants showed that 27/30 buff mutants (*BUF1*-gRNA1) and 26/30 buff mutants (*BUF1*-gRNA2) failed to produce a PCR product from the *BUF1* locus (Figure 3.1g, Supplementary Figure B.3 and Supplementary Table B.1). Therefore, the results are consistent between the two strains, and indicate the observations are isolate-independent.

We had anticipated that the majority of mutants would have a simple insertion mediated by the homologous sequence on the donor DNA, but this only occurred in a total of ~9% of the O-137 and Guy11 *buf1* mutants (13 Δ *buf1* mutants with ‘simple insertion’/142 Δ *buf1* mutants) (Supplementary Table B.1). Furthermore, the majority (~89%, 126 Δ *buf1* mutants with PCR negative/142 Δ *buf1* mutants) of repair events resulted in the inability to generate a PCR amplicon, suggesting a severe DNA alteration (Supplementary Table B.1). From these experiments, of the 142 Δ *buf1* mutants, we only recovered three Δ *buf1* (~2%) INDEL mutations (Figure 3.1f, g and Supplementary Table B.1).

Microhomology can mediate donor DNA integration

Given that a majority of our buff/*buf1* mutants did not have a simple insertion of donor DNA, we hypothesized that donor DNA lacking extended genome homology can be used to repair Cas12a-mediated DSB. That is, if the ~30 bp of homologous sequence had directed efficient

integration of the donor DNA, we would have observed a higher frequency of simple insertions. To test this hypothesis, we again targeted the *BUF1* locus using Cas12a-RNP and the two gRNAs targeting, but supplied a *HYG* donor DNA that lacked flanking sequences homologous to the *BUF1* locus (no-homology *HYG*) (Figure 3.2a). Transforming strain O-137, we obtained 74 and 53 *HYG*^R transformants, from *BUF1*-gRNA1 and *BUF1*-gRNA2, respectively. We could confirm 58 of 74 (~78% for *BUF1*-gRNA1) and 44 of 53 (~83 for *BUF1*-gRNA2) *HYG*^R transformants had a mutation at *BUF1*. When we considered these two guides together, we found that ~73% (74/102) of the edited strains were buff colored and produced no PCR product, ~19% (19/102) were buff color and had a simple insertion of DNA and ~6% (6/102) were wild-type color but had an intron INDEL from gRNA1, ~3% (3/102) contained INDEL from gRNA2 (Figure 3.2b, c, f, Supplementary Figure B.4, 5, 7a, b and Supplementary Table B.2). We additionally transformed strain O-137 with the two RNPs simultaneously (*BUF1*-gRNA1 and *BUF1*-gRNA2) and the no-homology *HYG* DNA donor (Figure 3.2d), which resulted in a 73% editing frequency (44 *Δbuf1* mutants/60 *HYG*^R transformants), where 70% (31/44) produced no PCR product, 25% (11/44) had a simple insertion and 5% (2/44) has INDEL mutations (Figure 3.2d, e, f, Supplementary Figure B.6a, b, 7c and Supplementary Table B.2). To test whether the above observation is isolate-specific, both gRNAs were used to repeat the experiments in the Guy11 strain producing a ~72% editing efficiency (64 *Δbuf1* mutants/89 total *HYG*^R transformants), where 52% (33/64) produced no PCR product, 44% (28/64) had a simple insertion, and three of the 89 *HYG*^R transformants had an INDEL at *BUF1* (Figure 3.2g, Supplementary Figure B.8, 10 and Supplementary Table B.2). For the dual RNP transformation ~76% (26 *Δbuf1* mutants/34 *HYG*^R transformants) had a mutation, where 81% (21/26) showed the PCR negative genotype and 19% (5/26) had a simple insertion (Figure 3.2g and Supplementary Figure B.9a, b and Supplementary Table B.2). Through

Sanger sequencing randomly selected simple insertion transformants generated with *BUF1*-gRNA1, we found frequent ~2bp microhomology between no-homology *HYG* DNA donor and *BUF1* locus at the integration junction (Figure 3.2h). Additionally, to rule out that our observations are dependent on the *HYG* donor DNA and hygromycin selection, we performed the same experiments with a different donor DNA sequence coding for resistance to the drug G418 (Geneticin). We again recovered ~44% (32/72) G418 resistant (G418^R) transformants with the buff mutant phenotype and found that 72% (23 /32) were PCR negative and 22% (7/32) had PCR amplification that suggested a simple insertion. We also recovered two transformant (2/32) that carried a *BUF1* INDEL from guide2 through tandem duplication or INDEL (Supplementary Figure B.11a, b, c, d, g, h and Supplementary Table B.2).

From these experiments using a donor DNA with no-homologous sequence, we conclude that i) donor DNA does not require extended homologous sequence to resolve Cas12a-mediated DSB; ii) non-homologous donor DNA can integrate at DSB sites at a reasonably high frequency (26% across all experiments, 70/268) ; iii) more severe DNA alterations resulting in no PCR products are common (68% across experiments, 182/268); iv) INDELS were not common from these experimental conditions (6% across experiments, 16/268); v) targeted DNA mutation at *BUF1* is dependent on the Cas12a RNP complex, as 70 transformants obtained using no-homology *HYG* or *G418* donor DNA alone (i.e., in the absence of Cas12a RNP) did not cause the buff phenotype and among 50/70 that were PCR tested, none showed distinguishable *BUF1* DNA mutations (Figure 3.2f, g and Supplementary Figure B.6c, d, e, 9c, d, 11e, f, g and Supplementary Table B.2).

Long-read sequencing and *de novo* assembly resolve genotypes following Cas12a-mediated DSB repair

We sought to further understand what DNA mutation occurred in the roughly 70% of buff mutants that failed to produce a PCR product at the *BUF1* locus. Eight O-137 derived buff mutants from the Cas12a RNP and no-homology *HYG* DNA donor transformation were selected for high-molecular-weight DNA extraction, Nanopore sequencing and *de novo* assembly (Δ buf1#2, -#4, -#5, -#6 from rep 1 in Figure 3.2c, Δ buf1#10 from Figure 3.2e, Δ buf1#1, -#5, -#13 from rep 4 in Supplementary Figure B.4c). All eight sequenced strains displayed the mutant buff color, where seven produced no PCR product, and one transformant (Δ buf1#2 Figure 3.2c) produced the ~3.1 kb PCR product we inferred to be a simple insertion of donor DNA. The eight strains were sequenced to an average depth of 52x and yielded highly contiguous assemblies (average N50 of 3.29 Mb) (Supplementary Table B.3). These assemblies allowed for base pair resolution interrogation of *BUF1* DNA alterations. Consistent with PCR genotyping, the transformant thought to have a simple donor DNA insertion (Δ buf1#2 Figure 3.2c) indeed had an almost full copy of the hygromycin coding sequence (1,328 bp) plus an additional hygromycin fragment (140 bp) at the Cas12a cut site (16 bp after the PAM sequence) based on the long-read assembly (Figure 3.3a, mutant 1). The insertion was nearly scar free, with the junction sequence showing 2 and 3 bp of microhomology at the 5' and 3' ends respectively, and only two base pairs were deleted at the Cas12a endonuclease site (Supplementary Figure B.12). The results from the other seven assemblies were grouped into one of three categories, namely, large insertion, large deletion, and deletion plus insertion (Figure 3.3b, c, d). Two mutants (Δ buf1#5, -#6 from rep 1 in Figure 3.2c) each contained large insertions of concatemer *HYG* donor sequence, including promoter, coding, and terminator sequences, totaling 10 and 17 kb insertions respectively (Figure 3.3b, mutant 2, 3).

Not all the *HYG* DNA fragments were intact, and the coding sequences were in both the forward and reverse orientation (Figure 3.3b). The insertion junction for both large insertion mutations had 2 bp of microhomology at the 5' end, and no homology at the 3' end, where in one mutant there was an error-free insertion at the locus, while the other mutant had a 17 bp deletion at the 3' end (Supplementary Figure B.12).

Assemblies from two *Δbuf1* mutants (Rep1-*Δbuf1*#4 and Rep4-*Δbuf1*#5) identified that same ~21 kb deletion around the *BUF1* locus, where *BUF1* and eight additional genes were all deleted (Figure 3.3c, mutant 4, 5). Interestingly, the assemblies indicated the deletions took place between similar flanking non-LTR retrotransposons, which appear to be nested insertions of a LINE element, termed MGL, inserted into a hybrid LINE element termed MINE (Fudal *et al.* 2005; Hamer *et al.* 1989) (Supplementary Figure B.13). The deletions suggest that homology between the two elements was used to resolve the break, potentially through the SSA pathway, resulting in a single retrotransposon copy and the 21 kb deletion (Figure 3.3c, mutant 4, 5, Supplementary Figure B.13). The *HYG* coding sequence for these two mutants was identified at independent loci on other chromosomes (Supplementary Figure B.14a, b). In mutant 4, the *HYG* insertion was a large concatemer of ~20 kb, while the other deletion mutant had two *HYG* copies inserted (Supplementary Figure B.14a, b). To confirm the assembly-based deletion, a ~6.7 kb PCR product that spanned the break resolution junction (MGL/MINE) was amplified in mutant 4, which failed to amplify a product in the wild-type as expected (Supplementary Figure B.15).

The remaining three mutants (Rep4-*Δbuf1*#1, Rep1-*Δbuf1*#10 and Rep4-*Δbuf1*# 13) had both *BUF1* locus deletions, ranging in size from 3 to 11 kb on either the 5' or 3' side of the Cas12a endonuclease site, along with large insertions of concatemer donor DNA (Figure 3.3d, mutant 6,7,8). The assemblies for mutants 7 and 8 did not completely resolve the *BUF1* locus in a single

assembled contig. Here, we identified two contigs in both mutant assemblies with sequence homology to the *BUF1* locus, which were joined and analyzed. Interestingly, these results indicated the insertion of additional genomic DNA from other regions of the genome, including the insertion of coding sequences (Figure 3.3d). As noted for the other mutation types, we also observed microhomology (2-3 bp) at three of the four integration junction sites from mutants 7 and 8 (Supplementary Figure B.16). In order to validate these mutation outcomes and support the assembly results independently, Nanopore long-reads were mapped to the O-137 reference. Here we find a correspondence between the assembly-based deletions and a loss of read-coverage to the O-137 genome (Figure 3.3e). These results support that large deletion and deletion plus insertion mutants lost DNA corresponding to the *BUF1* locus consistent with the *de novo* assemblies. Given that more than half of the identified buff mutants across our experiments failed to produce PCR products, we were interested to use the assembly identified genotypes (large insertion, large deletion and deletion plus insertion) to screen the previously PCR negative transformants. For this, we designed primer pairs to amplify small fragments at the 5' upstream and 3' downstream of the *BUF1* locus (Supplementary Figure B.17a). Our results showed that 64% (67/105) of O-137 derived PCR negative transformants with no-homology *HYG* DNA donor and *BUF1* RNP had a large insertion (i.e., both 5' and 3' amplified PCR products), 18% (19/105) had a large deletion (i.e., both 5' and 3' PCR failed) and 18% (19/105) had a deletion plus insertion (i.e., PCR amplified product at either 5' or 3' end) (Supplementary Figure B.4, 5b, c, 6 a, b, 17a,b, d and Supplementary Table B.2). We confirmed these results by also genotyping the transformants generated using the Guy11 derived strain, where we found that 81% (44/54) had a large insertion, 7% (4/54) had a large deletion and 11% (6/54) had a deletion plus insertion (Supplementary Figure B.8d, f, 9b, 17a, c, d and Supplementary Table B.2).

DNA repair outcomes following Cas12a editing differ between multiple loci

The *BUF1* locus has been characterized as unstable (Chumley and Valent 1990; Farman 2002), and we were interested to understand if the unexpected DSB repair outcomes found at *BUF1* were representative of other loci in the genome. Wild-type *M. oryzae* is sensitive to the drug FK506, while disruption of the corresponding receptor, *FKBP12*, causes insensitivity to FK506 (Supplementary Figure B.18) (Marroquin-Guzman and Wilson 2015; Yu *et al.* 2014). Therefore, we targeted *FKBP12* using Cas12a RNP and two separate gRNAs (*FKBP12*-guide1 and-guide2) and utilized sensitivity to FK506 to identify mutants (Figure 3.4a). In order to compare the results to those from *BUF1*, *FKBP12* editing included the no-homology *HYG* DNA donor and hygromycin selection (Figure 3.4b and Supplementary Table B.4). We obtained 84 *HYG*^R colonies, of which 33 (~39% editing efficiency) were insensitive to FK506 and presumably carried a mutation at the *FKBP12* locus, while none of the 20 no-homology *HYG* DNA alone transformants showed FK506 resistance (Figure 3.4b, Supplementary Figure B.19 and Supplementary Table B.4). The same PCR amplification strategy was used to genotype the FK506 insensitive mutants, and we found that ~30% (10/33) had a simple insertion, ~64% (21/33) had a large insertion, ~3% (1/33) had a deletion plus insertion and ~3% (1/33) contained INDEL (Figure 3.4a, c, Supplementary Figure B.19a, b, c, d and Supplementary Table B.4). None of the FK506 insensitive strains had PCR genotyping indicated that no large deletions took place (Figure 3.4a, c, Supplementary Figure B.19a, b, c and Supplementary Table B.4).

We additionally tested three other loci, one coding for an annotated plasma membrane iron permease (*FTR1*), which is 50 kb away from the *BUF1* locus, an apoplastic secreted protein *BAS4*, and an avirulence protein *AVRPI9*, both of which are presumed to help *M. oryzae* facilitate host infection (Dean *et al.* 2005; Mosquera *et al.* 2009; Wu *et al.* 2015). Two independent gRNAs were

designed for each of the three loci, which were transformed as Cas12a-RNPs with the no-homology *HYG* DNA donor (Figure 3.4b). From editing *FTR1*, 60 *HYG*^R transformants were selected and genotyped. This gene had a high editing efficiency at 60% (36/60), and of the strains carrying a mutation, we found that ~8% (3/36) had an INDEL, 50% (18/36) had a simple insertion, ~31% (11/36) had a large insertion, ~11% (4/36) had a deletion plus insertion and no large deletions were recovered (Figure 3.4b, c, Supplementary Figure B.20 and Supplementary Table B.4). The editing efficiency at *BAS4* was 25% (15 Δ *bas4* mutants/60 *HYG*^R transformants), where ~6% (1/15) of the mutants had an INDEL, ~46% (7/15) had a large insertion, ~6% (1/15) resulted from a large deletion, 40% (6/15) from a deletion plus insertion (Figure 3.4b, c, Supplementary Figure B.21 and Supplementary Table B.4). Interestingly, no simple insertions were recovered from the *BAS4* transformants using either of the two guides, despite simple insertion mutants being commonly found for the *BUF1* (19%), *FKBP12* (30%), and *FTR1* (50%) loci. Also of note, *BAS4* editing resulted in near half of the identified mutants having a deletion plus insertion or large deletion, which was not observed for other tested loci. The editing efficiency at the *AVRPI9* locus was ~30% (18 Δ *avrpi9* mutants/60 *HYG*^R transformants), and of these 18 mutants, nine resulted from a large insertion, six had a simple insertion, one contained deletion plus insertion and the other two lines carried INDELs (Figure 3.4b, c and Supplementary Figure B.22). While we did not exhaustively test gRNA at all loci, the results using two independent gRNA at each locus suggests that editing efficiency is not the same across the genome, which has been reported in other organisms (Creutzburg *et al.* 2020). An unexpected and novel finding, however, was that the spectrum of mutations resulting from DSB repair did not occur at equal proportions for the tested loci. Indeed, formal testing of the different edited loci into the five classes of DNA mutations indicated that the highly significant association between the loci and DNA mutation outcomes,

indicating a non-random association between repair mutation outcome and specific loci (Fisher's exact test, two-sided, p-value=0.0009995).

Among the loci we tested, *BAS4* showed the most dramatic mutation pattern, with more than half the mutations being large-scale deletions (i.e., large deletion or deletion plus insertion). In order to investigate the potential mechanisms of deletion at *BAS4* locus and compare the detailed mutational profiles to *BUF1*, we additionally sequenced two deletion plus insertion and one large deletion Δ *bas4* mutants with Nanopore long-read sequencing followed by *de-novo* assemblies (Δ *bas4*#2, -#3 and -#11 from rep1 in Supplementary Figure B.21a, Supplementary Table B.3). Consistent with the deletion plus insertion mutants sequenced at *BUF1* locus, we found the concatemer insertion of donor DNA up to 23 kb accompanying with *BAS4* locus deletions (Figure 3.5a and Supplementary Table B.3). Surprisingly, raw read coverage data suggested all the DNA sequences (~56kb) from telomere to *BAS4* targeting site have been deleted in these two deletion plus insertion mutants (Figure 3.5c). ~9kb additional genomic DNA insertion was found in one of these two mutants (Figure 3.5a, mutant2- Δ *bas4*#3). As observed at the *BUF1* locus, 3 bp and 6 bp microhomology shared between *BAS4* locus and donor DNA was also found in the repair junction (Figure 3.5a). Assembly and read coverage from the large deletion Δ *bas4* mutant both identified that ~17 kb deletion between two identical LTR retrotransposon *Copia*, which is similar to the large deletion observed at *BUF1*, even though the flanked repeated sequences are different (Figure 3.5b and c). ~21 kb concatemer donor DNA insertion was identified in a different chromosome (i.e., chr3) in this large deletion mutant. Sequencing and assemblies from the *Bas4* locus suggested that the subtelomeric distribution of *BAS4* (56 kb away from telomere) might contribute to the predominant deletion outcomes observed in this locus. In addition to the large DNA deletion between MGL/MINE elements at *BUF1*, a similar deletion between repetitive DNA

suggests that large deletions flanked by repeated DNA could be a more general repair outcome across the genome.

Locus-dependent DNA mutation frequency still occurs under a different editing scheme

Our initial editing experiments strongly favored donor DNA integration at the Cas12a-edited locus because of the induced DSB and selection on hygromycin. Given that this method was used for all the edited loci in this work, bias introduced by this experimental set-up should be similar among the loci, and not account for the observed differences. Nevertheless, to further show that the locus-dependent DNA mutation frequencies were not dependent on *HYG* insertion at the target locus, we devised a second editing scheme. Here, we developed an assay where the donor DNA for selection was targeted to a separate non-coding locus, termed second-site (SS), by transforming with two distinct RNPs at once (Figure 3.6a). This approach should capture DSB repair outcomes for the locus of interest (i.e., primary-site), without the need for donor DNA integration at the target locus, thereby decoupling donor DNA integration and selection from DSB repair outcomes. To target the *HYG* donor DNA to the second site, we added long DNA sequence (730 bp and 518 bp) homologous to the second-site targeted site to the *HYG* donor DNA, which has been shown to increase homology directed repair, and thereby increase the frequency at which *HYG* inserted at the second-site and not the primary site of interest (Figure 3.6a). Using this approach, we found high editing efficiency for the *BUF1* locus (89%, 25 $\Delta buf1$ mutants/28 *HYG*^R transformants) (Figure 3.6b and Supplementary Table B.5), which was similar to the earlier results obtained (~78% with *BUF1*-guide1). The proportion of the five types of DNA mutations at *BUF1* were 28% (7/25) INDELs, 16% (4/25) simple insertion, 24% (6/25) large insertion, 32% (8/25) large deletions, while no deletion plus insertions were detected (Figure 3.6c, Supplementary Figure

B.23 and Supplementary Table B.5). These editing experiments returned substantially more INDEL mutations at *BUF1* than initially observed, consistent with the donor DNA integrating at the SS locus and not requiring it for DSB repair. Interestingly, the majority of recovered mutants still indicated either donor DNA insertion or large DNA deletion at the *BUF1* locus. The editing efficiency for the *FKBP12* locus was low, with only one mutant recovered from 41 transformants (Figure 3.6b). The editing efficiency for this locus was also low for the single RNP assay, but it was not clear why it was even lower for this SS editing scheme. The one *Δfkbp12* mutant contained a large insertion at the locus (Figure 3.6c and Supplementary Figure B.24). For the *BAS4* locus, we obtained ~33% editing efficiency (8 *Δbas4* mutants/24 HYG^R transformants), of which, 25% (2/8) were INDELs, 12.5% (1/8) were large insertion, 37.5% (3/8) were large deletions, and 25% (2/8) were deletion plus insertions, (Figure 3.6b, c, Supplementary Figure B.25 and Supplementary Table B.5). Similar to initial single RNP assay, more than half of the *Δbas4* mutants were large deletion or deletion plus insertion mutations. The *FTR1* locus had a lower editing efficiency (9 *Δftr1* mutants/24 HYG^R transformants) for the second-site assay compared to the initial single RNP editing with gRNA1 (38% versus 83%, respectively). Despite the reduction in editing efficiency and increased INDELs, the proportion of mutation outcomes was quite similar between the two editing schemes, where the second-site assay showed ~44% (4/9) INDELs, ~22% (2/9) simple insertions, ~22% (2/9) large insertions, and ~11% (1/9) deletion plus insertion mutations (Figure 3.6b and c, Supplementary Figure B.26 and Supplementary Table B.5). In neither experiment did we recover large deletion mutants at *FTR1*. We additionally genotyped the second-site for each mutant strain and found PCR negative results for the second-site in most of cases (Supplementary Figure B.23, 24, 25 and 26). This suggested that the addition of long homologous sequences did direct the donor DNA to the second-site, but it did not provide precise homology

directed repair in the form of a single donor DNA insertion. As controls, the second-site was also edited with Cas12a SS-RNP and second-site flanked *HYG* DNA donor in the absence of the primary site RNP and similar PCR negative results were found (Supplementary Figure B.27a, b and Supplementary Table B.5). Also, second-site flanked *HYG* donor DNA was transferred alone, without RNPs, and *HYG*^R transformants did not show the buff mutant color or have FK506 resistance, nor did the transformants produce aberrant PCR products when amplifying the second-site locus, showing the mutations at these two primary sites, and the second-site were dependent on Cas12a-mediated DSB induction (Supplementary Figure B.27c, d and Supplementary Table B.5). We cannot completely rule out the possibility that difference of gRNAs we used caused the variation in the editing frequency and if there are any types of error-free, it is hard to be detected by the current methods. However, these results show that under a different editing scheme, which did not require donor DNA integration at the editing site of interest, we still observed a biased DNA mutation frequency following DSB repair at the tested loci.

Discussion

CRISPR-based genome engineering has accelerated functional genomic studies by providing a flexible platform to rapidly modify DNA and probe basic cell and molecular biology (Cong *et al.* 2013; Zetsche *et al.* 2015). Here we describe the development of an efficient and robust approach to modify specific loci in fungi using the Cas12a nuclease, delivered as an RNP. The use of RNPs for genome editing has the advantage of not requiring the integration and continual expression of the CRISPR-Cas system (Leisen *et al.* 2020; Wang *et al.* 2018; Wang and Coleman 2019). This can overcome cytotoxicity, a reported attribute of continual expression of Cas nucleases in some systems including fungi, and is especially helpful for asexually reproducing fungi, where the CRISPR-Cas DNA cannot easily be removed through crossing (Foster *et al.* 2018;

Jacobs *et al.* 2014). The strains generated in this study were made with a high success-rate, contain heritable mutations, and lack coding sequence of the CRISPR-Cas platform. We anticipate the described approach to deliver Cas12a RNPs to fungal protoplasts will be fungal species-agnostic and provides a rapid approach to generate gene disruption mutations, especially for recalcitrant loci (e.g., pathogen effectors) (Collemare *et al.* 2008; Selin *et al.* 2016).

Our results highlight an unresolved question at the interface of genome evolution and genome biology, which is how does hierarchy and cross-talk between endogenous DNA DSB repair pathways influence genome variation? Using a combination of PCR, Sanger sequencing and long-read sequencing-based assemblies, we show at base pair resolution how Cas12a-induced DNA DSBs can be variably repaired to generate a spectrum of DNA mutations. We observed INDELs, simple donor DNA insertions, large concatemer DNA insertions, large genomic deletions, and deletion plus insertion events, some of which resulted in drastic mutations at the targeted loci. It should be noted that error-free DSB repair might also exist in some unedited strain, while it is hard to distinguish them with the current techniques. It is difficult to determine the exact DNA repair pathway that resulted in each specific DNA mutation in the present work, however, there are mutation signatures of specific pathways that collectively suggest our mutation results were caused by at least three separate DNA repair pathways. This is based on the following i) we observed small insertion/deletion (INDEL) mutations with random sequence (i.e., no *cis*-template), and also insertions of tandem sequences, which are well characterized signature of C-NHEJ DSB repair dependent on Ku70/80 and Lig4 (Chang *et al.* 2017; Schimmel *et al.* 2017; Schimmel *et al.* 2019; Wolter *et al.* 2021); ii) we frequently observed microhomology between the genome and inserted donor DNA, highlighting the importance of sequence homology in DSB repair in our experiments. Microhomology was also found in the breakpoints of fused

chromosomes in the human fungal pathogen *Cryptococcus deuterogattii* (Schotanus and Heitman 2020). The majority of junction sequences from our editing experiments did not contain non-templated insertions as has been reported for NHEJ knock-in insertions (He *et al.* 2016), and the insertions did not share the same genomic ends, a sign of DNA end resection. These DNA mutation patterns are signatures of MMEJ, and similar to those reported using staggered Cas9 editing in mouse (Owens *et al.* 2019), but the possibility that NHEJ is also involved cannot be ruled out. It should be noted, the mechanism of so-called microhomology mediated end-joining (MMEJ) is not well characterized in filamentous fungi. The a-EJ pathway described to be dependent on DNA polymerase theta (Pol θ) (i.e., TMEJ) has not been described in fungi, and *M. oryzae* lacks a clear homolog containing the DNA polymerase domain (Supplementary Figure B. 28); iii) we identified large genomic deletions that resolved between two repetitive elements that reside variable lengths (i.e., 16 kb and 2 kb for MGL/MINE in *BUF1* locus and 11 kb and 4 kb for Copia in *BAS4* locus) away from the induced DSB sites,, requiring extensive DNA end resection and the use of long homologous sequence for resolution, hallmarks of SSA (Bhargava *et al.* 2016; Her and Bunting 2018). Similar large deletions between repetitive sequences have been reported in the protozoan parasite *Leishmania* (Zhang and Matlashewski 2019). The large deletion mutants we observed suggest the potential importance of repetitive elements and SSA in fungal genome repair and evolution, however, the SSA pathway in filamentous pathogens is also not well documented. Based on these DNA mutation signatures, we can conclude that multiple DNA repair pathways were used to resolve the Cas12a-induced DSBs.

Along with demonstrating that multiple DNA DSB repair pathways were used to resolve the Cas12a-induced DSBs, we observed that these pathways were used at unequal frequencies across the tested loci (i.e., locus-dependent frequencies). This suggests a locus-specific hierarchy

for DNA repair pathways, which to our knowledge has not been previously reported in fungi. It is unclear what factors may contribute to locus-dependent DNA repair preferences. Physical and genomic features such as chromatin structure, chromosome location, repetitive elements and the cell cycle have been reported to affect the outcome of DNA repair in other model systems (Ceccaldi *et al.* 2016; Gisler *et al.* 2019; Schep *et al.* 2021; Xue and Greene 2021). In *M. oryzae*, we found near 56 kb deletion from telomere to the *BAS4* locus, suggesting the subtelomeric distribution might affect DNA DSB outcomes. Demonstrating locus-dependent DNA repair is important because such a hierarchy would influence the evolutionary trajectory of different genomic regions. In natural systems, preferential repair of DSBs by different DNA repair pathways could create biased DNA variation and create genomic regions with accelerated evolution. There are numerous reports of compartmentalized genome evolution in filamentous pathogens, often referred to as two-speed genome evolution (Dong *et al.* 2015a), and variation for DNA DSB repair could be a major driver of this phenomena. Indeed, detailed analysis of DNA translocations and inversions between strains of *Verticillium dahliae*, a soil-borne wilt causing pathogen, found that chromosome re-arrangements co-localize with homologous sequence, often transposable element DNA (Faino *et al.* 2016). Repetitive DNA can serve as templates for homology-based repair, such as MMEJ or SSA. The occurrence of variable DNA repair reported here may influence adaptive genomic regions in *V. dahliae* that are associated with a unique chromatin profile (Cook *et al.* 2020; Schep *et al.* 2021). In *M. oryzae*, *AvrPi9* locus has proven to be a stable R-Avr interaction for breeding blast resistant rice (Qu *et al.* 2006; Wu *et al.* 2015). It will be interesting to further understand if DNA repair bias exists at the *AvrPi9* locus and contributes to the stability of this interaction. Also in *M. oryzae*, the presence of mini-chromosomes that are sequence variable

between strains, and contain duplicated sequences with core chromosomes, may be impacted by DNA repair variation (Peng *et al.* 2019).

It is unlikely the observations reported here are specific to *M. oryzae* or our experimental setup. Genome editing in other filamentous fungi, such as *Sclerotinia sclerotiorum*, *Aspergillus fumigatus* and *Trichoderma reesei*, reported unexpected genotyping results in which the target loci were PCR negative (Fuller *et al.* 2015; Hao and Su 2019; Li *et al.* 2018). Further TAIL-PCR and Illumina sequencing suggested that vector sequences and many uncharacterized sequences were inserted in the target loci in the case of *S. sclerotiorum* (Li *et al.* 2018). Those experiments used the Cas9 effector, DNA-based delivery system and different fungi, suggesting our results are not specific to Cas12a, RNP delivery or *M. oryzae* (Fuller *et al.* 2015; Li *et al.* 2018). Detailed study on mouse cell line suggested that large and complex DNA mutation is not dependent on transposition/delivery method and antibiotic selection (Kosicki *et al.* 2018). It should also be noted, large-scale on-target mutations created by CRISPR are easily missed without comprehensive genotyping or *de novo* assembly such as employed here, and are likely heavily under reported (Skryabin *et al.* 2020). Also, our repair outcomes are different than off-target editing often discussed for CRISPR research. Off-target editing refers to Cas nucleases binding and cutting DNA at unintended loci in the genome due to sequence homology (i.e., low-fidelity editing). Our results suggest that on-target DNA mutations following CRISPR DSB induction are varied and complex, consistent with recent reports in mammalian systems reporting extensive on-target mutations, including large deletions, complex rearrangements, and plasmid insertions (Adikusuma *et al.* 2018; Alanis-Lobato *et al.* 2021; Kosicki *et al.* 2018). These results underscore the need to further understand how CRISPR-based genome engineering interacts with endogenous

DNA repair mechanisms, and more importantly, how specific DNA repair mechanisms impact genome evolution in filamentous pathogens (Huang and Cook 2022a).

Materials and Methods

Fungal strains and incubation condition

M. oryzae field isolates O-137 (China) and Guy11 (French Guyana) were used as wild-types in this study (Bao *et al.* 2017; Giraldo *et al.* 2013). *Abuf1* mutants CP641 (gained from O-137 through spontaneous mutation) and CP281 (derived from weeping lovegrass pathogen 4091-5-8 through UV-mutagenesis) were used as a control in identifying the buff phenotype (Valent, unpublished data). JH7#1 and #2 are FK506 resistant mutants caused by a mutation in the *FKBP12* gene. The fungal cultures were maintained under light at 25 °C on OTA to observe mycelial color change (Huang and Cook 2022b). For high-molecular-weight DNA extraction and protoplast preparation, related mycelial plugs for different strains from OTA were cultured in liquid CM at 28°C, 120rpm for 3-4 days as described before (Zhang *et al.* 2021).

***In vitro* crRNA synthesis and LbCas12a RNP assembly**

Oligos including T7 promoter (taatagcactcactatagg), LbCas12a direct repeat (taatttctactaagtgtagat), and 23-nt target sequences (Supplementary Table B.6) were annealed and amplified to make the DNA template for *in vitro* RNA synthesis. HiScribe™ T7 High Yield RNA Synthesis Kit (New England BioLabs, catalog# E2040S) was used to make the crRNA/gRNA with the above prepared DNA template according to manufacturer's protocol. Monarch® RNA Cleanup Kit (New England BioLabs, catalog# T2050L) was used to purify synthesized gRNA after DNase I (RNase-free) treatment (New England BioLabs, catalog# M0303S). 5 µg purified

LbCas12a (New England BioLabs, catalog# M0653T or Integrated DNA Technologies, catalog# 10007923) was incubated with equal molar purified gRNA (near 0.5 µg) at 25°C for 15 mins for RNP assembly (Huang and Cook 2022b).

DNA donor preparation

pFGL821 (hygromycin selection, a gift from Dr. Naweed Naqvi; Addgene plasmid # 58223), and pFGL921 (G418 selection) (Zhang *et al.* 2021) were used as DNA templates for amplifying DNA donor with related primer pairs (Supplementary Table B.6). Long homologous sequences of SS flanked *HYG* DNA donor were amplified from O-137 genomic DNA and inserted into KpnI/XbaI and SalI/PstI sites in pFGL821 (Supplementary Table B.6). Phusion® High-Fidelity DNA Polymerase (New England BioLabs, catalog# M0530L) was used for DNA donor amplification.

Protoplast preparation and polyethylene glycol (PEG) mediated transformation

M. oryzae protoplast preparation was performed as described previously with minor modifications (Abdul *et al.* 2018). The fungal mycelium was filtered and dried through 2-layer 6_{1/2} in Disks Non Gauze Milk Filter papers, followed by addition of lysing solution (10 mg/mL Lysing Enzymes from *Trichoderma harzianum*, Sigma, catalog# L1412-10G, dissolved in 0.7M NaCl solution) and digestion at 30 °C with 70-80 rpm for 2-3 hours in the dark. After washing with 1xSTC (20% w/v Sucrose, 50mM Tris-Hcl pH=8.0, 50mM CaCl₂ dissolved in water), the concentration of released protoplasts was adjusted to 8x10⁶-5x10⁷ protoplasts/mL for transformation (Huang and Cook 2022b).

For protoplast transformation, RNP complexes (5 µg LbCas12a protein complexed with equal molar amount of crRNA/gRNA) and 3 µg DNA donor were mixed with 200 µl concentrated protoplast at room temperature for 20-25 mins. 1 mL 60% PEG solution (60% PEG4000, 20% w/v sucrose, 50 mM Tris-HCl pH=8.0, 50 mM CaCl₂) was added to above mixture and incubated at room temperature for 20-25 mins. This was followed by incubation with 5 mL TB3 liquid medium at 28°C, 90rpm, for 10-18 hours. After overnight incubation, the fungal cultures were mixed with 50mL molten (near 50-60°C) TB3 solid medium containing 100 µg/mL hygromycin (Corning, catalog# 45000-806) or 300 µg/mL G418 (VWR, catalog# 97064-358). The fungal medium suspension was poured into a plate (150 x 15 mm), dried and then overlaid and cultured with another 50mL molten TB3 solid medium plus 200 µg/mL hygromycin, or 600 µg/mL G418 in dark condition at 28°C for 5-7 days. Potential fungal transformants were picked and sub-cultured on CM, OTA or RPA for further phenotyping and genotyping (Huang and Cook 2022b). CM was supplemented with 1µg/mL FK506 (LC laboratories, catalog# NC0876958), for testing the sensitivity to FK506.

PCR genotyping

To test genotypes for the gene of interest, Q5® High-Fidelity DNA Polymerases (New England BioLabs, catalog# M0491) was used with 2 min extension time (technically up to 6 kb amplification) for the first-round genotyping with gene specific primer pairs (gene_F/R) (Supplementary Table B.6). 5' Upstream, 3' downstream regions and ACTIN were amplified with Taq DNA Polymerase (New England BioLabs, catalog# M0273) (Supplementary Table B.6) (Huang and Cook 2022b) .

High-molecular-weight DNA extraction and library preparation for Nanopore sequencing

The DNA extractions for long-read Nanopore sequencing were performed following the online protocol (<https://www.protocols.io/view/high-quality-dna-from-fungi-for-long-read-sequenci-k6qczdw>) (Wang *et al.* 2021a). g-Tube (Covaris, catalog# 520079) was used for shearing high-molecular-weight DNA into 20kb fragments followed by purification with AMPure XP beads (Beckman, catalog# NC9959336). Nanopore sequencing library preparation followed the Native barcoding genomic DNA (with EXP-NBD104, EXPNBD114, and SQK-LSK109) protocol (https://community.nanoporetech.com/protocols/native-barcoding-genomic-dna/checklist_example.pdf). Eleven barcoded DNAs from independent Cas12a edited strains were sequenced in three Nanopore MinION platforms for 72 hours.

Genome assembly and long-read mapping

Raw MinION fast5 files were transferred to fastq files by Guppy (version 3.4.4, <https://nanoporetech.com/nanopore-sequencing-data-analysis>) with the following parameters: `--disable_pings --compress_fastq --flowcell FLO-MIN106 --kit SQK-LSK109`. Adaptors were removed from basecalled reads by Porechop (version 0.2.4, <https://github.com/rrwick/Porechop>). Canu (version 1.9, 2.0 and 2.2.1, <https://github.com/marbl/canu>) (Koren *et al.* 2017) was used for *de novo* genome assemblies with the following parameters: `genomeSize=45m, minReadLength=1500`. Based on the flanking sequences, the contigs with *BUF1* flanking sequences (including #7, #8 assemblies from Figure 3.3d) were merged manually to scaffold based on the alignment result, when the contigs containing *BUF1* are not intact. The raw long-read mapping was performed with minimap2 (version 2.17-r941) (Li 2018). The mapping results were

adjusted with samtools (Li *et al.* 2009a) with following parameter: samtools view -q 60 to reduce the duplicated mapping and visualized with the Integrative Genomics Viewer (IGV) (version 2.5.0) (Robinson *et al.* 2011)

Synteny analysis

The synteny plots between sequenced *Abufl* and wild-type (O-137) were generated with Easyfig (version 2.2.5) (Sullivan *et al.* 2011) by with the following parameter: Blast option_Min. length 30, Max. e Value 0.001. Each individual synteny plot was modified with Adobe illustrator.

Statistical analysis

Fisher's exact test was performed in RStudio (Version 1.2.5001) with function (fisher.test).

Phylogenetic and domain analysis

The protein sequences of human Polq homologs were extracted from NCBI Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) and FungiDB (<https://fungidb.org/fungidb/app/>). The sequences used for analysis included MGG_15295 (*M. oryzae*), NCU07411 (*Neurospora crassa*), FGRAMPH1_01G23295 (*Fusarium graminearum*), NP_955452.3 (*Homo sapiens*), NP_084253.1 (*Mus musculus*), NP_524333.1 (*Drosophila melanogaster*), NP_498250.3 (*Caenorhabditis elegans*), AT4G32700.2 (*Arabidopsis thaliana*), XP_015619406 (*Oryza sativa*) and Pp3c5_12930V3.1 (*Physcomitrella patens*). The Neighbor-joining tree was made with MEGA X (Kumar *et al.* 2018) with 1000 bootstrap value. TBtools (Chen *et al.* 2020) was used visualized the domain structure gained from pfam (<http://pfam.xfam.org/>).

Data availability

The *de novo* assemblies of Cas12a-edited strains from Nanopore sequencing have been deposited in the WGS Genome of the National Center for Biotechnology Information (NCBI) under the BioProject accession No. PRJNA753862. The base called Nanopore reads after adapter removal have also been deposited under the same BioProject accession No. PRJNA753862 with the Short Read Archive (SRA) accession No. SRR15459267 to SRR15459274 and No. SRR20662591 to SRR20662593.

Funding

The research was funded by the United State Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA) awards no. 2018-67013-28492 to D.E.C. and no. 2017-67013-26525 to B.V., and the National Science Foundation Division of Molecular and Cellular Biosciences –Systems and Synthetic Biology award no. 1936800 to D.E.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Contribution number 22-059-J from the Kansas Agricultural Experiment Station.

Acknowledgments

The authors would like to thank Dr. Qiang Wang (Auburn University), Dr. Sanzhen Liu and Dr. Huakun Zheng (Kansas State University) for helpful discussion during the preparation of the manuscript. J.H. would like to thank Haolang Jiang (Fujian Agriculture and Forestry University) for helpful suggestions for fungal transformation and Melinda Dalby, Nathan Ryan and Dr. Velazhahan Rethinasamy (Kansas State University) for technical support. The authors thank computational support from the Beocat High-Performance Computing cluster at Kansas State University.

Author contributions

D.E.C. and J.H. conceived and designed the project; J.H. performed the experiments with help from D.R. P.S. and T.S.; J.H. and W.Z. performed the Nanopore sequencing. J.H. assembled and annotated the genomes in conjunction with D.E.C.; J.H., B.V. and D.E.C. analyzed the experiments; All authors contributed to writing the manuscript.

Competing interests

The authors declare no competing interests

Figures

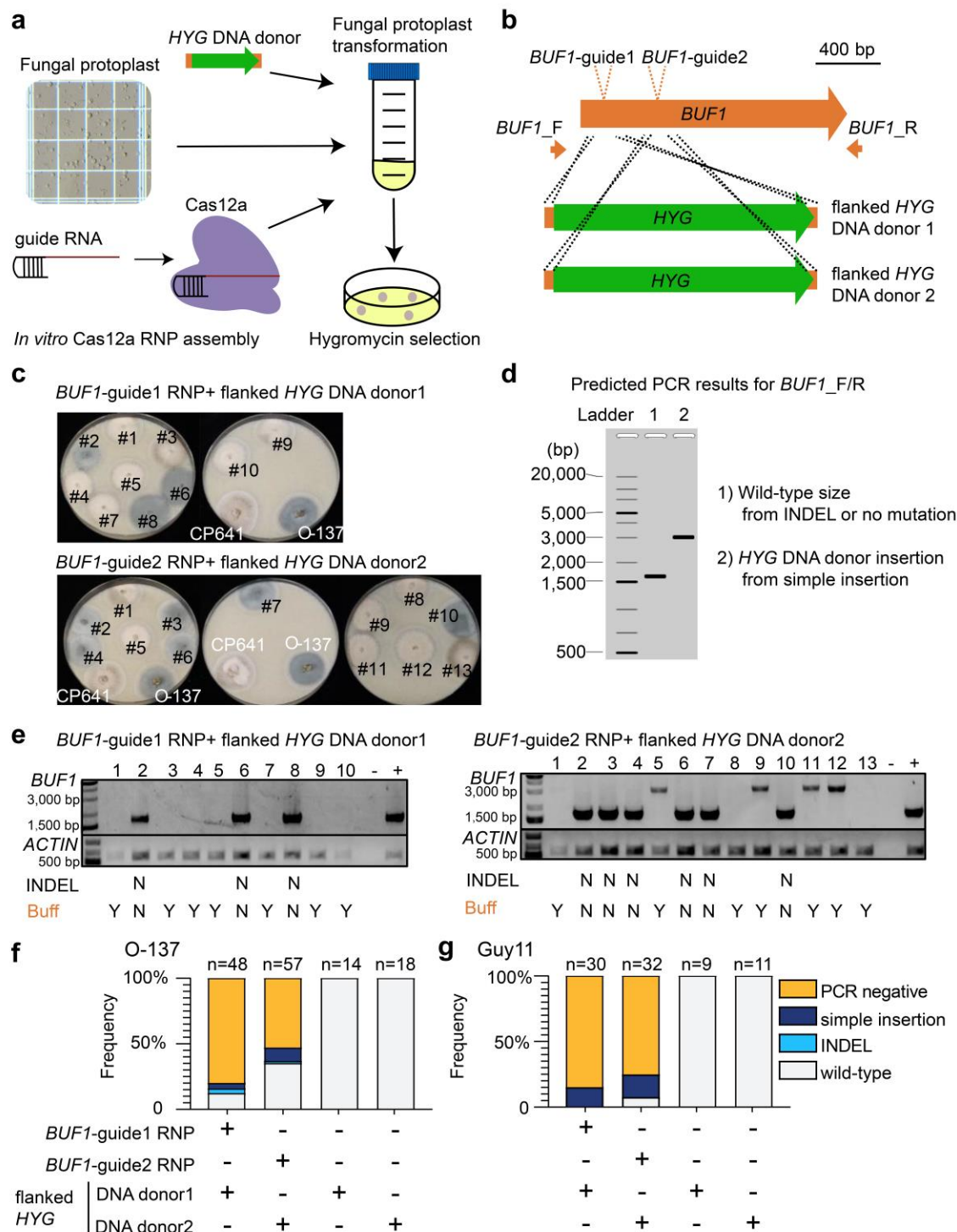


Figure 3.1. Cas12a RNP mediated editing results in unexpected genotyping at *BUF1* locus

(a) Schematic diagram of CRISPR-Cas12a RNP mediated genome editing through protoplast transformation in *M. oryzae*. (b) Illustration of two *BUF1* (MGG_02252, 70-15 MG8 annotation) gRNA design. Green rectangles indicate two different *HYG* DNA donor with flanked sequence homologous to the *BUF1* locus shown in orange. The location of PCR primer pair used for genotyping is shown (*BUF1*-F/R). (c) Phenotypes of hygromycin resistant (*HYG*^R) transformants plated on OTA. The wild-type O-137 is shown (dark grey hyphae) and a previously characterized $\Delta buf1$ in O-137 (CP641) showing the buff phenotype. Individual transformed colonies showing wild-type and buff color hyphae are shown labeled with numbers (d) Diagram of expected results following PCR amplification from transformants. Ladder indicates the molecular weight ladder to determine product size, lane 1 (1) shows the expected size product for wild-type or small INDEL *BUF1* amplification, and lane 2 (2) shows the expected size product for *BUF1* amplification where a single or multiple copy(s) of the *HYG* donor was inserted. (e) Genotyping results for the strains presented in the (c), the wild-type like PCR products from *BUF1* locus were purified and Sanger sequenced to detect potential INDELs. INDEL N indicates there were no INDELs observed after sequencing. Buff Y indicates the strain displayed the buff mutant color, while Buff N indicates wild-type phenotype. Lane (-) indicates negative control (water) and (+) a positive control O-137 genomic DNA used for PCR amplification. A separate PCR to amplify a portion of *ACTIN* was used as a DNA extraction control. (f and g) The frequency summary of DNA DSB repair outcomes in O-137 and Guy11. The number of independent PCR validated transformants (x) is listed (n=x). The mutation outcomes were concluded from three (O-137) or two biological (Guy11) replications.

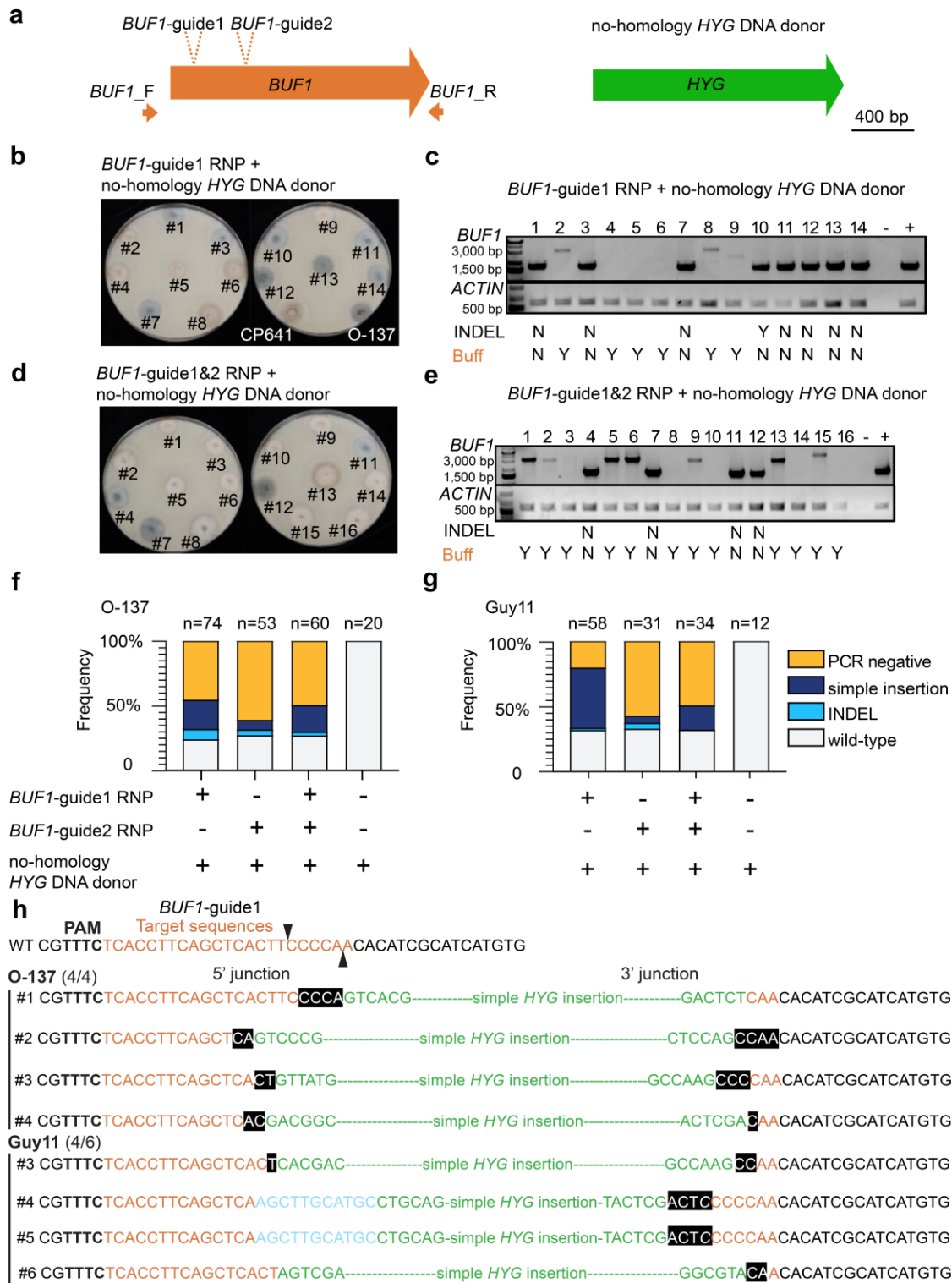


Figure 3.2. Microhomologous sequences are associated with donor DNA integrates at Cas12a target site

(a) No-homology *HYG* DNA donor with *BUF1*-guide1 or/and guide2 RNP were used for protoplast transformation. (b and d) Phenotyping result of the *HYG*^R transformants from single or dual *BUF1* RNP targeting assays. The strains were plated on OTA for phenotyping. CP641 is a positive control (*Δbuf1*) for buff color hyphae, O-137 is the wild-type isolate used in the experiment. Individual transformed colonies showing wild-type and buff color hyphae are shown labeled with numbers. (c and e) Genotyping results for the strains presented in (b and d), the wild-type like PCR products from *BUF1* locus were purified and Sanger sequenced to detect potential INDELS. Image labels are the same as described for Figure 3.1. (f) The frequency of DSB repair outcomes in O-137. The number of independent transformants (x) is listed (n=x). The assays for *BUF1*-guide1 were counted from four independent rounds of transformation (rep1 to rep4), the rest of assays were performed three times independently in O-137. (g) The frequency of DSB repair outcomes in Guy11. The number of independent transformants (x) is listed (n=x). The assays for *BUF1*-guide1 were counted from three independent rounds of transformation (rep1 to rep3), the rest of assays were performed twice independently in Guy11. (h) DNA sequence at the Integration junction of simple insertion mutants from *BUF1*-guide1. The ratio to the right of the strain name is the number of mutants with at least 1 bp microhomology at the integration junction from the randomly selected simple insertion mutants Sanger sequenced. Bold letters indicate PAM sequences; orange letters indicate target sequences; black triangles highlight the potential Cas12a cut site. Green sequences are from *HYG* DNA donor; White letters in black boxes highlight microhomology (i.e., shared sequence) between the *BUF1* locus and donor DNA; blue letters are sequence insertions of unknown source. Italicized letters indicate SNP.

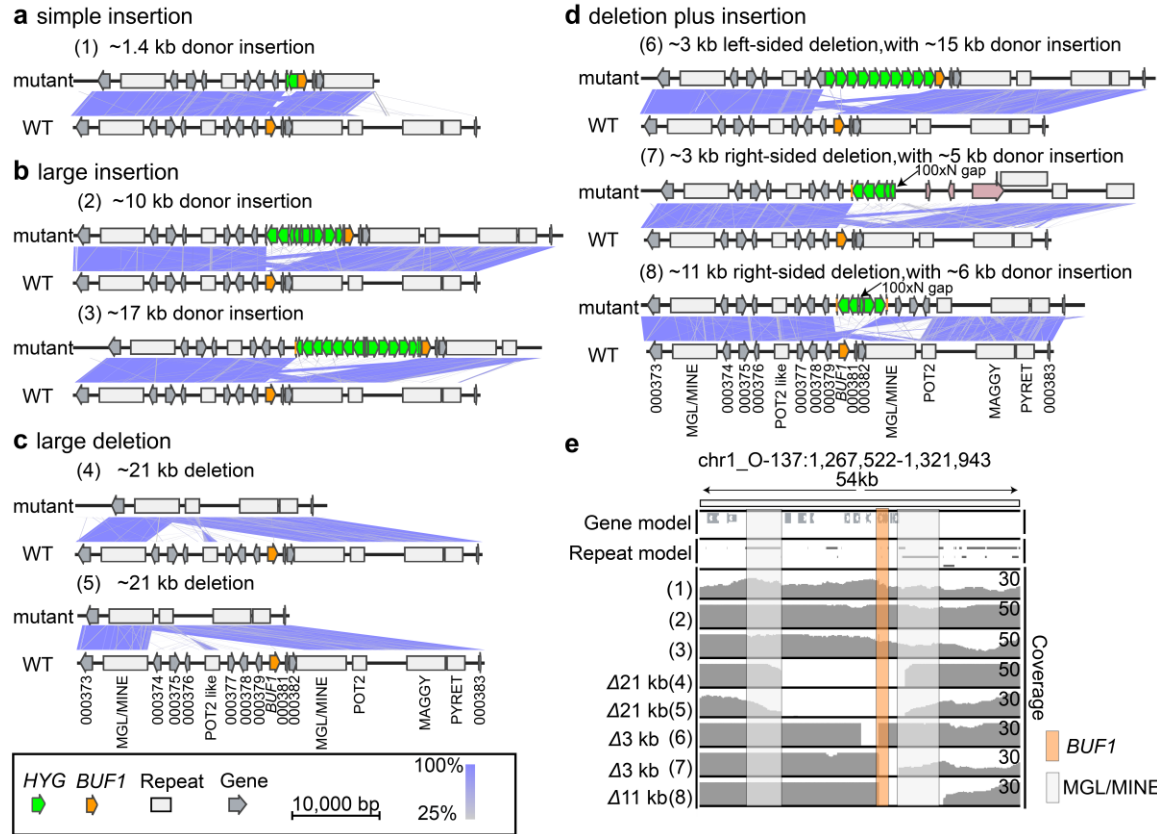


Figure 3.3. Long-read assemblies reveal three non-canonical error-prone DNA repair outcomes

Microsynteny between assembled edited strains (top) and the wild-type (bottom) indicated by purple connecting bands ranging from 25 to 100% similarity as indicated by the color scale. Individual assemblies were classified into one of four mutation classes (a-d). (a) simple donor insertion, (b) large donor DNA insertion, (c) large DNA deletion, (d) DNA deletion plus donor insertion, where two assemblies were incomplete and required the identification of another contig to complete the locus. The merged region is indicated with an arrow and label 100xN gap. The pink labelled genes in (7) indicated three genes inserted from a *trans* locus. (e) Coverage of Nanopore long-read mapping to the *BUF1* locus. Reads were mapped to the O-137 wild-type genome, where reads were filtered to remove low quality mapping (MAPQ < 60). The *BUF1* gene is highlighted with as a vertical orange box while the flanking repetitive DNA (MGL/MINE) are highlighted with a light grey box. Symbols are shown in the key to indicate the *BUF1* coding sequence (orange arrow), *HYG* coding sequence (green arrow), other coding sequences (grey arrow), and annotated repetitive DNA (grey box). Mutants labeled (1), (2), (3), (4), (5), (6), (7) and (8) indicate Rep1-*Δbuf1*#2, Rep1-*Δbuf1*#5, Rep1-*Δbuf1*#6, Rep1-*Δbuf1*#4, Rep4-*Δbuf1*#5, Rep4-*Δbuf1*#1, Rep1-*Δbuf1*#10 and Rep4-*Δbuf1*#13 mutants respectively (Supplementary Table B.3).

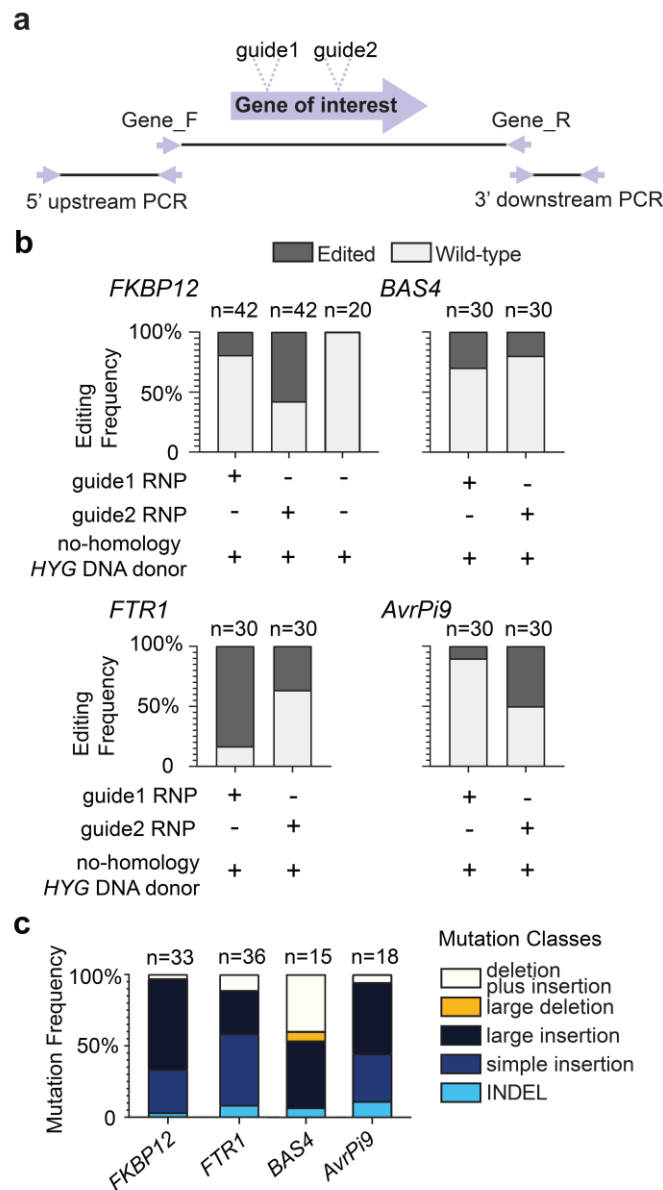
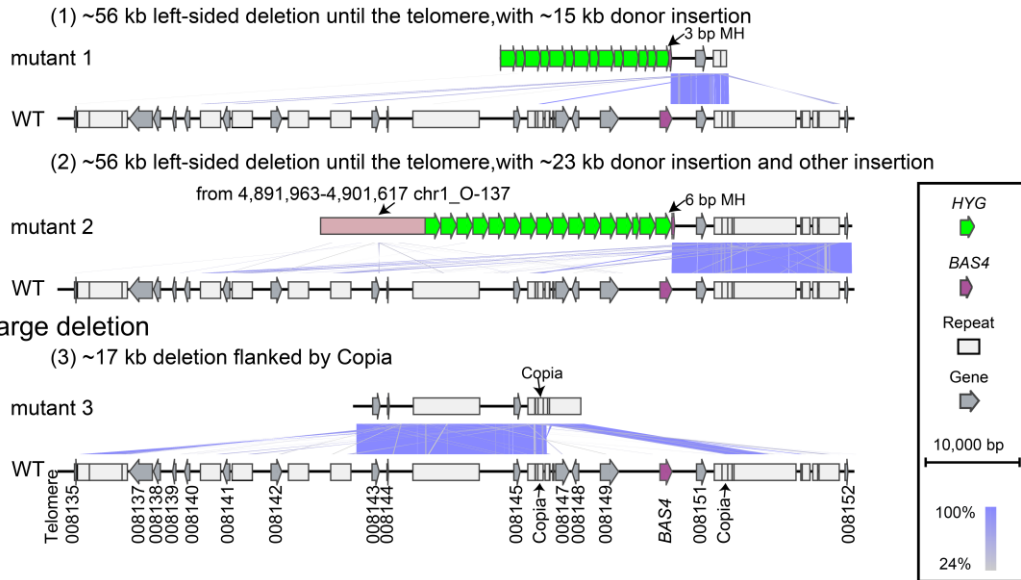


Figure 3.4. DNA repair outcomes following Cas12a editing differ among multiple loci

(a) Schematic illustration for CRISPR-Cas12a editing experiments. Two different guides were tested for each locus. Gene_F/R primer pairs were used for a first round of genotyping to determine the type of mutation present in a transformant. PCR negative strains were subsequently genotyped using the 5' upstream and 3' downstream primer pairs per gene to further genotype the occurrence of large-scale DNA alternations. (b) The editing frequencies at *FKBP12* (MGG_06035), *FTR1* (MGG_02158), *BAS4* (MGG_10914), and *AVRPI9* (MGG_12655). The MGG numbers based on the 70-15 assembly are provided for clarity (c) The mutation frequencies at *FKBP12*, *FTR1*, *BAS4* and *AVRPI9* are shown and indicated by

the color key to the right. For (b) the number of independent transformants genotyped (x) is listed (n=x), while (c) lists the number of independent mutants from (b). All the transformations were performed twice independently.

a deletion plus insertion



b large deletion

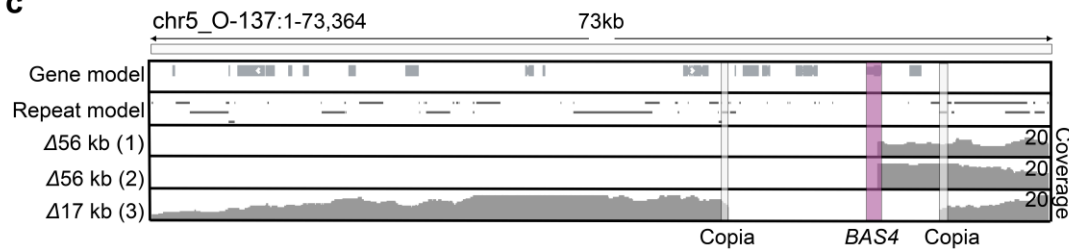


Figure 3.5. Dramatic DNA deletions at subtelomeric *BAS4* locus

Microsynteny between assembled *bas4* edited strains (top) and the wild-type (bottom) indicated by purple connecting bands ranging from 24 to 100% similarity as indicated by the color scale. Three $\Delta bas4$ mutants with two mutations classes (a) deletion plus insertion and (b) large deletion were sequenced and displayed. The pink labelled segment in (2) indicated sequences inserted from different chromosome (c) Coverage of Nanopore long-read mapping from telomere to the *BAS4* locus. Reads were mapped to the O-137 wild-type genome, where reads were filtered to remove low quality mapping (MAPQ < 60). The *BAS4* gene is highlighted with as a vertical purple box while the flanking repetitive DNA (Copia) are highlighted with a light grey box. Symbols are shown in the key to indicate the *BAS4* coding sequence (purple arrow), *HYG* coding sequence (green arrow), other coding sequences (grey arrow), and annotated repetitive DNA (grey box). Mutants labelled as (1), (2) and (3) indicate $\Delta bas4\#2$, $\#3$ and $\#11$ mutants from rep1 in Supplementary Figure B.21a, respectively (Supplementary Table B.3).

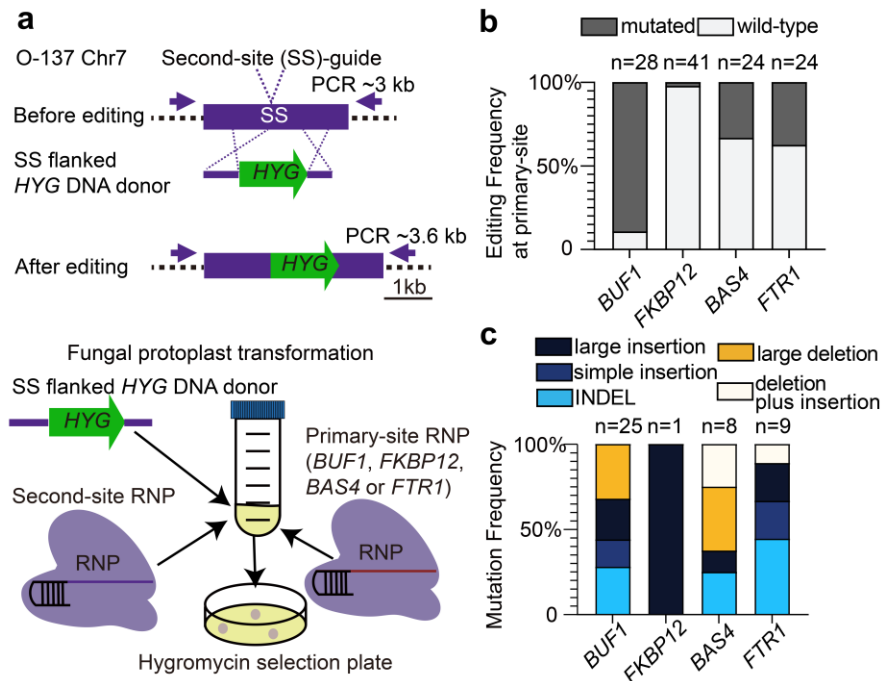


Figure 3.6. Distinct DSB repair patterns following donor insertion at a second integration site

(a) Schematic illustration of second-site (SS) insertion assay. The *HYG* coding sequence was directed to a second site through long homologous sequence located in a non-coding region on chromosome7 (Chr7:1,521,898–1,524,866). Fungal protoplasts were simultaneously transformed with two RNPs, one targeting the second-site and the other targeting the locus of interest. (b) The editing frequencies for the four tested loci following *HYG* selection. The proportion of mutated to wild-type transformants is indicated by dark and light grey shading respectively. (c) The frequency of distinct DNA repair outcomes for the four tested loci, colored according to the respective DNA mutation class. The number of independent transformants (b) and independent mutants (c) is listed (n=x). All the transformations were performed twice independently.

Chapter 4 - CRISPR-Cas12a ribonucleoprotein-mediated gene editing in the plant pathogenic fungus *Magnaporthe oryzae*

Note: This content from this chapter has been published in STAR Protocols (10.1016/j.xpro.2021.101072). Jun Huang established the protocol under the supervision and guidance of David E. Cook.

Summary

Magnaporthe oryzae (synonym of *Pyricularia oryzae*) is a plant pathogenic fungus that causes devastating blast disease in many crops including rice and wheat. As a model to study the interaction between host and pathogen, gene replacements through homologous recombination (HR) repair have been extensively used for functional genomic studies. However, the general efficiency of HR repair can be low in filamentous fungi, and laborious. Here, we provide a detailed protocol for efficient gene editing by inserting donor DNA into a region of interest following Cas12a ribonucleoprotein (RNP) mediated DNA double-strand break in *M. oryzae*.

For complete details on the use and execution of this protocol, please refer to (Huang *et al.* 2021).

Before You Begin

CRITICAL: To avoid contamination, all the steps involving fungal protoplasts and cultures require a sterile work environment, such as a biosafety cabinet. This includes using sterilized materials and tools.

Fungal culture

Timing: 1-3 weeks

1. Begin your starter culture on plates from a stock. For *M. oryzae*, sterile scissor can be used to remove embedded mycelium from filter paper to oatmeal agar (OTA).
2. Incubate the culture in a growth chamber at 25 °C under light for 1-2 weeks.
3. Cut 5-8 small mycelial plugs (1-2 mm side length) from OTA plate with a sterilized toothpick. Move the mycelial plugs into 100 mL liquid complete medium (CM) and grow at 28 °C, 120 rpm for approximately 2-4 days in the dark.

CRITICAL: Check the fungal status daily to avoid melanized mycelia (i.e., mycelium turning black), which negatively affects the efficiency to obtain fungal protoplasts.

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Oatmeal Agar	BD	Cat#DF0552-17-3
Sucrose	Fisher	Cat#S5-12
Yeast extract	BD	Cat#21750
Casamino acid	VWR	Cat#J851-500G
Yeast Nitrogen Base Without Amino Acids and Ammonium Sulfate	Fisher	Cat#DF0335-15-9

Ammonium nitrate	Sigma	Cat#221244-500G
L- (+) - Asparagine	Fisher	Cat#AAB2147322
Glucose	Sigma	Cat#G8270-100G
Agarose, low gelling temperature	Sigma	Cat#A9414-5G
di-sodium hydrogen phosphate	Sigma	Cat#1065851000
Bacteriological Agar	IBI scientific	Cat#IB49171
Hygromycin B in solution	Corning	Cat#45000-806
G418 solution	VWR	Cat#97064-358
Bialophos/glufosinate ammonium	Crescent Chemical	Cat#C140300
FK506	LC laboratories	Cat#NC0876958
Agarose	VWR	Cat#0710-500G
LbCas12a protein	NEB	Cat#M0653T
Lysing Enzymes from <i>Trichoderma harzianum</i>	Sigma	Cat#L1412-10G
Sodium chloride	Fisher	Cat#S271-3
PEG4000	Fluka	Cat#81240
Potassium chloride	Sigma	Cat#P-4504
Chloroform	Sigma	Cat#C2432-1L
Isopropanol	Fisher	Cat#A416-4
Ethanol	Fisher	Cat#BP2818-500
Tris-HCl, pH 8.0	Thermo Fisher	Cat#AM9856
Calcium chloride	Sigma	Cat#C8106-1KG
GeneRuler 1 kb Plus DNA Ladder	Thermo Fisher	Cat#SM1331
Critical Commercial Assays		
Phusion® High-Fidelity DNA Polymerase	NEB	Cat#M0530L
Q5® High-Fidelity DNA Polymerases	NEB	Cat#M0491L
Taq DNA Polymerase with Standard Taq Buffer	NEB	Cat#M0273L
HiScribe™ T7 High Yield RNA Synthesis Kit	NEB	Cat#E2040S
Monarch® RNA Cleanup Kit	NEB	Cat#T2050L
DNase I (RNase-free)	NEB	Cat#M0303S
Deoxynucleotide (dNTP) Solution Mix	NEB	Cat#N0447L
Wizard SV Gel and PCR Clean-Up system	Promega	Cat#A9282
Experimental Models: Organisms/Strains		
<i>M. oryzae</i> field isolate O-137	Dr. Barbara Valent (Kansas State Univ)	N/A
Oligonucleotides		

Forward oligo for LbCas12a-crRNA universal primer: CCCTAATACGACTCACTATAGGTAATTTCTACTAAGT GTAGAT	(Huang <i>et al.</i> 2021)	N/A
Reverse oligo for <i>BUF1</i> -guide1: TTGGGGAAGTGAGCTGAAGGTGAATCTACACTTAGT AGAAATTA	(Huang <i>et al.</i> 2021)	N/A
Reverse oligo for <i>BUF1</i> -guide2: GAACTGGCCACGGGTGTTGATGGATCTACACTTAGT AGAAATTA	(Huang <i>et al.</i> 2021)	N/A
Forward primer to amplify no-homology <i>HYG</i> DNA donor: AGAGTCGAGTATATTGAAGG	(Huang <i>et al.</i> 2021)	N/A
Reverse primer to amplify no-homology <i>HYG</i> DNA donor: CCCAGTCACGACGTTGTAAA	(Huang <i>et al.</i> 2021)	N/A
Forward primer to genotyping the gene region of <i>BUF1</i> : GCATCCATTCCCATTTCATCC	(Huang <i>et al.</i> 2021)	N/A
Reverse primer to genotyping the gene region of <i>BUF1</i> : GGGCAGCAACTGTCTCCATT	(Huang <i>et al.</i> 2021)	N/A
Forward primer to amplify the 5' upstream region of <i>BUF1</i> : GTTGGCACGGCACCTCTGTA	(Huang <i>et al.</i> 2021)	N/A
Reverse primer to amplify the 5' upstream region of <i>BUF1</i> : CTGGGATCGCGTCATACTTG	(Huang <i>et al.</i> 2021)	N/A
Forward primer to amplify the 3' downstream region of <i>BUF1</i> : GACAGTTGCTGCCCAAAGAT	(Huang <i>et al.</i> 2021)	N/A
Reverse primer to amplify the 3' downstream region of <i>BUF1</i> : AATGAGAAGGAGCGAGGATG	(Huang <i>et al.</i> 2021)	N/A
Forward primer to amplify <i>ACTIN</i> : CCGTGACTTGACCGACTACCTGA	(Huang <i>et al.</i> 2021)	N/A
Reverse primer to amplify <i>ACTIN</i> : AGAGCGAGGCGAGAATGGAAC	(Huang <i>et al.</i> 2021)	N/A
Recombinant DNA		
pFGL821	Addgene	Cat#58223
pFGL921	(Zhang <i>et al.</i> 2021)	N/A
pBV9/pSM324	(Kankanala <i>et al.</i> 2007)	N/A
Software and Algorithms		

BLAST	FungiDB	https://fungidb.org/fungidb/app/search/transcript/UnifiedBlast
Other		
Membrane Filter, 0.22 µm pore size	Millex	Cat#SLGVM33RS
Plastic syringes	Thermo fisher	Cat#S7510-10
Toothpick	Fisher	Cat#S24554
Hemocytometer	Fisher	Cat#0267151B
Miracloth	Calbiochem	Cat#475855-1R
6 1/2 in Disks Non Gauze Milk Filter	KenAG	N/A
Water, Molecular Biology Grade	VWR	Cat#VWRL0201-0500
150 x 15 mm petri dishes	Fisher	Cat#FB0875714
60 x 15mm Petri dishes	MIDSCI	Cat#901
100 x 15mm Petri dishes	MIDSCI	Cat#900
50mL falcon tubes	MIDSCI	Cat#C50B
1.5mL microcentrifuge tubes	USA scientific	Cat#1615-5500
0.2mL PCR tubes	MIDSCI	Cat#AVTW-F
Funnels	Fisher	Cat#FB6015865
125ml flask	PYREX	N/A
200ml flask	PYREX	N/A
Parafilm	Sigma	Cat#P7793-1EA
NanoDrop spectrophotometer	Thermo fisher	N/A
2.0 ml impact resistant screw cap tube	USA scientific	Cat #1420-9600
1.0mm Silica Beads	Biospec	Cat #11079110z
Bead ruptor elite-Bead mill homogenizer	OMNI international	Cat#19-2141E
Filter paper	N/A	N/A

Materials and Equipment

Prepare medium and stock solution

Timing: 1 week

Oatmeal Agar (OTA)

Dissolve 72.5 g commercial OTA powder in 1 L of distilled water, mix well. Heat to near boiling using microwave for 1-2 min. Autoclave at 121 °C for 20 min in liquid cycle. Pour the sterilized OTA in 100 x 15 mm petri dish after autoclave, store at room temperature. The plate can be stored for up to 2 months.

CRITICAL: It is very important to pre-heat the OTA before autoclave steps.

Alternatives: Microwave can be replaced with water bath for pre-heating.

Complete medium (CM)

Reagent	Final Concentration	Amount
Sucrose	10 g/L	10 g
Yeast extract	6 g/L	6 g
Casamino acid	6 g/L	6 g
Agar	20 g/L	20 g
ddH ₂ O	n/a	up to 1000 mL
Total	n/a	1000 mL

Autoclave at 121 °C for 20 min in liquid cycle. Agar is not required for liquid CM.

Store at room temperature up to 6 months.

TB3 medium (TB3)

Reagent	Final Concentration	Amount
Sucrose	200 g/L	200 g
Yeast extract	6 g/L	6 g
Casamino acid	6 g/L	6 g
Agar	15 g/L	15 g
ddH ₂ O	n/a	up to 1000 mL
Total	n/a	1000 mL

Autoclave at 121 °C for 20 min in liquid cycle. Agar is not required for liquid TB3.

Store at room temperature up to 6 months.

Defined complex medium (DCM)

Reagent	Final Concentration	Amount
Yeast nitrogen base without amino acids and ammonium sulphate	1.7 g/L	1.7 g
Ammonium nitrate	1 g/L	1 g
Asparagine	2 g/L	2 g
Glucose	10 g/L	10 g
Agarose, low gelling temperature	20 g/L	20 g
ddH ₂ O	n/a	up to 1000 mL
Total	n/a	1000 mL

Adjust pH to 6 with di-sodium hydrogen phosphate (Na₂HPO₄). Autoclave at 121 °C for 15 min in liquid cycle. Store at room temperature up to 6 months.

1xSTC solution

Reagent	Final Concentration	Amount
Sucrose	20% w/v	80 g
1M Tris-Hcl pH=8.0	50 mM	20 mL
CaCl ₂ · 2H ₂ O	50 mM	2.94 g
ddH ₂ O	n/a	up to 400 mL
Total	n/a	400 mL

Autoclave at 121 °C for 20 min in liquid cycle. Store at 4 °C for up to 6 months

PTC solution

Reagent	Final Concentration	Amount
PEG4000	60% w/v	60 g
1x STC solution	n/a	up to 100 mL
Total	n/a	100 mL

Autoclave at 121 °C for 20 min in liquid cycle. Store at room temperature for up to 6 months.

If the solution is not completely dissolved before you use, microwave PTC solution to help dissolving.

0.7M NaCl solution

Dissolve 16.36 g NaCl in 400mL of distilled water. Autoclave at 121 °C for 20 min in liquid cycle.

Store at room temperature for up to 6 months.

Protoplast lysing solution

Reagent	Final Concentration	Amount
Lysing Enzymes from <i>Trichoderma harzianum</i>	10 mg/mL	0.3 g
0.7M NaCl solution	n/a	up to 30 mL
Total	n/a	30 mL

Filter sterilize using 0.22 µm filter. This solution needs to be freshly prepared on the date when you are ready to perform protoplast preparation.

1M KCl solution

Dissolve 29.8 g KCl in 400mL of distilled water. Autoclave at 121 °C for 20 min in liquid cycle.

Store at room temperature for up to 6 months.

Vector usages

pFGL821, pFGL921 and pBV9/pSM324 could be used for amplifying hygromycin, g418 and Bialaphos DNA donor, respectively. Please see more details in the Note of step 9.

Step-by-Step Method Details

Design oligos for LbCas12a guide RNA

Timing: 1 week

To demonstrate CRISPR-Cas12a editing in *M. oryzae*, we mutated the trihydroxynaphthalene reductase encoding gene termed *BUF1* (MGG_02252), which is involved in fungal melanin biosynthesis (Chumley and Valent 1990). The *buf1* deletion mutant showed buff color (i.e., tan/orange) compared with the grey/black mycelial color in the wild-type strain. In this step, we describe the procedure to design the Cas12a guide RNA targeting *BUF1*.

1. Download the sequence of gene of interest from FungiDB (<https://fungidb.org/fungidb/app>). For the *BUF1* gene, download the sequence for gene MGG_02252.
2. Search for the LbCas12a PAM sequences (5'-TTTV-3') in your gene of interest. Select a 23 bp DNA sequence downstream (3' end) of PAM as a targeting sequence (Figure 4.1).

In the case of *BUF1*, we designed two spacer/targeting sequences.

BUF1-targeting sequence#1: **TTT**CTCACCTTCAGCTCACTTCCCCAA

BUF1-targeting sequence#2: **TTT**ACCATCAACACCCGTGGCCAGTTC

First four **bold** letters are the PAM, the remaining 23 bp are the targeting sequences.

Note: For LbCas12a guide designing, the targeting sequences must on the same strand of PAM ('5-TTTV-3'). V indicates A, C or G, but not T. It has been found before that targeting sequences with GC-content (~40%-60%) would provide the highest editing activity for Cas12a (Kim *et al.* 2017). Avoiding designing targeting sequences with GC content <30% or >70% for the best performance (Kim *et al.* 2017).

3. Check the off-target effect by blasting the designed targeting sequences within FungiDB database (<https://fungidb.org/fungidb/app/search/transcript/UnifiedBlast>)
 - a. Choose 'genome' as Target Data Type
 - b. Choose 'blastn' as BLAST Program
 - c. Choose '*Pyricularia oryzae*70-15' as Target organism
 - d. Input 27 bp sequences (4 bp PAM + 23 bp targeting sequences)
 - e. If the blast result shows the other loci match first 4 bp PAM and at least 16 bp targeting sequences after PAM perfectly, it may cause off-target editing. It is best to discard this potential guide sequence. Otherwise, move forward with the targeting sequence design.

Note: *Magnaporthe oryzae* is the synonym of *Pyricularia oryzae*.

4. Design oligos for preparing the DNA template for *in vitro* guide RNA synthesis. Two oligos are designed to construct the DNA template.
 - a. Universal forward oligo including the T7 promoter and LbCas12a mature direct repeat (Figure 4.2).

- b. Target specific reverse oligo including the reverse complemented sequences of LbCas12a mature direct repeat and the 23 bp targeting sequences you designed for your gene of interest (Figure 4.2). Oligos designed for two guides targeting the *BUF1* locus, termed *BUF1*-guide1 and *BUF1*-guide2, are provided as examples (Figure 4.2).

Note: Do not include PAM sequence in your target reverse oligo.

Note: Various online tools have been developed for genome-wide CRISPR target design. Many are dependent on the organism of interest and must be checked for suitability.

Optional: If commercial guide RNA synthesis kit compatible with your Cas protein is available, that could be an alternative option for preparing gRNA. For example, if you use SpCas9 protein, the NEB EnGen® sgRNA Synthesis Kit, *S. pyogenes*, could be an alternative option for making gRNA (NEB, Cat#E3322V).

CRITICAL: If high-quality genome assembly of your isolate is not available, you can check the potential off-target effects of your guides with the commonly used 70-15 genome assembly (Dean *et al.* 2005).

***In vitro* guide RNA synthesis and RNP assembly**

Timing: 2 days

5. Generate DNA template for *in vitro* guide RNA synthesis. Set up the below PCR reaction with Phusion® High-Fidelity DNA Polymerase.

Reagent	Amount
5X Phusion HF Buffer	4 μ l
10mM dNTPs	0.8 μ l
Universal forward primer (100 μ M)	0.4 μ l
Target specific reverse primer (100 μ M)	0.4 μ l
Phusion DNA Polymerase	0.2 μ l
Molecular water	14.2 μ l
Total	20 μl

PCR Cycling Conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	95 °C	30 sec	1
Denaturation	95 °C	10 sec	35 cycles
Annealing	57 °C	10 sec	
Extension	72 °C	10 sec	
Final extension	72 °C	2 min	1
Hold	4 °C	2 min	

CRITICAL: Make sure to use the correct concentration of oligos in the reaction. Products from this step do not require purification. The quality of the amplification reaction can be checked by running 2 μ L of PCR product on a 2% agarose gel and visualization. The product should be 66 bp.

Optional: Ordering the top and bottom oligos with full sequences (i.e., T7 promoter, LbCas12a mature direct repeat and 23 bp targeting sequences), and annealing them as the DNA template would serve as an alternative option to avoid this PCR step.

6. Guide RNA *in vitro* transcription. Set up the transcription as below by using HiScribe™ T7 High Yield RNA Synthesis Kit (NEB, Cat#E2040S).

Reagent	Amount
10X Reaction Buffer	2 µl
ATP (100 mM)	2 µl
GTP (100 mM)	2 µl
CTP (100 mM)	2 µl
UTP (100 mM)	2 µl
Template DNA from preparation step	8 µl
T7 RNA Polymerase Mix	2 µl
Total	20 µl

Run the reaction in thermocycler with 37 °C for 8-18 hours.

Note: Based on our experience, the yield should be least 50 µg RNA from this step.

7. Purification of synthetic guide RNA

- a. Remove the DNA template by adding 1 µl (2 units) RNase-free DNase to the guide RNA reaction and incubate for 15 min at 37 °C in a thermocycler.

- b. Purify the DNase treated guide RNA through Monarch® RNA Cleanup Kit (NEB Cat#T2050L) followed manufacturer's protocol (<https://www.neb.com/protocols/2018/06/28/monarch-rna-cleanup-kit-protocol>).
 - c. Quantify the RNA quality by NanoDrop spectrophotometer. A 260/280 ratio of 2.0-2.2 following guide RNA transcription and purification indicates successful synthesis.
 - d. Use guide RNA immediately for RNP assembly or store at -80 °C for up to 2 months.
8. Assemble the RNP for one reaction in 0.2 mL PCR tube as indicated below.

Reagent	Final Concentration	Amount
LbCas12a protein (100 µM)	~ 5 ug/20 µL	0.34 µl
gRNA	~ 0.5 ug/20 µL	x µl
NEBuffer 2.1 Reaction Buffer (10X)	n/a	2 µl
Molecular water	n/a	up to (17.66-x) µl
Total	n/a	20 µl

The amount of gRNA depends on the concentration of gRNA (e.g., if the concentration is 1000 ng/µl, then use x=0.5 µl for the reaction).

Incubate the above reaction for 15 min at 25 °C using a thermocycler.

Note: Add the LbCas12a protein last to avoid protein precipitation (Fernandez *et al.* 2018).

Assemble the RNP immediately before fungal protoplast transformation.

Donor DNA preparation

Timing: 5-6 h

9. Amplify the no-homology *HYG* DNA donor as indicated below

Reagent	Amount
5X Phusion HF Buffer	10 µl
10mM dNTPs	1 µl
no-homology <i>HYG</i> DNA forward primer (10 µM)	2.5 µl
no-homology <i>HYG</i> DNA reverse primer (10 µM)	2.5 µl
Phusion DNA Polymerase	0.5 µl
pFGL821 (~5 ng/µl)	1 µl
Molecular water	32.5 µl
Total	50 µl

PCR Cycling Conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	98 °C	30 sec	1
Denaturation	98 °C	10 sec	30 cycles
Annealing	57 °C	30 sec	
Extension	72 °C	1 min	
Final extension	72 °C	10 min	1
Hold	4 °C	2min	

Note: Expected PCR amplification size is 1,466 bp. Each fungal protoplast transformation requires ~3 µg donor DNA in total. Adjust the number of above PCR reactions based on the total numbers of transformation you will run. We amplified the *HYG* DNA donor by using pFGL821 plasmid as DNA template. Other selectable markers (e.g., G418 and Bialaphos) also could be used, we amplified G418 DNA donor from pFGL921 plasmid and Bialaphos DNA donor from pBV9/pSM324 plasmid (Kankanala *et al.* 2007; Zhang *et al.* 2021). If the gene of interest could be used for selection directly (e.g., *FKBP12* or *UAR3/UAR5*) (Huang *et al.* 2021; Wang *et al.* 2018), donor DNA preparation step is not necessary.

10. Check the amplification quality by running 2 µl of PCR product on a 1% agarose gel.

Successful amplification should yield a single product of the expected size.

11. Purify the PCR reaction through Wizard SV Gel and PCR Clean-Up system (Promega).

The detailed purification step follows manufacturer's protocols ([https://www.promega.com/-/media/files/resources/protocols/technical-](https://www.promega.com/-/media/files/resources/protocols/technical-bulletins/101/wizard-sv-gel-and-pcr-clean-up-system-protocol.pdf?la=en)

[bulletins/101/wizard-sv-gel-and-pcr-clean-up-system-protocol.pdf?la=en](https://www.promega.com/-/media/files/resources/protocols/technical-bulletins/101/wizard-sv-gel-and-pcr-clean-up-system-protocol.pdf?la=en)). If there are more products from the amplification product as indicated by step 10, gel purification can be used to isolate the product of interest by using the same Wizard SV Gel and PCR Clean-Up system (Promega).

12. Quantify the DNA quality by NanoDrop spectrophotometer. Donor concentration near 200 ng/µl-350 ng/µl is preferred.

Note: Store the DNA donor in -20 °C freezer until future use.

***M. oryzae* protoplast preparation**

Timing: 5-6 h

13. Prepare the protoplast lysing solution freshly as described in “Prepare medium and stock solution” section.
14. Collect the mycelia from sub-cultured liquid CM through 2-layers of 6^{1/2} in Disks Non Gauze Milk Filter papers and funnel. Wash the collected mycelia with 5 mL 1xSTC solution.
15. Dry the collected mycelia with extra filter paper, transfer the dry mycelia to 100 mL flask with 25-30 mL protoplast lysing solution. Shake dry mycelia containing the flask at 30 °C, 70-80 rpm, for 2.5-3 hours under dark to digest the fungal cell wall and release protoplasts.

Note: When transferring the mycelium, manually separate the mass so it is not a single large mass. This will increase the accessibility to the lysing solution.

16. Check the digestion progress and protoplast quality using a hemocytometer under a dissecting microscope. Successful digestion should release many healthy protoplasts, while if many undigested mycelia are observed, extend the digestion time. However, excessive digestion can also cause protoplast collapse (Figure 4.3).
17. Collect the protoplast by filtering the above lysing reaction containing the fungal protoplasts, through 2-layer miracloth into pre-chilled, 50mL falcon tube. Wash the lysing debris left in the miracloth with 1xSTC again to collect all the protoplasts.
18. Centrifuge the 50mL falcon tube with protoplast at 4 °C, 5,000 rpm for 10 min.

Note: A visible pellet containing protoplasts should be seen.

19. Discard the supernatant into biohazard trash container. Gently resuspend the protoplast using 5-10 mL 1xSTC by pipetting or flicking.

20. Repeat the step 18.

21. Discard the supernatant into biohazard trash container, and gently resuspend the protoplast in 1mL 1xSTC. Quantify the protoplast concentration using a hemocytometer. Add additional 1xSTC to adjust the final concentration of protoplasts to 8×10^6 to 5×10^7 protoplasts/mL. Aliquot 200-250 μ L protoplast solution to individual 50 mL falcon tubes for further transformation.

Optional: The protoplasts can be frozen and stored by adding 7% v/v DMSO into the protoplast solution and freezing at -80°C . We recommend performing the protoplast transformation immediately after preparation for best results, but the frozen cultures can still be used for successful editing.

Note: This protocol is strain independent, and we have used it to make protoplasts for various isolates including O-137 and Guy11 (Bao *et al.* 2017).

***M. oryzae* protoplast transformation**

Timing: 2 days

22. Slowly add 20 μ L assembled RNP and 3 μ g of no-homology *HYG* DNA donor into the 50 mL falcon tube containing the aliquoted protoplast. Mix the solution by flicking gently. Place the tube at room temperature for 20-25 min.

Note: If your gene of interest can be used as selection directly, DNA donor is not required in step 22. For example, we have generated *fkbp12* (FK506 receptor) mutants by using the selection of drug FK506 without DNA donor successfully (Huang *et al.* 2021).

23. Slowly add 1,000 μ L PTC solution to the 50mL falcon tube. Mix the solution by flicking gently. Place the tube at room temperature for 20-25 min.

24. Slowly add 5 mL liquid TB3 medium into the 50 mL falcon tube. Mix the solution by flicking gently. Cover the 50 mL falcon tube lid with parafilm.

25. Shake 50 mL falcon tube at 28 °C, 90 rpm overnight (12-20 hours) in the dark.

26. Next day, unseal the 50 mL falcon tube and add melting TB3 medium (near 40 mL for one transformation, 50-60 °C) with final concentration of 100 μ g/mL hygromycin solution. Mix gently, then pour the mixture (overnight protoplast culture + medium) into 150 x 15 mm petri-dish. Wait for about 20 min until complete solidification.

Note: Protoplasts are fragile and require gentle handling. Other antibiotics corresponding to a different no-homology DNA donor can also be used. For example, we have also used G418 at a final concentration: 300 μ g/mL. Bialaphos also can be used at a final concentration: 25 μ g/mL in DCM instead of TB3 medium.

27. After the first selection layer has completely solidified, pour a second layer of melting TB3 medium (near 50 mL) with final concentration of 200 µg/mL hygromycin solution into the 150 x 15 mm petri-dish above the first layer. Wait for about 20 min until complete solidification.

Note: For G418 selection, the final concentration for the second layer is 600 µg/mL. Bialaphos also can be used at a final concentration: 100 µg/mL in DCM instead of TB3 medium.

28. Seal the 150 x 15 mm petri-dish with parafilm and incubate the plate upside down at 28 °C for 5-7 days under dark.

Note: Transformants using *M. oryzae* isolate O-137 can typically be seen and transferred at 5 days, while those from isolate Guy11 typically take 7 days. Other isolates will have to be empirically determined.

29. Pick the transformants growing in the top layer for further phenotyping and genotyping. This can be done using a sterile toothpick to move a piece of mycelium to a new plate containing the selection.

30. For *BUF1* deletion identification, we cultured the transformants to OTA to check the mycelial color change. For other genes of interest, transformants can be moved to CM with corresponding appropriate selection: 200 µg/mL hygromycin or 500 µg/mL G418. If you

use bialaphos selection (final concentration:100 µg/mL), select DCM instead of CM for screening.

Quick fungal DNA extraction for genotyping

Timing: 5-6 h

31. Prepare the screw cap tubes with 1.0 mm beads, 400 µl 1M KCL and 400 µl chloroform (added in hood).
32. Add fungal plug (near 1 cm diameter) into the above prepared tubes. (Performed in the biosafety cabinet hood)
33. Put the above tubes in homogenizer with power 4 x 6 cycles (30 s each cycle) to disrupt the fungal material.
34. Centrifuge 14,000 rpm x 10 min
35. Transfer 200 µL supernatant into new labelled tubes
36. Add 120 µL isopropanol in the tube, mix well to precipitate the DNA for at least 10 min (added in hood).
37. Centrifuge 14,000 rpm x 10 min.
38. Discard the supernatant and wash pellet with 500 µL 75% ethanol.
39. Centrifuge 14,000 rpm x 10 min.
40. Discard the ethanol and air dry for 10 min.
41. Add 50 µL sterilized water to dissolve the DNA pellet.

Note: We prefer to use 1 μ L of the dissolved DNA as template for PCR amplification. The DNA extracted by this quick extraction protocol is sufficient for PCR genotyping but should not be considered adequate for other uses such as next-generation sequencing or Southern blot.

PCR genotyping

42. We prefer to design three pairs of primer for genotyping; one primer pair amplifies within the gene of interest ($\sim$ 1-2 kb for PCR amplification, the RNP targeting sequence should be included in this amplifying region), and the other two primer pairs amplifying the 5' and 3' regions of the targeted sequence ($\sim$ 0.5 kb for PCR amplification, the reverse primer for 5' upstream amplification and the forward primer for 3' downstream amplification should locate closely (1-200 bp) to the start codon and stop codon of the gene of interest) (Figure 4.4).

43. Perform genotyping PCR followed the below program for the gene of interest. We used *BUF1* gene as an example.

Reagent	Amount
5X Q5 reaction buffer	5 μ l
10mM dNTPs	0.5 μ l
<i>BUF1</i> gene region forward primer (10 μ M)	1.25 μ l
<i>BUF1</i> gene region reverse primer (10 μ M)	1.25 μ l
Q5 High-Fidelity DNA Polymerase	0.25 μ l
DNA extracted from quick extraction protocol	1 μ l

Molecular water	15.75 µl
Total	25 µl

PCR Cycling Conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	98 °C	30 sec	1
Denaturation	98 °C	10 sec	30 cycles
Annealing	65 °C	30 sec	
Extension	72 °C	2 min	
Final extension	72 °C	2 min	1
Hold	4 °C	2min	

Note: The expected amplification size for wild-type is 1,648 bp, the size for mutants might vary based on the mutation pattern (Huang *et al.* 2021).

44. Perform genotyping PCR for 5'upstream, 3' downstream of *BUF1* and *ACTIN* (loading control) with below program

Reagent	Amount
10x standard Taq reaction buffer	2.5 µl
10mM dNTPs	0.5 µl
<i>BUF1</i> 5' upstream or downstream or <i>ACTIN</i> forward primer (10 µM)	0.5 µl

<i>BUF1</i> 5' upstream or downstream or <i>ACTIN</i> Reverse primer (10 µM)	0.5 µl
Taq DNA Polymerase	0.125 µl
DNA extracted from quick extraction protocol	1 µl
Molecular water	19.875 µl
Total	25 µl

PCR Cycling Conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	95 °C	30 sec	1
Denaturation	95 °C	10 sec	30 cycles
Annealing	See below note	30 sec	
Extension	72 °C	35 sec	
Final extension	68 °C	5 min	1
Hold	4 °C	2 min	

Note: Annealing temperature for *BUF1* 5' upstream or downstream is 54 °C, while it requires 57 °C for *ACTIN*. The expected amplification size for *BUF1* upstream, downstream and *ACTIN* is 311 bp, 458 bp and 503 bp, respectively.

Expected Outcomes

Gene disruption mutants can be generated by our method with an editing efficiency from ~50%-100% at *BUF1* locus (Figure 4.5A). Genotyping results suggested that large-scale insertion

and deletion is the major genetic events occurs in the *buf1* mutants (Figure 4.5B) Please see more detailed analysis and discussion in the original manuscript. (Huang *et al.* 2021).

Limitations

Due to locus-dependent and physiological variables (e.g., gRNA efficiency, chromatin feature, sub-telomere/telomere location, and cell cycle), gene-editing efficiency can vary between loci and experiments. (Ceccaldi *et al.* 2016; Gisler *et al.* 2019; Huang *et al.* 2021; Schep *et al.* 2021; Xue and Greene 2021). The use of two separate RNPs, targeting DNA regions within proximity to the desired location of editing may increase editing efficiency if single RNP transformation does not generate the desired mutation.

Troubleshooting

Problem 1:

Cannot get the DNA template for T7 *in vitro* transcription from step 5.

Potential Solution:

Double check the oligo concentration you used in your system, it requires 100 μ M for oligos, which is 10x higher than the normal PCR reaction.

Confirm the complementary sequences shared between universal forward primer and target specific reverse primer.

Problem 2:

gRNA quality and concentration are low from step 6-7.

Potential Solution:

The low gRNA could be owing to at least two factors:

(1) Inefficient gRNA synthesis

Run DNA template before T7 *in vitro* transcription in 2% agarose gel, this would help us confirm the quality template DNA. A strong band near 66 bp is expected.

Exactly follow the T7 *in vitro* transcription step described in this protocol (step 6). Mix all the reaction components well by pipetting before starting the reaction. Extend the reaction time to 18 hours for the maximum output gRNA.

(2) Inefficient gRNA purification

Exactly follow the purification step described in Monarch® RNA Cleanup Kit (NEB Cat#T2050L), store the gRNA correctly when they are ready. (Immediately use or 80 °C for long-term storage).

Problem 3:

Quality of generated Protoplast is bad from step 21.

Potential Solution:

(1) Avoid using the melanized mycelia for lysing enzyme digestion.

- (2) Prepare the lysing enzyme freshly.
- (3) Don't over digest (> 4 hours) the fungi.
- (4) Make sure you use miracloth rather than normal filter in filtering step after protoplasts are released.
- (5) Gently mix well the protoplasts before the hemocytometer qualification.
- (6) Keep released protoplast on the ice until you start transformation.

Problem 4:

Buff phenotype is not obvious from step 30.

Potential Solution:

We would suggest keeping potential buff transformants on OTA or RPA for a better distinguishable buff phenotype. We found that plating the buff strains on CM cannot sometimes produce the undistinguishable color change when compared to the wild-type strain.

Problem 5:

Low gene editing efficiency from step 30.

Potential Solution:

Many factors contribute to genome editing efficiency.

- (1) Guide design, if you are new for designing CRISPR gRNA, do an *in vitro* digestion assay followed the attached protocol (<https://international.neb.com/protocols/2017/12/19/in-vitro->

digestion-of-dna-with-engen-lba-cas12a-cpf1-neb-m0653), it will tell you whether your designed gRNA is compatible with Cas12a protein and the cutting efficiency.

(2) Fungal protoplast, use freshly prepared protoplast for transformation. We found that freshly prepared protoplast with correct concentration always showed better transformation efficiency than frozen stored protoplast.

(3) Temperature sensitivity, it has been shown before that LbCas12a is a temperature sensitive enzyme, with best editing efficiency at ~28 or higher °C (Malzahn *et al.* 2019). Therefore, it is necessary to main your lab temperature around 28 °C when you perform Cas12a transformation.

Acknowledgments

The authors would like to thank Dr. Barbara Valent and Melinda Dalby (Kansas State University) for sharing the resources and fungal isolates and David Rowe (Kansas State University) for critically reading prior to manuscript submission. The research was funded by the United State Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA) (award no. 2018-67013-28492), and the National Science Foundation Division of Molecular and Cellular Biosciences –Systems and Synthetic Biology (award no. 1936800) to D.E.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

J.H. executed the experiments and wrote the first draft of the manuscript, D.E.C supervised the study, and edited the manuscript.

Declaration of Interests

The authors declare no competing interests.

Figures

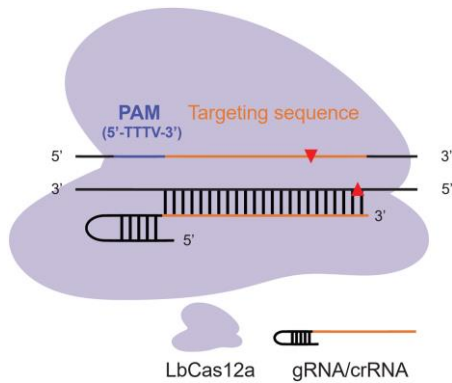


Figure 4.1. Schematic diagram of Cas12a cutting

Blue line highlights the PAM sequence, the orange line indicates the designed targeting sequence. Arrows indicate the cutting site (17th base after PAM in the non-targeted strand and 22nd base after PAM in the targeted strand) of Cas12a, which is represented by the purple shape. The gRNA contains a hairpin at the 5' end to facilitate Cas12a interaction.

		T7 promoter	LbCas12a mature direct repeat	Target sequence	
Universal forward oligo	5'	CCCTAATACGACTCACTATAGGTAATTTCTACTAAGTGTAGAT			3'
Target reverse oligo	3'		ATTAAAGATGATTACATCTANNNNNNNNNNNNNNNNNNNNNNN		5'
<i>BUF1</i> -guide1 reverse oligo	3'		ATTAAAGATGATTACATCTAAGTGGAAGTCGAGTGAAGGGGTT		5'
<i>BUF1</i> -guide2 reverse oligo	3'		ATTAAAGATGATTACATCTAGGTAGTTGTGGGCACCGGTCAAG		5'

Figure 4.2. Schematic example of oligo design

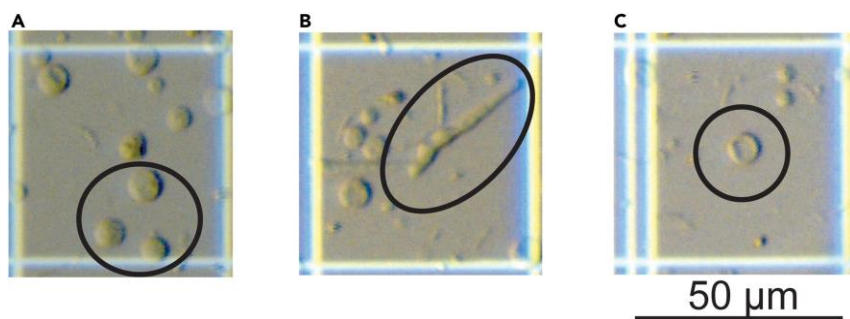


Figure 4.3. Different protoplast status after lysing enzyme digestion

(A) An example of healthy protoplasts, (B) undigested mycelium, and (C) collapsed protoplast each highlighted by black circles.

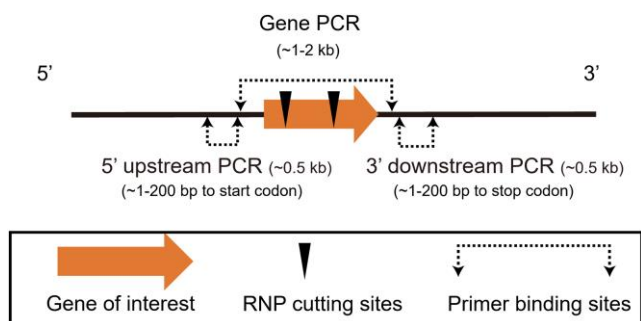


Figure 4.4. Schematic illustration of genotyping PCR primer pairs designing

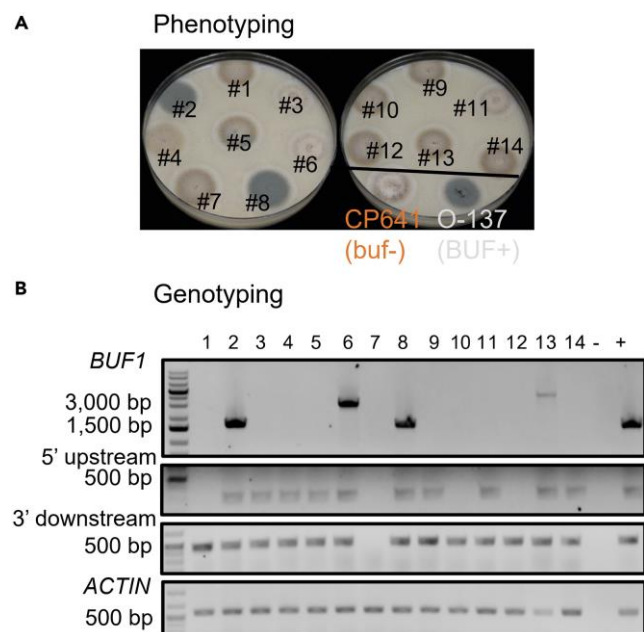


Figure 4.5. Representative phenotyping and genotyping outcomes at *BUF1* locus with *BUF1*-guide1 RNP and no-homology *HYG* DNA donor

(A) Δ *buf1* showed the buff/orange mycelial color in OTA. CP641 (*buf*-) derived from O-137 is the positive strain for buff mycelia. O-137 (*BUF*+) is the wild-type strain used in this assay. (B) DNA extracted from the transformants in (A) were used for genotyping with *BUF1*, *BU1* 5'upstream, *BUF1* 3' downstream and *Actin* primer pairs (loading control). – and + indicated the water (negative control) and O-137 genomic DNA (positive control) used in PCR reaction.

Chapter 5 - Re-establishing avirulence in a *Magnaporthe oryzae*-rice interaction using CRISPR-Cas12a based transposable element editing

Note: This work has not been published yet, Jun Huang performed most of the experiments, Veerendra Sharma conducted the Southern blotting.

Abstract

Rice blast disease caused by the filamentous fungus *Magnaporthe oryzae* is one of the most destructive plant diseases and threatens the global food security. Deploying rice cultivars with blast resistance genes is the most environmentally sustainable approach for disease prevention. To counter plant immunity, *M. oryzae* delivers effectors into host cells that can inhibit the immune response or alter host physiology to favor infection. In response, plants evolved receptor proteins to recognize effectors or their biochemical signatures, triggering an immune response and rendering the plant resistance. In this case, the effector gene is termed an avirulence gene as it renders the pathogen unable to cause disease. In natural populations of plant pathogenic fungi, genetic variation for effectors can result in the avoidance of resistance genes, imposing strong selection on both the host and pathogen. The insertion of transposable elements (TEs) into effector coding sequences has been frequently found in field isolates that are able to overcome plant receptor immunity (i.e., loss of avirulence). Additionally, TE insertions in proximity to, but not impacting the coding sequence are presumed to disrupt the effector's transcription, however, this hypothesis has not been fully tested. To assess the effect of cis-regulatory TE insertions for effector regulation, we used CRISPR-Cas12a ribonucleoprotein (RNP) editing to completely remove a

Pot3 DNA transposon insertion located upstream of the *AvrPiz-t* coding sequence in the field isolate Guy11. To determine the impact on *AvrPiz-t* regulation, a pathogenicity assay against a rice cultivar containing *Piz-t* was used to show that when the Pot3 TE was deleted from Guy11, it could no longer cause disease. These results demonstrate restored avirulence against *Piz-t* containing rice, suggesting that the Pot3 insertion interfered with transcriptional regulation of *AvrPiz-t*. Conceivably, such a DNA alteration can occur in natural populations, which might argue for the continued deployment of defeated R genes to avoid the re-activation of defeated effectors.

Introduction

To disable the plant immune response, plant pathogenic fungi deliver proteins or chemicals, termed effectors, to disrupt the cellular processes needed for host resistance to promote infection (Jones and Dangl 2006; Oliveira-Garcia and Valent 2015). However, some pathogen effectors can be recognized by corresponding plant receptors, generally termed resistance genes, which trigger the plant immune response (Cook *et al.* 2015). Plant resistance genes that recognize intracellular effectors encode nucleotide-binding domain leucine-rich repeat (NLR) receptors, which either directly or indirectly recognize a related effector (Jia *et al.* 2000; Park *et al.* 2012). For instance, direct interaction between the rice blast resistance gene *Pi-ta* and the corresponding effector gene *AVR*Pi-ta**, result in an incompatible interaction where the host is immune and the pathogen termed avirulent resulting from the expression of the avirulence gene (*AVR*) (Jia *et al.* 2000; Jones and Dangl 2006). For a different pair of *R-AVR* genes from rice and *M. oryzae*, *Pi9* and *AVR*Pi9**, direct interaction has not been found, and the detailed mechanism is unknown (Wu *et al.* 2015). To

overcome NLR immunity, multiple mechanisms to delete, alter or silence microbial avirulence genes have been reported to avoid host recognition (Chen *et al.* 2021; Yoshida *et al.* 2009).

The haploid filamentous fungus *M. oryzae* (synonym of *Pyricularia oryzae*), which causes blast disease in rice and wheat, is a threat to crop protection and food security (Bao *et al.* 2017; Peng *et al.* 2019). Based on long-read sequencing and high-quality assemblies, 10%~15% of *M. oryzae* genomes were annotated as repeated sequences (Bao *et al.* 2017; Peng *et al.* 2019). Among those repeated sequences, transposable elements (TEs) make up an important class of repetitive DNA that can move in the genome through direct and indirect mechanisms, and play an important role in fungal genome evolution (Bao *et al.* 2017; Faino *et al.* 2016; Seidl and Thomma 2017; Zhong *et al.* 2018). The movement of TEs also contributes to fungal pathogenicity variation by disrupting AVR genes. For instance, a Pot3 DNA transposon insertion in the presumed promoter of *AVRPiz-t*, *AVRPi-ta* and *AVRPib* lead the inserted *M. oryzae* strains to evade recognition by the respective rice resistance genes *Piz-t*, *Pi-ta* and *Pib* (Kang *et al.* 2001; Li *et al.* 2009b; Zhang *et al.* 2015a). Moreover, a genome-wide analysis revealed that higher Pot3 insertions frequency in the flanking region of secreted proteins than non-secreted proteins (Bao *et al.* 2017). In addition to DNA transposon, insertion of a short interspersed nuclear element, Mg-SINE, in the intron of *AVRPi9* from a sequential infection assay, allows the strain to overcome the resistance from *Pi9* (Wu *et al.* 2015).

It is not clear how proximal TE insertions mechanistically impact the function of AVR coding sequences allowing pathogens to overcome resistance. The generic hypothesis is that the TE insertions dysregulate transcription, but there have not been systematic studies to determine the mechanism by which transcription is altered. Two competing hypotheses are that the cis-regulatory logic (i.e., regulatory binding elements) are broken by the TE insertion, or that the

chromatin status of the locus is altered to prevent normal regulation. Previous efforts to address the impact of TE insertions on effector regulation relied on ectopic expression of coding sequences using native promoters, but this does not directly assess how the TE impacts transcriptional regulation at the native locus (Li *et al.* 2009b). The chromatin status and regulatory elements would be different between native- and ectopic loci. Allele swapping was used to create fungal strains without such described TE insertions, but the approach introduced a constitutively expressed selection cassette in proximity to the gene of interest, which might confound the observed impact on transcriptional activity (Meile *et al.* 2020; Wu *et al.* 2015). To our knowledge, efforts to create isogenic lines with or without the TE insertion at the native locus have not been reported in filamentous pathogens.

With the development and optimization of CRISPR-Cas toolkits, alternation of DNA sequences, including specific deletions and insertions, can be easily achieved in filamentous fungi (Cong *et al.* 2013; Foster *et al.* 2018; Leisen *et al.* 2020; Zetsche *et al.* 2015). In order to test the role of cis-TE insertion on transcriptional regulation of effectors, we deleted a Pot3 transposon insertion in the promoter of *avrpiz-t* from *Piz-t* virulent *M. oryzae* isolate Guy11. Using two different Cas12a RNPs to flank the region of interest, we generated *M. oryzae* transformants that contained the *avrpiz-t* sequence without the Pot3 insertion. However, this strategy had a flaw and cannot rule out the possible ectopic integration and expression of the *AVRPiz-t* donor DNA. Results from pathogenicity assays showed that the Cas12a edited strains could no longer infect *Piz-t* containing rice.

Results

To directly test hypotheses on how TE insertions in regulatory DNA alter AVR expression in *M. oryzae*, we wanted to delete the Pot3 insertion found in the promoter of *avrpiz-t* from the field isolate Guy11 without introducing a selectable marker in the proximity of *avrpiz-t*. To achieve this, we used the CRISPR-Cas12a editing system as a RNP complex that we previously showed works in *M. oryzae* (Huang and Cook 2022b). We predicted that removing the Pot3 sequence from the upstream coding sequence of *avrPiz-t* would restore avirulence on rice plants expressing the corresponding NLR receptor *Piz-t* (Li *et al.* 2009b). We generated the Pot3 deletion mutants through co-transferring 3 different components: (1) dual RNP complexes targeting the upstream and downstream region of the Pot3 insertion; (2) a homologous recombination template for repairing the DNA double strand breaks caused by Cas12a; (3) a hygromycin (*HYG*) DNA donor lacking homology to *M. oryzae* genome sequences. The hygromycin DNA donor was presumed to be inserted in the genome randomly and used for selection (Figure 5.1a and 5.2). After two independent (A and B) fungal transformations and PCR screening, we found two potential Pot3 deletion mutations with approximately 1.8 kb smaller PCR amplified products compared with wild-type Guy11, which is consistent with the expected deletion of Pot3 (1,863 bp) (Figure 5.1a and b). To further confirm the deletion of Pot3 in the promoter, PCR products were Sanger sequenced from the two transformants. The sequencing results confirmed the complete deletion of Pot3 in both transformants ($\Delta pot3\#A11$ and $\Delta pot3\#B8$) (Figure 5.1c). The mutant, $\Delta pot3\#B8$, also contained an extra 8 bp deletion in the Cas12a downstream cutting site, likely the result of DNA DSB repair (Figure 5.1 c). To assess the accurate integration of repair template, Southern blotting was performed with Guy11, $\Delta pot3\#A11$, $\Delta pot3\#B8$ and P131, which contains an intact *AVRPiz-t* promoter and coding sequence (Xue *et al.* 2012). The Southern blotting results showed a ~3.9kb

band from wild-type Guy11, and a ~2 kb band from P131, consistent with the former having the Pot3 insertion and not the latter (Figure 5.1d and 5.3). Similar to the results from P131, a ~ 2 kb band was also detected in $\Delta pot3\#B8$, but additional signal was also detected in the upper panel of the gel (Figure 5.1d), possibly indicating additional integration of the DNA segment at other loci or the accumulation of the undigested DNA. Surprisingly, only a faint band near 3.9 kb was detected from the $\Delta pot3\#A11$ transformant, which is conflicted with the PCR amplification, and may be due to a more complex DNA integration than expected (Figure 5.1b, d and 5.3).

In order to test the transcriptional impact on *AVRPiz-t* when Pot3 was deleted, we performed pathogenicity tests using Guy11, $\Delta pot3\#A11$, $\Delta pot3\#B8$, P131 and 0.25% gelatin (negative control) to infect a susceptible rice cultivar, YT16 lacking the *Piz-t* resistance gene, and a resistant cultivar, Toride 1 containing the *Piz-t* resistance gene (Park *et al.* 2012; Zhou *et al.* 2006). We found that all four strains, including the two Pot3 deletion strains, produced typical blast lesions on YT16 (Figure 5.1e), indicating the Pot3 deletion does not generally affect virulence. As a negative control, 0.25% gelatin did not produce any lesion on the rice leaves (Figure 5.1e). However, when infecting Toride 1, only Guy11 infected plants showed minor blast symptoms, albeit with less virulence compared to YT16 infection (Figure 5.1e). The other three strains, P131 and the two Pot3 deletion mutants cannot infect Toride 1 (Figure 5.1e). The observation that the two Pot3 edited strains could no longer infect Toride 1 suggests that *AVRPiz-t* was now expressed and recognized. A further assay assessing the expression of *AVRPiz-t* would provide a more direct and quantitative evidence for the restored transcription. These results show that replacing the *AVRPiz-t* promoter and coding sequence at the native locus, lacking the Pot3 insertion, might restore the expected *R-AVR* interaction. However, the current results still cannot

determine if this restoration is due to repairing specific cis-regulatory DNA sequences or by altering the chromatin of the locus due to removal of Pot3.

Discussions

The availability of complete genome assemblies and high-quality repeat annotations has reinforced the importance of TEs in fungal genome stability and evolution (Bao *et al.* 2017; Faino *et al.* 2016). Moreover, taking advantage of these sequencing resources, the role of TEs in other biological process has also been explored. For instance, TEs have been identified as cis-regulators (e.g., enhancer and silencer) for transcriptional control in model systems (Sundaram and Wysocka 2020). However, it is not clear what role TEs play in transcriptional regulation of filamentous pathogen genes. We took advantage of a well-known Pot3 insertion in the presumed promoter of *avrpiz-t* to investigate its role in transcriptional regulation. The role of Pot3 insertion in the expression of *AVRPiz-t* can be easily monitored by testing the pathogenicity against *Piz-t* resistant cultivar. Our pathogenicity data suggested that resistance interaction between *AVRPiz-t* and *Piz-t* might be restored by removal of Pot3 cis-insertion by CRISPR-Cas12a genome editing. We cannot distinguish between our hypotheses (i.e., broken regulatory binding elements vs TE mediated chromatin silencing) based on current pathogenicity results. We also note that the two mechanisms are not necessarily mutually exclusive. Restored avirulence is expected for both, and further experiments, such as assessing H3K27me3 levels or chromatin status, are required to test these hypotheses.

When we designed these experiments, we overlooked the impact of ectopic insertion of swap DNA donor that contains the intact *AVRPiz-t* sequence, which could restore the avirulence phenotype. We are not sure whether there are additional ectopic insertions of swap DNA donor

from our current Southern results. Consequently, we cannot fully rule out the importance of swap DNA donor in the change of pathogenicity. A further Southern blotting with a different strategy for restriction enzyme digestion (i.e., one cutting within the sequence from swap DNA donor, another one cutting away from the swap DNA donor) might be helpful to test the ectopic insertion. However, this potential ectopic insertion has no connection to testing our hypotheses. Additional experiments to develop a DNA-free editing system, such as targeting endogenous locus for selection (e.g., *FKBP12* or *URA3/5*) (Huang and Cook 2022b) or using transient DNA donor for selection (e.g., telomere vectors) are needed (Leisen *et al.* 2020).

Our results suggested that individual TE manipulation can be achieved by applying dual CRISPR-Cas12a editing. In order to test the function of TE insertion systematically, it is necessary to develop a platform to target multiple TEs simultaneously and test their impact on expression individually (e.g., single-cell sequencing). Taken together, our result suggested an unexplored role of cis-insertion of TE in fungal transcriptional regulation.

Methods and materials

Fungal strains

M. oryzae field isolates Guy11 (French Guyana) was used as wild-type in this study (Bao *et al.* 2017). *M. oryzae* field isolate P131 (Japan) was used *Piz-t* pathogenicity assay (Xue *et al.* 2012).

***In vitro* crRNA synthesis and LbCas12a RNP assembly**

Oligos containing 23 nt target sequences were used for preparing DNA template for *in vitro* RNA synthesis. crRNAs were synthesized with HiScribe™ T7 High Yield RNA Synthesis Kit (New England BioLabs, catalog# E2040S) with above DNA template according to manufacturer's

protocol, followed by purification with Monarch® RNA Cleanup Kit (NEB, catalog# T2050L). 5 µg purified LbCas12a (NEB, catalog# M0653T) was incubated with equal molar purified crRNA at 25°C for 15mins for RNP assembly (Huang and Cook 2022b).

Polyethylene glycol (PEG) mediated transformation

M. oryzae protoplast preparation was performed as described previously (Huang and Cook 2022b). For protoplast transformation, dual RNP complexes (5 µg LbCas12a protein complexed with equal molar amount of crRNA) and donor DNA (near 2-3 µg, amplified from pBV1167) were mixed with 200 µl concentrated protoplasts at room temperature for 20 minutes. 1 mL 60% PEG solution (60% PEG4000, 20% sucrose, 50 mM Tris-HCl pH=8.0, 50 mM CaCl₂) was added to above mixture and incubated at room temperature for 20 mins. The mixture of DNA-protoplasts-PEG was followed by incubated with 5 mL TB3 liquid medium at 28°C, 90rpm, for 10-18 hours. Overnight fungal culture was mixed with 50mL molten (near 50-60°C) TB3 solid medium plus 100 µg/mL hygromycin which poured in plate (150 x 15 mm) until dried then followed by overlaid with another 50 mL molten TB3 solid medium plus 200 µg/mL hygromycin, and cultured in dark condition at 28°C for 5-7 days.

Pathogenicity assay

Two-three weeks old rice plants (YT16 (lacking Piz-t) and Toride-1 (containing Piz-t)) were used for pathogenicity test. 1×10^5 conidia from wild-type isolates Guy11, P131 and Pot3 deletion strains were collected from 2 weeks old OTA plates through two layers of miracloth with 0.25% gelatin. The symptom investigations were performed five days after inoculation. Three biological replicates were performed with similar results.

Southern blotting

Electrophoresed *Bam*HI digested DNA (Figure 5.3) was run on 0.8% agarose gel and a smear in the lanes was observed. The gel was treated with depurination solution (0.25N-HCl) for 10 min until the blue dye color changing to yellow. The gel was washed with distilled water for 5 min and treated with denaturation solution (0.5M NaOH and 1.5M NaCl) for 45 min. The change of dye color into blue was found. Following washing with distilled water for 5 min. the gel was treated with neutralization solution (0.5M Tris-HCl and 1.5 M NaCl, pH 7.2) twice for 30 min and 45 min. The gel was washed with distilled water for 5 min and transferred to 10X SSC (3.0 M NaCl and 0.3M sodium citrate). The gel size was measured and the membrane and three Whatman papers of same size were prepared and dipped in 10X SSC. The gel was transferred onto the Whatman paper wick upside down. The membrane was moved on top of the gel using forceps. The sides of the gel were covered using parafilm to allow the transfer only through the gel. Three Whatman papers were put on top of the membrane and then stack of papers. On top of the stack, 500g weight (e.g., book) was added. Transfer step would be proceeded for overnight. Next day, the set up was carefully removed and the lanes on the blot were marked using a pencil. The membrane was washed with 6X SSC for 5 min and crosslinked using UV transilluminator for 1 min on each side. After that, the membrane was baked at 80°C for 1 hr. The protocol of DIG labelling kit for pre-hybrid and hybrid was followed at 42°C and membrane was visualized via Odyssey® FC Imaging System. (The Southern blotting protocol was kindly provided by Dr. Veerendra Sharma, KSU)

Figures

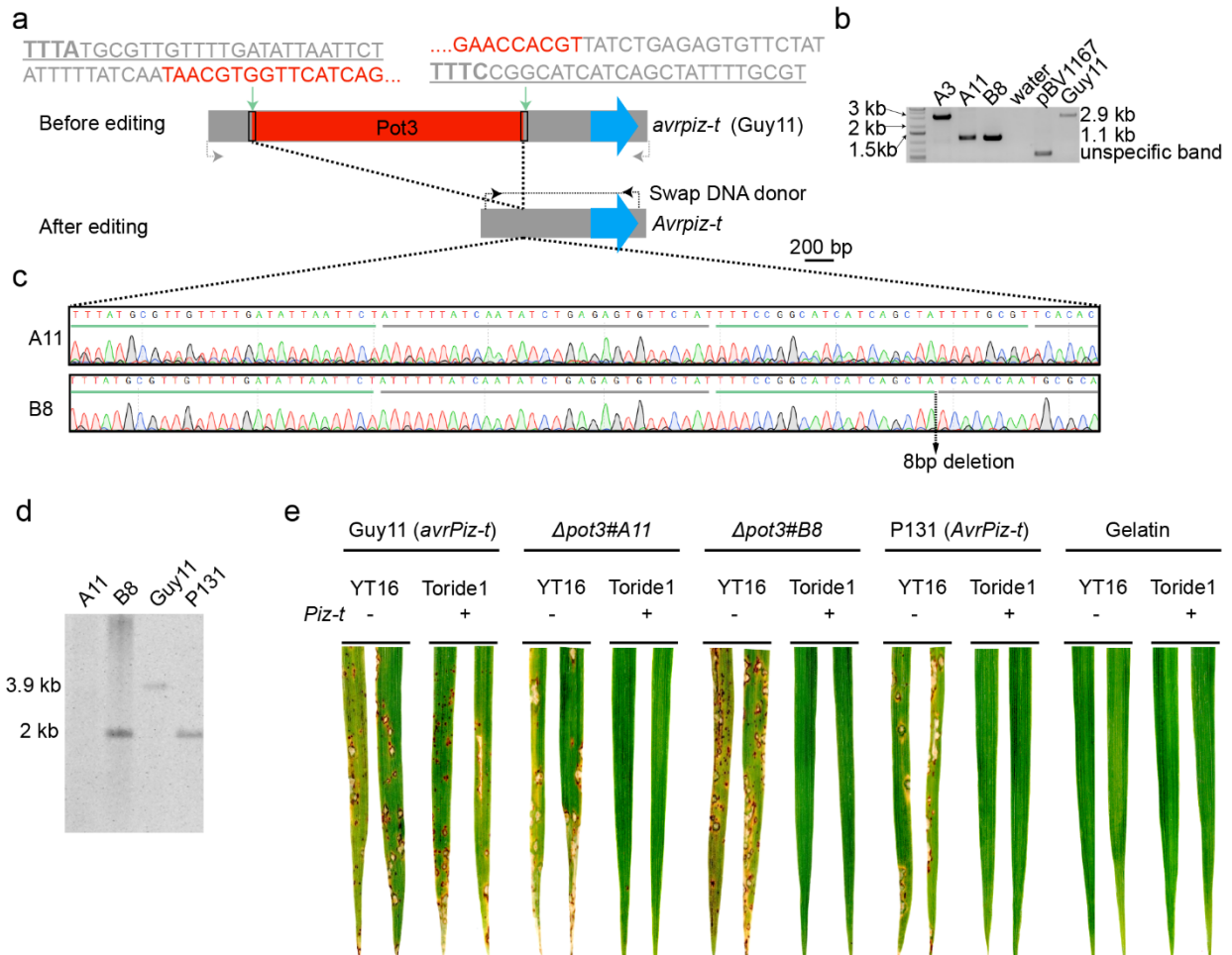


Figure 5.1. Deletion Pot3 in the promoter of *avrPiz-t* restores the avirulence to *Piz-t* rice

(a) Illustration diagram of *avrPiz-t* sequence before and after editing. Sequences above the *avrPiz-t* are the targeting region including 4 bp PAM sequence (bold grey and underlined letters), 23 bp targeting sequence (grey and underlined letters), neighboring sequence (grey letters) and Pot3 DNA transposable sequence (red letters). Swap DNA donor was used as repair template for homologous recombination, which was amplified from pBV1167 as the primer pair indicated by black arrows. **(b)** Genotyping PCR was performed with the primer pair indicated as grey arrows

in (a), A3, A11, B8 are the representative transformants after hygromycin selection. An unspecific band was amplified from pBV1167 with grey genotyping primer pair. **(c)** Sanger sequencing results of the A11 and B8's amplified products from (b), 8bp deletion was found in the potential Cas12a downstream cutting site in B8. **(d)** Southern blotting analysis of the integration of swap DNA donor and integrated number in A11 and B8. **(e)** YT16 (lacking *Piz-t*) and Toride 1 (*Piz-t* donor) cultivars were inoculated with Guy11, $\Delta pot3$ #A11, $\Delta pot3$ #B8, P131 and 0.25% gelatin solution for pathogenicity assays.

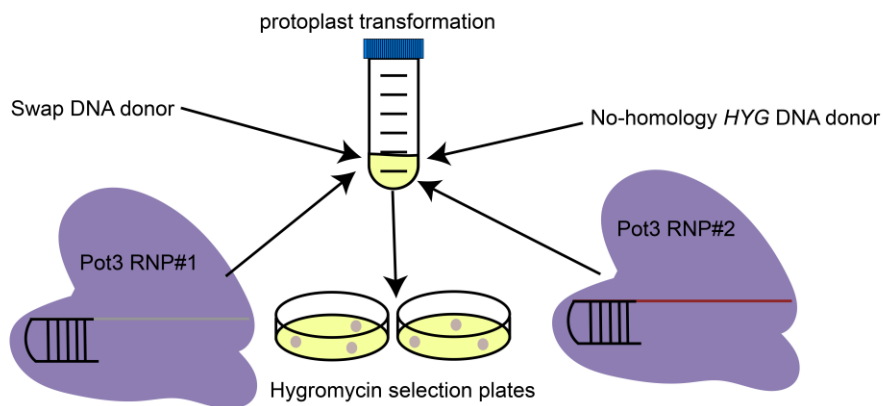


Figure 5.2. Illustration diagram of 3 components used in fungal transformation

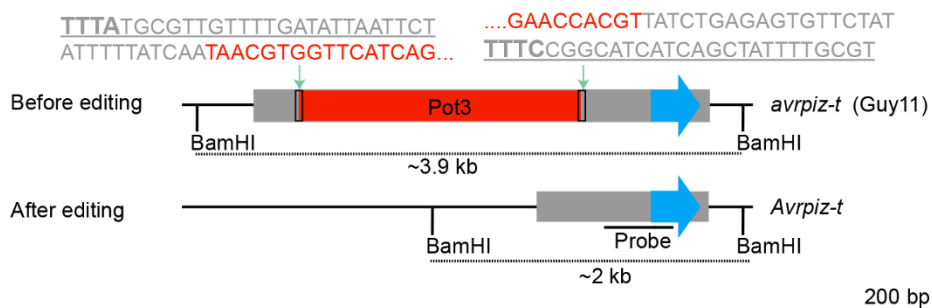


Figure 5.3. Strategy of Southern blotting confirmation

*Bam*HI was the restriction enzyme used for digesting the fungal genomics DNA. Probe matches *AvrPiz-t* promoter and coding sequences was designed. Expected detected size for CRISPR editing is ~3.9kb, the size would change to ~2kb after editing.

Chapter 6 - Development of FuNCaP, a method for isolating fungal nuclei from *in-planta* infected sample

Note: This chapter has been not published yet. Jun Huang performed all the experiments under the supervision and guidance of David E. Cook.

Abstract

The filamentous phytopathogenic fungus *Magnaporthe oryzae*, the causal agent of severe rice and wheat blast diseases, leads to significant crop yield loss. In addition to its threat to agriculture, *M. oryzae* also serves as a model system to study host-pathogen interaction. Significant research effort has yielded a great understanding of blast biology, but there are many unknown aspects of *M. oryzae* infection. For instance, effectors are silent *in vitro* but highly expressed during *in planta* growth, and how this strict transcriptional regulation is achieved during host infection remains unknown. One challenge to investigating the genomics of *in planta* fungal growth is caused by the relatively limited fungal material that can be extracted compared to the host plant tissue.

In order to advance our understanding of fungal transcriptional regulation during *in planta* infection, we proposed to develop a method termed fungal nuclei capture from plants (FuNCaP). The approach was to biotinylate fungal nuclei that could be isolated from infected tissue by affinity purification. The approach to enrich fungal nuclei was based on the method developed in *Arabidopsis* to isolate tissue-specific nuclei. Theoretically, purified, and enriched fungal nuclei can be captured by this method to serve as the initial materials for further sequencing. Here, we present an approach and proof of concept for labeled nuclei that can be used for future work to dissect *in planta* chromatin and epigenomic profiles.

Introduction

Magnaporthe oryzae is a globally important plant pathogen with different host adapted pathotypes able to infect rice, wheat and multiple monocots. (Bao *et al.* 2017; Peng *et al.* 2019; Zhong *et al.* 2016). Yield losses caused by rice blast disease can reach up to 100% in some paddies (Dean *et al.* 2012). There is extensive knowledge about fungal growth and development during *in vitro* growth, including conidium germination, appressorium formation and penetration (Kershaw *et al.* 2019; Oses-Ruiz *et al.* 2021; Sakulkoo *et al.* 2018; Wilson and Talbot 2009; Xu and Hamer 1996). However, transcriptional regulation responding to fungal development and virulence within host cells is poorly understood. For instance, *M. oryzae* forms filamentous and bulbous invasive hyphae within plant cells, but the genes responsible for this development are not clear (Kankanala *et al.* 2007; Khang *et al.* 2010). The underlying signals or genes involved in the formation of the biotrophic interfacial complex (BIC) structure are also not understood (Khang *et al.* 2010). Our lab previously found that histone modification dynamics at H3K27 are associated with altered gene expression between *in vitro* and *in planta* growth (Zhang *et al.* 2021). However, we were not able to dissect the whole-genome distribution of these histone modifications during *in planta* growth due to limited amounts of fungal materials that could be harvested during infection. A major limitation to studying gene regulation for *in planta* induced genes is related to the difficulty in collect fungal material from infected plant host. This is because that majority of extracted material using traditional homogenization techniques is from plant cells, and resulting data from dual-species sequencing mainly represents the host response (Kawahara *et al.* 2012). This highlights the necessity of developing a method to enrich fungal material for investigating fungal transcriptional control during host infection. Previously, enrichment of fungal RNA, up to ~20% of total sequenced tissue, was achieved by trimming away uninfected leaf sheath materials

(Mosquera *et al.* 2009). In many animal systems, cell sorting has been applied for the purpose of isolating specific cell types, but it requires specialized equipment, significant starting material, and to our knowledge, has not been applied to isolate filamentous fungal material from mixed species samples.

In order to address this problem, the Henikoff lab developed and optimized a method called INTACT (isolation of nuclei tagged in specific cell types) for profiling the epigenome of hair and non-hair cell types from the *Arabidopsis* root epidermis. This system requires two major components- i) an expressed *Escherichia coli* biotin ligase BirA in the organism; ii) an expressed nuclear envelope protein under control of a cell type specific promoter fused with biotin ligase recognition peptide (BLRP) and a visual marker (e.g., GFP). The tagged nuclear envelope protein under the designed promoter labels cell type-specific nuclei. Following nuclear biotinylation mediated by BirA, the biotinylated nuclei from specific cell type can be immunopurified by streptavidin beads (i.e., affinity purification) and the extracted material would be used for further sequencing (e.g., ChIP-seq for detecting histone modifications and ATAC-seq for dissecting chromatin accessibility). The INTACT system has been applied to a variety of eukaryotes for cell type specific sequencing, such as *Arabidopsis*, *Oryza sativa*, *Caenorhabditis elegans* and *Drosophila melanogaster* (Deal and Henikoff 2010; Deal and Henikoff 2011; Reynoso *et al.* 2018; Steiner *et al.* 2012). A similar strategy was adapted for the isolation of mitochondria in specific plant cell types, termed IMTACT (Boussardon *et al.* 2020). Based on the INTACT method, we designed our FuNCaP approach with the goal of isolating fungal nuclei from infected plant samples. The major steps of FuNCaP includes I) labelling the fungal nuclei within the infected tissues; II) releasing the labelled fungal nuclei from infected materials via grinding; III) purifying

the biotinylated fungal nuclei by streptavidin dynabeads pull down; IV) sequencing on the harvested DNA or RNA from isolated nuclei (Figure 6.1).

Our results suggested that *M. oryzae* nuclei can be biotinylated successfully via the interaction between BirA and tagged nuclear envelope protein. However, we were not able to enrich the fungal nuclei based on current transformants and protocol, which might need further optimization.

Results and Discussions

As the substrate for the *BirA*-mediated biotinylation, an appropriate nuclear targeting protein is required for the downstream nuclei tagging and isolation. The nuclear targeting fusion (NTF) proteins used in previously described INTACT approaches include *Arabidopsis* RAN GTPASE ACTIVATING PROTEIN 1 (RanGAP1) and *C. elegans* NPP-9 (RANBP2/NUP358 homolog) (Deal and Henikoff 2010; Steiner *et al.* 2012). Neither proteins are present in *M. oryzae*. In order to find the suitable fungal NTF for our FuNCaP method, we first searched the annotated or functional validated nuclear envelope proteins from the model fungi *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Five potential nuclear envelope proteins with a corresponding homolog in budding or fission yeast were identified in *M. oryzae* (Figure 6.2a). These five candidates have different localization in the nuclear pore complex (NPC) (Knockenbauer and Schwartz 2016). Gle1, Gle2 and Nsp1 localize in the cytoplasmic region, while Ndc1 and Pom34 reside in trans-membrane nucleoporins (Figure 6.2a). Theoretically, all five proteins can be accessed by streptavidin beads on the exterior of the nucleus.

To express and biotinylate these nuclear envelope proteins in *M. oryzae*, we synthesized a DNA fragment to express the transgenes. One was a *M. oryzae* codon optimized *BirA* fused with

3x HA tag, and under transcriptional control of the RP27 constitutive promoter. The other open reading frame was a human codon optimized EGFP linked with *BirA* recognition and biotinylating sequence (a 45 bp Avitag), also transcriptionally controlled by the RP27 constitutive promoter. Except for Pom34, the other four candidates were successfully inserted in between RP27 promoter and EGFP in the destination vector pFunag27H via restriction enzyme-based cloning (Figure 6.2b). The empty vector pFunag27H and the cloned vector inserted with *GLE1*, *GLE2*, *NSP1* and *NDC1* were further ectopically inserted into Guy11 genome via *Agrobacterium tumefaciens*-mediated transformation (ATMT). The transformant with the empty vector pFunag27H was used as a control strain to rule out the potential unspecific localization and enrichment due to the empty vector.

As predicted, transformants expressing the candidate NTF proteins showed clear nuclear envelope localization in conidium (Figure 6.2c). Further rice leaf sheath assays suggested the same pattern for the transformants expressing either Ndc1 or Gle2 (Figure 6.2d). Transformants expressing free EGFP did not show specific nuclear envelope localization but displayed a general cytoplasmic distribution (Figure 6.2c and d). To confirm that the nuclear envelope-GFP fusions were biotinylated, we performed immunoprecipitation using streptavidin beads, followed by western blotting using an antibody against EGFP. We expected to see the western blotting signal near the size of fused protein (i.e., selected nuclear envelope protein + EGFP + Avitag). As a negative control, we did not observe any GFP signal due to lack of BirA and tagged nuclear envelope protein in the wild-type sample (Figure 6.2e). An expected band near 29.4 kDa (EGFP+Avitag) was found in two independent free-GFP samples, which indicated the success of BirA-mediated EGFP biotinylation (Figure 6.2e). Only transformants expressing Gle2 showed the correct size of GFP signal (i.e., 69.2kDa Gle2+EGFP+Avitag). Either no signal (Gle1, expected

90.1kDa) or incorrect size of signals (Nsp1, expected 95.4 kDa and Ndc1, expected 102.1 kDa) was found in the rest of the transformants (Figure 6.2e). Thus, we moved forward with Gle2 expressing transformants along with free GFP sample for the nuclei isolation. With some minor modifications on established INTACT protocol (Bajic *et al.* 2018), we incubated the mixed sample from infected materials with streptavidin beads for fungal nuclei enrichment. Strong binding between fungal nuclei and streptavidin beads is an indicator of successful fungal nuclei enrichment, while such binding should not be found in the transformant expressing free GFP. However, we did not find closer binding between beads and potential nuclei in Gle2 expressing transformants compared to free GFP transformants (Figure 6.2f). Following DNA extraction, we did not detect obvious difference between Gle2 and free-GFP samples (Gle2-18.1 ng DNA vs free GFP 16.95 ng DNA), which suggested that fungal nuclei were not enriched.

The poor fungal nuclei enrichment might be due to the following: I) inefficient BirA-biotinylation system. When we cloned Gle1, Gle2, Nsp1 and Ndc1 into the empty vector pFunag27H, all the NTF protein expressing lines showed the clear nuclear envelop localization. However, we only detected correct size of fusion protein for Gle2 after streptavidin pull down. This result suggested that some unknown factors might disrupt the normal biotinylation process. An alternative option would be using GFP-beads to isolate the GFP-labelled fungal nuclei instead of using streptavidin beads. A similar assay has been done to isolate nuclei from *Drosophila* neurons (Henry *et al.* 2012); II) inefficient nuclei isolation and biotinylation step. We modified the nuclei isolation based on original INTACT method, while it will be necessary to adapt nuclei isolation protocol and provide a better input for streptavidin beads pulling down. Moreover, adding additional exogenous biotin into the pull-down solution might be also helpful for promoting biotinylation activity (Cho *et al.* 2020).

Experimental Procedures

Fungal strains

M. oryzae wild-type Guy11 (French Guyana) was used as the recipient strain for all the fungal transformation in this study (Bao *et al.* 2017).

DNA constructions and transformation

The DNA sequence including pRP27+ fungal codon optimized BirA+3xHA and pRP27+multiple cloning site (MCS) + human codon optimized EGFP + Avitag was synthesized and ordered from GenScript (<https://www.genscript.com/>). This DNA fragment was further cloned into an ATMT compatible vector (i.e., pFGL821), and termed pFunag27H. *GLE1*, *GLE2*, *NSP1*, *POM34* and *NDC1* were amplified with gene specific primer pairs and cloned into pFunag27H with the available MCS. ATMT transformations were performed as described in chapter 2.

Microscopy

Conidia and invasion hyphae from 24-36 h post infection were used for detecting GFP signal with Zeiss Axioplan 2 IE MOT microscope.

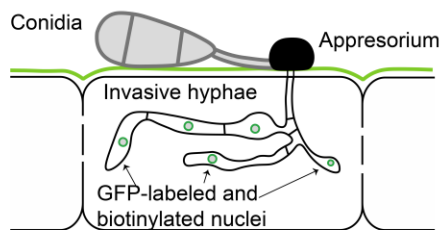
Pull down by streptavidin beads

Pure fungal or infected materials were ground in liquid nitrogen by using mortar/pestle for 5 mins to break the tissue and resuspended in 10 mL FNPB buffer (20 mM MOPS-pH 7, 20 mM NaCL, 45 mM KCL, 0.5 mM spermidine, 0.2 mM spermine, 1% proteinase inhibitor, 250 mM sucrose and 40% glycerol). Following filtering through 40 µm cell strainer, sample was spin down at 3,000 g for 15 min at 4 °C. After centrifugation, the supernatant was discarded and the pellet was resuspended with 1mL FNPBI buffer (20 mM MOPS-pH 7, 20 mM NaCL, 45 mM KCL, 0.5 mM spermidine, 0.2 mM spermine, 1% proteinase inhibitor, 50 mM sucrose and 5% glycerol). The resuspension was slowly rotated with 50 µL washed Dynabeads™ M-280 Streptavidin at 4 °C for

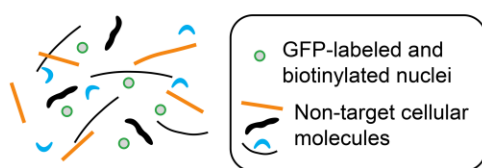
30 min. Following rotation, sample was moved to a 15 ml falcon tube and complemented with additional 14 mL FNPBlt buffer (20 mM MOPS-pH 7, 20 mM NaCL, 45 mM KCL, 0.5 mM spermidine, 0.2 mM spermine, 1% proteinase inhibitor, 0.1% Triton-x 100). The total 15 mL solution was incubated in the magnetic rack at 4 °C for 20 min. After incubation, the supernatant was removed, the pellet was resuspended with another 14 mL of FNPBlt buffer, slowly rotated for 30s and put inside the magnetic rack at 4°C for 20 min. The supernatant was discarded, and the pellet was resuspended with 100 µL FNPB buffer, this final solution was used for further microscopy confirmation and DNA extraction and quantification. DAPI (final concentration 1 µg/mL) was used for staining nuclei.

Figures

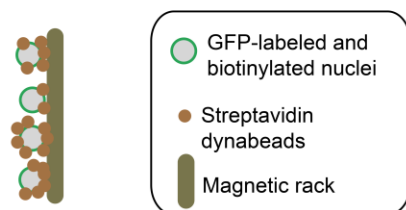
1. Infect rice with *M. oryzae* expressing GFP-labeled and biotinylated nuclei



2. Freeze and grind the infected leaf materials



3. Capture the labelled nuclei by streptavidin dynabeads pull down



4. High-throughput sequencing to identify the fungal regulatory mechanisms

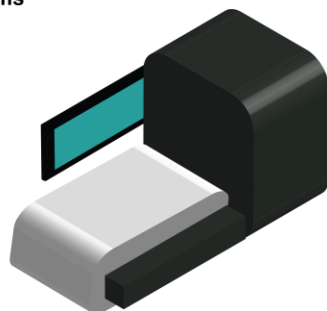


Figure 6.1. Workflow of FuNCaP

Step1. Nuclear labelled strain is used to infect rice leaf sheath samples; **Step2.** infected sample is ground to release the fungal nuclei; **Step3.** fungal nuclei are captured by streptavidin dynabeads; **Step4.** harvested nuclei are used for high-throughput sequencing for detecting *in planta* regulatory mechanisms.

free-GFP sample was used as control. **(e)** Confirmation of nuclear biotinylation by using anti-GFP antibody against streptavidin pull down samples. **(f)** Microscopy checking the final products of FuNCaP enrichment. DAPI was used for staining the potential nuclei in the solution. Transformant expressing free GFP was used for determining the specificity of fungal nuclei capture.

References

- Abdul W, Aliyu SR, Lin L *et al.* Family-Four Aldehyde Dehydrogenases Play an Indispensable Role in the Pathogenesis of *Magnaporthe oryzae*. *Front Plant Sci* 2018;**9**: 980.
- Adikusuma F, Piltz S, Corbett MA *et al.* Large deletions induced by Cas9 cleavage. *Nature* 2018;**560**: E8-E9.
- Ah-Fong AMV, Boyd AM, Matson MEH *et al.* A Cas12a-based gene editing system for *Phytophthora infestans* reveals monoallelic expression of an elicitor. *Mol Plant Pathol* 2021;**22**: 737-52.
- Ahnesorg P, Smith P, Jackson SP. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* 2006;**124**: 301-13.
- Ahrabi S, Sarkar S, Pfister SX *et al.* A role for human homologous recombination factors in suppressing microhomology-mediated end joining. *Nucleic Acids Res* 2016;**44**: 5743-57.
- Al-Minawi AZ, Saleh-Gohari N, Helleday T. The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic Acids Res* 2008;**36**: 1-9.
- Al Abdallah Q, Ge W, Fortwendel JR. A Simple and Universal System for Gene Manipulation in *Aspergillus fumigatus*: *In Vitro*-Assembled Cas9-Guide RNA Ribonucleoproteins Coupled with Microhomology Repair Templates. *mSphere* 2017;**2**.
- Alagoz M, Katsuki Y, Ogiwara H *et al.* SETDB1, HP1 and SUV39 promote repositioning of 53BP1 to extend resection during homologous recombination in G2 cells. *Nucleic Acids Res* 2015;**43**: 7931-44.
- Alanis-Lobato G, Zohren J, McCarthy A *et al.* Frequent loss of heterozygosity in CRISPR-Cas9-edited early human embryos. *Proc Natl Acad Sci U S A* 2021;**118**.
- Allen F, Crepaldi L, Alsinet C *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol* 2018, DOI 10.1038/nbt.4317.
- Anand R, Ranjha L, Cannavo E *et al.* Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol Cell* 2016;**64**: 940-50.
- Anzalone AV, Randolph PB, Davis JR *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;**576**: 149-57.
- Aparicio T, Baer R, Gautier J. DNA double-strand break repair pathway choice and cancer. *DNA Repair (Amst)* 2014;**19**: 169-75.
- Aravind L, Walker DR, Koonin EV. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 1999;**27**: 1223-42.
- Argueso JL, Westmoreland J, Mieczkowski PA *et al.* Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc Natl Acad Sci U S A* 2008;**105**: 11845-50.
- Arnoult N, Correia A, Ma J *et al.* Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature* 2017;**549**: 548-52.
- Audebert M, Salles B, Calsou P. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J Biol Chem* 2004;**279**: 55117-26.
- Aylon Y, Liefshitz B, Kupiec M. The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *EMBO J* 2004;**23**: 4868-75.
- Aymard F, Bugler B, Schmidt CK *et al.* Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* 2014;**21**: 366-74.
- Ayrappetov MK, Gursoy-Yuzugullu O, Xu C *et al.* DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *Proc Natl Acad Sci U S A* 2014;**111**: 9169-74.
- Bai W, Zhu G, Xu J *et al.* The 3'-flap endonuclease XPF-ERCC1 promotes alternative end joining and chromosomal translocation during B cell class switching. *Cell Rep* 2021;**36**: 109756.

- Bajic M, Maher KA, Deal RB. Identification of Open Chromatin Regions in Plant Genomes Using ATAC-Seq. *Methods Mol Biol* 2018;**1675**: 183-201.
- Bao J, Chen M, Zhong Z *et al.* PacBio Sequencing Reveals Transposable Elements as a Key Contributor to Genomic Plasticity and Virulence Variation in *Magnaporthe oryzae*. *Mol Plant* 2017;**10**: 1465-8.
- Belmont AS. Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. *Cold Spring Harb Perspect Biol* 2021, DOI 10.1101/cshperspect.a041268.
- Bertrand C, Thibessard A, Bruand C *et al.* Bacterial NHEJ: a never ending story. *Mol Microbiol* 2019;**111**: 1139-51.
- Betermier M, Bertrand P, Lopez BS. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* 2014;**10**: e1004086.
- Bhargava R, Onyango DO, Stark JM. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet* 2016;**32**: 566-75.
- Bicocca VT, Ormsby T, Adhvaryu KK *et al.* ASH1-catalyzed H3K36 methylation drives gene repression and marks H3K27me2/3-competent chromatin. *Elife* 2018;**7**.
- Bohnert HU, Fudal I, Diah W *et al.* A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. *Plant Cell* 2004;**16**: 2499-513.
- Bonnet J, Lindeboom RGH, Pokrovsky D *et al.* Quantification of Proteins and Histone Marks in *Drosophila* Embryos Reveals Stoichiometric Relationships Impacting Chromatin Regulation. *Dev Cell* 2019;**51**: 632-44 e6.
- Borde V, de Massy B. Programmed induction of DNA double strand breaks during meiosis: setting up communication between DNA and the chromosome structure. *Curr Opin Genet Dev* 2013;**23**: 147-55.
- Boulton SJ, Jackson SP. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res* 1996a;**24**: 4639-48.
- Boulton SJ, Jackson SP. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J* 1996b;**15**: 5093-103.
- Boussardon C, Przybyla-Toscano J, Carrie C *et al.* Tissue-specific isolation of *Arabidopsis*/plant mitochondria - IMTACT (isolation of mitochondria tagged in specific cell types). *Plant J* 2020;**103**: 459-73.
- Boyd JB, Golino MD, Shaw KE *et al.* Third-chromosome mutagen-sensitive mutants of *Drosophila melanogaster*. *Genetics* 1981;**97**: 607-23.
- Boyd JB, Sakaguchi K, Harris PV. mus308 mutants of *Drosophila* exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease. *Genetics* 1990;**125**: 813-9.
- Brader G, Compant S, Vescio K *et al.* Ecology and Genomic Insights into Plant-Pathogenic and Plant-Nonpathogenic Endophytes. *Annu Rev Phytopathol* 2017;**55**: 61-83.
- Brambati A, Barry RM, Sfeir A. DNA polymerase theta (Poltheta) - an error-prone polymerase necessary for genome stability. *Curr Opin Genet Dev* 2020;**60**: 119-26.
- Cannavo E, Cejka P. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature* 2014;**514**: 122-5.
- Cannavo E, Johnson D, Andres SN *et al.* Regulatory control of DNA end resection by Sae2 phosphorylation. *Nat Commun* 2018;**9**: 4016.
- Cao Z, Yin Y, Sun X *et al.* An Ash1-Like Protein MokMT2H Null Mutant Is Delayed for Conidium Germination and Pathogenesis in *Magnaporthe oryzae*. *Biomed Res Int* 2016;**2016**: 1575430.
- Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol* 2016;**26**: 52-64.

- Chakrabarti AM, Henser-Brownhill T, Monserrat J *et al.* Target-Specific Precision of CRISPR-Mediated Genome Editing. *Mol Cell* 2019;**73**: 699-713 e6.
- Chan SH, Yu AM, McVey M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet* 2010;**6**: e1001005.
- Chandramouly G, Liao S, Rusanov T *et al.* Poltheta promotes the repair of 5'-DNA-protein crosslinks by microhomology-mediated end-joining. *Cell Rep* 2021;**34**: 108820.
- Chang HHY, Pannunzio NR, Adachi N *et al.* Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* 2017;**18**: 495-506.
- Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* 2017;**58**: 235-63.
- Chen C, Chen H, Zhang Y *et al.* TBtools - an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 2020, DOI 10.1016/j.molp.2020.06.009.
- Chen H, Lisby M, Symington LS. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol Cell* 2013a;**50**: 589-600.
- Chen H, Raffaele S, Dong S. Silent control: microbial plant pathogens evade host immunity without coding sequence changes. *FEMS Microbiol Rev* 2021, DOI 10.1093/femsre/fuab002.
- Chen H, Shu H, Fang Y *et al.* Genome-wide analyses of histone modifications and chromatin accessibility reveal the distinct genomic compartments in the Irish potato famine pathogen *Phytophthora infestans*. *bioRxiv* 2022, DOI 10.1101/2022.02.18.480484: 2022.02.18.480484.
- Chen J, Zheng W, Zheng S *et al.* Rac1 is required for pathogenicity and Chm1-dependent conidiogenesis in rice fungal pathogen *Magnaporthe grisea*. *PLoS Pathog* 2008;**4**: e1000202.
- Chen S, Songkumarn P, Venu RC *et al.* Identification and characterization of in planta-expressed secreted effector proteins from *Magnaporthe oryzae* that induce cell death in rice. *Mol Plant Microbe Interact* 2013b;**26**: 191-202.
- Chen X, Niu H, Chung WH *et al.* Cell cycle regulation of DNA double-strand break end resection by Cdk1-dependent Dna2 phosphorylation. *Nat Struct Mol Biol* 2011;**18**: 1015-9.
- Cho KF, Branon TC, Udeshi ND *et al.* Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nat Protoc* 2020;**15**: 3971-99.
- Choi Y-E, Shim W-B. Enhanced homologous recombination in *Fusarium verticillioides* by disruption of FvKU70, a gene required for a non-homologous end joining mechanism. *The Plant Pathology Journal* 2008;**24**: 1-7.
- Choquer M, Robin G, Le Pecqueur P *et al.* Ku70 or Ku80 deficiencies in the fungus *Botrytis cinerea* facilitate targeting of genes that are hard to knock out in a wild-type context. *FEMS Microbiol Lett* 2008;**289**: 225-32.
- Chu VT, Weber T, Wefers B *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* 2015;**33**: 543-8.
- Chujo T, Scott B. Histone H3K9 and H3K27 methylation regulates fungal alkaloid biosynthesis in a fungal endophyte-plant symbiosis. *Mol Microbiol* 2014;**92**: 413-34.
- Chuma I, Isobe C, Hotta Y *et al.* Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. *PLoS Pathog* 2011;**7**: e1002147.
- Chumley FG, Valent B. Genetic analysis of melanin-deficient, nonpathogenic mutants of *Magnaporthe grisea*. *Mol Plant-Microbe Interact* 1990;**3**: 135-43.
- Citarelli M, Teotia S, Lamb RS. Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. *BMC Evol Biol* 2010;**10**: 308.
- Clouaire T, Legube G. A Snapshot on the Cis Chromatin Response to DNA Double-Strand Breaks. *Trends Genet* 2019;**35**: 330-45.

- Clouaire T, Rocher V, Lashgari A *et al.* Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. *Mol Cell* 2018;**72**: 250-62 e6.
- Coleman JJ, Rounsley SD, Rodriguez-Carres M *et al.* The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet* 2009;**5**: e1000618.
- Collemare J, Pianfetti M, Houille AE *et al.* *Magnaporthe grisea* avirulence gene ACE1 belongs to an infection-specific gene cluster involved in secondary metabolism. *New Phytol* 2008;**179**: 196-208.
- Collemare J, Seidl MF. Chromatin-dependent regulation of secondary metabolite biosynthesis in fungi: is the picture complete? *FEMS Microbiol Rev* 2019;**43**: 591-607.
- Cong L, Ran FA, Cox D *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;**339**: 819-23.
- Connolly LR, Smith KM, Freitag M. The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. *PLoS Genet* 2013;**9**: e1003916.
- Cook DE, Kramer HM, Torres DE *et al.* A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen. *Elife* 2020;**9**.
- Cook DE, Mesarich CH, Thomma BP. Understanding plant immunity as a surveillance system to detect invasion. *Annu Rev Phytopathol* 2015;**53**: 541-63.
- Covo S. Genomic Instability in Fungal Plant Pathogens. *Genes (Basel)* 2020;**11**.
- Creutzburg SCA, Wu WY, Mohanraju P *et al.* Good guide, bad guide: spacer sequence-dependent cleavage efficiency of Cas12a. *Nucleic Acids Res* 2020;**48**: 3228-43.
- Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog* 2012;**8**: e1002608.
- Cromie GA, Connelly JC, Leach DR. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol Cell* 2001;**8**: 1163-74.
- Cullot G, Boutin J, Toutain J *et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat Commun* 2019;**10**: 1136.
- D'Ambrosio U, Alonso-Lifante MP, Barros K *et al.* B-chrom: a database on B-chromosomes of plants, animals and fungi. *New Phytol* 2017;**216**: 635-42.
- da Silva Ferreira ME, Kress MR, Savoldi M *et al.* The akuB(KU80) mutant deficient for nonhomologous end joining is a powerful tool for analyzing pathogenicity in *Aspergillus fumigatus*. *Eukaryot Cell* 2006;**5**: 207-11.
- Daley JM, Gaines WA, Kwon Y *et al.* Regulation of DNA pairing in homologous recombination. *Cold Spring Harb Perspect Biol* 2014;**6**: a017954.
- Daley JM, Wilson TE. Rejoining of DNA double-strand breaks as a function of overhang length. *Mol Cell Biol* 2005;**25**: 896-906.
- de Jonge R, Bolton MD, Kombrink A *et al.* Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res* 2013;**23**: 1271-82.
- de Jonge R, van Esse HP, Maruthachalam K *et al.* Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc Natl Acad Sci U S A* 2012;**109**: 5110-5.
- de Sena-Tomas C, Yu EY, Calzada A *et al.* Fungal Ku prevents permanent cell cycle arrest by suppressing DNA damage signaling at telomeres. *Nucleic Acids Res* 2015;**43**: 2138-51.
- Deal RB, Henikoff S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev Cell* 2010;**18**: 1030-40.
- Deal RB, Henikoff S. The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat Protoc* 2011;**6**: 56-68.

- Dean R, Van Kan JA, Pretorius ZA *et al.* The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol* 2012;**13**: 414-30.
- Dean RA, Talbot NJ, Ebbole DJ *et al.* The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 2005;**434**: 980-6.
- DeFalco TA, Zipfel C. Molecular mechanisms of early plant pattern-triggered immune signaling. *Mol Cell* 2021;**81**: 3449-67.
- Deng SK, Gibb B, de Almeida MJ *et al.* RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat Struct Mol Biol* 2014;**21**: 405-12.
- Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**: 15-21.
- Doksani Y, de Lange T. The role of double-strand break repair pathways at functional and dysfunctional telomeres. *Cold Spring Harb Perspect Biol* 2014;**6**: a016576.
- Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev* 2015a;**35**: 57-65.
- Dong Y, Li Y, Zhao M *et al.* Global genome and transcriptome analyses of *Magnaporthe oryzae* epidemic isolate 98-06 uncover novel effectors and pathogenicity-related genes, revealing gene gain and loss dynamics in genome evolution. *PLoS Pathog* 2015b;**11**: e1004801.
- Dutheil JY, Mannhaupt G, Schweizer G *et al.* A Tale of Genome Compartmentalization: The Evolution of Virulence Clusters in Smut Fungi. *Genome Biol Evol* 2016;**8**: 681-704.
- Escribano-Diaz C, Orthwein A, Fradet-Turcotte A *et al.* A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol Cell* 2013;**49**: 872-83.
- Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 2007;**8**: 286-98.
- Fagerlund RD, Staals RH, Fineran PC. The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biol* 2015;**16**: 251.
- Faino L, Seidl MF, Shi-Kunne X *et al.* Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res* 2016;**26**: 1091-100.
- Fang Y, Coelho MA, Shu H *et al.* Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. *PLoS Genet* 2020;**16**: e1008646.
- Farman ML. Meiotic deletion at the BUF1 locus of the fungus *Magnaporthe grisea* is controlled by interaction with the homologous chromosome. *Genetics* 2002;**160**: 137-48.
- Fattah F, Lee EH, Weisensel N *et al.* Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. *PLoS Genet* 2010;**6**: e1000855.
- Ferenczi A, Pyott DE, Xipnitou A *et al.* Efficient targeted DNA editing and replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-stranded DNA. *Proc Natl Acad Sci U S A* 2017;**114**: 13567-72.
- Fernandez J, Orth K. Rise of a Cereal Killer: The Biology of *Magnaporthe oryzae* Biotrophic Growth. *Trends Microbiol* 2018;**26**: 582-97.
- Fernandez JP, Vejnar CE, Giraldez AJ *et al.* Optimized CRISPR-Cpf1 system for genome editing in zebrafish. *Methods* 2018;**150**: 11-8.
- Fnu S, Williamson EA, De Haro LP *et al.* Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc Natl Acad Sci U S A* 2011;**108**: 540-5.
- Foster AJ, Martin-Urdiroz M, Yan X *et al.* CRISPR-Cas9 ribonucleoprotein-mediated co-editing and counterselection in the rice blast fungus. *Sci Rep* 2018;**8**: 14355.
- Fouche S, Oggenfuss U, Chanclud E *et al.* A devil's bargain with transposable elements in plant pathogens. *Trends Genet* 2022;**38**: 222-30.

- Frantzeskakis L, Kracher B, Kusch S *et al.* 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* 2018;**19**: 381.
- Frantzeskakis L, Kusch S, Panstruga R. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol* 2019;**20**: 3-7.
- Freitag M. Histone Methylation by SET Domain Proteins in Fungi. *Annu Rev Microbiol* 2017;**71**: 413-39.
- Friedberg EC. DNA damage and repair. *Nature* 2003;**421**: 436-40.
- Friedman N, Rando OJ. Epigenomics and the structure of the living genome. *Genome Res* 2015;**25**: 1482-90.
- Fudal I, Bohnert HU, Tharreau D *et al.* Transposition of MINE, a composite retrotransposon, in the avirulence gene ACE1 of the rice blast fungus *Magnaporthe grisea*. *Fungal Genet Biol* 2005;**42**: 761-72.
- Fuller KK, Chen S, Loros JJ *et al.* Development of the CRISPR/Cas9 System for Targeted Gene Disruption in *Aspergillus fumigatus*. *Eukaryot Cell* 2015;**14**: 1073-80.
- Gadgil R, Barthelemy J, Lewis T *et al.* Replication stalling and DNA microsatellite instability. *Biophys Chem* 2017;**225**: 38-48.
- Galazka JM, Freitag M. Variability of chromosome structure in pathogenic fungi--of 'ends and odds'. *Curr Opin Microbiol* 2014;**20**: 19-26.
- Garcia V, Phelps SE, Gray S *et al.* Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature* 2011;**479**: 241-4.
- Giraldo MC, Dagdas YF, Gupta YK *et al.* Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nat Commun* 2013;**4**: 1996.
- Gisler S, Goncalves JP, Akhtar W *et al.* Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nat Commun* 2019;**10**: 1598.
- Gong F, Chiu LY, Cox B *et al.* Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination. *Genes Dev* 2015;**29**: 197-211.
- Gotz S, Garcia-Gomez JM, Terol J *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 2008;**36**: 3420-35.
- Grimme JM, Honda M, Wright R *et al.* Human Rad52 binds and wraps single-stranded DNA and mediates annealing via two hRad52-ssDNA complexes. *Nucleic Acids Res* 2010;**38**: 2917-30.
- Guo X, Zhong D, Xie W *et al.* Functional Identification of Novel Cell Death-inducing Effector Proteins from *Magnaporthe oryzae*. *Rice (N Y)* 2019;**12**: 59.
- Haas BJ, Kamoun S, Zody MC *et al.* Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 2009;**461**: 393-8.
- Haber JE. DNA repair. Gatekeepers of recombination. *Nature* 1999;**398**: 665, 7.
- Haber JE. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics* 2012;**191**: 33-64.
- Habig M, Lorrain C, Feurtey A *et al.* Epigenetic modifications affect the rate of spontaneous mutations in a pathogenic fungus. *Nat Commun* 2021;**12**: 5869.
- Hahne F, Ivanek R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol* 2016;**1418**: 335-51.
- Hamer JE, Farrall L, Orbach MJ *et al.* Host species-specific conservation of a family of repeated DNA sequences in the genome of a fungal plant pathogen. *Proc Natl Acad Sci U S A* 1989;**86**: 9981-5.
- Hamperl S, Cimprich KA. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell* 2016;**167**: 1455-67.
- Hao Z, Su X. Fast gene disruption in *Trichoderma reesei* using in vitro assembled Cas9/gRNA complex. *BMC Biotechnol* 2019;**19**: 2.

He X, Tan C, Wang F *et al.* Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res* 2016;**44**: e85.

Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 2014;**15**: 585-98.

Henry GL, Davis FP, Picard S *et al.* Cell type-specific genomics of *Drosophila* neurons. *Nucleic Acids Res* 2012;**40**: 9691-704.

Her J, Bunting SF. How cells ensure correct repair of DNA double-strand breaks. *J Biol Chem* 2018;**293**: 10502-11.

Hilton IB, D'Ippolito AM, Vockley CM *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 2015;**33**: 510-7.

Hogg M, Sauer-Eriksson AE, Johansson E. Promiscuous DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res* 2012;**40**: 2611-22.

Horikoshi N, Sharma D, Leonard F *et al.* Pre-existing H4K16ac levels in euchromatin drive DNA repair by homologous recombination in S-phase. *Commun Biol* 2019;**2**: 253.

Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;**157**: 1262-78.

Huang J, Cook DE. The contribution of DNA repair pathways to genome editing and evolution in filamentous pathogens. *FEMS Microbiol Rev* 2022a, DOI 10.1093/femsre/fuac035.

Huang J, Cook DE. CRISPR-Cas12a ribonucleoprotein-mediated gene editing in the plant pathogenic fungus *Magnaporthe oryzae*. *STAR Protoc* 2022b;**3**: 101072.

Huang J, Rowe D, Zhang W *et al.* CRISPR-Cas12a induced DNA double-strand breaks are repaired by locus-dependent and error-prone pathways in a fungal pathogen. *bioRxiv* 2021, DOI 10.1101/2021.09.08.459484: 2021.09.08.459484.

Huertas P, Cortes-Ledesma F, Sartori AA *et al.* CDK targets Sae2 to control DNA-end resection and homologous recombination. *Nature* 2008;**455**: 689-92.

Hung PJ, Johnson B, Chen BR *et al.* MRI Is a DNA Damage Response Adaptor during Classical Non-homologous End Joining. *Mol Cell* 2018;**71**: 332-42 e8.

Hwang WY, Fu Y, Reyon D *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 2013;**31**: 227-9.

Inoue H, Schroeder AL. A new mutagen-sensitive mutant in *Neurospora*, mus-16. *Mutat Res* 1988;**194**: 9-16.

Inoue Y, Vy TTP, Yoshida K *et al.* Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science* 2017;**357**: 80-3.

Ioannidi EI, Yarnall MTN, Schmitt-Ulms C *et al.* Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases. *bioRxiv* 2021, DOI 10.1101/2021.11.01.466786: 2021.11.01.466786.

Irieda H, Inoue Y, Mori M *et al.* Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. *Proc Natl Acad Sci U S A* 2019;**116**: 496-505.

Jackson SP. Sensing and repairing DNA double-strand breaks. *Carcinogenesis* 2002;**23**: 687-96.

Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature* 2009;**461**: 1071-8.

Jacobs JZ, Ciccaglione KM, Tournier V *et al.* Implementation of the CRISPR-Cas9 system in fission yeast. *Nat Commun* 2014;**5**: 5344.

Jamieson K, Rountree MR, Lewis ZA *et al.* Regional control of histone H3 lysine 27 methylation in *Neurospora*. *Proc Natl Acad Sci U S A* 2013;**110**: 6027-32.

Janevska S, Baumann L, Sieber CMK *et al.* 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* 2018;**208**: 153-71.

- Jha DK, Strahl BD. An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair. *Nat Commun* 2014;**5**: 3965.
- Jia Y, McAdams SA, Bryan GT *et al*. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J* 2000;**19**: 4004-14.
- Jones JD, Dangl JL. The plant immune system. *Nature* 2006;**444**: 323-9.
- Joubert PM, Krasileva KV. The extrachromosomal circular DNAs of the rice blast pathogen *Magnaporthe oryzae* contain a wide variety of LTR retrotransposons, genes, and effectors. *bioRxiv* 2021, DOI 10.1101/2021.10.12.464130: 2021.10.12.464130.
- Kahmann R, Basse C. REMI (restriction enzyme mediated integration) and its impact on the isolation of pathogenicity genes in fungi attacking plants. *European Journal of Plant Pathology* 1999;**105**: 221-9.
- Kamper J, Kahmann R, Bolker M *et al*. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 2006;**444**: 97-101.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**: 27-30.
- Kang S, Lebrun MH, Farrall L *et al*. Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. *Mol Plant Microbe Interact* 2001;**14**: 671-4.
- Kankanala P, Czymmek K, Valent B. Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. *Plant Cell* 2007;**19**: 706-24.
- Kawahara Y, Oono Y, Kanamori H *et al*. Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS One* 2012;**7**: e49423.
- Kent T, Chandramouly G, McDevitt SM *et al*. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol* 2015;**22**: 230-7.
- Kent T, Mateos-Gomez PA, Sfeir A *et al*. Polymerase theta is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *Elife* 2016;**5**.
- Kershaw MJ, Basiewicz M, Soanes DM *et al*. Conidial Morphogenesis and Septin-Mediated Plant Infection Require Smo1, a Ras GTPase-Activating Protein in *Magnaporthe oryzae*. *Genetics* 2019;**211**: 151-67.
- Khang CH, Berruyer R, Giraldo MC *et al*. Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. *Plant Cell* 2010;**22**: 1388-403.
- Khang CH, Park SY, Rho HS *et al*. Filamentous Fungi (*Magnaporthe grisea* and *Fusarium oxysporum*). *Methods Mol Biol* 2006;**344**: 403-20.
- Kim HK, Song M, Lee J *et al*. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat Methods* 2017;**14**: 153-9.
- Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* 2019;**20**: 207-20.
- Knockenbauer KE, Schwartz TU. The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell* 2016;**164**: 1162-71.
- Knoll A, Fauser F, Puchta H. DNA recombination in somatic plant cells: mechanisms and evolutionary consequences. *Chromosome Res* 2014;**22**: 191-201.
- Kojic M, Yang H, Kostrub CF *et al*. The BRCA2-interacting protein DSS1 is vital for DNA repair, recombination, and genome stability in *Ustilago maydis*. *Mol Cell* 2003;**12**: 1043-9.
- Koo DH, Molin WT, Saski CA *et al*. Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*. *Proc Natl Acad Sci U S A* 2018;**115**: 3332-7.
- Koole W, van Schendel R, Karambelas AE *et al*. A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun* 2014;**5**: 3216.

- Koren S, Walenz BP, Berlin K *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017;**27**: 722-36.
- Kosicki M, Allen F, Steward F *et al.* Cas9-induced large deletions and small indels are controlled in a convergent fashion. *Nat Commun* 2022;**13**: 3422.
- Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* 2018;**36**: 765-71.
- Kothe GO, Kitamura M, Masutani M *et al.* PARP is involved in replicative aging in *Neurospora crassa*. *Fungal Genet Biol* 2010;**47**: 297-309.
- Kramara J, Osia B, Malkova A. Break-Induced Replication: The Where, The Why, and The How. *Trends Genet* 2018;**34**: 518-31.
- Krieger LM, Mladenov E, Soni A *et al.* Disruption of Chromatin Dynamics by Hypotonic Stress Suppresses HR and Shifts DSB Processing to Error-Prone SSA. *Int J Mol Sci* 2021;**22**.
- Kumar R, Ichihashi Y, Kimura S *et al.* A High-Throughput Method for Illumina RNA-Seq Library Preparation. *Front Plant Sci* 2012;**3**: 202.
- Kumar S, Stecher G, Li M *et al.* MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 2018;**35**: 1547-9.
- Lamm N, Rogers S, Cesare AJ. Chromatin mobility and relocation in DNA repair. *Trends Cell Biol* 2021;**31**: 843-55.
- Langner T, Harant A, Gomez-Luciano LB *et al.* Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLoS Genet* 2021;**17**: e1009386.
- Lau-Corona D, Bae WK, Hennighausen L *et al.* Sex-biased genetic programs in liver metabolism and liver fibrosis are controlled by EZH1 and EZH2. *PLoS Genet* 2020;**16**: e1008796.
- Lax C, Navarro-Mendoza MI, Perez-Arques C *et al.* Stable and reproducible homologous recombination enables CRISPR-based engineering in the fungus *Rhizopus microsporus*. *Cell Rep Methods* 2021;**1**: 100124.
- Lee H, Kim JS. Unexpected CRISPR on-target effects. *Nat Biotechnol* 2018;**36**: 703-4.
- Lee K, Ji JH, Yoon K *et al.* Microhomology Selection for Microhomology Mediated End Joining in *Saccharomyces cerevisiae*. *Genes (Basel)* 2019;**10**.
- Lee K, Lee SE. *Saccharomyces cerevisiae* Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. *Genetics* 2007;**176**: 2003-14.
- Leisen T, Bietz F, Werner J *et al.* CRISPR/Cas with ribonucleoprotein complexes and transiently selected telomere vectors allows highly efficient marker-free and multiple genome editing in *Botrytis cinerea*. *PLoS Pathog* 2020;**16**: e1008326.
- Lemaitre C, Grabarz A, Tsouroula K *et al.* Nuclear position dictates DNA repair pathway choice. *Genes Dev* 2014;**28**: 2450-63.
- Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature* 1978;**276**: 565-70.
- Li G, Jain R, Chern M *et al.* The Sequences of 1504 Mutants in the Model Rice Variety Kitaake Facilitate Rapid Functional Genomic Studies. *Plant Cell* 2017;**29**: 1218-31.
- Li G, Zhou X, Xu JR. Genetic control of infection-related development in *Magnaporthe oryzae*. *Curr Opin Microbiol* 2012;**15**: 678-84.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**: 3094-100.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**: 1754-60.
- Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009a;**25**: 2078-9.
- Li J, Fokkens L, Conneely LJ *et al.* Partial pathogenicity chromosomes in *Fusarium oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environ Microbiol* 2020;**22**: 4985-5004.

- Li J, Zhang Y, Zhang Y *et al.* Introduction of Large Sequence Inserts by CRISPR-Cas9 To Create Pathogenicity Mutants in the Multinucleate Filamentous Pathogen *Sclerotinia sclerotiorum*. *mBio* 2018;**9**.
- Li W, Wang B, Wu J *et al.* The *Magnaporthe oryzae* avirulence gene AvrPiz-t encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene Piz-t. *Mol Plant Microbe Interact* 2009b;**22**: 411-20.
- Liang L, Deng L, Nguyen SC *et al.* Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Res* 2008;**36**: 3297-310.
- Liang M, Zhang S, Dong L *et al.* Label-Free Quantitative Proteomics of Lysine Acetylome Identifies Substrates of Gcn5 in *Magnaporthe oryzae* Autophagy and Epigenetic Regulation. *mSystems* 2018;**3**.
- Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 2010;**79**: 181-211.
- Lim G, Chang Y, Huh WK. Phosphoregulation of Rad51/Rad52 by CDK1 functions as a molecular switch for cell cycle-specific activation of homologous recombination. *Sci Adv* 2020;**6**: eaay2669.
- Liu M, Zhang W, Xin C *et al.* Global detection of DNA repair outcomes induced by CRISPR-Cas9. *Nucleic Acids Res* 2021;**49**: 8732-42.
- Liu R, Chen L, Jiang Y *et al.* Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. *Cell Discov* 2015;**1**: 15007.
- Liu S, Lin G, Ramachandran SR *et al.* Rapid mini-chromosome divergence among fungal isolates causing wheat blast outbreaks in Bangladesh and Zambia. *bioRxiv* 2022, DOI 10.1101/2022.06.18.496690: 2022.06.18.496690.
- Loeillet S, Palancade B, Cartron M *et al.* Genetic network interactions among replication, repair and nuclear pore deficiencies in yeast. *DNA Repair (Amst)* 2005;**4**: 459-68.
- Longmuir S, Akhtar N, MacNeill SA. Unexpected insertion of carrier DNA sequences into the fission yeast genome during CRISPR-Cas9 mediated gene deletion. *BMC Res Notes* 2019;**12**: 191.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**: 550.
- Luria SE, Delbruck M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* 1943;**28**: 491-511.
- Ma JL, Kim EM, Haber JE *et al.* Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol* 2003;**23**: 8820-8.
- Ma LJ, van der Does HC, Borkovich KA *et al.* Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 2010;**464**: 367-73.
- Mali P, Yang L, Esvelt KM *et al.* RNA-guided human genome engineering via Cas9. *Science* 2013;**339**: 823-6.
- Malzahn AA, Tang X, Lee K *et al.* Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and *Arabidopsis*. *BMC Biol* 2019;**17**: 9.
- Mansour WY, Rhein T, Dahm-Daphi J. The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. *Nucleic Acids Res* 2010;**38**: 6065-77.
- Mansour WY, Schumacher S, Roskopf R *et al.* Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. *Nucleic Acids Res* 2008;**36**: 4088-98.
- Mara K, Charlot F, Guyon-Debast A *et al.* POLQ plays a key role in the repair of CRISPR/Cas9-induced double-stranded breaks in the moss *Physcomitrella patens*. *New Phytol* 2019;**222**: 1380-91.

- Marroquin-Guzman M, Wilson RA. GATA-Dependent Glutaminolysis Drives Appressorium Formation in *Magnaporthe oryzae* by Suppressing TOR Inhibition of cAMP/PKA Signaling. *PLoS Pathog* 2015;**11**: e1004851.
- Mateos-Gomez PA, Gong F, Nair N *et al*. Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature* 2015;**518**: 254-7.
- Mathiasen DP, Lisby M. Cell cycle regulation of homologous recombination in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* 2014;**38**: 172-84.
- Mehta A, Haber JE. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol* 2014;**6**: a016428.
- Meile L, Peter J, Puccetti G *et al*. Chromatin Dynamics Contribute to the Spatiotemporal Expression Pattern of Virulence Genes in a Fungal Plant Pathogen. *mBio* 2020;**11**.
- Mentlak TA, Kombrink A, Shinya T *et al*. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell* 2012;**24**: 322-35.
- Meyer D, Fu BX, Heyer WD. DNA polymerases delta and lambda cooperate in repairing double-strand breaks by microhomology-mediated end-joining in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 2015;**112**: E6907-16.
- Milholland B, Dong X, Zhang L *et al*. Differences between germline and somatic mutation rates in humans and mice. *Nat Commun* 2017;**8**: 15183.
- Miller D, Reynolds GE, Mejia R *et al*. Subtelomeric regions in mammalian cells are deficient in DNA double-strand break repair. *DNA Repair (Amst)* 2011;**10**: 536-44.
- Mimitou EP, Symington LS. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J* 2010;**29**: 3358-69.
- Mojica FJM, Diez-Villasenor C, Garcia-Martinez J *et al*. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* 2009;**155**: 733-40.
- Moller M, Habig M, Freitag M *et al*. Extraordinary Genome Instability and Widespread Chromosome Rearrangements During Vegetative Growth. *Genetics* 2018;**210**: 517-29.
- Moller M, Schotanus K, Soyer JL *et al*. Destabilization of chromosome structure by histone H3 lysine 27 methylation. *PLoS Genet* 2019;**15**: e1008093.
- Moller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol* 2017;**15**: 771.
- Monroe JG, Srikant T, Carbonell-Bejerano P *et al*. Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* 2022;**602**: 101-5.
- Mosquera G, Giraldo MC, Khang CH *et al*. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as Biotrophy-associated secreted proteins in rice blast disease. *Plant Cell* 2009;**21**: 1273-90.
- Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol* 2010;**11**: 196-207.
- Munoz MC, Laulier C, Gunn A *et al*. RING finger nuclear factor RNF168 is important for defects in homologous recombination caused by loss of the breast cancer susceptibility factor BRCA1. *J Biol Chem* 2012;**287**: 40618-28.
- Muraki K, Han L, Miller D *et al*. Processing by MRE11 is involved in the sensitivity of subtelomeric regions to DNA double-strand breaks. *Nucleic Acids Res* 2015;**43**: 7911-30.
- Nagai S, Dubrana K, Tsai-Pflugfelder M *et al*. Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science* 2008;**322**: 597-602.
- Nambiar TS, Baudrier L, Billon P *et al*. CRISPR-based genome editing through the lens of DNA repair. *Mol Cell* 2022;**82**: 348-88.
- Nick McElhinny SA, Snowden CM, McCarville J *et al*. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* 2000;**20**: 2996-3003.

- Niehaus EM, Studt L, von Bargen KW *et al.* Sound of silence: the beauvericin cluster in *Fusarium fujikuroi* is controlled by cluster-specific and global regulators mediated by H3K27 modification. *Environ Microbiol* 2016;**18**: 4282-302.
- Ninomiya Y, Suzuki K, Ishii C *et al.* Highly efficient gene replacements in *Neurospora* strains deficient for nonhomologous end-joining. *Proc Natl Acad Sci U S A* 2004;**101**: 12248-53.
- Niu Q, Song Z, Tang K *et al.* A histone H3K4me1-specific binding protein is required for siRNA accumulation and DNA methylation at a subset of loci targeted by RNA-directed DNA methylation. *Nat Commun* 2021;**12**: 3367.
- Nodvig CS, Nielsen JB, Kogle ME *et al.* A CRISPR-Cas9 System for Genetic Engineering of Filamentous Fungi. *PLoS One* 2015;**10**: e0133085.
- Oliveira-Garcia E, Valent B. How eukaryotic filamentous pathogens evade plant recognition. *Curr Opin Microbiol* 2015;**26**: 92-101.
- Orbach MJ, Farrall L, Sweigard JA *et al.* A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell* 2000;**12**: 2019-32.
- Oses-Ruiz M, Cruz-Mireles N, Martin-Urdiroz M *et al.* Appressorium-mediated plant infection by *Magnaporthe oryzae* is regulated by a Pmk1-dependent hierarchical transcriptional network. *Nat Microbiol* 2021;**6**: 1383-97.
- Owens DDG, Caulder A, Frontera V *et al.* Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res* 2019;**47**: 7402-17.
- Pai CC, Deegan RS, Subramanian L *et al.* A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. *Nat Commun* 2014;**5**: 4091.
- Pannunzio NR, Watanabe G, Lieber MR. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *J Biol Chem* 2018;**293**: 10512-23.
- Park CH, Chen S, Shirsekar G *et al.* The *Magnaporthe oryzae* effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* 2012;**24**: 4748-62.
- Pasini D, Malatesta M, Jung HR *et al.* 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* 2010;**38**: 4958-69.
- Paulsen T, Malapati P, Shibata Y *et al.* MicroDNA levels are dependent on MMEJ, repressed by c-NHEJ pathway, and stimulated by DNA damage. *Nucleic Acids Res* 2021;**49**: 11787-99.
- Pellegrino S, Michelena J, Teloni F *et al.* Replication-Coupled Dilution of H4K20me2 Guides 53BP1 to Pre-replicative Chromatin. *Cell Rep* 2017;**19**: 1819-31.
- Peng Z, Oliveira-Garcia E, Lin G *et al.* Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. *PLoS Genet* 2019;**15**: e1008272.
- Pennisi E. Armed and dangerous. *Science* 2010;**327**: 804-5.
- Pfister SX, Ahrabi S, Zalmas LP *et al.* SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell Rep* 2014;**7**: 2006-18.
- Pfluger J, Wagner D. Histone modifications and dynamic regulation of genome accessibility in plants. *Curr Opin Plant Biol* 2007;**10**: 645-52.
- Pham KT, Inoue Y, Vu BV *et al.* MoSET1 (Histone H3K4 Methyltransferase in *Magnaporthe oryzae*) Regulates Global Gene Expression during Infection-Related Morphogenesis. *PLoS Genet* 2015;**11**: e1005385.
- Pickar-Oliver A, Gersbach CA. The next generation of CRISPR-Cas technologies and applications. *Nat Rev Mol Cell Biol* 2019;**20**: 490-507.
- Port F, Starostecka M, Boutros M. Multiplexed conditional genome editing with Cas12a in *Drosophila*. *Proc Natl Acad Sci U S A* 2020;**117**: 22890-9.

- Pradillo M, Knoll A, Oliver C *et al.* Involvement of the Cohesin Cofactor PDS5 (SPO76) During Meiosis and DNA Repair in *Arabidopsis thaliana*. *Front Plant Sci* 2015;**6**: 1034.
- Qu S, Liu G, Zhou B *et al.* The broad-spectrum blast resistance gene Pi9 encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multigene family in rice. *Genetics* 2006;**172**: 1901-14.
- Quiroz D, Zhao K, Carbonell-Bejerano P *et al.* Relationship between biased mutagenesis and H3K4me1-targeted DNA repair in plants. *bioRxiv* 2022, DOI 10.1101/2022.05.28.493846: 2022.05.28.493846.
- Raffaele S, Farrer RA, Cano LM *et al.* Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 2010;**330**: 1540-3.
- Ramirez F, Dundar F, Diehl S *et al.* deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;**42**: W187-91.
- Rando OJ, Ahmad K. Rules and regulation in the primary structure of chromatin. *Curr Opin Cell Biol* 2007;**19**: 250-6.
- Ray Chaudhuri A, Nussenzweig A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat Rev Mol Cell Biol* 2017;**18**: 610-21.
- Ray S, Singh PK, Gupta DK *et al.* Analysis of *Magnaporthe oryzae* Genome Reveals a Fungal Effector, Which Is Able to Induce Resistance Response in Transgenic Rice Line Containing Resistance Gene, Pi54. *Front Plant Sci* 2016;**7**: 1140.
- Reynoso MA, Pauluzzi GC, Kajala K *et al.* Nuclear Transcriptomes at High Resolution Using Retooled INTACT. *Plant Physiol* 2018;**176**: 270-81.
- Ricchetti M, Dujon B, Fairhead C. Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast. *J Mol Biol* 2003;**328**: 847-62.
- Ridenour JB, Moller M, Freitag M. Polycomb Repression without Bristles: Facultative Heterochromatin and Genome Stability in Fungi. *Genes (Basel)* 2020;**11**.
- Robinson JT, Thorvaldsdottir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**: 24-6.
- Rodgers K, McVey M. Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol* 2016;**231**: 15-24.
- Roerink SF, van Schendel R, Tijsterman M. Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res* 2014;**24**: 954-62.
- Roper M, Ellison C, Taylor JW *et al.* Nuclear and genome dynamics in multinucleate ascomycete fungi. *Curr Biol* 2011;**21**: R786-93.
- Rosler SM, Kramer K, Finkemeier I *et al.* The SAGA complex in the rice pathogen *Fusarium fujikuroi*: structure and functional characterization. *Mol Microbiol* 2016;**102**: 951-74.
- Roth DB, Wilson JH. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol Cell Biol* 1986;**6**: 4295-304.
- Rouxel T, Grandaubert J, Hane JK *et al.* Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat Commun* 2011;**2**: 202.
- Ryder LS, Dagdas YF, Kershaw MJ *et al.* A sensor kinase controls turgor-driven plant infection by the rice blast fungus. *Nature* 2019;**574**: 423-7.
- Ryu T, Spatola B, Delabaere L *et al.* Heterochromatic breaks move to the nuclear periphery to continue recombinational repair. *Nat Cell Biol* 2015;**17**: 1401-11.
- Saitoh H, Fujisawa S, Mitsuoka C *et al.* Large-scale gene disruption in *Magnaporthe oryzae* identifies MC69, a secreted protein required for infection by monocot and dicot fungal pathogens. *PLoS Pathog* 2012;**8**: e1002711.
- Sakulkoo W, Oses-Ruiz M, Oliveira Garcia E *et al.* A single fungal MAP kinase controls plant cell-to-cell invasion by the rice blast fungus. *Science* 2018;**359**: 1399-403.

- Sallmyr A, Tomkinson AE. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *J Biol Chem* 2018;**293**: 10536-46.
- Sanchez-Vallet A, Fouche S, Fudal I *et al.* The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. *Annu Rev Phytopathol* 2018;**56**: 21-40.
- Saredi G, Huang H, Hammond CM *et al.* H4K20me0 marks post-replicative chromatin and recruits the TONSL-MMS22L DNA repair complex. *Nature* 2016;**534**: 714-8.
- Sasaki T, Lynch KL, Mueller CV *et al.* Heterochromatin controls gammaH2A localization in *Neurospora crassa*. *Eukaryot Cell* 2014;**13**: 990-1000.
- Saur IML, Panstruga R, Schulze-Lefert P. NOD-like receptor-mediated plant immunity: from structure to cell death. *Nat Rev Immunol* 2021;**21**: 305-18.
- Schep R, Brinkman EK, Leemans C *et al.* Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell* 2021;**81**: 2216-30 e10.
- Schiestl RH, Zhu J, Petes TD. Effect of mutations in genes affecting homologous recombination on restriction enzyme-mediated and illegitimate recombination in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1994;**14**: 4493-500.
- Schimmel J, Kool H, van Schendel R *et al.* Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *EMBO J* 2017;**36**: 3634-49.
- Schimmel J, van Schendel R, den Dunnen JT *et al.* Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining. *Trends Genet* 2019;**35**: 632-44.
- Schotanus K, Heitman J. Centromere deletion in *Cryptococcus deuterogattii* leads to neocentromere formation and chromosome fusions. *Elife* 2020;**9**.
- Schotanus K, Soyer JL, Connolly LR *et al.* Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. *Epigenetics Chromatin* 2015;**8**: 41.
- Schrempf A, Slyskova J, Loizou JI. Targeting the DNA Repair Enzyme Polymerase theta in Cancer Therapy. *Trends Cancer* 2021;**7**: 98-111.
- Schuettengruber B, Bourbon HM, Di Croce L *et al.* Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 2017;**171**: 34-57.
- Schuster M, Kahmann R. CRISPR-Cas9 genome editing approaches in filamentous fungi and oomycetes. *Fungal Genet Biol* 2019;**130**: 43-53.
- Schwammle V, Aspalter CM, Sidoli S *et al.* Large scale analysis of co-existing post-translational modifications in histone tails reveals global fine structure of cross-talk. *Mol Cell Proteomics* 2014;**13**: 1855-65.
- Scully R, Panday A, Elango R *et al.* DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* 2019;**20**: 698-714.
- Sebastian R, Oberdoerffer P. Transcription-associated events affecting genomic integrity. *Philos Trans R Soc Lond B Biol Sci* 2017;**372**.
- Seidl MF, Cook DE, Thomma BP. Chromatin Biology Impacts Adaptive Evolution of Filamentous Plant Pathogens. *PLoS Pathog* 2016;**12**: e1005920.
- Seidl MF, Thomma B. Transposable Elements Direct The Coevolution between Plants and Microbes. *Trends Genet* 2017;**33**: 842-51.
- Selin C, de Kievit TR, Belmonte MF *et al.* Elucidating the Role of Effectors in Plant-Fungal Interactions: Progress and Challenges. *Front Microbiol* 2016;**7**: 600.
- Semighini CP, Savoldi M, Goldman GH *et al.* Functional characterization of the putative *Aspergillus nidulans* poly(ADP-ribose) polymerase homolog PrpA. *Genetics* 2006;**173**: 87-98.
- Sfeir A, Symington LS. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci* 2015;**40**: 701-14.

- Shan Q, Wang Y, Li J *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 2013;**31**: 686-8.
- Shao Z, Davis AJ, Fattah KR *et al.* Persistently bound Ku at DNA ends attenuates DNA end resection and homologous recombination. *DNA Repair (Amst)* 2012;**11**: 310-6.
- Shima N, Hartford SA, Duffy T *et al.* Phenotype-based identification of mouse chromosome instability mutants. *Genetics* 2003;**163**: 1031-40.
- Shimizu M, Nakano Y, Hirabuchi A *et al.* RNA-Seq of in planta-expressed *Magnaporthe oryzae* genes identifies MoSVP as a highly expressed gene required for pathogenicity at the initial stage of infection. *Mol Plant Pathol* 2019;**20**: 1682-95.
- Sijbers AM, de Laat WL, Ariza RR *et al.* *Xeroderma pigmentosum* group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* 1996;**86**: 811-22.
- Skryabin BV, Kummerfeld DM, Gubar L *et al.* Pervasive head-to-tail insertions of DNA templates mask desired CRISPR-Cas9-mediated genome editing events. *Sci Adv* 2020;**6**: eaax2941.
- Snelders NC, Rovenich H, Thomma B. Microbiota manipulation through the secretion of effector proteins is fundamental to the wealth of lifestyles in the fungal kingdom. *FEMS Microbiol Rev* 2022, DOI 10.1093/femsre/fuac022.
- Sone T, Takeuchi S, Miki S *et al.* Homologous recombination causes the spontaneous deletion of AVR-Pia in *Magnaporthe oryzae*. *FEMS Microbiol Lett* 2013;**339**: 102-9.
- Soyer JL, Balesdent MH, Rouxel T *et al.* To B or not to B: a tale of unorthodox chromosomes. *Curr Opin Microbiol* 2018;**46**: 50-7.
- Sperschneider J, Dodds PN, Gardiner DM *et al.* Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol* 2018;**19**: 2094-110.
- Srinivas US, Tan BWQ, Vellayappan BA *et al.* ROS and the DNA damage response in cancer. *Redox Biol* 2019;**25**: 101084.
- Stahl FW. The Holliday junction on its thirtieth anniversary. *Genetics* 1994;**138**: 241-6.
- Stam R, Munsterkotter M, Pophaly SD *et al.* A New Reference Genome Shows the One-Speed Genome Structure of the Barley Pathogen *Ramularia collo-cygni*. *Genome Biol Evol* 2018;**10**: 3243-9.
- Stavnezer J, Guikema JE, Schrader CE. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 2008;**26**: 261-92.
- Steenwyk JL, Opulente DA, Kominek J *et al.* Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PLoS Biol* 2019;**17**: e3000255.
- Stefanovie B, Hengel SR, Mlcouskova J *et al.* DSS1 interacts with and stimulates RAD52 to promote the repair of DSBs. *Nucleic Acids Res* 2020;**48**: 694-708.
- Steinberg G, Perez-Martin J. *Ustilago maydis*, a new fungal model system for cell biology. *Trends Cell Biol* 2008;**18**: 61-7.
- Steiner FA, Talbert PB, Kasinathan S *et al.* Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res* 2012;**22**: 766-77.
- Strecker J, Ladha A, Gardner Z *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 2019;**365**: 48-53.
- Studt L, Rosler SM, Burkhardt I *et al.* Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*. *Environ Microbiol* 2016;**18**: 4037-54.
- Sturzenegger A, Burdova K, Kanagaraj R *et al.* DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *J Biol Chem* 2014;**289**: 27314-26.
- Suganuma T, Workman JL. Crosstalk among Histone Modifications. *Cell* 2008;**135**: 604-7.
- Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;**27**: 1009-10.

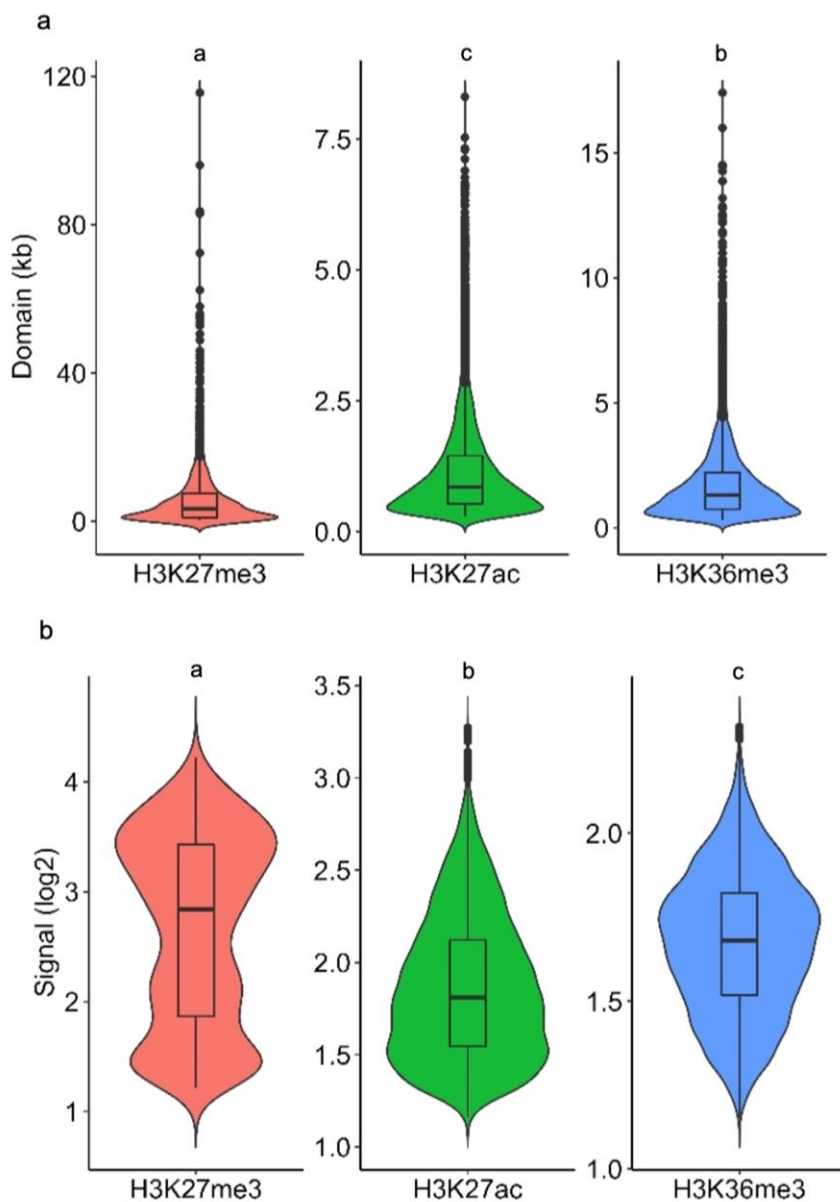
- Sun Z, Zhang Y, Jia J *et al.* H3K36me3, message from chromatin to DNA damage repair. *Cell Biosci* 2020;**10**: 9.
- Sundaram V, Wysocka J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos Trans R Soc Lond B Biol Sci* 2020;**375**: 20190347.
- Suzuki K, Tsunekawa Y, Hernandez-Benitez R *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* 2016;**540**: 144-9.
- Symington LS. End resection at double-strand breaks: mechanism and regulation. *Cold Spring Harb Perspect Biol* 2014;**6**.
- Tanaka S, Schweizer G, Rossel N *et al.* Neofunctionalization of the secreted Tin2 effector in the fungal pathogen *Ustilago maydis*. *Nat Microbiol* 2019;**4**: 251-7.
- Tang X, Lowder LG, Zhang T *et al.* A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat Plants* 2017;**3**: 17018.
- Tao Z, Gao P, Liu HW. Studies of the expression of human poly(ADP-ribose) polymerase-1 in *Saccharomyces cerevisiae* and identification of PARP-1 substrates by yeast proteome microarray screening. *Biochemistry* 2009;**48**: 11745-54.
- Thyme SB, Schier AF. Polq-Mediated End Joining Is Essential for Surviving DNA Double-Strand Breaks during Early Zebrafish Development. *Cell Rep* 2016;**15**: 707-14.
- Tie F, Banerjee R, Stratton CA *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development* 2009;**136**: 3131-41.
- Todd RT, Wikoff TD, Forche A *et al.* Genome plasticity in *Candida albicans* is driven by long repeat sequences. *Elife* 2019;**8**.
- Torres DE, Oggenfuss U, Croll D *et al.* Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews* 2020.
- Truong LN, Li Y, Shi LZ *et al.* Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc Natl Acad Sci U S A* 2013;**110**: 7720-5.
- Tsouroula K, Furst A, Rogier M *et al.* Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol Cell* 2016;**63**: 293-305.
- Tyler BM, Tripathy S, Zhang X *et al.* *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 2006;**313**: 1261-6.
- Valent B. The Impact of Blast Disease: Past, Present, and Future. *Methods Mol Biol* 2021;**2356**: 1-18.
- Valent B, Farrall L, Chumley FG. *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics* 1991;**127**: 87-101.
- van Kregten M, de Pater S, Romeijn R *et al.* T-DNA integration in plants results from polymerase-theta-mediated DNA repair. *Nat Plants* 2016;**2**: 16164.
- van Schendel R, Roerink SF, Portegijs V *et al.* Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. *Nat Commun* 2015;**6**: 7394.
- Villalba F, Collemare J, Landraud P *et al.* Improved gene targeting in *Magnaporthe grisea* by inactivation of MgKU80 required for non-homologous end joining. *Fungal Genet Biol* 2008;**45**: 68-75.
- Villarreal DD, Lee K, Deem A *et al.* Microhomology directs diverse DNA break repair pathways and chromosomal translocations. *PLoS Genet* 2012;**8**: e1003026.
- Vitor AC, Huertas P, Legube G *et al.* Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox. *Front Mol Biosci* 2020;**7**: 24.
- Wacker T, Helmstetter N, Wilson D *et al.* Two-speed genome expansion drives the evolution of pathogenicity in animal fungal pathogens. *bioRxiv* 2021, DOI 10.1101/2021.11.03.467166: 2021.11.03.467166.
- Walker JR, Corpina RA, Goldberg J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 2001;**412**: 607-14.

- Wang C, Milgate AW, Solomon PS *et al.* The identification of a transposon affecting the asexual reproduction of the wheat pathogen *Zymoseptoria tritici*. *Mol Plant Pathol* 2021a;**22**: 800-16.
- Wang C, Zhang S, Hou R *et al.* Functional analysis of the kinome of the wheat scab fungus *Fusarium graminearum*. *PLoS Pathog* 2011;**7**: e1002460.
- Wang Q, Cobine PA, Coleman JJ. Efficient genome editing in *Fusarium oxysporum* based on CRISPR/Cas9 ribonucleoprotein complexes. *Fungal Genet Biol* 2018;**117**: 21-9.
- Wang Q, Coleman JJ. Progress and Challenges: Development and Implementation of CRISPR/Cas9 Technology in Filamentous Fungi. *Comput Struct Biotechnol J* 2019;**17**: 761-9.
- Wang Y, Wang M, Djekidel MN *et al.* eccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature* 2021b;**599**: 308-14.
- Wang Y, Wu J, Kim SG *et al.* *Magnaporthe oryzae*-Secreted Protein MSP1 Induces Cell Death and Elicits Defense Responses in Rice. *Mol Plant Microbe Interact* 2016;**29**: 299-312.
- Waters CA, Strande NT, Wyatt DW *et al.* Nonhomologous end joining: a good solution for bad ends. *DNA Repair (Amst)* 2014;**17**: 39-51.
- Wilson RA, Talbot NJ. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nat Rev Microbiol* 2009;**7**: 185-95.
- Woloshuk CP, Shim WB. Aflatoxins, fumonisins, and trichothecenes: a convergence of knowledge. *FEMS Microbiol Rev* 2013;**37**: 94-109.
- Wolter F, Schindele P, Beying N *et al.* Different DNA repair pathways are involved in single-strand break-induced genomic changes in plants. *Plant Cell* 2021, DOI 10.1093/plcell/koab204.
- Wood RD, Doublet S. DNA polymerase theta (POLQ), double-strand break repair, and cancer. *DNA Repair (Amst)* 2016;**44**: 22-32.
- Wray J, Williamson EA, Singh SB *et al.* PARP1 is required for chromosomal translocations. *Blood* 2013;**121**: 4359-65.
- Wright WD, Shah SS, Heyer WD. Homologous recombination and the repair of DNA double-strand breaks. *J Biol Chem* 2018;**293**: 10524-35.
- Wu J, Kou Y, Bao J *et al.* Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. *New Phytol* 2015;**206**: 1463-75.
- Xiong X, Du Z, Wang Y *et al.* 53BP1 promotes microhomology-mediated end-joining in G1-phase cells. *Nucleic Acids Res* 2015;**43**: 1659-70.
- Xu JR, Hamer JE. MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus *Magnaporthe grisea*. *Genes Dev* 1996;**10**: 2696-706.
- Xue C, Greene EC. DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. *Trends Genet* 2021;**37**: 639-56.
- Xue M, Yang J, Li Z *et al.* Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet* 2012;**8**: e1002869.
- Yadav V, Sun S, Coelho MA *et al.* Centromere scission drives chromosome shuffling and reproductive isolation. *Proc Natl Acad Sci U S A* 2020;**117**: 7917-28.
- Yamada T, Makimura K, Hisajima T *et al.* Enhanced gene replacements in Ku80 disruption mutants of the dermatophyte, *Trichophyton mentagrophytes*. *FEMS Microbiol Lett* 2009;**298**: 208-17.
- Yang H, Howard M, Dean C. Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC. *Curr Biol* 2014;**24**: 1793-7.
- Yi M, Valent B. Communication between filamentous pathogens and plants at the biotrophic interface. *Annu Rev Phytopathol* 2013;**51**: 587-611.
- Yildirim G, Sperschneider J, Malar CM *et al.* Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytol* 2022;**233**: 1097-107.

- Yoshida K, Saitoh H, Fujisawa S *et al.* Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* 2009;**21**: 1573-91.
- Yu F, Gu Q, Yun Y *et al.* The TOR signaling pathway regulates vegetative development and virulence in *Fusarium graminearum*. *New Phytol* 2014;**203**: 219-32.
- Yuan W, Xu M, Huang C *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem* 2011;**286**: 7983-9.
- Zelensky AN, Schimmel J, Kool H *et al.* Inactivation of Pol theta and C-NHEJ eliminates off-target integration of exogenous DNA. *Nat Commun* 2017;**8**: 66.
- Zentner GE, Henikoff S. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* 2013;**20**: 259-66.
- Zetsche B, Gootenberg JS, Abudayyeh OO *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015;**163**: 759-71.
- Zetsche B, Heidenreich M, Mohanraju P *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* 2017;**35**: 31-4.
- Zhang C, Meng X, Wei X *et al.* Highly efficient CRISPR mutagenesis by microhomology-mediated end joining in *Aspergillus fumigatus*. *Fungal Genet Biol* 2016a;**86**: 47-57.
- Zhang H, Zheng X, Zhang Z. The *Magnaporthe grisea* species complex and plant pathogenesis. *Mol Plant Pathol* 2016b;**17**: 796-804.
- Zhang S, Liang M, Naqvi NI *et al.* Phototrophy and starvation-based induction of autophagy upon removal of Gcn5-catalyzed acetylation of Atg7 in *Magnaporthe oryzae*. *Autophagy* 2017a;**13**: 1318-30.
- Zhang S, Wang L, Wu W *et al.* Function and evolution of *Magnaporthe oryzae* avirulence gene AvrPib responding to the rice blast resistance gene Pib. *Sci Rep* 2015a;**5**: 11642.
- Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications - writers that read. *EMBO Rep* 2015b;**16**: 1467-81.
- Zhang W, Corwin JA, Copeland D *et al.* Plastic Transcriptomes Stabilize Immunity to Pathogen Diversity: The Jasmonic Acid and Salicylic Acid Networks within the *Arabidopsis/Botrytis* Pathosystem. *Plant Cell* 2017b;**29**: 2727-52.
- Zhang W, Corwin JA, Copeland DH *et al.* Plant-necrotroph co-transcriptome networks illuminate a metabolic battlefield. *Elife* 2019;**8**.
- Zhang W, Huang J, Cook DE. Histone modification dynamics at H3K27 are associated with altered transcription of in planta induced genes in *Magnaporthe oryzae*. *PLoS Genet* 2021;**17**: e1009376.
- Zhang WW, Matlashewski G. Single-Strand Annealing Plays a Major Role in Double-Strand DNA Break Repair following CRISPR-Cas9 Cleavage in *Leishmania*. *mSphere* 2019;**4**.
- Zhang X, Clarenz O, Cokus S *et al.* Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol* 2007a;**5**: e129.
- Zhang Y, Hefferin ML, Chen L *et al.* Role of Dnl4-Lif1 in nonhomologous end-joining repair complex assembly and suppression of homologous recombination. *Nat Struct Mol Biol* 2007b;**14**: 639-46.
- Zhang Y, Jasin M. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat Struct Mol Biol* 2011;**18**: 80-4.
- Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;**9**: R137.
- Zhao S, Guo Y, Sheng Q *et al.* Advanced heat map and clustering analysis using heatmap3. *Biomed Res Int* 2014;**2014**: 986048.
- Zhao X, Wei C, Li J *et al.* Cell cycle-dependent control of homologous recombination. *Acta Biochim Biophys Sin (Shanghai)* 2017;**49**: 655-68.
- Zheng YM, Lin FL, Gao H *et al.* Development of a versatile and conventional technique for gene disruption in filamentous fungi based on CRISPR-Cas9 technology. *Sci Rep* 2017;**7**: 9250.

- Zhong Z, Chen M, Lin L *et al.* Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *ISME J* 2018;**12**: 1867-78.
- Zhong Z, Norvinyeku J, Chen M *et al.* Directional Selection from Host Plants Is a Major Force Driving Host Specificity in *Magnaporthe* Species. *Sci Rep* 2016;**6**: 25591.
- Zhou B, Qu S, Liu G *et al.* The eight amino-acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specificity to *Magnaporthe grisea*. *Mol Plant Microbe Interact* 2006;**19**: 1216-28.
- Zschenker O, Kulkarni A, Miller D *et al.* Increased sensitivity of subtelomeric regions to DNA double-strand breaks in a human cancer cell line. *DNA Repair (Amst)* 2009;**8**: 886-900.
- Zuccaro MV, Xu J, Mitchell C *et al.* Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell* 2020;**183**: 1650-64 e15.

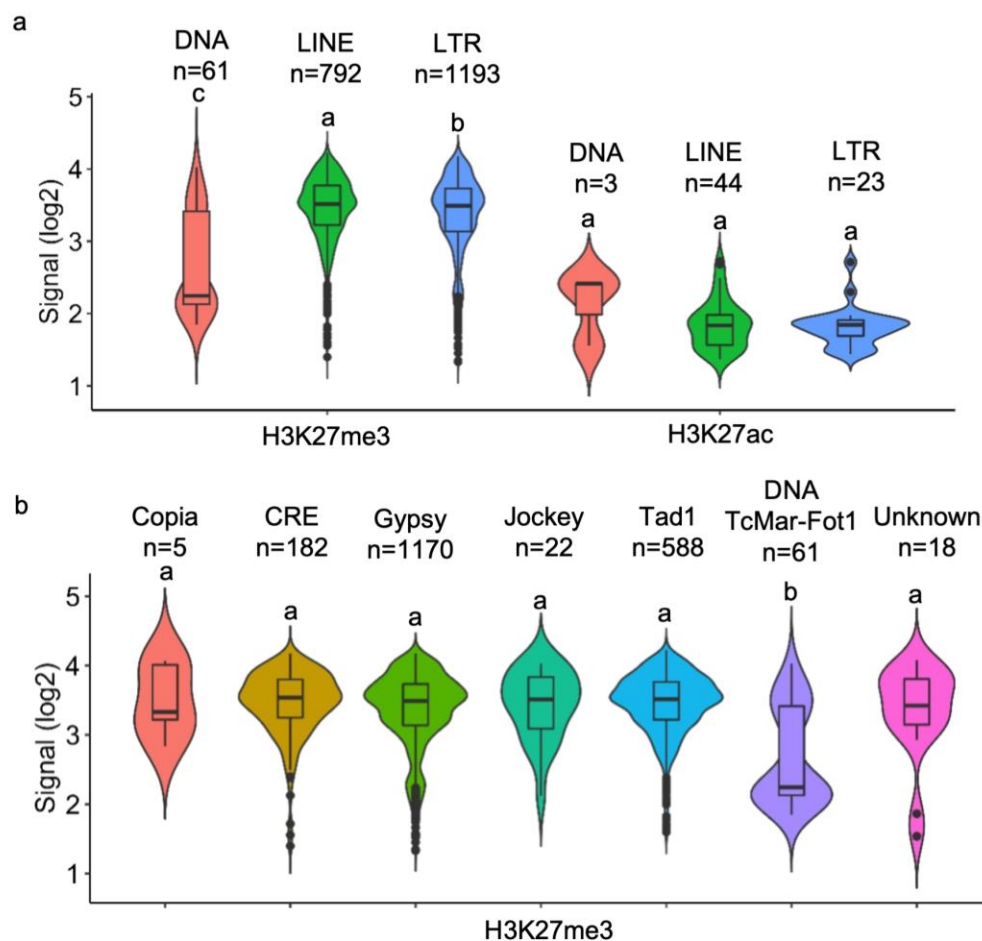
Appendix A - Supplementary Figures and Tables of Chapter 2



Supplementary Figure A.1

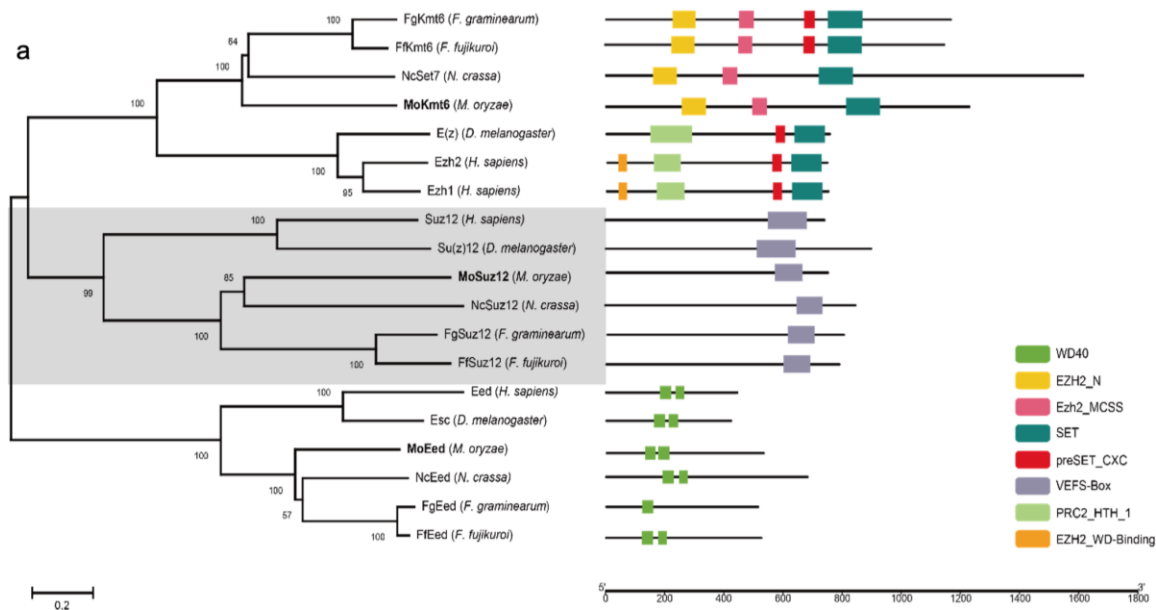
Distribution of H3K27me3, H3K27ac, and H3K36me3 in *M. oryzae* Guy11. (A) and (B)

Violin plots illustrating genome-wide distribution of domains (A) and ChIP signals (B) of H3K27me3, H3K27ac, and H3K36me3 in Guy11 wide type growing under *in vitro* complete medium. Letters above the violin plots indicates the significance.



Supplementary Figure A.2

Distribution of histone modifications at H3K27 across diverse types of transposable elements in *M. oryzae*. (A) and (B) ChIP signals of H3K27me3 and H3K27ac from MACS2 peak calling across transposable elements (TE) families (A) and subfamilies (B). ChIP-Seq data were collected from *M. oryzae* Guy11 wild type growing under *in vitro* complete medium. The number of TEs for each group is shown above each violin plot. Letters above the violin plots indicate the significant difference among groups based on ANOVA and Tukey's HSD test.

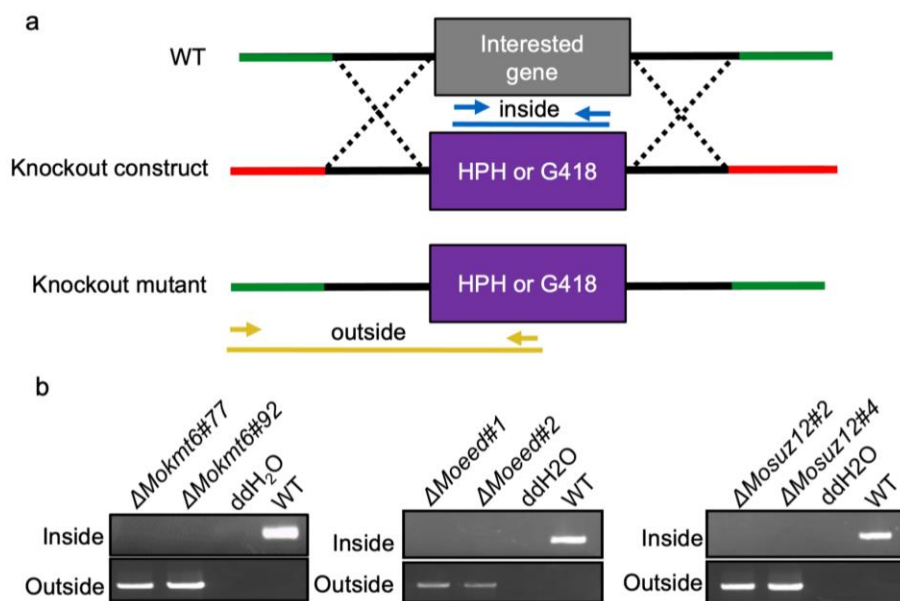


Supplementary Figure A.3

Phylogenetic and homology matrix analysis of PRC2 core components in different

organisms. (A) Neighbor-join tree of selected PRC2 core components generated by MEGA X with 1000 bootstrap replications. The protein domains are predicated by SMART and visualized by TBtools. The sequences used for analysis included *MoKmt6* (MGG_00152), *MoSuz12* (MGG_03169) and *MoEed* (MGG_06028) from *M. oryzae*; *FgKmt6* (FGSG_15795), *FgSuz12* (FGSG_04321), and *FgEed* (FGSG_15909) from *F. graminearum*; *FfKmt6* (FFUJ_00719), *FfSuz12* (FFUJ_09784) and *FfEed* (FFUJ_12272) from *F. fujikuroi*; *NcSet7* (NCU07496), *NcSuz12* (NCU05460), and *NcEed* (NCU05300) from *N. crassa*; *Ezh2* (NP_001190176.1), *Ezh1* (NP_001308008.1), *Suz12* (NP_056170.2) and *Eed* (AAC23685) from *H. sapiens*; *E(z)* (NP_001137932.1), *Su(z)12* (NP_730465.1), and *Esc* (NP_477431.1) from *D. melanogaster*. (B)

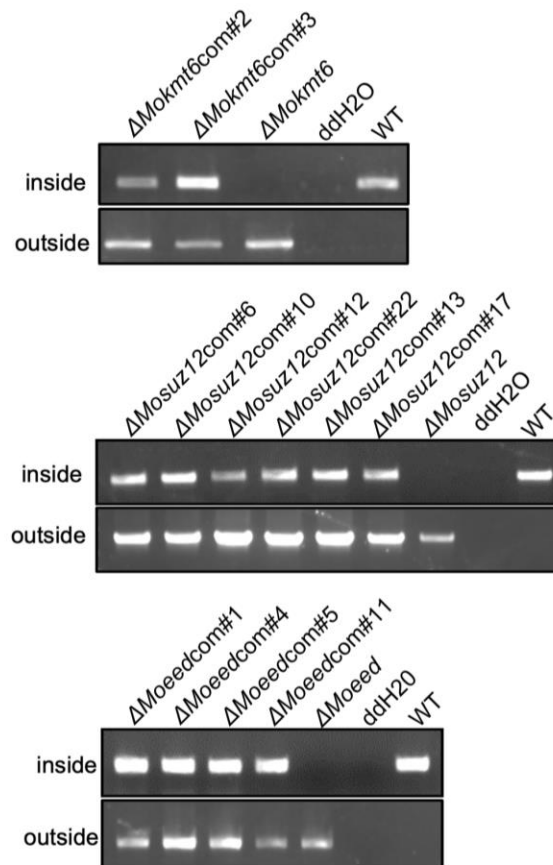
Homology matrix analysis of *M. oryzae* PRC2 core components with PRC2 homologs in different organisms as described above. Numbers indicate protein coverage and identity.



Supplementary Figure A.4

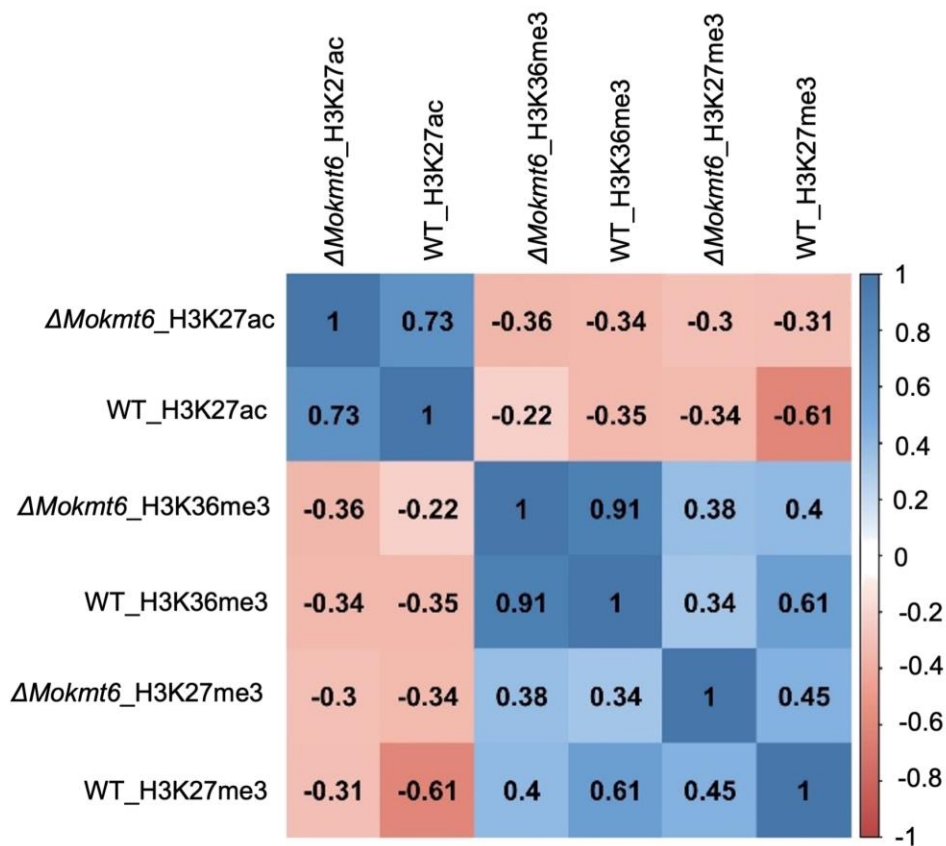
Homologous recombination strategy was used for generating deletion mutants. (A)

Homologous recombination was used for gene knockout. ± 1 kb gene flanking region (black lines) was amplified for targeting gene of interest. The gene coding region was replaced with hygromycin resistant cassette (HPH) for single knockout or geneticin resistant cassette (G418) for double knockout. Inside primer pair (blue arrow) was used for testing the presence/absence of gene of interest in transformants. Outside primer pair (yellow arrow) was used for testing whether there is correct resistant cassette integration in transformants. Green lines indicate upstream and downstream sequences of interested gene and red lines indicate upstream and downstream sequences of the resistant cassette. (B) Δ Mokmt6, Δ Moeed, and Δ Mosuz12 were confirmed by PCR amplifications with inside and outside primers.



Supplementary Figure A.5

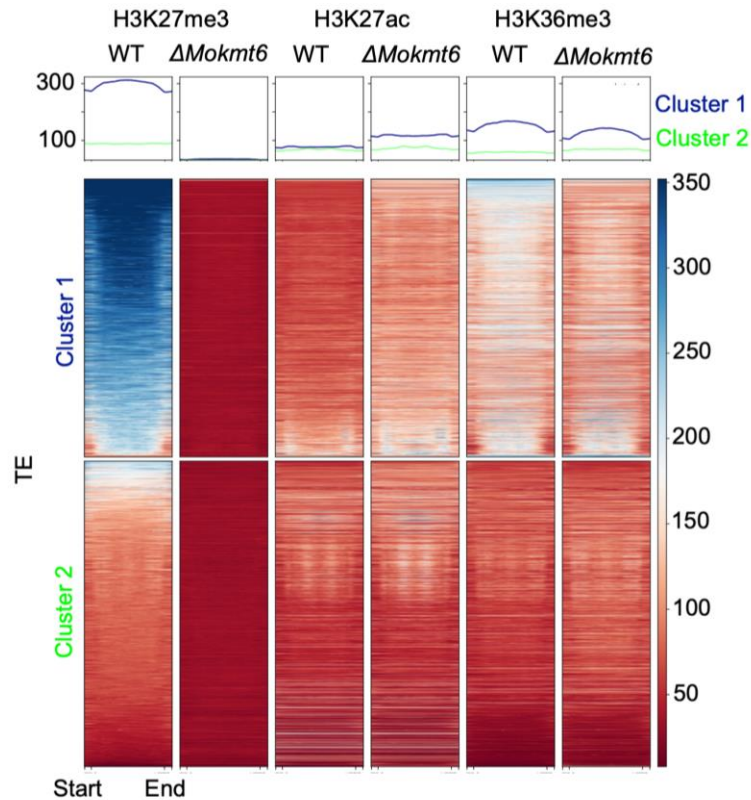
The complemented strains were screened by PCR. The complementation for *ΔMokmt6*, *ΔMosuz12* and *Δmoeed* was screened by inside and outside primers. Amplifications with inside primers suggested the success in re-introducing the original gene back to the deletion mutants.



Supplementary Figure A.6

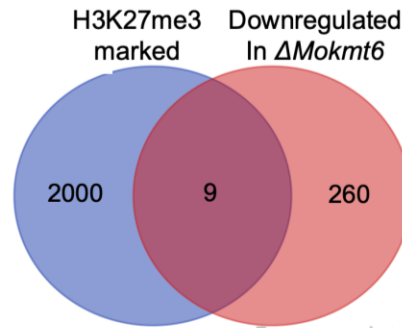
Loss of H3K27me3 leads to dynamics of H3K27ac and H3K36me3 at whole genome level.

Heatmap illustrating pairwise Pearson correlation coefficient of ChIP signals of H3K27me3, H3K27ac, H3K36me3 between Guy11 wild type (WT) and mutant $\Delta Mokmt6$ lacking H3K27me3.



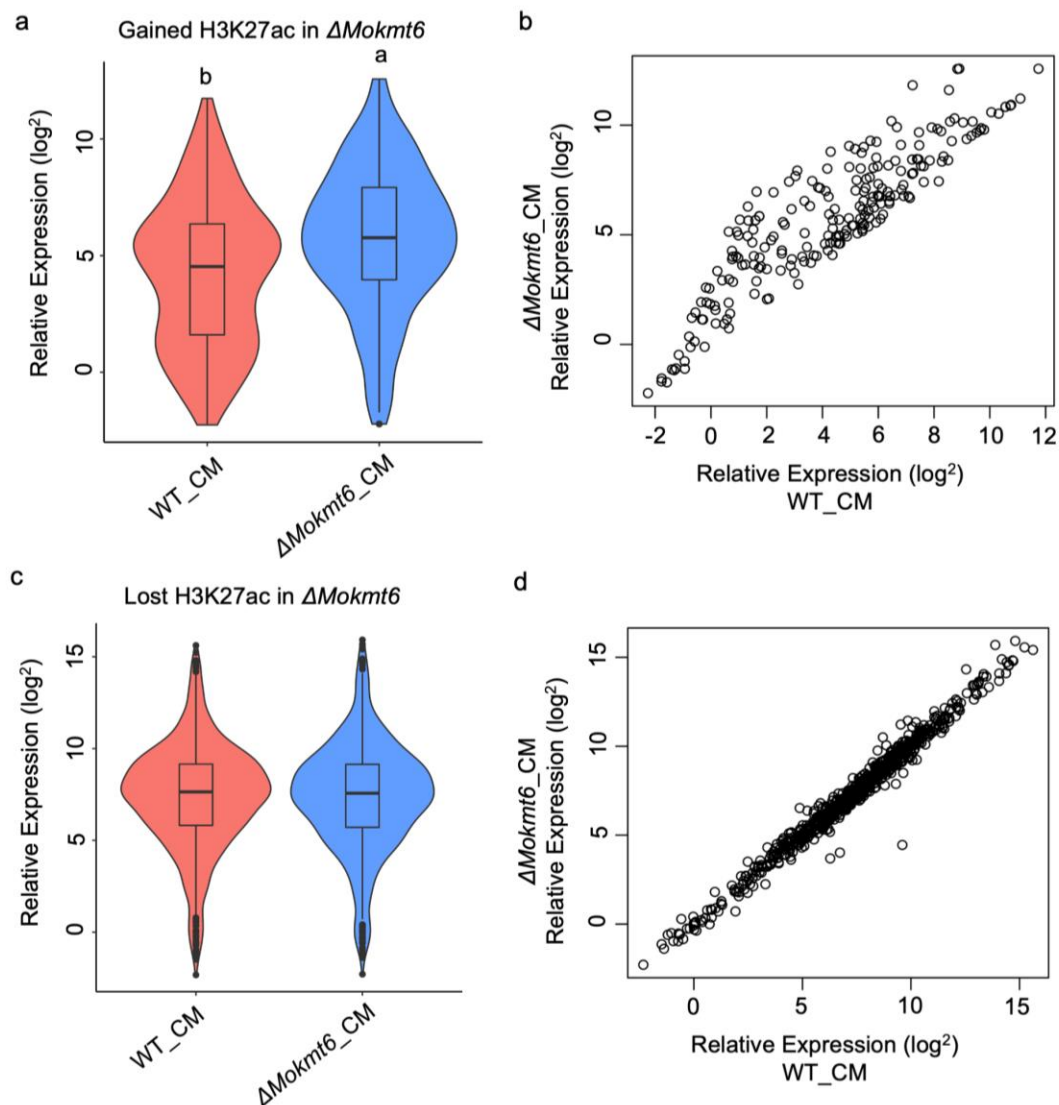
Supplementary Figure A.7

Loss of H3K27me3 leads to re-distribution of H3K27ac and H3K36me3. Heat maps visualizing profiles of ChIP signals for H3K27me3, H3K27ac and H3K36me3 across transposable elements (TEs) in Guy11 wild type (WT) and mutant $\Delta Mokmt6$. Two clusters of TEs are generated by unsupervised k-means analysis and ranked according to their H3K27me3 enrichment from wild type grown *in vitro* complete medium.



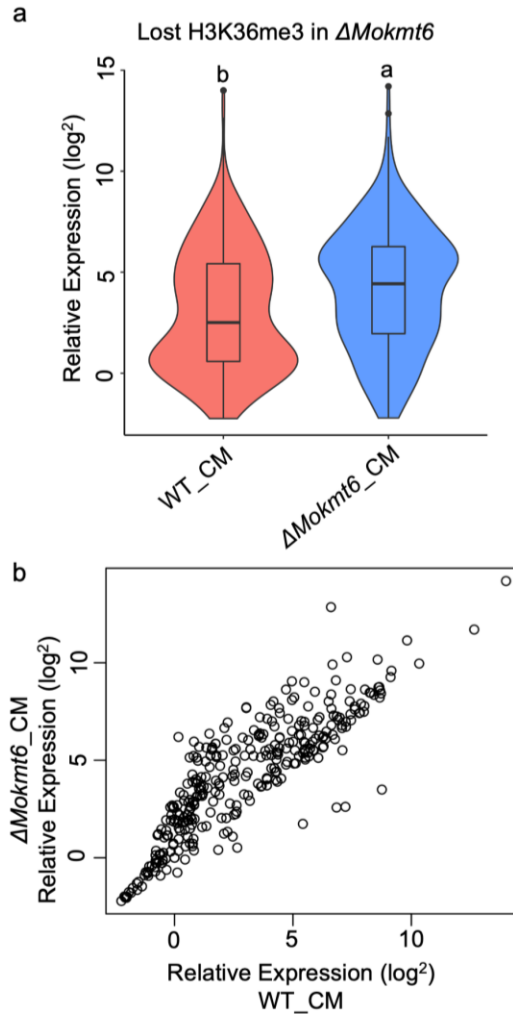
Supplementary Figure A.8

The majority of downregulated genes in $\Delta Mokmt6$ are not marked by H3K27me3. Venn diagram showing the overlap of genes downregulated in $\Delta Mokmt6$ and marked by H3K27me3 in Guy11 wild type.



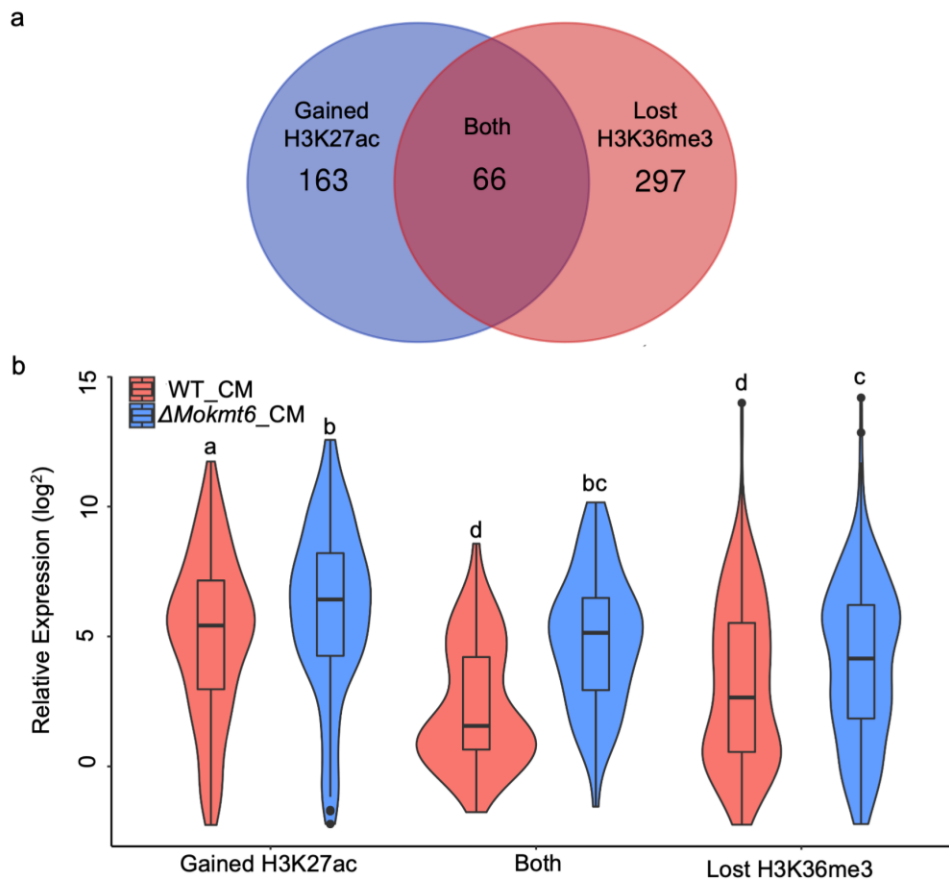
Supplementary Figure A.9

Expression profiles in wild type and $\Delta Mokmt6$ for genes that gained and lost H3K27ac in $\Delta Mokmt6$. (A) Violin plot and (B) scatter plot showing gene expression profiles during *in vitro* complete medium (CM) growth in wild type and $\Delta Mokmt6$ for 218 expressed genes that gained H3K27ac in $\Delta Mokmt6$. (C) Violin plot and (D) scatter plot showing gene expression profiles during *in vitro* complete medium (CM) growth in wild type and $\Delta Mokmt6$ for 794 expressed genes that lost H3K27ac in $\Delta Mokmt6$. Letters above the violin plots indicate the significant difference among groups based on ANOVA and Tukey's HSD test.



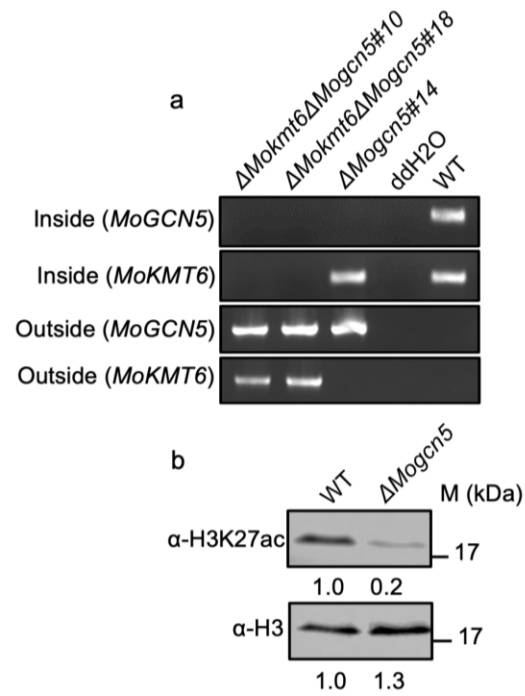
Supplementary Figure A.10

Expression profiles in wild type and $\Delta Mokmt6$ for genes that lost H3K36me3 in $\Delta Mokmt6$ mutant. (A) Violin plots and (B) scatter plot showing the expression profiles during *in vitro* complete medium (CM) growth in wild type and $\Delta Mokmt6$ for 321 expressed genes that lost H3K36me3 in $\Delta Mokmt6$. Letters above the violin plots indicate the significant difference among groups based on ANOVA and Tukey's HSD test.



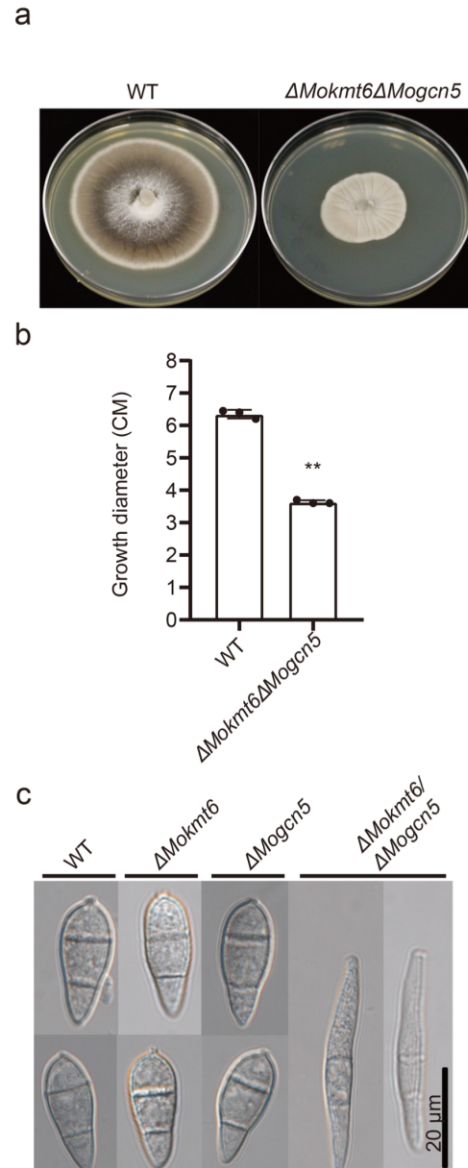
Supplementary Figure A.11

Comparison of gene expression profiles under a combination of genotypes and growth conditions. (A) Venn diagram showing the number and overlap of genes that only gained H3K27ac, only lost H3K36me3 or had both changes occur in the Δ Mokmt6 mutant compared to wild type under *in vitro* complete medium (CM) growth. (B) Violin plots for the RNA-seq expression of wild type (WT_CM) and Δ Mokmt6 (Δ Mokmt6_CM) during *in vitro* complete medium (CM) growth for the gene sets from (A). Letters above the violin plots indicate the significant difference among groups based on ANOVA and Tukey's HSD test.



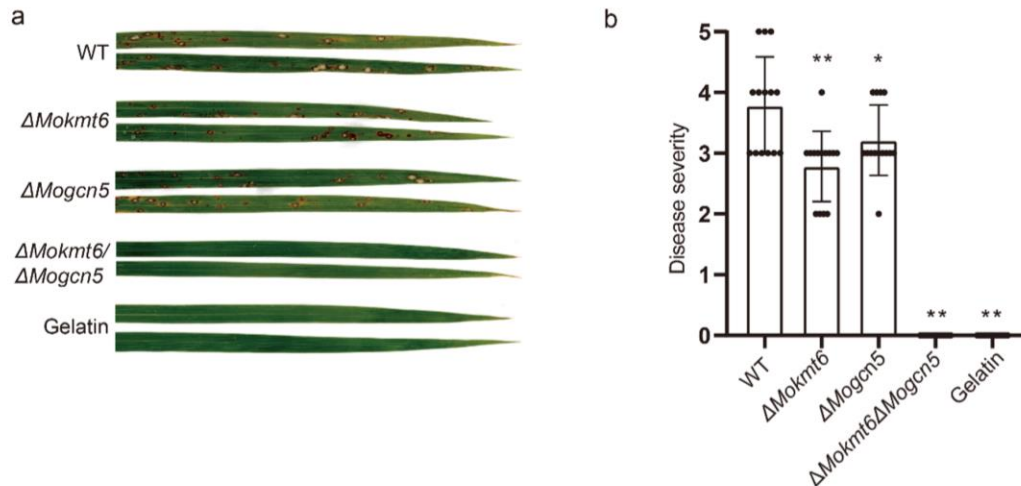
Supplementary Figure A.12

MoGcn5 is involved in mediating majority of H3K27ac in *M. oryzae*. (A) Δ *Mogcn5* and Δ *Mokmt6* Δ *Mogcn5* were confirmed by PCR amplification with inside and outside primers. (B) The anti-H3K27ac was used for detecting the global level of H3K27ac in the samples and anti-H3 was used as loading control. Signal intensities were measured by ImageJ.



Supplementary Figure A.13

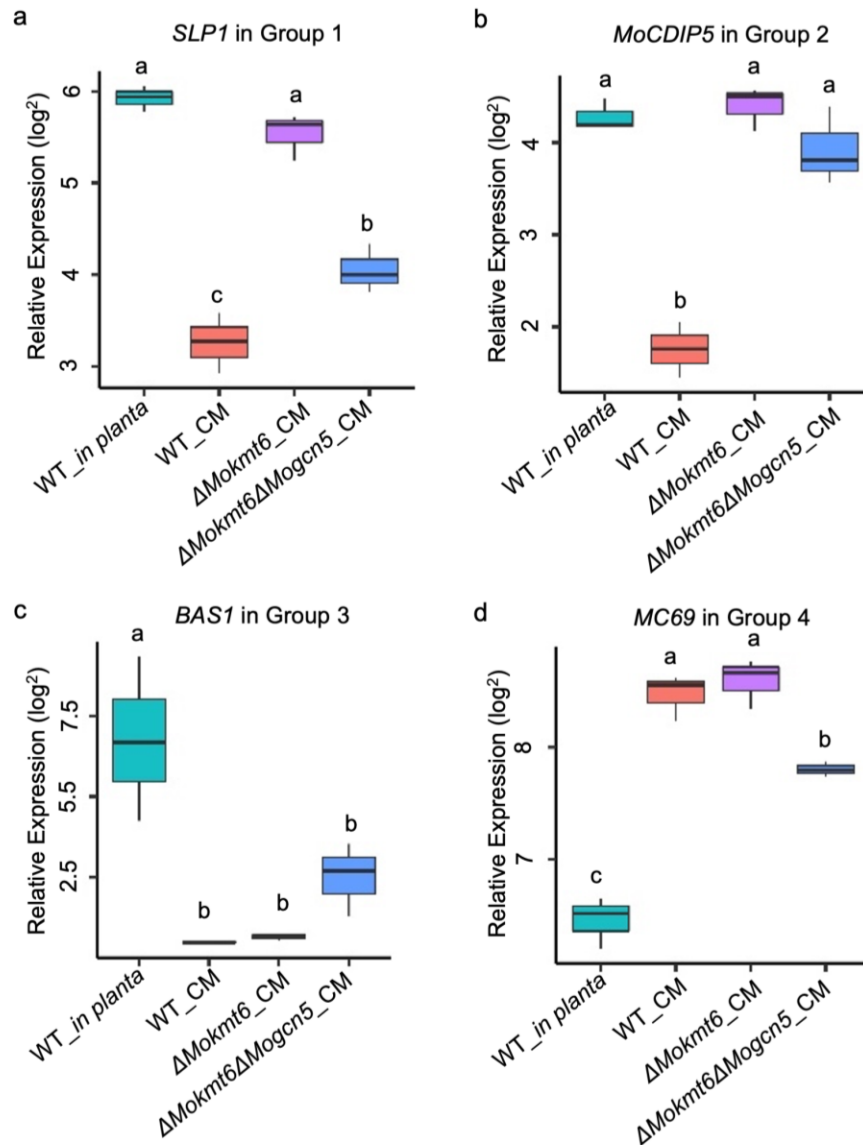
The double mutant Δ Mokmt6 Δ Mogcn5 has abnormal growth and conidia morphology. (A) Colony morphology of wild type and Δ Mokmt6 Δ Mogcn5 on complete medium agar (CM) at 12 days. (B) Colony diameters measured at 12 days. The growth rates were determined to be significantly different (**, $p < 0.01$) between wild type and Δ Mokmt6 Δ Mogcn5 compared using student's t-test. (C) Representative conidial morphology of wild type (WT), Δ Mokmt6, Δ Mogcn5, and Δ Mokmt6 Δ Mogcn5 collected after growth on rice polish agar. Bar = 20 μ m.



Supplementary Figure A.14

Single mutants $\Delta Mokmt6$, $\Delta Mogcn5$ caused reduced symptoms on rice, while

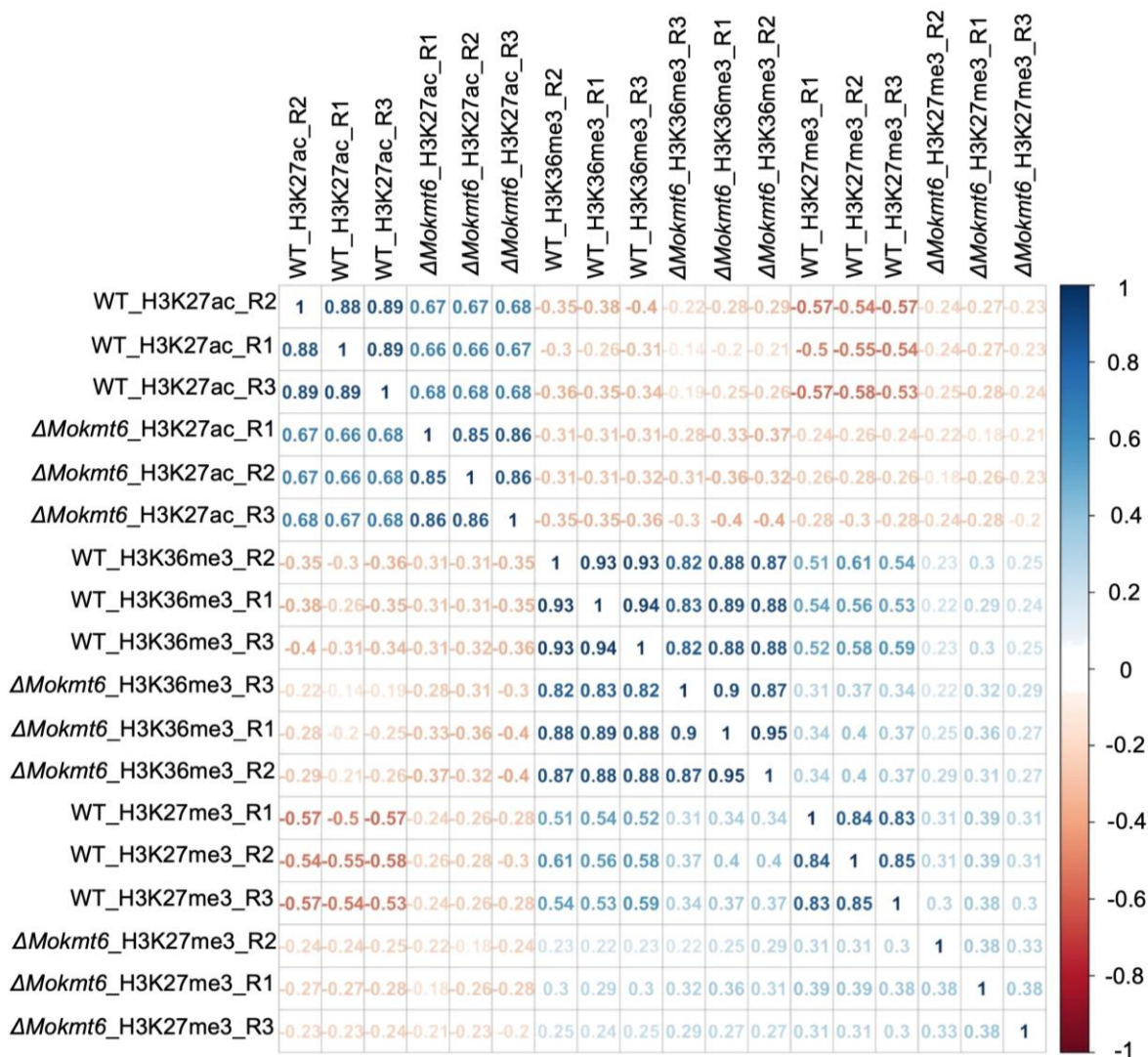
$\Delta Mokmt6\Delta Mogcn5$ fails to cause disease. (A) Representative blast lesions of rice leaves sprayed with wild type (WT), $\Delta Mokmt6$, $\Delta Mogcn5$, $\Delta Mokmt6\Delta Mogcn5$ and 0.25% gelatin (control). The photos were taken at 7 days post inoculation and show typical blast lesions during a compatible interaction. (B) Bar plots showing quantitative analysis of 14 independent infected rice leaves (n=14) from two biological experiments, classified as described in (Valent *et al.* 1991). The disease severity rating for each treatment was compared to the wild type infection and student's t-test was used to determine statistically significant differences. **, $p < 0.01$; *, $p < 0.05$.



Supplementary Figure A.15

Distinct gene expression pattern of effectors in H3K27me3-dependent and -independent manners. Expression of effector gene *SLP1* (A), *MoCDIP5* (B), *BAS1* (C), and *MC69* (D).

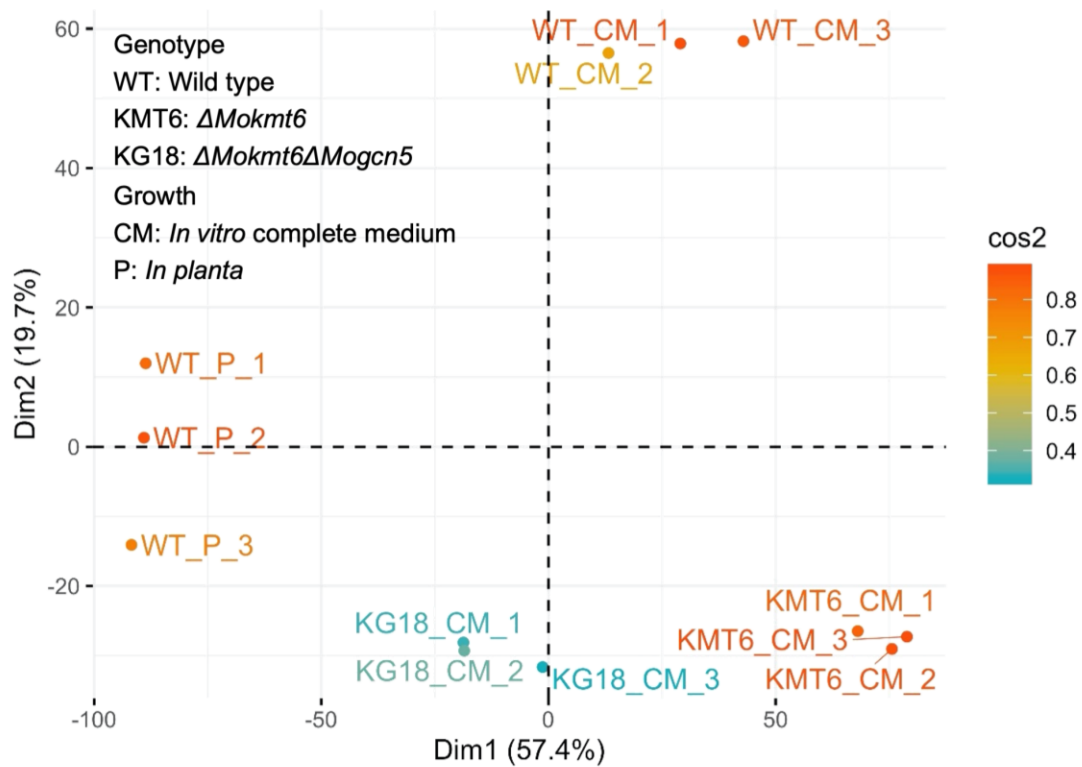
RNA-seq data are collected from wild type (WT) *M. oryzae* strain Guy11 growing under *in vitro* complete medium (CM) and *in planta*, and two mutants Δ *Mokmt6* and Δ *Mokmt6* Δ *Mogcn5* growing in CM. The mutant Δ *Mokmt6* lacks H3K27me3 and the double mutant Δ *Mokmt6* Δ *Mogcn5* lacks H3K27me3 and majority of H3K27ac. Letters above the violin plots indicate the significant difference among groups based on ANOVA and Tukey's HSD test.



Supplementary Figure A.16

Reproducibility of ChIP-seq data across biological replicates by pair-wise Pearson

correction analysis. ChIP-seq experiment were conducted on *M. oryzae* Guy11 wild type (WT) and the mutant Δ Mokmt6 with absence of H3K27me3 growing in complete medium (CM). R represents biological replicates.



Supplementary Figure A.17

Principal component analysis on RNA-Seq dataset. RNA-Seq experiment was conducted on RNAs extracted from *M. oryzae* Guy11 wild type (WT) growing under *in vitro* complete medium (CM) and *in planta*, and two mutants $\Delta Mokmt6$ and $\Delta Mokmt6 \Delta Mogcn5$ growing in complete medium. The mutant $\Delta Mokmt6$ lacks H3K27me3 and the double mutant $\Delta Mokmt6 \Delta Mogcn5$ lacks H3K27me3 and majority of H3K27ac.

Supplementary Table A.1

Summary of ChIP-seq data from experiments conducted on wild type (WT) *M. oryzae*, strain Guy11 and mutant *ΔMokmt6* grown under *in vitro* complete medium (CM).

Supplementary Table A.2

Location and information of H3K27me3 ChIP peaks called from MACS2.

Supplementary Table A.3

Location and information of H3K27ac ChIP peaks called from MACS2.

Supplementary Table A.4

Gene Ontology for H3K27me3 marked genes identified for *M. oryzae* strain Guy11 grown *in vitro* complete medium.

Supplementary Table A.5

Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of H3K27me3 marked genes.

Supplementary Table A.6

Primers used in this study.

Supplementary Table A.7

Contingency table for chi-square test of association between predicted effectors and H3K27me3 domains.

Supplementary Table A.8

Contingency table for chi-square test of association between predicted effectors induced during infection and H3K27me3 domains during complete medium growth.

Supplementary Table A.9

High-confidence characterized effectors, avirulence and *in planta* induced genes with H3K27me3 status.

Supplementary Table A.10

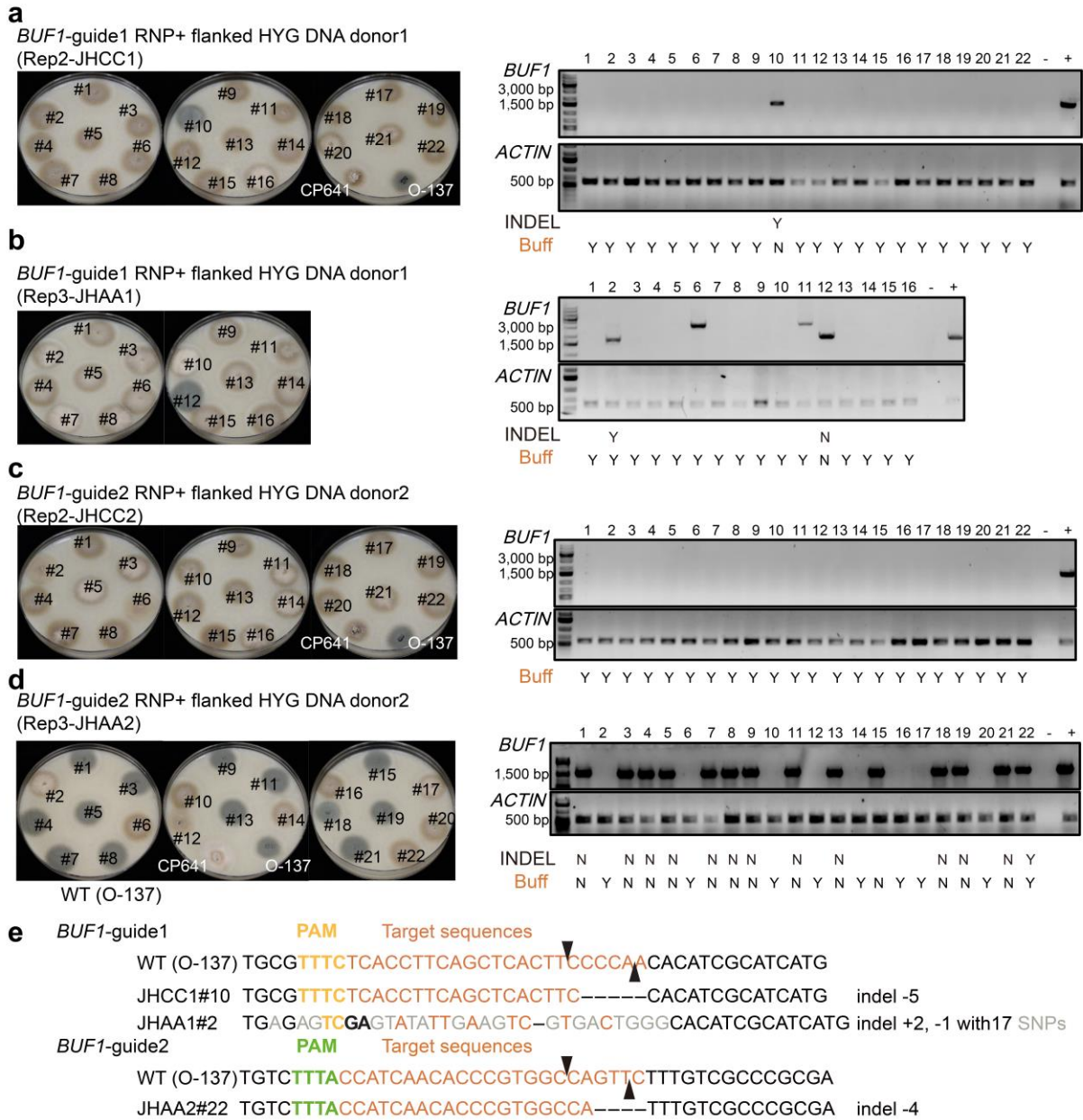
Contingency table for chi-square test of association between high-confidence characterized effectors and H3K27me3 domains.

Supplementary Table A.11

Contingency table for chi-square test of association between high-confidence characterized effectors and transcriptional induction in *ΔMokmt6*.

Appendix B - Supplementary Figures and Tables of Chapter 3

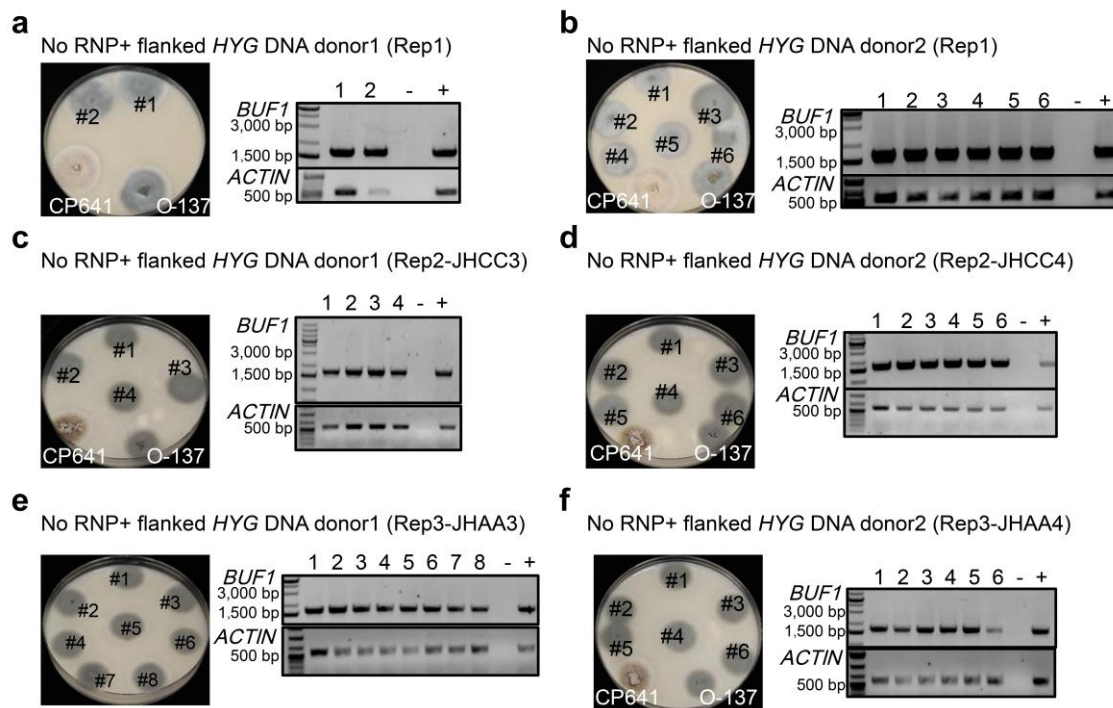
Supplementary data



Supplementary Figure B.1 Cas12a RNP combined with short homology donor DNA cause efficient gene deletion at *BUF1* locus in the field isolate O-137.

(a, b, c, and d) After protoplast transformations, hygromycin resistant transformants from replications 2 and 3 were plated in OTA for phenotyping and genotyping. DNA of transformants was extracted for genotyping with *BUF1*_F/R primer pair. *ACTIN* amplification served as the loading control. CP641 derived from O-137, is the control strain for showing the buff phenotype,

O-137 is the wild-type isolate used in these assays. - and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR amplification, respectively. (e) The detailed sequences change of the mutants with INDELs are presented. Yellow (*BUF1*-guide1)/green (*BUF1*-guide2) and orange letters highlight PAM sequence and target sequence, respectively. Black triangles around the target sequences highlight the predicted Cas12a cut site. Grey letters indicate the observed SNPs.



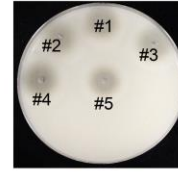
Supplementary Figure B.2 Flanked *HYG* DNA donor alone cannot cause the buff mutation.

(a, c and e) Hygromycin resistant transformants from three biological replications generated with flanked *HYG* DNA donor1 alone were plated on OTA for phenotyping. CP641 derived from O-137, is the control strain for showing the buff phenotype, O-137 is the wild-type isolate used in these assays. Genotyping result with *BUF1* primers shows the indistinguishable size of amplifications with WT (+). Water (-) was used as the negative control. *ACTIN* amplification served as the loading control. Similar assays performed with flanked *HYG* DNA donor2 are showed in panel (b, d and f).

a *BUF1*-guide1 RNP+ flanked *HYG* DNA donor1 (Rep1)



b No RNP+ flanked *HYG* DNA donor1 (Rep1)



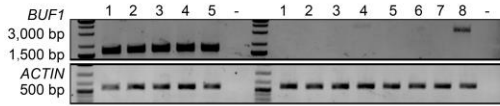
c *BUF1*-guide2 RNP+ flanked *HYG* DNA donor2 (Rep1)



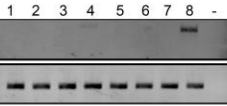
d No RNP+ flanked *HYG* DNA donor2 (Rep1)



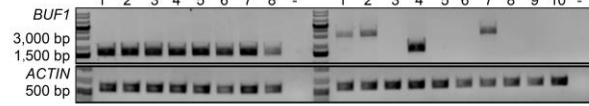
e No RNP+ flanked *HYG* DNA donor1 (Rep1)



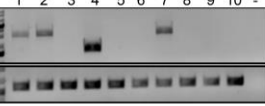
BUF1-guide1 RNP+ flanked *HYG* DNA donor1 (Rep1)



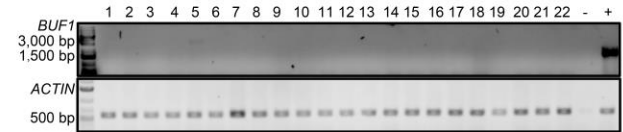
f No RNP+ flanked *HYG* DNA donor2 (Rep1)



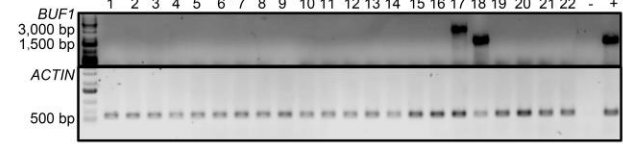
BUF1-guide2 RNP+ flanked *HYG* DNA donor2 (Rep1)



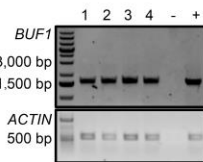
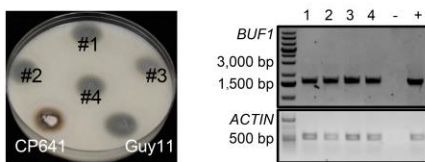
g *BUF1*-guide1 RNP+ flanked *HYG* DNA donor1 (Rep2-JHBB1)



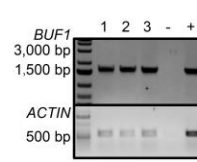
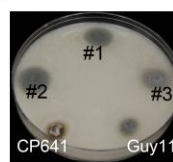
h *BUF1*-guide2 RNP+ flanked *HYG* DNA donor2 (Rep2-JHBB2)



i No RNP+ flanked *HYG* DNA donor1 (Rep2-JHBB3)



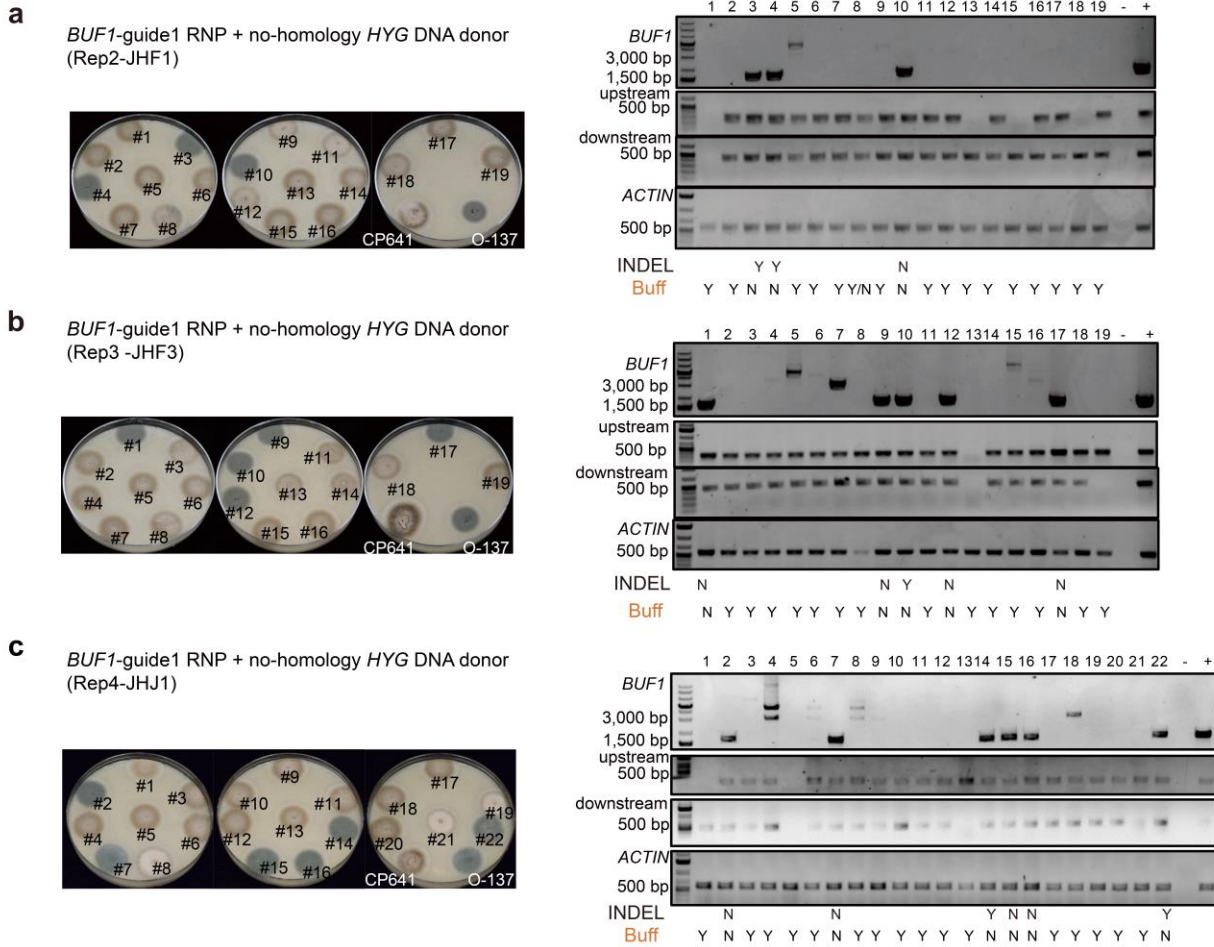
j No RNP+ flanked *HYG* DNA donor2 (Rep2-JHBB4)



Supplementary Figure B.3 Cas12a RNP combined with short homology donor DNA cause efficient gene deletion at *BUF1* locus in the field isolate Guy11.

(a, c, g, h, i and j) After protoplast transformations, hygromycin resistant transformants from biological replications were plated in RPA for phenotyping. (e, f, g, h, i and j) DNA of selected transformants were extracted for genotyping with *BUF1*_F/R primer pair. *ACTIN* amplification served as the loading control. CP281 (a weeping lovegrass strain) or CP641 (derived from O-137) showing the buff phenotype, Guy11 is the wild-type isolate used in these assays. - and +

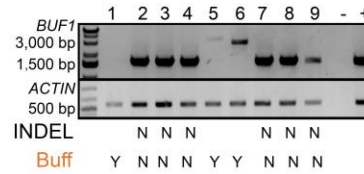
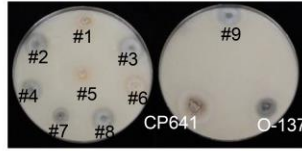
indicate the negative control (water) and positive control (Guy11 genomic DNA) for PCR amplification, respectively.



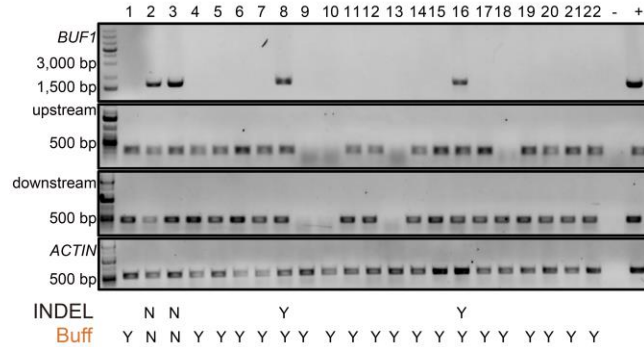
Supplementary Figure B.4 No-homology donor DNA alone with *BUF1*-guide1 RNP causes buff mutation.

(a, b and c) Hygromycin resistant transformants with *BUF1*-guide1 RNP and no-homology *HYG* DNA donor generated from three biological assays (rep 2 to 4) were placed on OTA for testing the mycelial color change. CP641 is the control for buff color change, O-137 is the wild-type isolate used in the experiment. Genotyping results are shown in the right panel. The wild-type-like PCR products were purified and Sanger sequenced to detect the potential INDELs. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. Buff Y indicates the strains used for genotyping were buff, while Buff N indicates wild-type phenotype. - and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR amplification, respectively. Upstream and downstream primer pairs were used for detecting the existence of large-scale locus alternation. *ACTIN* amplification served as the loading control.

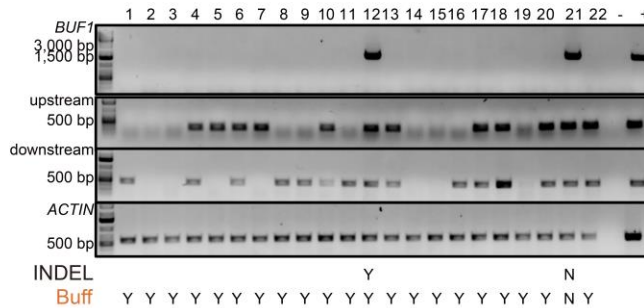
a *BUF1*-guide2 RNP + no-homology *HYG* DNA donor (Rep1)



b *BUF1*-guide2 RNP + no-homology *HYG* DNA donor (Rep2-JHEE2)

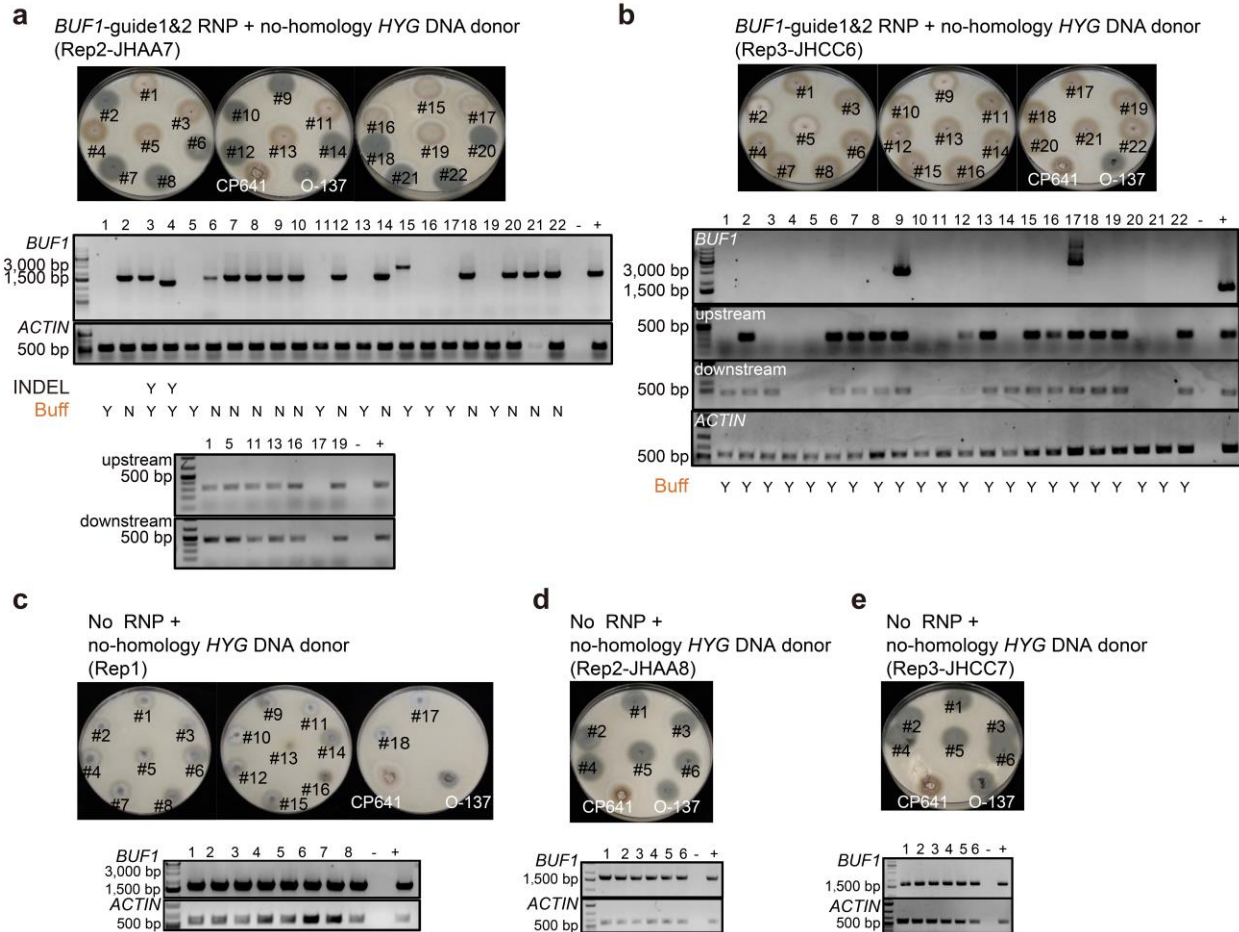


c *BUF1*-guide2 RNP + no-homology *HYG* DNA donor (Rep3-JHCC5)



Supplementary Figure B.5 No-homology donor DNA alone with *BUF1*-guide2 RNP causes buff mutation.

(a, b and c) Hygromycin resistant transformants with *BUF1*-guide2 RNP and no-homology *HYG* DNA donor generated from three biological assays (rep 1 to 3) were placed on OTA for testing the mycelial color change. CP641 is the control for buff color change, O-137 is the wild-type isolate used in the experiment. Genotyping results are shown in the right panel. The wild-type-like PCR products were purified and Sanger sequenced to detect the potential INDELs. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. Buff Y indicates the strains used for genotyping were buff, while Buff N indicates wild-type phenotype. - and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR amplification, respectively. Upstream and downstream primer pairs were used for detecting the existence of large-scale locus alternation. *ACTIN* amplification served as the loading control.



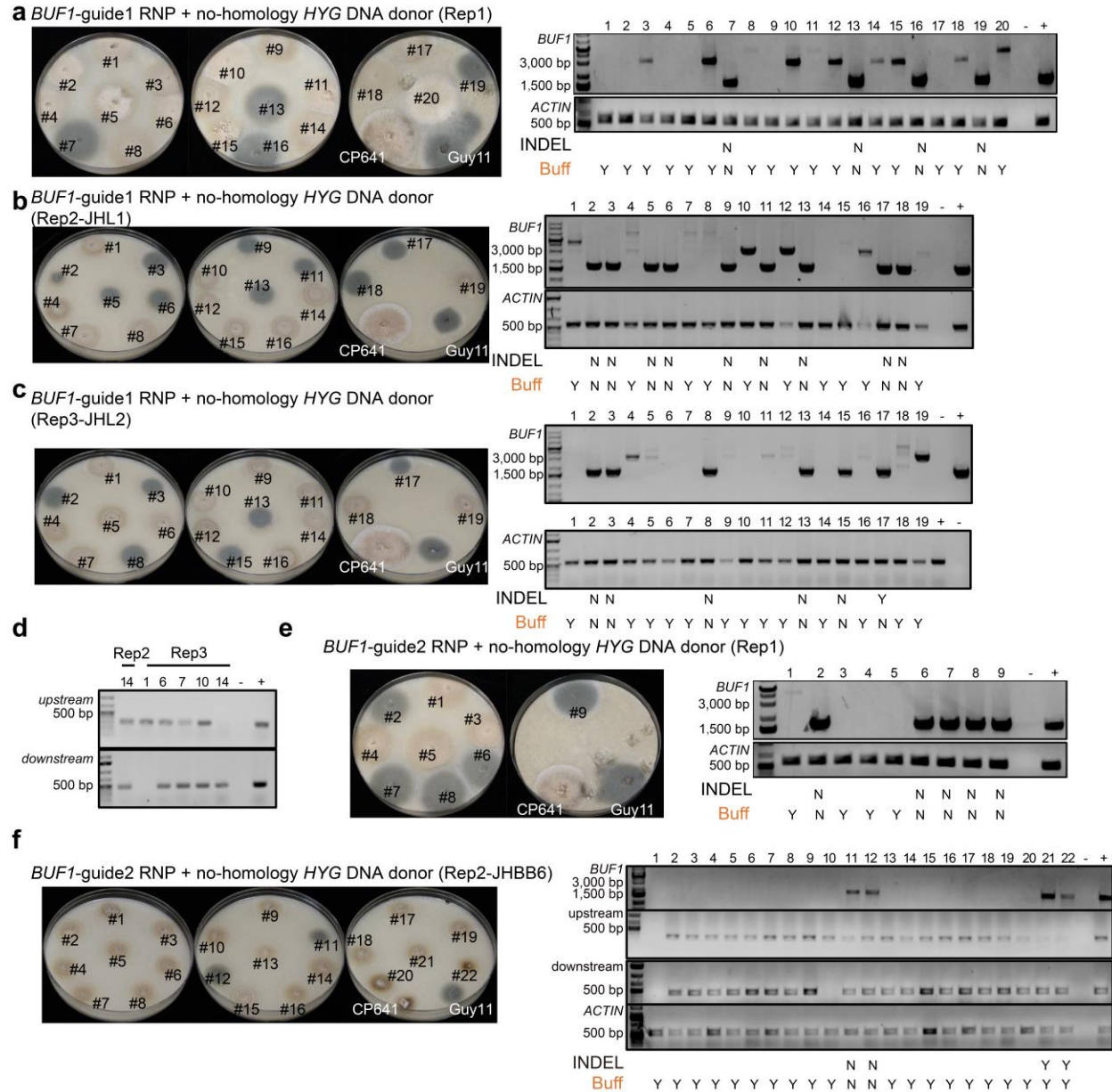
Supplementary Figure B.6 No-homology donor DNA alone with dual *BUF1*- RNP causes buff mutation, while DNA donor alone cannot cause the buff mutation.

(a and b) Hygromycin resistant transformants with *BUF1*-guide2 RNP and no-homology *HYG* DNA donor generated from two biological assays (rep 2 and 3) were placed on OTA for testing the mycelial color change. CP641 is the control for buff color change, O-137 is the wild-type isolate used in the experiment. Genotyping results are shown in the bottom panel. The wild-type-like PCR products were purified and Sanger sequenced to detect the potential INDELs. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. Buff Y indicates the strains used for genotyping were buff, while Buff N indicates wild-type phenotype. - and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR amplification, respectively. Upstream and downstream primer pairs were used for detecting the existence of large-scale locus alternation. *ACTIN* amplification served as the loading control. Similar assays performed with no-homology *HYG* DNA alone are showed in panel (c, d and e).

a <i>BUF1-guide1</i>		PAM	Target sequences	
WT (O-137)	TGCG	TTTCTCACCTTCAGCTCACTT	CCCCAACACATCGCATCATG	
Rep1				
Δ <i>BUF1</i> #10	TGCG	TTTCTCACCTTCAGCTCACTT	CCC <u>CAA</u> CAACACATCGCATCATG	indel +3 bp tandem duplication
Rep2-JHF1				
Δ <i>BUF1</i> #3	TGCG	TTTCTCACCTTCAGCTCACTT	---CCAACACATCGCATCATG	indel -2 bp
Δ <i>BUF1</i> #4	TGCG	TTTCTCACCTTCAGCTCA	----CCCAACACATCGCATCATG	indel -4 bp
Rep3-JHF3				
Δ <i>BUF1</i> #10	TGCG	TTTCTCACCTTCAGCTCA	-----CCAACACATCGCATCATG	indel -5 bp
Rep4-JHJ1				
Δ <i>BUF1</i> #14	TGCG	TTTCTCACCTTCAGCTCACTT	---CCCAACACATCGCATCATG	indel -1 bp
Δ <i>BUF1</i> #22	TGCG	TTTCTCACCTTCAGCTCACTT	---CCAACACATCGCATCATG	indel -2 bp
b <i>BUF1-guide2</i>		PAM	Target sequences	
WT (O-137)	TGTC	TTTACCATCAACACCCGTGGC	AGTTCTTTGTCGCCCGCGA	
Rep2-JHEE2				
Δ <i>buf1</i> #8	TGTC	TTTACCATCAACACCCGTG	---AGTTCTTTGTCGCCCGCGA	indel -3 bp
Δ <i>buf1</i> #16	TGTC	TTTACCATCAACACCCGT	-----GTTCTTTGTCGCCCGCGA	indel -5 bp
Rep3-JHCC5				
Δ <i>buf1</i> #12	TGTC	TTTACCATCAACACCCGTGGT	C-GTTCTTTGTCGCCCGCGA	indel -1 with 1 SNP
c <i>BUF1-guide1&2</i>				
Rep2-JHAA7				
Δ <i>buf1</i> #3	TGTC	TTTACCATCAACACCCGT	-----GTTCTTTGTCGCCCGCGA	indel -5 bp in guide2 region
Δ <i>buf1</i> #4	TGCG	TTTCTCACCTTCAGCTCACTTC	--deletion---TTTGTGCGCCCGCGA	deletion -365 bp in between guide1&2 region

Supplementary Figure B.7 INDEL are detected in editing strains from O-137.

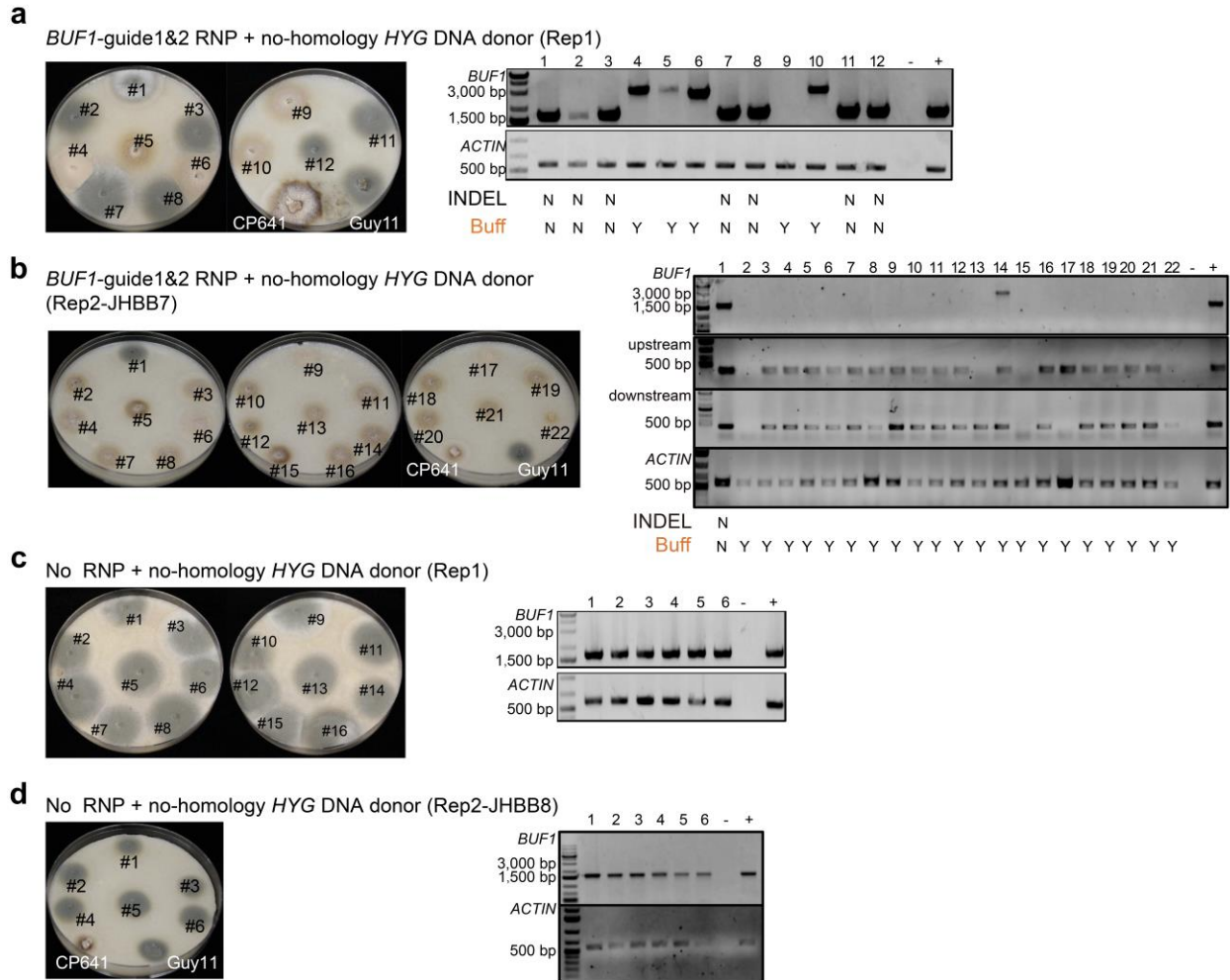
Yellow (guide1)/green (guide2) and orange letters highlight PAM sequence and target sequence, respectively. Black triangles around the target sequences highlight the potential Cas12a cut site. Underlined letters indicate the potential template for tandem duplication. The INDEL mutations found with *BUF1*-guide1, *BUF1*-guide2 and dual targeting are presented in (a), (b) and (c), respectively.



Supplementary Figure B.8 Donor DNA without homology sequences integrates at Cas12a target site in the genetic background of Guy11.

(a, b and c) Hygromycin resistant transformants with *BUF1*-guide1 RNP and no-homology *HYG* DNA donor from three biological assays (rep 1 to 3) were plated on OTA for phenotyping. CP641 is a control for buff color change, Guy11 is the wild-type isolate used in the experiment. *BUF1_F/R* primer pair was used for genotyping. *ACTIN* amplification was served as the loading control. - and + indicate the negative control (water) and positive control (Guy11 genomic DNA) for PCR amplification, respectively. The wild-type-like PCR products were purified and Sanger sequenced to detect the INDELs. INDEL N indicates there were no INDELs observed after

sequencing, otherwise labelled as INDEL Y. Buff Y indicates the strains used for genotyping were buff, while Buff N indicates wild-type phenotype. (d) Upstream and downstream primer pairs were used for detecting the existence of large-scale locus alternation. Similar layouts for different transformants are presented in (e and f) transformants with *BUF1*-guide2 RNP and no-homology *HYG* DNA donor from two independent biological replications (rep1 and 2).



Supplementary Figure B.9 No-homology donor DNA alone with dual *BUF1*- RNP causes buff mutation, while DNA donor alone cannot cause the buff mutant in the genetic background of Guy11.

(a and b) Hygromycin resistant transformants with dual *BUF1*-guide1/2 RNPs and no-homology *HYG* DNA donor generated from two biological assays (rep 1 and 2) were placed on OTA for testing the mycelial color change. CP641 is the positive control for buff color change, Guy11 is the wild-type isolate used in the experiment. Genotyping results are shown in the right panel. The wild-type-like PCR products were purified and Sanger sequenced to detect the potential INDELs. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. Buff Y indicates the strains used for genotyping were buff, while Buff N indicates wild-type phenotype. - and + indicate the negative control (water) and positive control (Guy11 genomic DNA) for PCR amplification, respectively. Upstream and downstream primer pairs were used for detecting the existence of large-scale locus alternation. *ACTIN* amplification

served as the loading control. Similar assays performed with no-homology *HYG* DNA alone are showed in panel (c and d).

BUF1-guide1 PAM Target sequences

WT (Guy11) TGCCTTTCTCACCTTCAGCTCACTTCCCAACACATCGCATCATG

Rep3-JHL2

ΔBUF1#17 TGCCTTTCTCACCTTCAGCTCA-----CCCAACACATCGCATCATG indel -4 bp

BUF1-guide2 PAM Target sequences

WT (Guy11) TGTCTTTACCATCAACACCCGTGGCCAGTTCTTTGTCGCCCGCGA

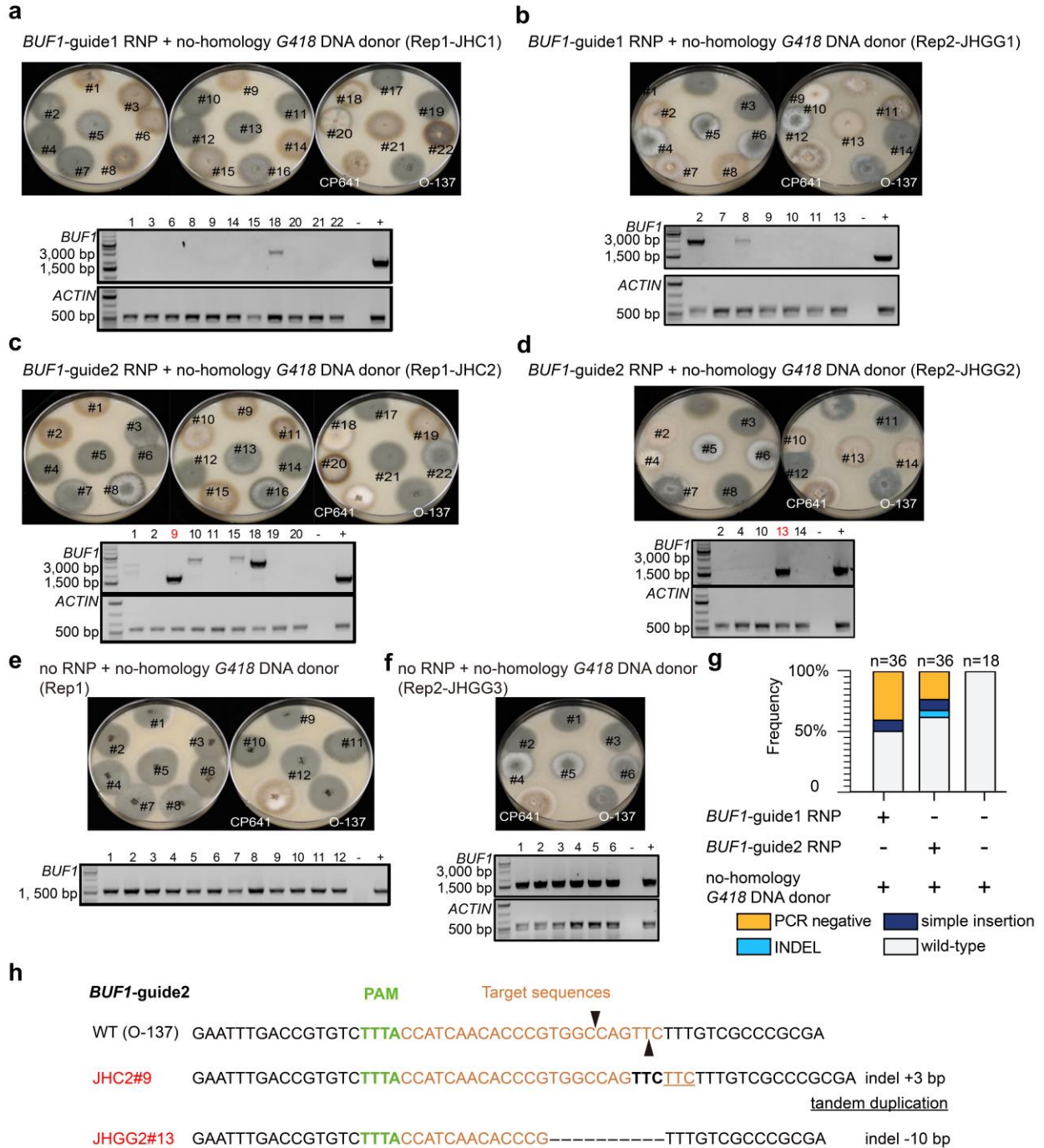
Rep2-JHBB6

Δbuf1#21 TGTCTTTACCATCAACACCCGTG-----TGTGCGCCCGCGA indel -10 bp

Δbuf1#22 TGTCTTTACCATCAACACCCGT-----GTTCTTTGTCGCCCGCGA indel -5 bp

Supplementary Figure B.10 INDEL are detected in editing strains from Guy11.

Yellow (guide1), Green (guide2) and orange letters highlight PAM sequence and target sequence, respectively. Black triangles around the target sequences highlight the predicted Cas12a cut site.



Supplementary Figure B.11 Non-canonical DNA repair at *BUF1* locus is donor and antibiotic independent.

(a and b) *G418* resistant transformants with *BUF1*-guide1 RNP and no-homology *G418* DNA donor from two independent biological replications (rep1 and 2) were plated on OTA for phenotyping. Buff mutants were used for DNA extraction and further genotyping with *BUF1*_F/R primer pair. *ACTIN* amplification served as the loading control. Same assays but with

BUF1-guide2 RNP and no-homology *G418* DNA donor or donor alone are presented in (c and d) and (e and f), respectively. CP641 is a control for buff color change, O-137 is the wild-type isolate used in the experiment. - and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR amplification, respectively. (g) The frequency summary of DNA DSB repair outcomes in O-137. The number of independent PCR validated transformants (x) is listed (n=x). The mutation outcomes were concluded from two biological replications. (h) DNA from JHC2#9 from (c) and JHGG2#13 from (d) as highlighted by red were subjected to Sanger sequencing and aligned with wild-type sequence. Underlined letters indicate the potential template for tandem duplication.

WT (O-137)

PAM Target sequences
TGC GTTCTCACCTTCAGCTCACTTCCCAACACATCGCATCATGTG

(1) Rep1- Δ buf1#2

TGC GTTCTCACCTTCAGCTCACTTGTGTTATG --~1.4 kb *HYG* insertion-GCCAAGCCCAACACATCGCATCATGTG

(2) Rep1- Δ buf1#5

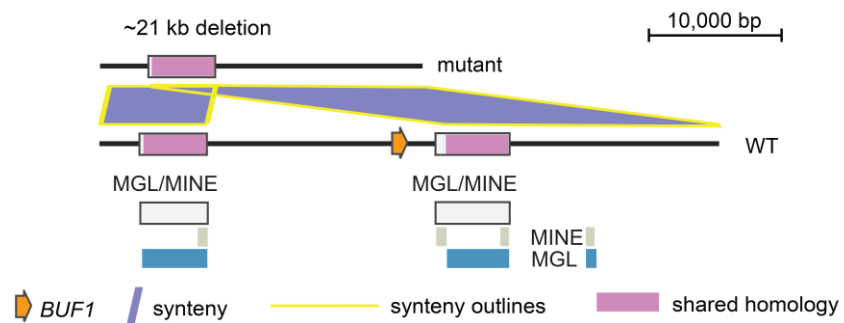
TGC GTTCTCACCTTCAGCTCACTTGGCCCTC --~10 kb *HYG* insertion-GAAATACCCCAACACATCGCATCATGTG

(3) Rep1- Δ buf1#6

TGC GTTCTCACCTTCAGCTCACTTGCACAGTT-----~17 kb *HYG* insertion-----GCATCGTGTG

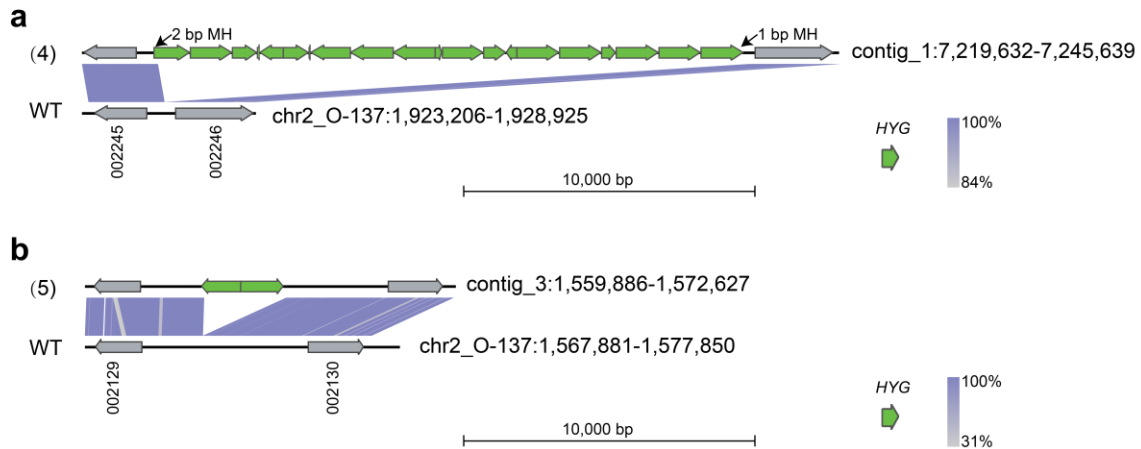
Supplementary Figure B.12 *HYG* DNA donor shares microhomology with original *BUF1* DNA sequences in the sequenced simple and large insertion mutants.

2-3 bp microhomology shared between *HYG* DNA donor and original *BUF1* DNA sequences are highlighted by white letters with black boxes.



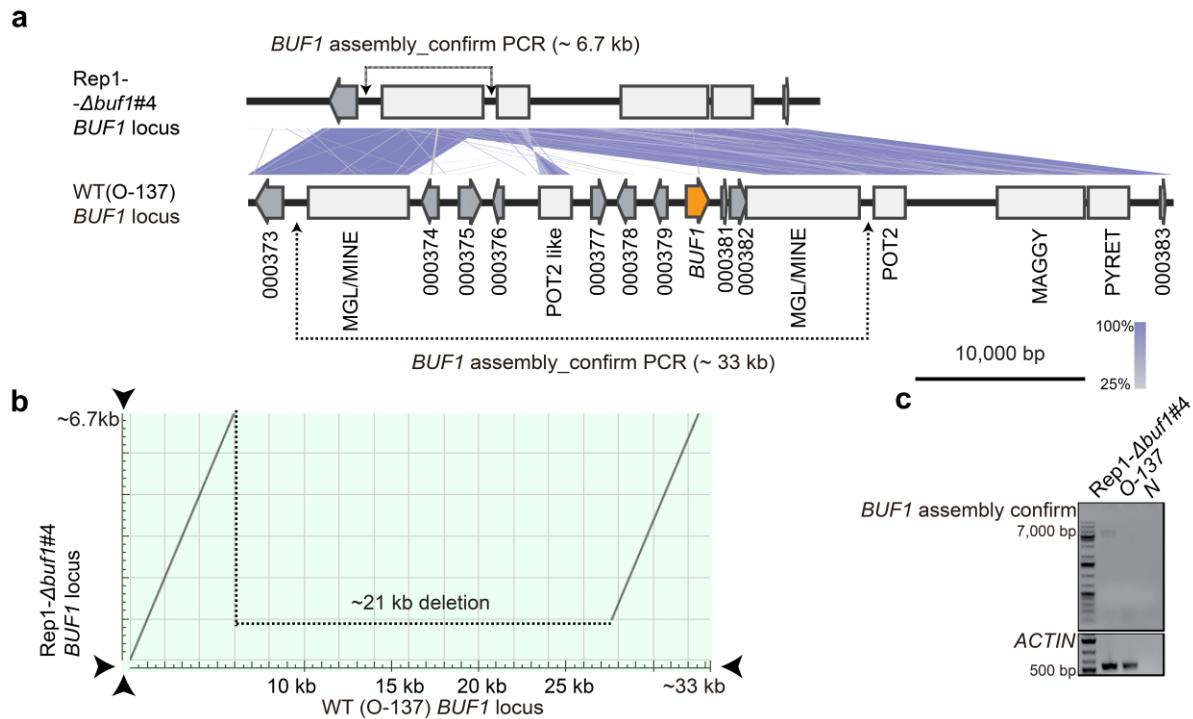
Supplementary Figure B.13 Composition of MGL/MINE in the large deletion mutant.

Presumably the sequence similarity between two hybrid MGL/MINE was used to resolve the break resulting in a single retrotransposon copy and the 21 kb deletion.



Supplementary Figure B.14 Insertion site of no-homology *HYG* DNA donor in the large deletion mutants.

(a) (b) Synteny illustration of insertion site of no-homology *HYG* DNA donor in the large deletion mutants ((4) Rep1-*Δbuf1*#4 and (5) Rep4-*Δbuf1*#5 from Fig. 3c) are presented. Note: there is a short contig (contig_73, (20,048bp)) in Rep4-*Δbuf1*#5 showing the same *HYG* DNA donor insertion site, but with 5 copies of no-homology *HYG* DNA donor inserted.



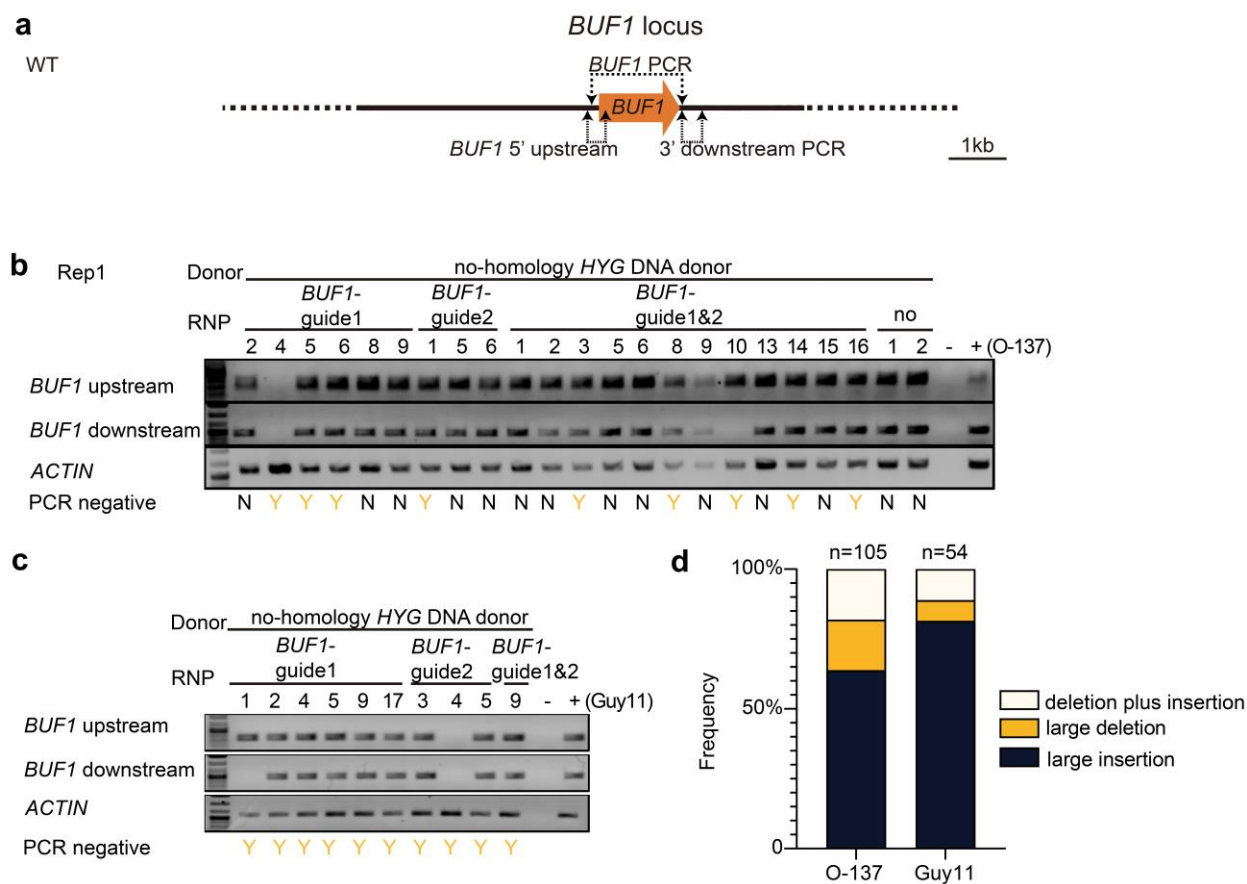
Supplementary Figure B.15 Confirmation of accurate assembly for the large deletion mutant by PCR amplification.

(a) Illustration of *BUF1* locus in wild-type (O-137) and Rep1- Δ buf1#4 ((4) from Fig. 3c). *BUF1* assembly_confirm PCR primers were used to test the accuracy of assembly. (b) The BLAST result of *BUF1* locus in two above-described strains (query: *BUF1* locus in Rep1- Δ buf1#4, (4) from Fig. 3c, subject: *BUF1* locus in wild-type (O-137), the black arrow indicates where the *BUF1* assembly_confirm PCR primers locate. (c) PCR amplification confirms the large deletion in Rep1- Δ buf1#4, N indicates the negative control (water).

WT (O-137) PAM Target sequences
 CGTGACTGGCGCTGGTGCGTTTCTACCTTCAGCTCACTTCCCCACACATCGCATCATGTG
 (7) Rep1-*Δbuf1*#10
 CGTGACTGGCGCTGGTGCGTTTCTACCTTCAGCTCACTTTGCCCT ~5.6 kb HYG insertion-TTTGAC100xN gap
 (8) Rep4-*Δbuf1*#13
 CGTGACTGGCGCTGGTGCGTTTCTCACCCAGTCA~6.4 kb HYG insertion-TGACTGCCLGVGCV

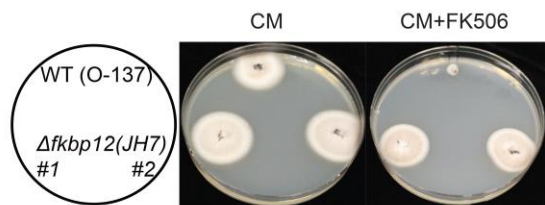
Supplementary Figure B.16 *HYG* DNA donor shares microhomology with original *BUF1* DNA sequences in the deletion plus insertion mutants.

2-3 bp microhomology shared between *HYG* DNA donor and original *BUF1* DNA sequences are highlighted with white letters under black boxes. Deletion plus insertion mutants ((7) Rep1-*Δbuf1*#10 and (8) Rep4-*Δbuf1*#13 from Fig. 3d) are presented.



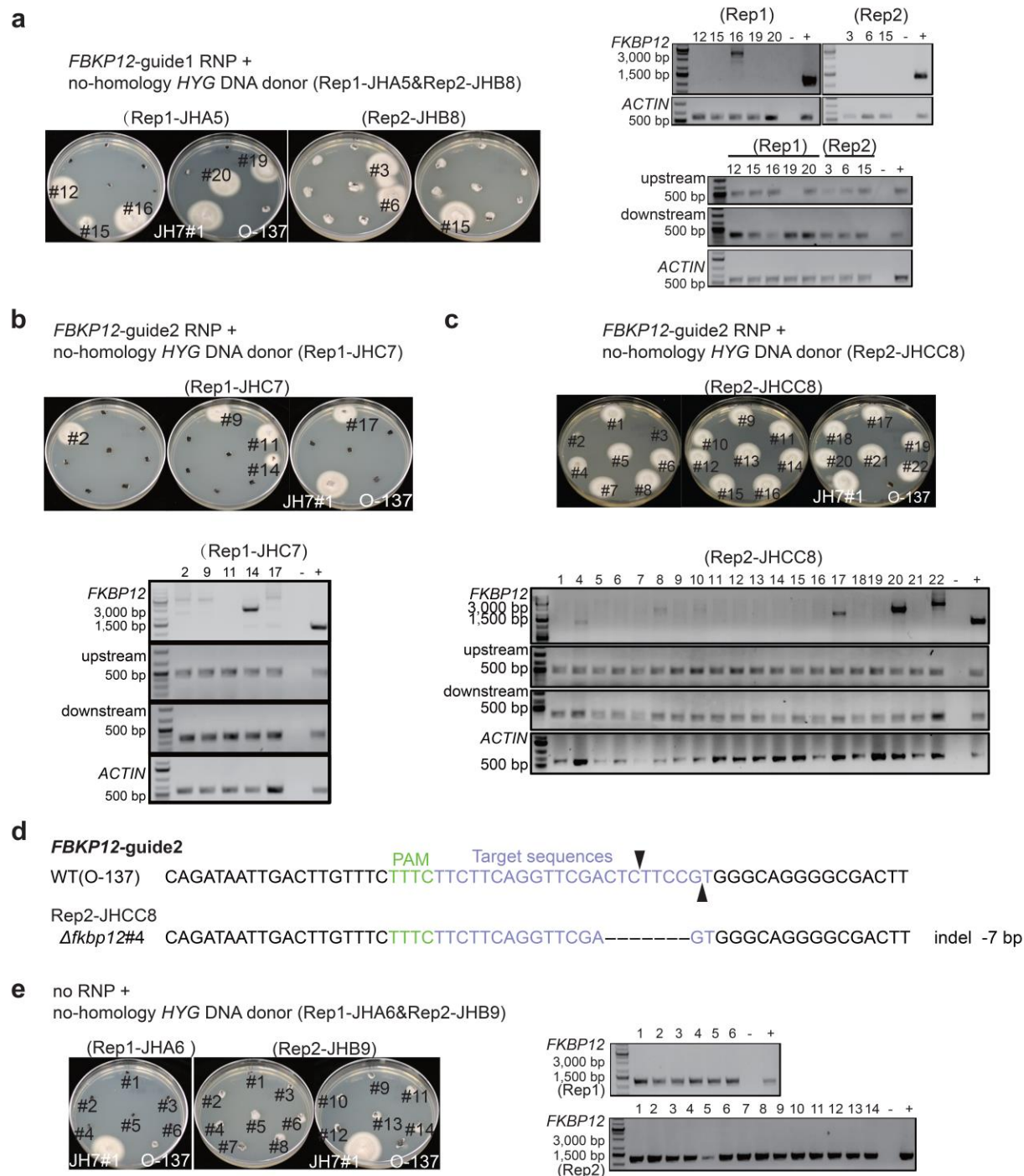
Supplementary Figure B.17 PCR genotyping confirms three non-canonical error-prone DNA repair patterns.

(a) Illustration of *BUF1* locus in the wild-type. (b, c) 5' upstream and 3' downstream regions of *BUF1*, as highlighted in (a) were amplified for testing the occurrence of large insertions, large deletions, or deletion plus insertions at the *BUF1* locus. Black line above the lanes indicates the DNA donor and RNP used for generating the strains. PCR negative Y indicates that PCR negative genotyping result was found in the strain with *BUF1* PCR, otherwise, labelled as PCR negative N. *ACTIN* amplification served as the loading control. – and + indicate the negative control (water) and positive control (O-137 or Guy11 genomic DNA) for PCR amplification, respectively. (d) The frequency for different non-canonical DNA repair outcomes in O-137 or Guy11 with no-homology *HYG* DNA donor and *BUF1* RNPs. The O-137 data were calculated from at least three biological replications, while the Guy11 data were from at least two biological replications. The number (x) of independent transformants used for counting is listed 'n=x'.



Supplementary Figure B.18 $\Delta fkbp12$ is insensitive to FK506

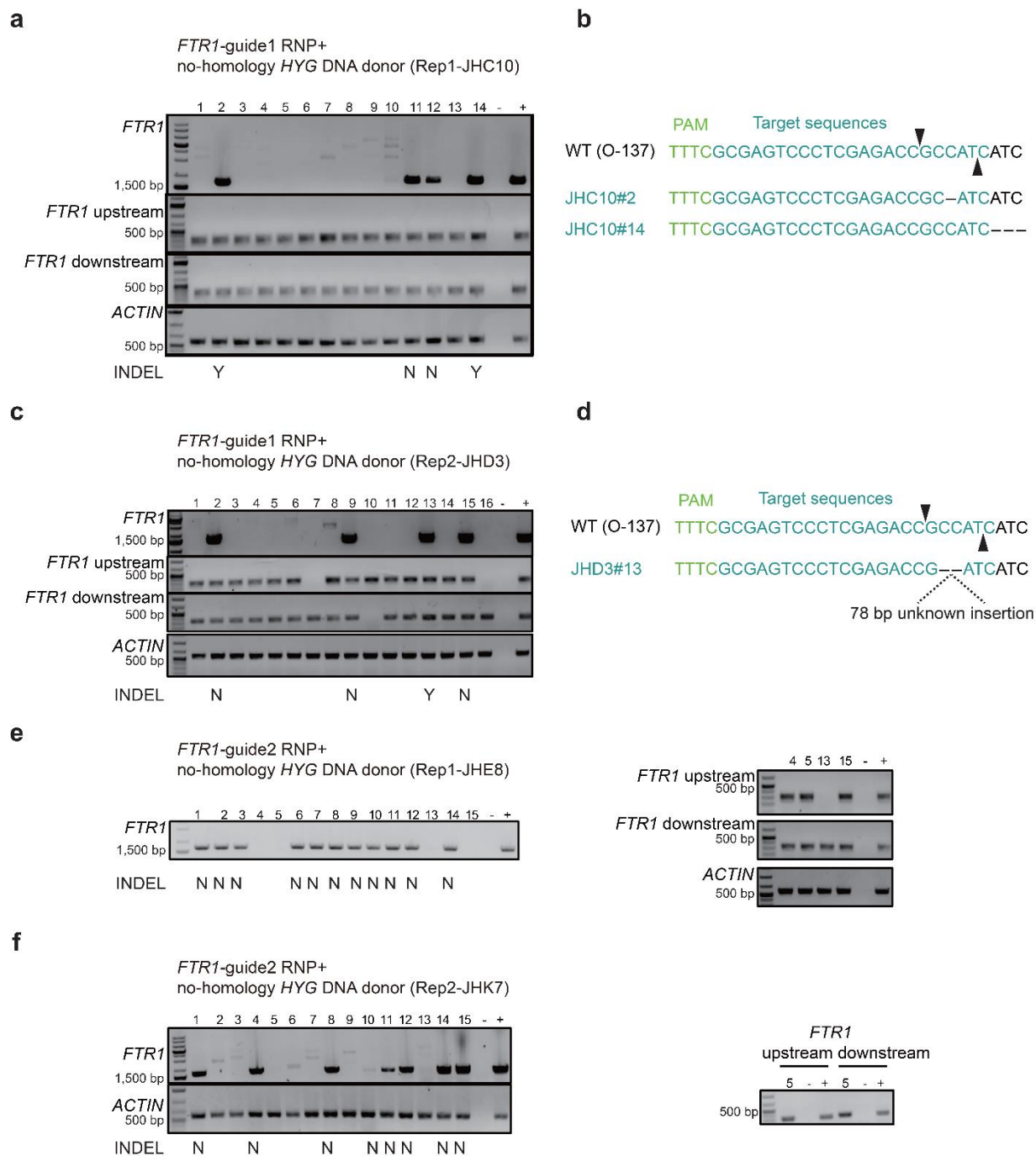
Representative phenotyping result of $\Delta fkbp12$. CM supplemented with 1 $\mu\text{g/mL}$ FK506 was used for testing the sensitivity of distinct strains to FK506.



Supplementary Figure B.19 *FBKP12* RNP along with donor DNA generate FK506 insensitive strains.

(a) Hygromycin resistant transformants with *FBKP12*-guide1 RNP along with no-homology *HYG* DNA donor were plated on CM supplemented with 1 μ g/mL FK506 to test the drug sensitivity. JH7#1 is an *Δfkbp12* mutant used as a positive control for FK506 insensitivity, O-137 is the wild-type isolate used in the assay. Two independent transformation assays were

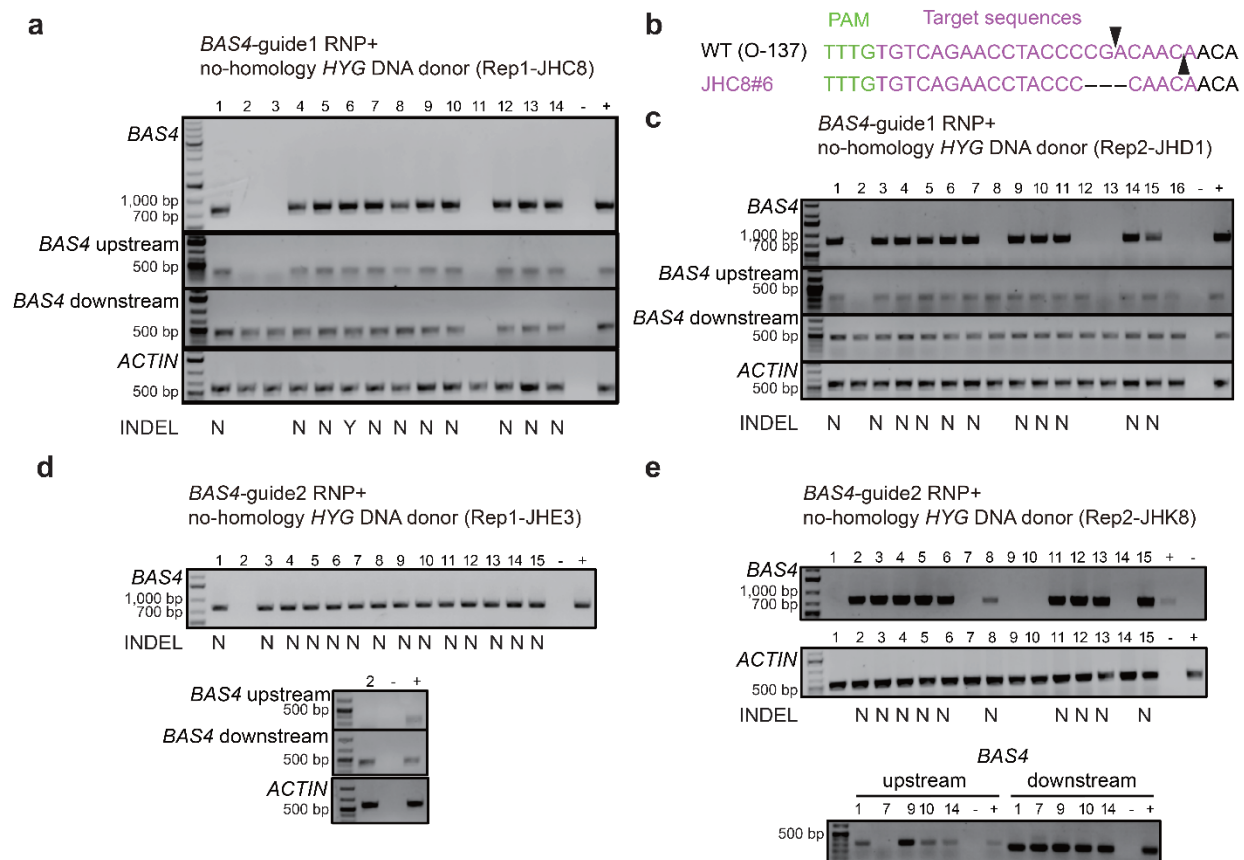
performed (JHA5 and JHB8). The FK506 insensitive strains were used for further genotyping with primer pairs spanning *FKBP12* coding sequence, 5' upstream and 3' downstream. *ACTIN* was used as a control for proper DNA extraction. – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction. Similar layout but for the hygromycin resistant transformants with *FKBP12*-guide2 RNP along with no-homology *HYG* DNA donor from two biological replications (JHC7 and JHCC8) were showed in (b and c). (d) The mutational profile of the INDEL mutant from *FKBP12*-guide2 RNP targeting assay is presented. (e) Hygromycin resistant transformants with no-homology *HYG* DNA donor alone from two biological replications were plated for phenotyping and further genotyping with the similar strategy used in (a, b, and c).



Supplementary Figure B.20 Genotyping for transformants from *FTR1* RNPs and no-homology *HYG* DNA donor.

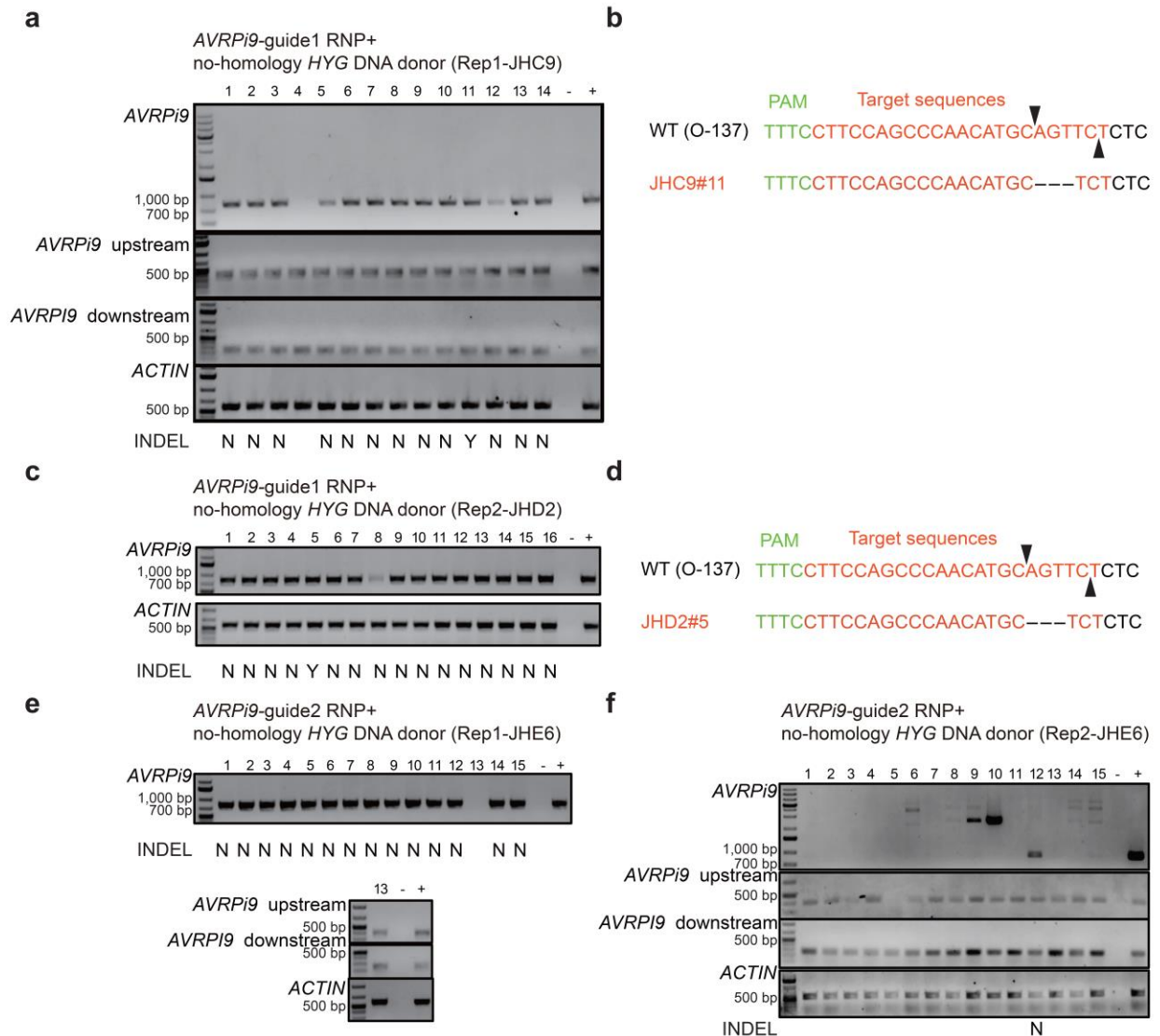
(a) PCR genotyping results for hygromycin resistant transformants from *FTR1*-guide1 RNP with no-homology *HYG* DNA donor. The wild-type-like PCR products were purified and Sanger sequenced to detect the potential INDELs. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. (b) The sequence alignment between WT and

INDEL strains from (a). Biological replication result for same guide1 is placed in (c) and (d). (e and f) Replicated PCR genotyping results for hygromycin resistant transformants from *FTR1*-guide2 RNP with no-homology *HYG* DNA donor. – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction.



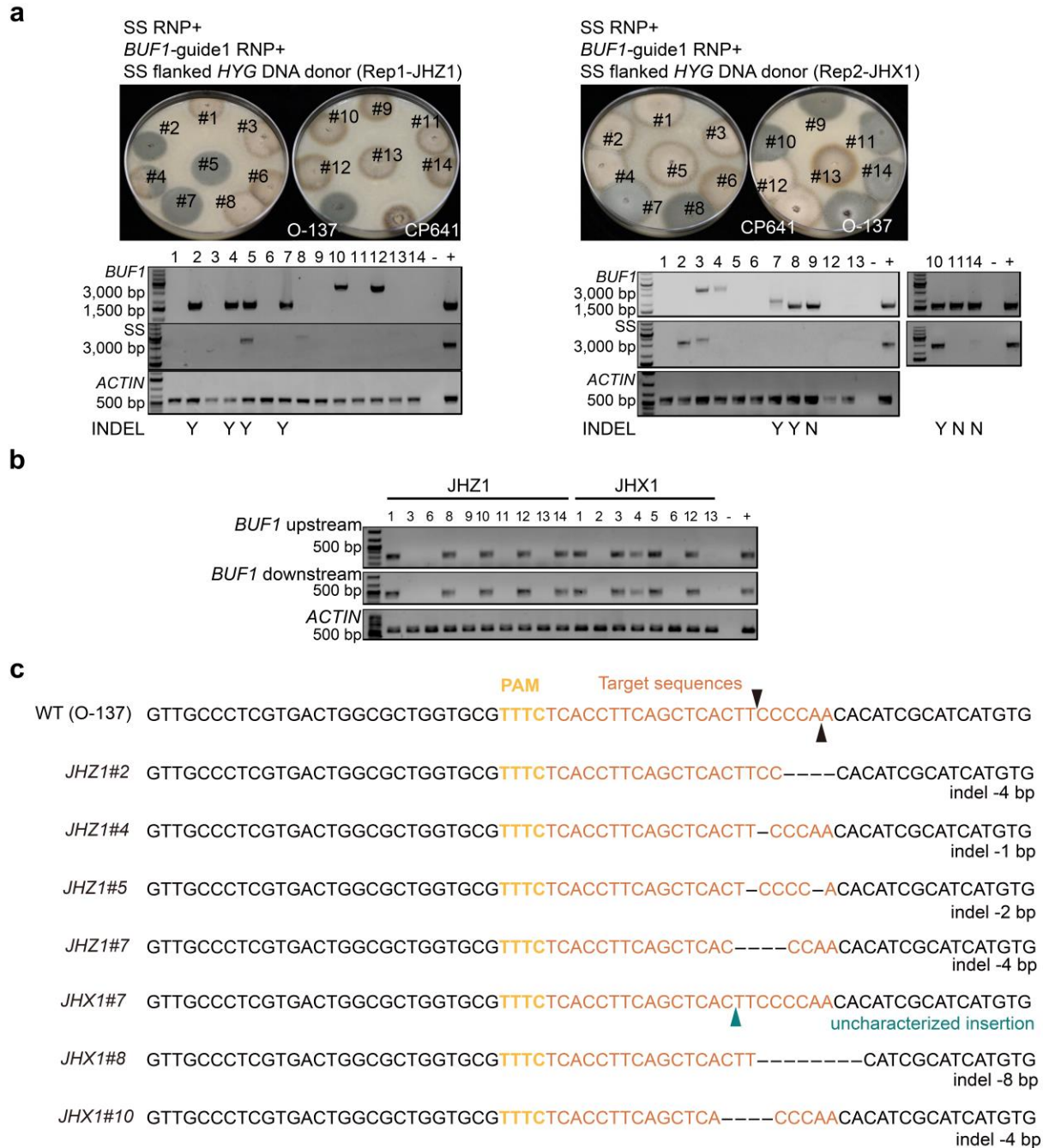
Supplementary Figure B.21 Genotyping for transformants from *BAS4* RNPs and no-homology *HYG* DNA donor.

(a) PCR genotyping results for hygromycin resistant transformants from *BAS4*-guide1 RNP and no-homology *HYG* DNA donor. The wild-type-like PCR products were purified and Sanger sequenced to detect the potential INDEL. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. (b) The sequence alignment between WT and INDEL strains from (a). Independent replication for same guide1 were shown in (c). (d and e) Replicated PCR genotyping results for hygromycin resistant transformants from *BAS4*-guide2 RNP and no-homology *HYG* DNA donor. – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction.



Supplementary Figure B.22 Genotyping for transformants from AVR*Pi*9 RNPs and no-homology *HYG* DNA donor.

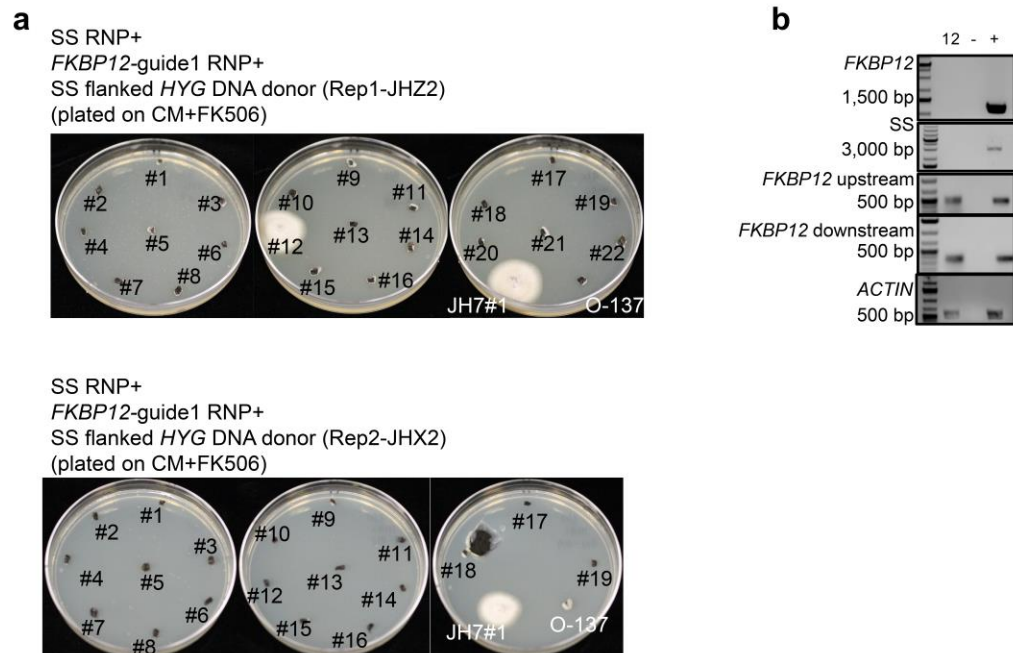
(a) PCR genotyping results for hygromycin resistant transformants from AVR*Pi*9-guide1 RNP and no-homology *HYG* DNA donor. The wild-type-like PCR products were purified and Sanger sequenced to detect the potential INDELs. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. (b) The sequence alignment between WT and INDEL strains from (a). Biological replication result for same guide1 were placed in (c) and (d). (e and f) Replicated PCR genotyping results for hygromycin resistant transformants from AVR*Pi*9-guide2 RNP and no-homology *HYG* DNA donor. – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction.



Supplementary Figure B.23 Genotyping and phenotyping of *BUF1* transformants from the second-site targeting assay.

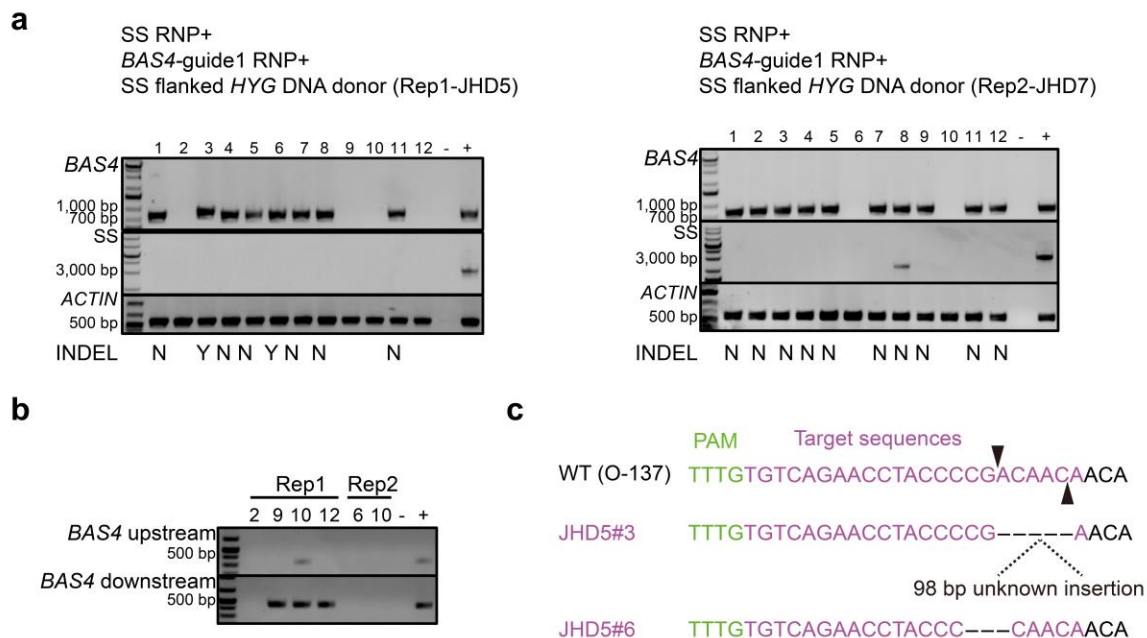
(a) Hygromycin resistant transformants were plated on OTA for phenotyping the mycelial color change. DNA was extracted and used for genotyping with primary-site (*BUF1*), second-site (SS) and *ACTIN* (control). The wild-type-like PCR products from primary-site were purified and Sanger sequenced to detect the potential INDELS. INDEL N indicates there were no INDELS observed after sequencing, otherwise labelled as INDEL Y. CP641 derived from O-137, is

positive strain for showing buff phenotype, O-137 is the wild-type isolate used in these assays. The assays were repeated twice independently (rep1-JHZ1 and rep2-JHX1). (b) PCR negative strains from *BUF1* locus in (a) were further characterized with 5' *BUF1* upstream and 3' downstream primers. – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction. (c) The sequence alignment between wild-type and the strains with INDEL at *BUF1* targeting region identified from (a) is showed.



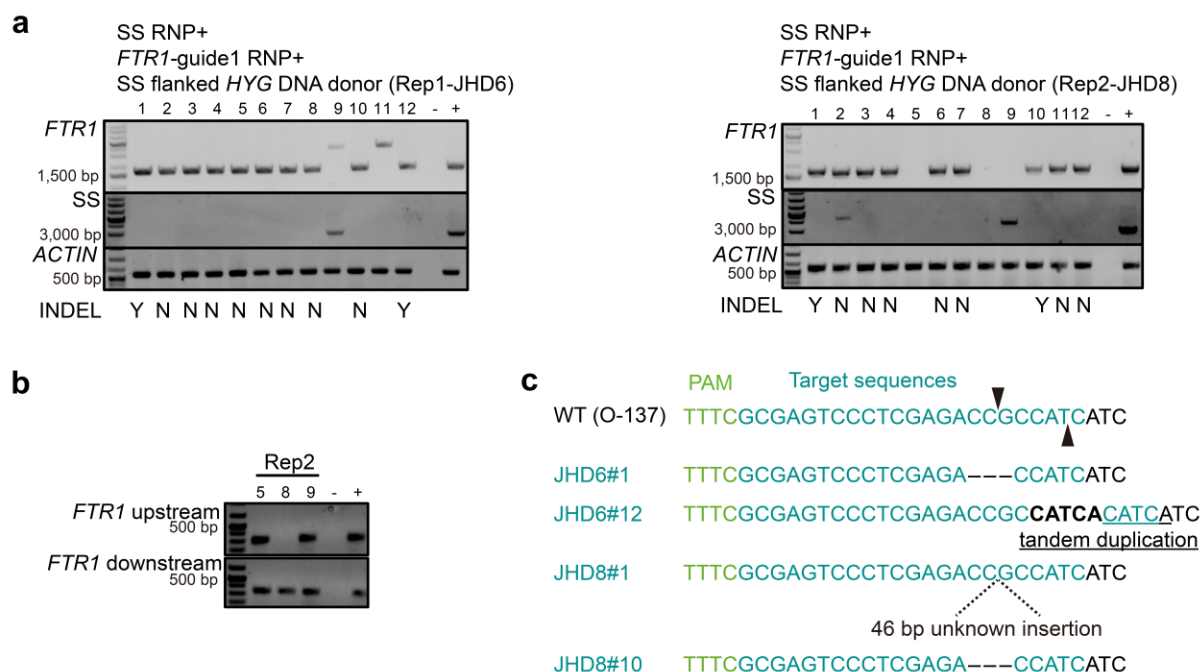
Supplementary Figure B.24 Genotyping and phenotyping of *FKBP12* transformants from the second-site targeting assay.

(a) Hygromycin resistant transformants were plated on CM+FK506 for phenotyping the FK506 resistance. The assays were repeated twice (rep1- JHZ2 and rep2-JHX2). JH7#1 derived from O-137 is a control strain for FK506 resistance, O-137 is the wild-type used in these assays. (b) DNA from the FK506 resistant strain (JHZ2#12) was extracted and used for genotyping with primary-site (*FKBP12*), second-site (SS), *FKBP12* upstream (5')/ downstream (3') and *ACTIN* (control). – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction.



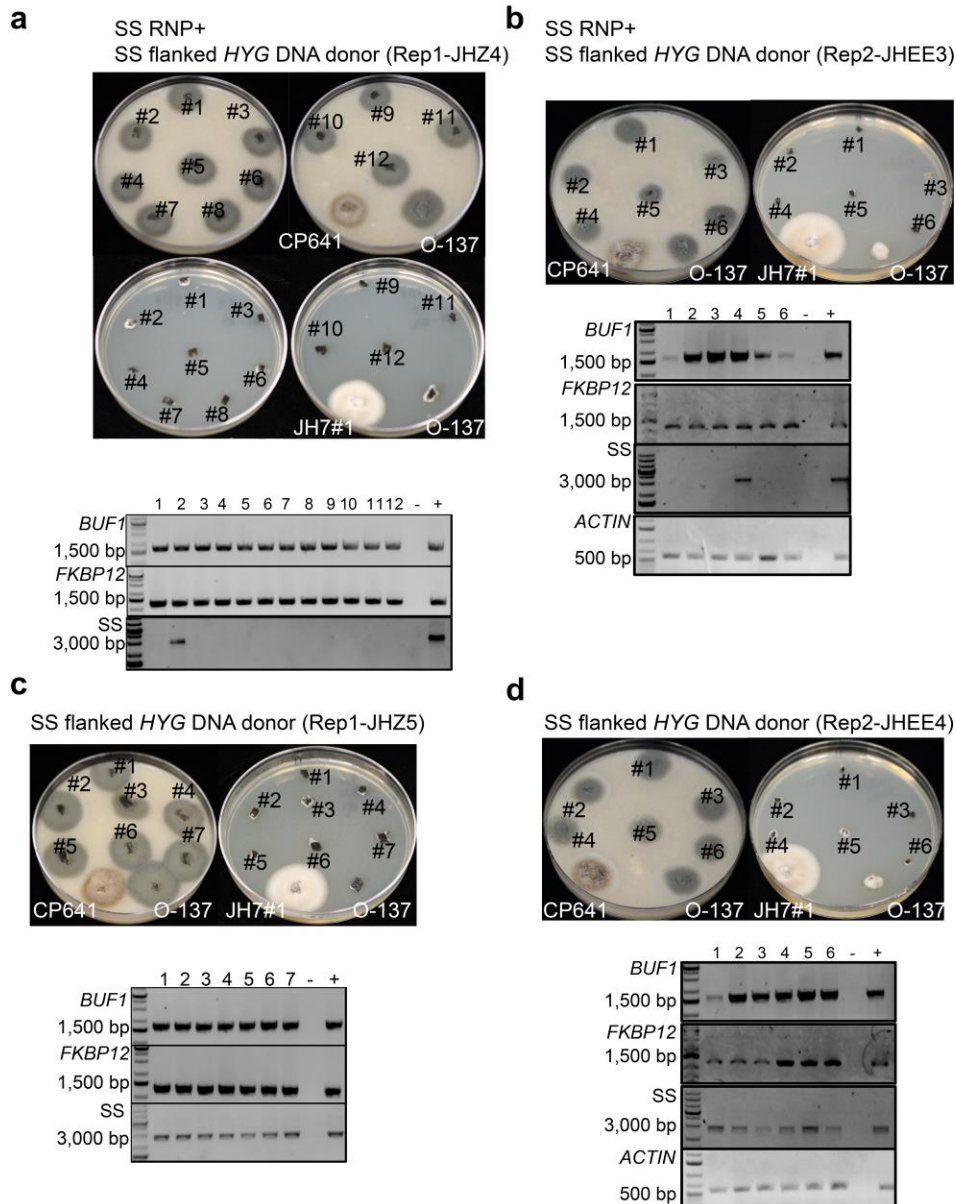
Supplementary Figure B.25 Genotyping of *BAS4* transformants from the second-site targeting assay.

(a) Hygromycin resistant transformants were used for DNA extraction and further genotyping with primary-site (*BAS4*), second-site (SS) and *ACTIN* (control). The wild-type-like PCR products from primary-site were purified and Sanger sequenced to detect the potential INDELs. INDEL N indicates there were no INDELs observed after sequencing, otherwise labelled as INDEL Y. The assays were repeated twice (rep1-JHD5 and rep2-JHD7). (b) PCR negative strains from *BAS4* locus in (a) were further characterized with 5' *BAS4* upstream and 3' downstream primers. – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction. (c) The alignment between wild-type and the strains with INDEL at *BAS4* targeting region identified from (a) is showed.



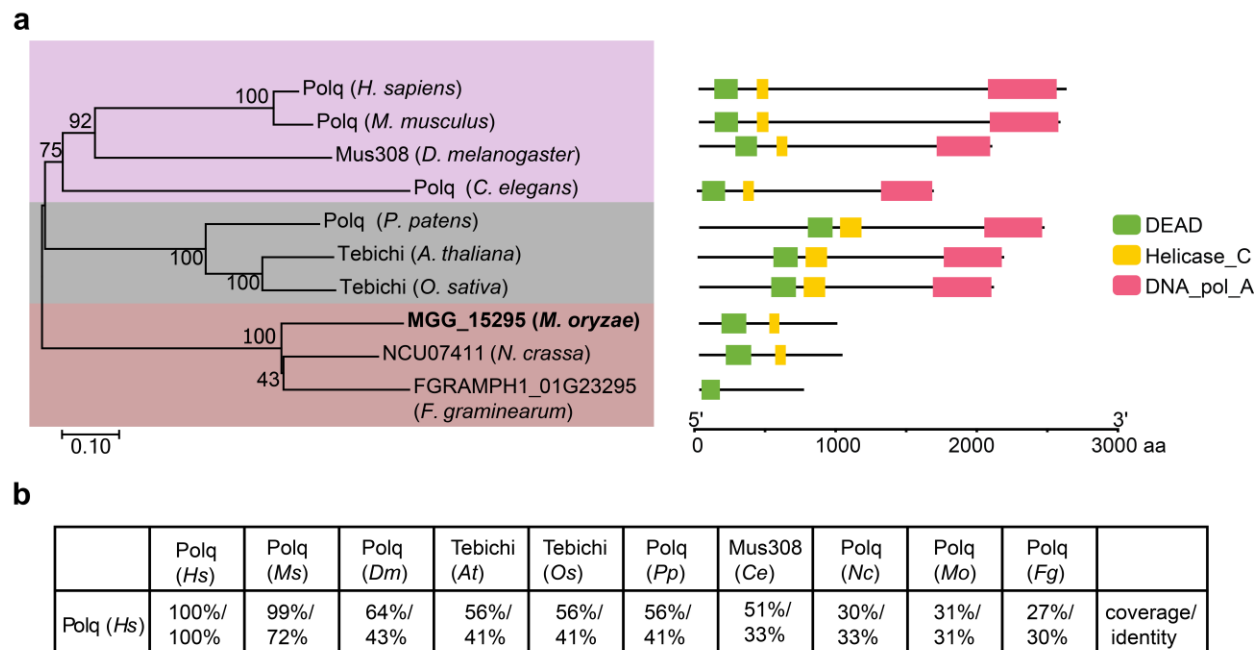
Supplementary Figure B.26 Genotyping of *FTR1* transformants from the second-site targeting assay.

(a) Hygromycin resistant transformants were used for DNA extraction and further genotyping with primary-site (*FTR1*), second-site (SS) and *ACTIN* (control). The wild-type-like PCR products from primary-site were purified and Sanger sequenced to detect the INDEL. INDEL N indicated there were no INDELs observed after sequencing, otherwise labeled as INDEL Y. The assays were repeated twice (rep1-JHD6 and rep2-JHD8). (b) PCR negative strains from *FTR1* locus in (a) were further characterized with *FTR1* 5' upstream and 3' downstream primers. – and + indicate the negative control (water) and positive control (O-137 genomic DNA) for PCR reaction. (c) The sequence alignment between wild-type and the strains with INDEL at *FTR1* targeting region identified from (a) is showed. Underlined letters indicated the potential template for tandem duplication.



Supplementary Figure B.27 The second-site RNP along with the second-site donor or donor alone serve as control for the second-site targeting assay.

(a and b) Replicated hygromycin resistant transformants from Second-site RNP along with second-site (SS) donor were plated on OTA and CM+FK506 for phenotyping, and further genotyping with primary-site (*BUF1* and *FKBP12*), second-site (SS). (c and d) Replicated phenotyping and genotyping results for hygromycin resistant transformants from SS donor alone are presented, with similar layout as (a).



Supplementary Figure B.28 Phylogenetic and sequence similarity analysis of *POLQ* homologs.

(a) Neighbor-joining tree of selected *POLQ* homologs generated by MEGA X with 1000 bootstrap replications. Domain distribution of each protein is shown in the right panel (b) Homology matrix analysis of selected *POLQ* in different organisms. Numbers indicate protein coverage and identity.

Supplementary Table B.1 Source data for flanked DNA donor assay with *BUF1*.

Supplementary Table B.2 Source data for no-homology DNA donor assay with *BUF1*.

Supplementary Table B.3 Statistic summary for eleven long-read *Δbuf1* and *Δbas4* assemblies.

Supplementary Table B.4 Source data for no-homology DNA donor assay with *FKBP12*, *FTR1*, *BAS4* and *AVRPI9*.

Supplementary Table B.5 Source data for second site assay with *BUF1*, *FKBP12*, *FTR1* and *BAS4*.

Supplementary Table B.6 The primers used in this study.

Appendix C - Supplementary Table of Chapter 5

Supplementary Table C.1 The primers used in this study.

Appendix D - Supplementary Table of Chapter 6

Supplementary Table D.1 The primers used in this study.