University of Massachusetts Amherst

DISCOVERING MECHANISMS DRIVING ADAPTIVE EVOLUTION IN THE CROSS-KINGDOM FUNGAL PATHOGEN FUSARIUM OXYSPORUM

Item Type	Dissertation (Open Access)
Authors	Ayhan, Dilay Hazal
DOI	10.7275/24509379
Download date	2025-10-30 21:47:04
Link to Item	https://hdl.handle.net/20.500.14394/18577

DISCOVERING MECHANISMS DRIVING ADAPTIVE EVOLUTION IN THE CROSS-KINGDOM FUNGAL PATHOGEN *FUSARIUM OXYSPORUM*

A Dissertation Presented

by

DİLAY HAZAL AYHAN

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2021

Molecular and Cellular Biology

© Copyright by Dilay Hazal Ayhan 2021

All Rights Reserved

DISCOVERING MECHANISMS DRIVING ADAPTIVE EVOLUTION IN THE CROSS-KINGDOM FUNGAL PATHOGEN *FUSARIUM OXYSPORUM*

A Dissertation Presented

by

DİLAY HAZAL AYHAN

Approved as to style and content by:

_____

Li-Jun Ma, Chair

_____

Ana Caicedo, Member

_____

Patrick Flaherty, Member

_____

Antonio Di Pietro, Member

_____

Thomas J. Maresca, Program Director

Molecular and Cellular Biology Program

DEDICATION

To my family

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Li-Jun Ma, for her immense support through my doctorate program. She provided the opportunity for me to grow as a scientist and also encouraged me to be better in all aspects of life.

I would give my gratitude to the other committee members: Dr. Antonio Di Pietro, Dr. Ana Caicedo, and Dr. Patrick Flaherty whose support, encouragements, and critical insights were very valuable. Dr. Di Pietro was also an irreplaceable collaborator for my research.

I am grateful to the past and current members of Ma Lab for the nurturing environment they created. I want to specifically thank Vista Sohrab, Taylor Aguiar, Katherine Rickelton, Rua'a Wahhas, Sarah Gregory for their help in my studies.

I would like to thank Dr. Cristina López Díaz, for being an excellent research partner; Dr. Eric Pearlman and Dr. Serena Abbondante for their valuable collaborations; Ravi Ranjan of Genomics Resource Lab, MGHPCC, and JGI for their technical and experimental supports.

I want to thank the Molecular and Cellular Biology Graduate Program and Biochemistry and Molecular Biology Department for providing me with the opportunity and the funding to study.

Additionally, I would like to thank Dr. Erdal Toprak, my undergraduate advisor, who helped me to enter graduate studies, and introduced me to microbial evolution. Finally, I would like to thank my family for their unconditional love and support.

ABSTRACT

DISCOVERING MECHANISMS DRIVING ADAPTIVE EVOLUTION IN THE CROSS-KINGDOM FUNGAL PATHOGEN *FUSARIUM OXYSPORUM*

SEPTEMBER 2021

DİLAY HAZAL AYHAN, B.S., SABANCI UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Li-Jun Ma

Fusarium oxysporum is a cross-kingdom pathogenic fungus that can cause vascular wilt disease in many economically important plants and local or disseminated infections in humans. Although it lacks a sexual stage in its life cycle, *F. oxysporum* can adapt to a wide range of hosts because of accessory chromosomes (ACs) which are enriched in host-specific genes and repeat content. This dissertation investigates the mechanisms that drive the adaptive evolution in the cross-kingdom pathogen *F. oxysporum* using comparative genomics and an experimental evolution approach. The first chapter compares phenotypes and genomes of a plant pathogenic isolate *F. oxysporum* f. sp. *lycopersici* 4287 (Fol4287) and a human pathogenic isolate *F. oxysporum* MRL8996. Fol4287 and MRL8996 differ in both morphology and AC gene and repeat content. The second chapter analyzes the Fol4287 populations generated by evolution experiments through whole-genome sequencing. Many copy number variations, single nucleotide variations, insertions deletions, and transposable element insertion variations (TIV) were detected. The final chapter compares the adaptive mutation mechanisms in Fol4287 and MRL8996 through experimental evolution. While transposons are the major cause of variation in both strains, the active transposons are different and encoded in their ACs.

For Fol4287 a DNA transposon, Hormin is highly active while in MRL8996, a short interspersed nuclear element, Foxy5, had the highest activity. The dynamic chromosomes are also different in both strains: in Fol4287 ACs and chromosome 13 have many copy number variations while in MRL8996 ACs are relatively stable while chromosome 12 is unstable. In addition, the Velvet complex, which is a major regulator of growth and mutated in multiple independent populations has an important role in the adaptation of *F. oxysporum*. Many interesting evolutionary events are also observed. In conclusion, the accessory chromosomes provide evolutionary hotspots for *F. oxysporum.*

CONTENTS

ACKNOWLEDGEMENTS v

ABSTRACT vi

LIST OF TABLES xii

LIST OF FIGURES xiii

LIST OF SYMBOLS AND ABBREVIATIONS xvi

INTRODUCTION 1

CHAPTER 1 COMPARATIVE STUDY OF TWO *F. OXYSPORUM* STRAINS WITH A PLANT OR A HUMAN HOST 4

1.1 Introduction 4

1.2 Methods 8

1.2.1 Fungal strains and growth conditions 8

1.2.2 Phenotyping of isolates 8

1.2.3 Genome assembly of Fol4287 10

1.2.4 Genome assembly of MRL8996 11

1.2.5 Read coverages and genomic features 11

1.2.6 Repeat Analysis 12

1.2.7 Genome sequence comparison 13

1.2.8 Chromosome separation using contour-clamped homogeneous electric field electrophoresis 13

1.2.9 Minichromosome confirmation in Fol4287 14

1.2.10 Gene annotations and gene ontology term enrichment analysis 15

1.3 Results 17

1.3.1 Morphological characterization of fungal isolates 17

1.3.2 Genome Assemblies 20

1.3.3 Genome comparisons 30

1.4 Discussion 41

1.4.1 Phenotypic characterization revealed unique niche adaptation 41

1.4.2 The genotypic comparison revealed distinct sets of accessory chromosomes enriched for genes that could be linked to the functional adaptation 42

1.4.3 Strategies to correlate genotypic variations with phenotypic differences 44

CHAPTER 2 EXPERIMENTAL EVOLUTION OF *F. OXYSPORUM* F. SP. *LYCOPERSICI* 4287 45

2.1 Introduction 45

2.2 Methods 52

2.2.1 Short-term evolution experiment in vivo and in vitro (performed by Cristina López Díaz) 52

2.2.2 Whole-genome sequencing of ancestor and passaged populations 53

2.2.3 Mapping and variant calling 53

2.2.4 Genomic Features 57

2.2.5 RNA sequencing 58

2.2.6 TE sequence alignment 59

2.3 Results 60

2.3.1 Surveying genotypic variation over short-term experimental evolution via whole-genome sequencing....60
2.3.2 Signatures of selection observed in evolving populations....73
2.4 Discussion....81
2.4.1 The role of DNA transposons....81
2.4.2 Genetic parallelism and predictability of evolution....82
2.4.3 Other events commonly observed in evolution experiments....83
2.4.4 Perspectives and future directions....84
CHAPTER 3 EXPERIMENTAL EVOLUTION OF *F. OXYSPORUM* MRL8996....86
3.1 Introduction....86
3.2 Methods....89
3.2.1 Short-term evolution experiment....89
3.2.2 Phenotyping....90
3.2.3 Whole genome sequencing....90
3.2.4 Variant Calling....91
3.2.5 Gene ontology (GO) enrichment analysis....93
3.2.6 Annotation of Foxy5....93
3.2.7 Other statistics....93
3.3 Results....95
3.3.1 Short-term evolution experiment set-up....95
3.3.2 Changed phenotype over the short-term evolution experiment....96
3.3.3 Surveying genotypic variation in STEE by whole genome sequencing....100

3.3.4 Transposon activity in STEE 110
3.3.5 Weak selection in MRL8996 evolving populations 114
3.3.6 Mutation patterns....116
3.4 Discussion....119
3.4.1 Loss of chromosome 12 119
3.4.2 Transposon activity 120
3.4.3 Selection pressure....121
3.4.4 Genotype-phenotype map 121
CONCLUSION 123
APPENDIX A....125
APPENDIX B....145
BIBLIOGRAPHY....150

LIST OF TABLES

Table 1.1 The assembly statistics for Fol4287 and MRL8996 genome assemblies 21

Table 1.2 Candidate centromeric sequences of Fol4287 23

Table 1.3 The predicted centromere sequences of MRL8996 genome.....29

Table 1.4 The chromosome sizes (bp) in the Fol4287 and MRL8996 genome assemblies 31

Table 1.5 Alignment metrics of the pair-wise comparisons in Figure 1.7B 31

Table 1.6 Enriched gene ontology terms of Fol4287 and MRL8996 accessory chromosomes.....33

Table 1.7 The transposable element compositions of Fol4287 and MRL8996 38

Table 2.1. The resequencing of ancestor (WT) and sequencing of the final evolved populations of Fol4287. 61

Table 2.2. The number of filtered mutations in final populations 66

Table 2.3 The total number of transposon insertion variations by the active transposable elements in the final populations.....71

Table 2.4 The resequencing of intermediate populations of P2, Y3, and M4 populations 73

Table 3.1 Whole-genome shotgun sequencing and mapping metrics for short-term evolution experiment final populations and the ancestors. 101

Table 3.2 The number of filtered mutations in the final populations 108

Table 3.3 The frequency of mutations observed in the final populations.....112

LIST OF FIGURES

Figure 1.1 MRL8996 grows faster than Fol4287 in regular growth media and is better adapted to the higher temperature.....18

Figure 1.2 Fol4287 has more tolerance to osmotic and cell wall stress than MRL8996 .19

Figure 1.3 Candidate centromere sequences of Fol4287 genome sequence.....23

Figure 1.4 Read coverage distributions of Fol4287 genome 25

Figure 1.5 Verification of the minichromosome 27

Figure 1.6 The genome assembly of MRL8996 29

Figure 1.7 Comparison of Fol4287 and MRL8996 genomes.....30

Figure 1.8 Enriched gene ontology terms of Fol4287 and MRL8996 accessory chromosomes (AC).....32

Figure 1.9 Transposon abundances in the Fol4287 and MRL8996 genomes 39

Figure 2.1. The short-term experimental evolution of Fol4287 51

Figure 2.2 TEfinder: a pipeline to detect TE insertion events 56

Figure 2.3 Chromosomal distribution of the mutation events 62

Figure 2.4 The copy number variations (CNVs) in the final populations are presented as read depth over the whole genome of Fol4287.....64

Figure 2.5 Genome size changes of final evolved populations 65

Figure 2.6 Mutation summary of final populations.....67

Figure 2.7 Allele frequency (AF) distributions of the final evolved populations 68

Figure 2.8 Histone marker levels estimated of genomic loci with variations 70

Figure 2.9 Active transposable elements in short-term evolution experiment.....72

Figure 2.10 The mutation dynamics of the plant-passaged P2 population 75

Figure 2.11 The mutation dynamics of the plate-passaged Y3 (A) and M4 (B) populations 76

Figure 2.12 The number of new mutations emerged in single passages in P2, Y3, and M4 populations 77
Figure 2.13 Mutations related to the Velvet complex 79
Figure 3.1 Short-term evolution experiments with the clinical isolate MRL8996 95
Figure 3.2 The growth rates of the final populations and the ancestors in their evolving environments 97
Figure 3.3 Phenotypes of the evolved populations in various conditions 98
Figure 3.4 The growth of the selected final populations and their ancestors in stress conditions 99
Figure 3.5 Chromosomal distribution of the mutation events 102
Figure 3.6 The read coverage distribution of MRL8996 chromosomes 103
Figure 3.7 The read coverage distribution of Fol4287 chromosomes 104
Figure 3.8 The genome size change of the evolved populations with respect to their ancestors 107
Figure 3.9 Number of SNVs and INDELs (A) and TIVs (B) in final populations 109
Figure 3.10 Distribution of allele frequencies (AF) of the mutations in the evolution experiments 110
Figure 3.11 The ratios of the mutations in coding sequences (CDS), UTR or intronic regions, 1 kb up or downstream of the genes, or more than 1 kb distance to the genes in filtered mutations 111
Figure 3.12 Total filtered number of transposon insertion events by transposable element (TE) across the final populations 112
Figure 3.13 Comparison of Foxy superfamily transposable elements Foxy3 and Foxy5 113
Figure 3.14 The normalized expressions of MRL8996 transposable elements (TEs) in CM media pH 5.0 at 28°C and in CM media pH 7.4 at 34°C 114

Figure 3.15 The allele frequency distributions of the detected variations. 115

Figure 3.16 List of Genes mutated across more than one evolved sample 117

LIST OF SYMBOLS AND ABBREVIATIONS

AC	Accessory chromosome
CC	Core chromosome
CM	Complete medium
CNV	Copy number variation
CR	Congo Red
Fol4287	*F. oxysporum* f. sp. *lycopersici* 4287
FOSC	*Fusarium oxysporum* species complex
INDEL	Small insertion or deletion
LINE	Long interspersed nuclear element
MM	Minimal medium
MRL8996	*F. oxysporum* MRL8996
PDA	Potato dextrose agar medium
PDB	Potato dextrose broth medium
SINE	Short interspersed nuclear element
SNV	Single nucleotide variation
STEE	Short term evolution experiment
TE	Transposable element
TIV	Transposon insertion variation
YPD	Yeast extract peptone dextrose medium

INTRODUCTION

Fusarium oxysporum species complex (FOSC) is a group of soil-borne filamentous fungi (classification: Ascomycota/Sordariomycetes/Hypocreales/Nectriaceae). Although they are found ubiquitously, some strains of F. oxysporum are pathogenic and can cause vascular wilt disease in many plant species as well as can infect animals and cause local or systemic mycosis [1,2].

FOSC collectively can infect more than 120 plants including banana, melon, cotton, tomato, and many other agriculturally or horticulturally important crops [3,4]. In a 2012 survey carried out by plant pathologists, *F. oxysporum* ranked 5th among scientifically and economically important plant pathogens [5]. Each plant pathogenic strain of FOSC can infect one or a few hosts. Isolates of the same host are grouped into the same *formae speciales* (f. sp.) [6]. For example, while *F. oxysporum* f. sp. *lycopersici* strains infect tomato plant, *F. oxysporum* f. sp. *cubense* strains infect banana plant.

F. oxysporum is also an opportunistic pathogen in animals [7]. It can cause localized and persistent skin or nail lesions, and infect the eye causing blindness [8–10]. In severe cases, it can cause disseminated fusariosis in immunodeficient people [11]. Its clinical importance is on the rise because of its resistance to most antifungal drugs [12–14].

The life cycle of *F. oxysporum* consists of asexual spore production in the form of microconidia, macroconidia, or chlamydospores and vegetative growth, hyphae, following the germination of spores. In contrast to its close relatives, such as *F. graminearum* or *F. solani*, *F. oxysporum* does not have a known sexual reproduction cycle. Although it lacks this important resource of genetic diversity through sexual reproduction, FOSC has adapted to many different host environments successfully [1,2].

Its wide range of host adaptability was linked to the presence of conditionally dispensable chromosomes [15,16]. Different strains of FOSC have different sets of these chromosomes, also called lineage-specific (LS), supernumerary, or accessory chromosomes (AC). The strains of the same *forma specialis*, even if they are polyphyletic, have similar LS chromosomes, suggesting horizontal transfer as their origin [15,17,18].

The genome of *F. oxysporum* contains more repeats than its close relatives *F. verticillioides* or *F. graminearum*. While only 0.14% and 0.03% of the *F. verticillioides* and *F. graminearum* genomes are transposable elements (TEs), respectively, this ratio is 3.98% for the *F. oxysporum* f. sp. *lycopersici* 4287 genome and ACs contain 74% of the total TE content while having the size of only 6% of the total genome [15].

FOSC employs many pathways to overcome plant defenses and cause disease [2]. General pathogenicity genes include mitogen-activated protein kinase (MAPK) pathway genes [19], G-protein signaling proteins [20], and Velvet complex components [21–23]. Mutations on these genes generally result in decreased pathogenicity. There are also host-specific pathogenicity genes such as toxins [24], fungal enzymes [25,26], and secreted effectors [27]. Although the pathogenicity of clinical isolates to animals is not fully understood, infection of animals by plant-pathogenic *F. oxysporum* is observed, and similar genes are found to be related to the infection of the host [22,24,28].

DNA transposons such as miniature impala elements and Helitrons are found to be associated with pathogenicity [29–31]. However, the pathogenicity-associated TEs in plant-pathogenic *F. oxysporum* were absent in a clinical isolate, and instead, different AT-rich short repeats are found [32].

Many host-specific pathogenicity-related genes are carried on ACs, and the transfer of the virulence genes carrying AC to a non-pathogenic strain results in pathogenicity, as

demonstrated in gained virulence to tomato plant of previously non-pathogen *F. oxysporum* strain by transfer of an AC [15,16].

Previously published work from the Ma Lab suggests differences both genomically and phenotypically for the plant pathogen and human-infecting *F. oxysporum* isolates [15,32]. However, efforts are needed to bring them together and test under precisely controlled conditions to investigate the cross-kingdom pathogenicity.

To understand the genetic base for the establishment and adaptation of the cross-kingdom pathogenesis of *F. oxysporum*, my dissertation will compare a plant pathogenic and a human pathogenic strain, from a genomic, phenotypic, and evolutionary perspective.

CHAPTER 1 COMPARATIVE STUDY OF TWO *F. OXYSPORUM* STRAINS WITH A PLANT OR A HUMAN HOST

1.1 Introduction

F. oxysporum presents a devastating problem for agriculture and poses serious threats to human health. To understand the mechanisms of its adaptation, comparative approaches were utilized to examine plant and animal-associated *F. oxysporum* strains in order to gain an understanding of their convergent and divergent host-adaptive strategies.

F. oxysporum f. sp. *lycopersici* strain 4297 (Fol4287), one of the best-characterized *F. oxysporum* strains, was selected as a representative of plant pathogenic *F. oxysporum* strains. The strain was originally isolated in 1984 from a diseased tomato plant (*Solanum lycopersicum)* by Dr. Javier Tello, Instituto National de Semillas y Plantas de Vivero, Spain [33,34]. The strain was deposited in multiple public strain repositories including Fungal Genetics Stock Center (FGSC 9935), NCAUR/USDA (NRRL 34936), and CBS-KNAW (CBS 123668). Since then, it has been adopted by the international *Fusarium* research community as a representative strain of the *F. oxysporum* species complex (FOSC) [1,2,5,6]

The genome of Fol4287 was first generated in 2010 using Sanger sequencing methods and an improved genome assembly incorporating PacBio and Illumina sequencing technologies was published in 2018 [15,35]. A comparative study of this reference genome with those of closely related species identified lineage-specific chromosomes (also called dispensable [16], accessory [36,37], B [38], or supernumerary [39] chromosomes) that are rich in transposons and genes related to pathogenicity.

In this milestone study, Fol4287 has been found to contain 4 accessory chromosomes (ACs) which were enriched in repeat content and genes related to host-specificity. Its genome effectively demonstrated the compartmentalization of the *F. oxysporum* genome.

Transferring these ACs between strains of *F. oxysporum* had led to converting a nonpathogenic strain into a pathogen [15,40,41]. Together, these studies shed light on the structural and functional partitioning of the *F. oxysporum* genome into core chromosomes (CCs) and accessory chromosomes (ACs), which provides a novel means of dissecting fungal pathogenesis.

The pathogenicity of Fol4287 is caused by the effectors, enzymes, and secondary metabolites. The coevolution of plant and fungus goes through an arms race between the host and the pathogen [42]. The generalized pathogen-associated molecular pattern (PAMP) trigger immunity protects the tomato plant from a non-pathogenic *F. oxysporum.* The pathogenic *F. oxysporum* inhibits PAMP-triggered immunity via effectors such as Avr2 which defines the race 1 of tomato plant virulence. However, these effectors can be recognized by leucine-rich repeat proteins I-2 and neutralize the effectors into avirulence factors leading to the effector-triggered immunity to the race 1 fungus. Finally, race 2 emerges when a different effector is acquired by the fungus which circumvents the effector-triggered immunity and causes the disease [6,43,44]. The virulence factors in the race 2 *F. oxysporum* f. sp. *lycopersici* 4287 strain are the secreted-in-xylem (SIX) genes that are located in the accessory chromosome 14 and thought to be originated through horizontal gene transfer [45].

The vascular wilt disease caused by Fol4287 and other *F. oxysporum* f. sp. *lycopersici* strains can be devastating since the tomato plant's economic importance. In 2012 45% yield losses were observed in India [46].

The Ma Lab had sequenced diverse tomato pathogenic strains tracing the evolving process from race 1 to race 2 and race 3 (unpublished). Collectively, this strain represents a good model for in-depth comparative study.

F. oxysporum MRL8996, a *Fusarium* keratitis strain, was selected as a representative of the human pathogen [47]. The strain was originally isolated in 2006 from the cornea of a patient with contact lens-associated multistate outbreak fungal keratitis at Cleveland Clinic Foundation in Ohio, USA [48]. The strain is also available at NCAUR/USDA (NRRL 47514). The strain clusters together with other human pathogenic isolates belong to the clade FOSC 3-a [34,49].

Keratitis caused by fungal pathogens is one of the major causes of cornea infections in the developing world [50] and the *Fusarium* species were reported as the most common causative agents of fungal keratitis in India [51,52], China [53,54], South Africa [55], and Brazil [56]. The first reported keratitis case of *Fusarium* was in 1964 [57] while the first reported *F. oxysporum* case was in 1989 [58]. Since then, there have been 42 papers cited with *F. oxysporum* in association with keratitis in NCBI PubMed.

Fusarium spp. were also responsible for fungal keratitis among contact lens wearers, as illustrated by the 2005/06 contact lens associated *Fusarium* keratitis outbreak [8,49,59]. An investigation led by the Centers for Disease Control (CDC) *Fusarium* investigation team reported 318 cases of *Fusarium* keratitis in the United States, of which 94% involved soft contact lens wear [47]. The outbreak was also reported in Britain, France, Hong Kong, and Singapore [60–63].

Because the pathogenic *Fusarium* spp. are resistant to many antifungal drugs, controlling these infections is challenging [64,65]. Therefore, *Fusarium* keratitis is the leading cause of blindness among fungal keratitis patients, causing over one million new cases of blindness annually patients [50,56,66]. This is why it is important to expand our understanding of this pathogen and develop new strategies to manage the disease.

Since no genomic information was available for any eye-infecting strains of *F. oxysporum*, the genome assembly for MRL8996 was produced using the same sequence technologies and computational strategies as Fol4287 strain to facilitate effective comparative study [32]. A murine model of *Fusarium* keratitis was established at our collaborator Eric Pearlman's Lab at the University of California Irvine for *in vivo* studies [9].

These two strains allowed me to compare the pathogenicity of two different host models and dissect genetic bases for their distinct host adaptation. This chapter includes *de novo* genome assemblies, comparative genomics, and phenotyping.

1.2 Methods

1.2.1 Fungal strains and growth conditions

In this study mainly two *Fusarium oxysporum* strains were used: *Solanum lycopersicum* (tomato plant) pathogenic *F. oxysporum* f. sp. *lycopersici* 4287 or NRRL 34936 provided by Antonio Di Pietro from the University of Córdoba, Spain. This strain was from the same stock that was sequenced in 2010 via Sanger sequencing [15]. FOSC 3-a strain MRL8996 or NRRL 47514 was isolated from a patient with contact lens-associated fungal keratitis at Cleveland Clinic Foundation and provided by Eric Pearlman, University of California Irvine, USA [48]. Another clinical isolate NRRL 32931 which was isolated from the blood of a leukemia patient was sometimes used as a comparison [32].

The conidia of the fungal strains were stored at -80°C ultra-freezer in 25-50% glycerol for the long term and propagated from their stocks in Potato Dextrose Broth media (PDB, BD Difco, USA) or Potato Sucrose Broth (PSB: 25% w/v boiled potato and 0.5% w/v of sucrose) for at least 4 days in a 28°C shaking incubator. The conidia were collected by filtration of the liquid fungal culture via 2-layer EMD Millipore Miracloth and washed with sterile water. The filtrate was then centrifuged, and the spores were resuspended in sterile water to the desired concentration. The spore concentrations were determined using a hemocytometer.

1.2.2 Phenotyping of isolates

1.2.2.1 Media preparation and experiment

To observe the morphological differences square petri dishes with the following growth media were prepared:

i. Yeast Extract Peptone Dextrose (YPD) Agar: 0.3% w/v yeast extract, 1% w/v Bactpeptone, 2% w/v dextrose, 1.5% w/v agar)

ii. Modified Czapek-Dox (minimal media, MM) Agar: 0.1% w/v KH_2PO_4, 0.05% w/v $MgSO_4 \cdot 7H_2O$, 0.05% w/v KCl, 0.02% v/v Trace Elements Solution, 0.2% w/v $NaNO_3$, 3% w/v sucrose, and 2% w/v Agar; Trace Elements Solution: 5% w/v citric acid, 5% w/v $ZnSO_4 \cdot 6H_2O$, 1% w/v $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$, 0.25% w/v $CuSO_4 \cdot 5H_2O$, 0.05% w/v $MnSO_4$, 0.05% w/v H_3BO_3, 0.05% w/v $Na_2MoO_4 \cdot 2H_2O$)

iii. YPD agar with 6.5% v/v 0.1 M citric acid and 43.5% v/v 0.2 M Na_2HPO_4 for pH 7.4

iv. MM agar with 5.0% v/v 0.1 M citric acid and 40.0% v/v 0.2 M Na_2HPO_4 for pH 7.4

v. PDA with 0.6 M NaCl for osmotic stress

vi. PDA with 1 mM H_2O_2 for oxidative stress

vii. YPD agar with 1 mg/ml Congo Red for cell wall stress

Isolates were cultured in PSB for 4 days at 28°C, and conidia were filtered as described above. The 2 µl of spore suspensions with 5×10^6, 5×10^5, and 5×10^4 conidia/ml concentrations were inoculated with three replicates on the same media plates. The plates were incubated at 28°C or 34°C for 3 days and photographed once a day.

1.2.2.2 Data analysis

The growth rates of each replicate and dilution were calculated as the slope of the growth curve for 3 days.

The 5-way ANOVA test with the linear model was performed for the following groups: strains, temperatures, media, dilutions, and replicates followed by the multi comparison with strains, temperatures, and media groups to identify the significant difference in MatLab [67].

To quantify the temperature adaptation, the three values of the 10^3 inoculation replicates of 34°C incubated samples were divided by the three values of the 10^3 inoculation

replicates of 28°C incubated samples (total of 9 values). A two-way student's *t*-test was performed between Fol4287 and MRL8996 in MatLab.

Similarly, to quantify the effect of stress tolerance, the three values of the 10^3 inoculation replicates of samples in media with stress component were divided by the three values of the 10^3 inoculation replicates of YPD samples (total of 9 values) followed by a two-way student's *t*-test was performed between Fol4287 and MRL8996 in MatLab.

1.2.3 Genome assembly of Fol4287

The genomic DNA extraction for Fol4287 was performed by Cristina López Díaz using the CTAB extraction method [68]. The DNA library for short reads was prepared with an average 400 bp insert size at Tufts University Core Facilities with Illumina HiSeq 2500 Platform (2×71 cycles). In addition, the genomic DNA was sequenced using the PacBio RS II system at Yale Center for Genome Analysis, USA. The raw reads were quality checked using FastQC version 0.11.5 [69]. The raw reads were deposited to NCBI SRA (under the BioProject PRJNA450629).

The initial assembly was generated via SPAdes version 3.9.1, combining raw Illumina and PacBio reads with default parameters [70]. Quiver in SMRT Analysis (version 2.2.0) was used to polish the assembly based on the PacBio reads [71]. Further polishing was performed by mapping the Illumina reads to the assembly using BWA version 0.7.12 with options `'-t 8 -a -M'` [72]. FreeBayes version v0.9.10-3-g47a713e with options `'-C 10 -p 1 -F 0.7 -u -n 1 -m 30 -q 20 --read-snp-limit 5 --min-coverage 5'` was used to identify base variants between the reads and the assembly [73]. Highly confident variant sites were used to correct the assembly using a FASTAeditWithVCF.m (available at github.com/d-ayhan/tools) which replaces the alternate allele with reference allele at a given position in the fasta file.

Structural variant (SV) callers, GRIDSS version 1.4.1 and Sniffles version 1.0.8, were used with default options to identify the SVs in the initial assembly [74,75]. All identified SVs were inspected manually to ensure accuracy. High-confidence merges/splits were integrated into the assembly. The improved assembly was then quality checked by remapping. This process was repeated until no future correction could be identified.

The assembly was compared to the 2010 reference sequence to identify the CCs and ACs chromosomes (more detail in 1.2.7). The genome was deposited to GenBank (GCA_003315725) [35].

1.2.4 Genome assembly of MRL8996

The genomic DNA of MRL8996 was extracted and sequenced using the Illumina NextSeq 500 platform in the University of Massachusetts Amherst Genomics Resource Laboratory, USA, and using the PacBio RS II platform in Yale Center for Genome Analysis, USA. The read qualities were assessed with FastQC v0.11.5 [69]. The raw reads were deposited to NCBI SRA (under the BioProject PRJNA554890).

The assembly of MRL8996 was performed in the same way as Fol4287 with the difference of scaffolding which was performed using scaffolding.m (available at github.com/d-ayhan/tools) to semi-automate the process. The final assembly was manually inspected for any scaffolding mistakes using aligned reads, and the contigs smaller than 1 kb were removed. The contigs were categorized into CCs and ACs by comparing them to the Fol4287 genome (more detail in 1.2.7). The assembly was deposited to GenBank (GCA_009746015.1) [32].

1.2.5 Read coverages and genomic features

The short reads were mapped to the assemblies via BWA mem with options `'-T 64 -M -a -v 2 -L 10'`. The alignments were cleaned, sorted, mates fixed, and duplicates

marked with Picard (version 2.0.1) and Samtools (version 1.3) [76,77]. The read coverages and the gene densities using the genome annotations in 1.2.10 were calculated with BEDtools genomecov version 2.25.0 in single-base resolution [78]. The median read coverage and gene density over 10 kb windows were calculated and plotted using a custom MatLab script. GC content is also calculated and plotted in 10 kb windows with MatLab.

The possible centromeric sequences were identified by (1) zero gene density, (2) low GC%, (3) change in the read coverage in 20 to 100 kb regions.

1.2.6 Repeat Analysis

For both genome assemblies, the repeats were *de novo* identified and classified using RepeatModeller version 1.0.11 [79]. A previously curated TE library with 69 TEs was included in the repeat sequence database. Usearch with option `'-id 0.75'` was used to cluster the sequences around centroid repeat sequences and the annotations were fixed manually [80]. RepeatMasker version 4.0.7 was used to mask and annotate the genome assemblies [81].

TECNEstimator, a bioinformatics pipeline to quantify copy numbers of the repeat sequences using WGS short reads was used to estimate read counts (available at https://github.com/d-ayhan/TECNEstimator). For Fol4287 Illumina HiSeq 2500 reads (1.2.3) were used. For MRL8996, the NovaSeq 6000 reads generated for the later study were used (A0 sample described in 3.2.3). The counts were normalized to the median read coverages of the samples. For both genomes, in a sequence cluster, only one representative sequence that has the highest copy number in the cluster was selected for downstream analyses.

1.2.7 Genome sequence comparison

Pairwise genome comparisons between the genomes with the masked repeats were performed using MUMmer version 3.22 nucmer with minimum alignment length 500 bp and otherwise default options [82]. The total alignment lengths are calculated by summation of aligned reference sequences in the MUMmer output. Average sequence identity is calculated by the following formula: $I_W = \sum_s L_i \times I_i / \sum_s L_i$ where I_W is weighted sequence identity, s is aligned sequences, L_i is the length of aligned reference sequence, and I_i is the percent sequence identity of aligned sequences.

Fol4287 assembly was aligned to reference Fol4287 genome and the contigs belong to CCs were placed in scaffolds and the contigs belong to ACs and unmapped scaffolds in the reference genome were labeled accordingly. MRL8996 assembly was aligned to Fol4287 genome and contigs belongs to CCs were oriented and labeled accordingly. Contigs aligned to ACs or unmapped sequences were categorized as AC.

The alignments of Supercontig_1.20 of NRRL 32931 and MRL8996 S008 and S027 contigs were performed by Mauve (version 20150226) [83].

1.2.8 Chromosome separation using contour-clamped homogeneous electric field electrophoresis

10^8 conidia of Fol4287, MRL8996, and NRRL 32931 strains were inoculated in 40 ml YPD broth media and incubated for 7 hours at 28°C, 80 rpm. Mycelia were harvested in EMD Millipore Miracloth by centrifuge at 4°C at 4000 rpm for 5 min. It was washed with first sterile water, then with 1.2 M KCl, and finally resuspended in the protoplasting solution (25 mg/ml Driselease, 5 mg/ml Lysing Enzyme, and 1.2 M KCl) and incubated overnight at 30°C, 80 rpm. Protoplasts were centrifuged at 4°C, 2100 rpm, for 15 min and resuspended in 10 ml STC buffer (1 M sorbitol, 10 mM Tris-HCl pH 7.4, and 10 mM $CaCl_2$)

slowly to wash and centrifuged again. The supernatant was removed by pouring out and the protoplasts were resuspended in STE buffer (2% SDS, 0.5 M Tris pH 8.1, and 10 mM EDTA) to 2×10^8 conidia/ml and incubated in 50°C for 10 min and mixed 1:1 ratio with 1.2% w/v BioRad Certified Low Melt Agarose. The mixture was transferred to CHEF Disposable Plug Molds and stored at 4°C for the long term.

The contour-clamped homogeneous electric field (CHEF) electrophoresis was performed by Shunsuke Kotera at Tokyo University of Agriculture and Technology, Japan. The CHEF electrophoresis was run for 260 h in 0.5× TBE (Tris/Borate/EDTA) buffer with 1.0% agarose gel at 1.5 V, 4°C.

1.2.9 Minichromosome confirmation in Fol4287

CHEF electrophoresis was performed using Fol4287 by Lucía Gómez Gil and Cristina López Díaz at the University of Córdoba, Spain. The smallest band around 1 Mb size was excised and genomic DNA was re-isolated. The paired-end sequencing was performed by Novogene Company Limited, UK via Illumina NovaSeq 6000 platform with 2×151 cycles. The reads were mapped to Fol4287 assembly via BWA mem version 0.7.15 with options `'-t 8 -M -a -v 1'` [72]. The alignments were cleaned, sorted, mates fixed, and duplicates marked with Picard version 2.0.1 and Samtools version 1.4.1 [76,77]. The read coverages were calculated at single-base resolution via BEDtools version 2.26.0 genomecov. Coverages over 10 kb windows were calculated and normalized first by the overall median coverage then the background was subtracted using normalized Fol4287 whole-genome sequencing read coverage.

To get a complete sequence of the minichromosome, the sequences with high read mapping were joined into a single pseudochromosome and PacBio reads of Fol4287 whole-genome sequencing mapped to that reference sequence via BWA bwasw version

0.7.15 with default options [84]. The mapped reads, then, extracted and assembled via Canu version 1.5 with options ‘genomeSize=1m useGrid=false correctedErrorRate=0.105’ [85]. Minichromosome sequence was reconstructed from the assembly contigs and the contigs from the Fol4287 genome assembly as shown in **Figure 1.5**C. 100 bp gaps added between continuous sections. When sequences overlapped, Fol4287 genome sequences were used, instead of the new contig sequences.

Repeat sequences were identified using RepeatMasker version 4.0.7 and the repeats database generated above and the transposons shorter than 100 bp were filtered out. The same filters were applied for the whole WT genome.

To check the SNPs that might have occurred during the emergence of the minichromosome, GATK Mutect2 and FilterMutectCalls (version 4.1.4.1) were used [86].

1.2.10 Gene annotations and gene ontology term enrichment analysis

The MRL8996 genome was annotated by US DOE Joint Genome Institute with the transcriptome evidence provided by us (see Methods 2.2.5 for detail of RNA sequencing) [87,88].

Because there were minimal changes in gene contents of the Fol4287 genome that is recently sequenced and the reference genome that was sequenced in 2010 [15], the annotations were lifted from the reference genome with the assembly accession ASM14995v2 using the Genome Remapping Service in NCBI (available at https://www.ncbi.nlm.nih.gov/genome/tools/remap) to map the previous annotations to the new assembly.

The AC genes were identified as the genes located in AC contigs of MRL8996 and AC and unmapped contigs of Fol4287.

The orthologous genes were identified using Orthofinder version 2.3.3 with default options [89]. Gene ontology (GO) terms of Fol4287 and MRL8996 genomes were downloaded from JGI Mycocosm (JGI specific genome identifiers: Fusox2 and FoxMRL8996, respectively) [87]. GO_term_enrichment_analysis.m script (available at https://github.com/d-ayhan/tools) which utilizes the hypergeometric cumulative distribution function was used to analyze the GO-term enrichments in ACs.

1.3 Results

1.3.1 Morphological characterization of fungal isolates

1.3.1.1 The Fusarium keratitis strain MRL8996 is better adapted to an elevated temperature

The two selected *F. oxysporum* strains, plant pathogenic Fol4287 and keratitis isolate MRL8996, were grown under the same conditions using either Minimal Media (MM) or complete medium Yeast Peptone Dextrose media (YPD). In addition to the native medium pH (pH ~5), both medium types were adjusted with citric acid-sodium phosphate buffer to pH 7.4, the physical pH of human blood. These 4 different type plates were further tested under two temperature treatments at 28°C the optimal laboratory incubation temperature for *F. oxysporum* or 34°C, to mimic the temperature of the human cornea (**Figure 1.1**A) [90].

MRL8996 has a significantly higher growth rate than Fol4287 in each tested condition at 28°C and 34°C (**Figure 1.1**B). While both strains have inhibited growth at an elevated temperature, the inhibition observed in MRL8996 is less than in Fol4287 (**Figure 1.1**C). MRL8996 grows at 73.9%, 73%, 43.1%, 76.7% of the growth rates at 28°C in MM, MM pH7.4, YPD, and YPD pH7.4 at 34°C; while Fol4287 grows at 20.2%, 28.5%, 32.3%, and 51.0% of its growth rate at 28°C, respectively.

Clearly, the human pathogenic strain MRL8996 grows faster than 4287 in both 28°C and 34°C and it is better adapted to an elevated temperature. The same behavior was observed in a separate experiment (Appendix B1).

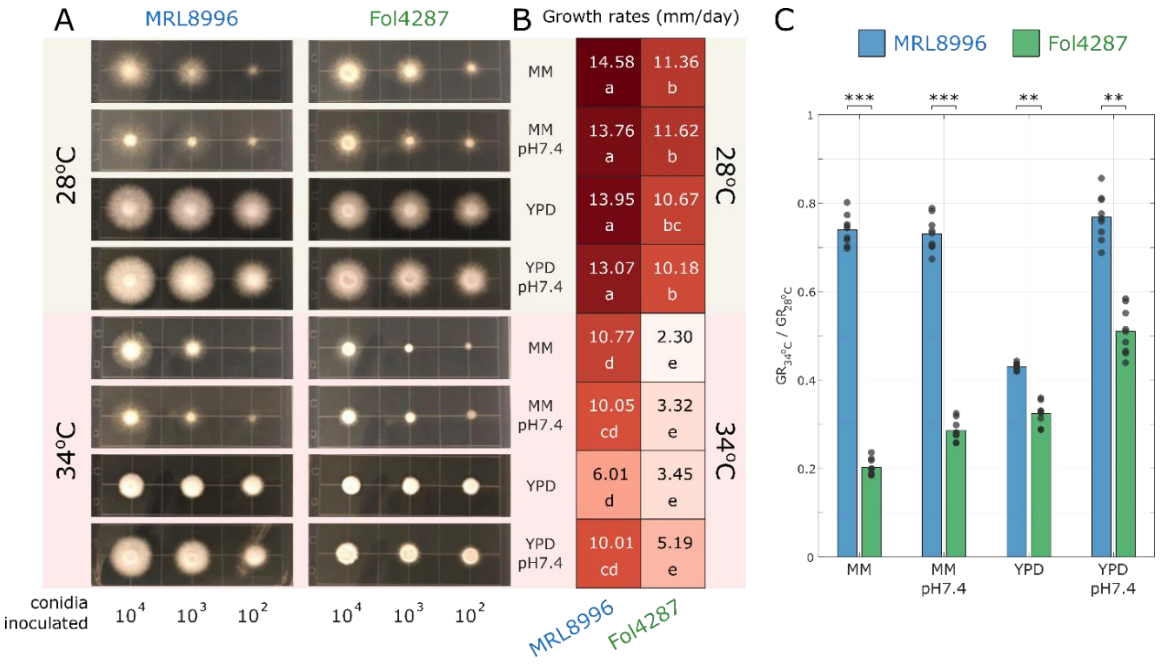


Figure 1.1 MRL8996 grows faster than Fol4287 in regular growth media and is better adapted to the higher temperature. (A) The colony morphologies of MRL8996 and Fol4287 on minimal media (MM, modified Czapek-Dox Agar, pH ~5), MM with pH adjusted to 7.4 with citric acid-sodium phosphate buffer (MM pH7.4), Yeast Extract Dextrose Peptone Agar (YPD, pH ~5), or in YPD with pH adjusted to 7.4 with citric acid-sodium phosphate buffer (YPD pH7.4). The inoculated conidia counts, from left to right: 10^4, 10^3, 10^2. Plates were incubated at 28°C (yellow background) or 34°C (pink background). The pictures are representative of three replicates and were taken at 2-day post-inoculation. (B) The hyphal growth rates of the MRL8996 (left) and Fol4287 (right) colonies were calculated as mm per day. The numbers are average growth rates over three days of three replicates of the 10^3 inoculations. The letters represent the significant differences when analyzed with N-way ANOVA using strains, temperatures, media, dilutions, and replicates groups. The *p*-values with the linear model for the groups are 3.0×10^{-35}, 2.0×10^{-54}, 6.7×10^{-3}, 0.26, and 0.51, respectively. The same letter represents no significant difference of variance. (C) The ratios of the growth rates at 34°C and 28°C ($GR_{34°C}/GR_{28°C}$) as a representation of temperature adaptation. Black data points indicate each growth rate value at 34°C normalized to each growth rate value at 28°C, and the bar graph shows the mean values. ** $10^{-9} < p\text{-value} < 10^{-5}$; *** *p*-value < 10^{-9}, calculated by two-sample *t*-test.

1.3.1.2 The phytopathogenic Fol4287 is more tolerant to osmatic and cell wall stress conditions

To test if the two strains behave different in some stress conditions, their hyphal growth rates were also measured under high salt stress (0.6 M of NaCl), oxidative stress (1 mM H_2O_2), and cell wall stress (1 mg/mL Congo Red) and compared to YPD, base media for cell wall stress and the closest tested medium for others (**Figure 1.2**). Similar to the observation in MM and YPD media, both fungal strains grow slower under elevated

temperature at 34°C (**Figure 1.2**B) and the keratitis strain had better adaptation to elevated temperature (**Figure 1.2**D).

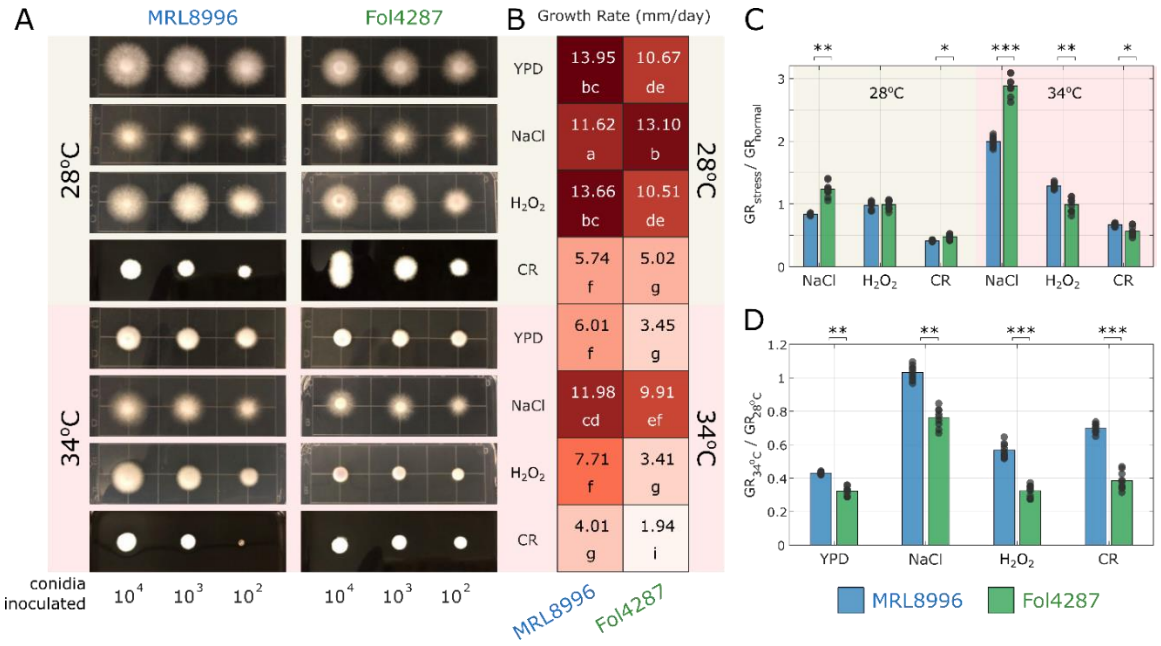


Figure 1.2 Fol4287 has more tolerance to osmotic and cell wall stress than MRL8996. (A) The colony morphologies of MRL8996 and Fol4287 on YPD, Potato Dextrose Agar (PDA) with 0.6 M Sodium Chloride for osmotic stress (NaCl), PDA with 1 mM Hydrogen Peroxide for oxidative stress (H_2O_2), and YPD with 1 mg/mL Congo Red for cell wall stress (CR). The inoculated conidia counts from left to right: 10^4, 10^3, 10^2. Plates were incubated at 28°C (yellow background) or 34°C (pink background). The pictures are representative of three replicates and were taken at 2-day post-inoculation. (B) The hyphal growth rates of the MRL8996 (left) and Fol4287 (right) colonies were calculated as mm per day. The numbers are average growth rates over three days of three replicates of the 10^3 inoculations. The letters represent the significant differences when analyzed with N-way ANOVA using strains, temperatures, media, dilutions, and replicates groups. The *p*-values with the linear model for the groups are 1.7×10^{-12}, 3.3×10^{-30}, 2.3×10^{-30}, 0.5, and 0.8, respectively. The same letter represents no significant difference of variance. (C) The ratios of mean rates of growth in stress conditions (GR_{stress}) and in YPD (GR_{normal}) as a representation of stress tolerance at 28°C (yellow background), and 34°C (pink background). Black data points indicate each growth rate value in a stress condition normalized to each growth rate value in YPD, and the bar graph shows the mean values. (D) The ratios of the growth rates at 34°C and 28°C ($GR_{34°C}/GR_{28°C}$) as a representation of temperature adaptation. Black data points indicate each growth rate value at 34°C normalized to each growth rate value at 28°C, and the bar graph shows the mean values. * *p*-value < 0.05; ** *p*-value < 10^{-5}; *** *p*-value < 10^{-9}, calculated by two-sample *t*-test.

When encountered with stress, the two strains display complex behaviors. Fol4287 grows faster in the osmatic stress condition than in the complete medium condition. This effect increases when it is coupled with temperature (**Figure 1.2**B-C).

While MR8996 has 16.7%, 2.12%, and 58.8% reduction in stress conditions with respect to YPD, Fol4287 had a 22.7% increase in osmotic stress and smaller growth reduction in oxidative and cell wall stresses (1.5% and 52.9%, respectively) (**Figure 1.2**B-C).

It is a 99.4% and 28.4% increase for MRL8996 in osmatic and oxidative stresses, respectively, and only a 33.3% reduction for cell wall stress when compared to growth in YPD. For Fol4287 the growth increase in osmatic stress is much more pronounced with a 187.6% increase. And while oxidative stress does not change the growth rate than YPD at 34°C, cell wall stress reduces it by 43.8%.

The three stress conditions reduce the temperature stress effect that was observed in YPD as MRL8996's growth rate increased by 3.1% in osmotic stress and reduced only 43.5% and 30.3% in oxidative and cell wall stresses, respectively; Fol4287's growth rate was only reduced by 24.3% in osmatic stress in 34°C. On the other hand, oxidative and cell wall stresses do not have an additional effect on the growth inhibition in Fol4287 (67.5% and 61.4% growth reduction, respectively) (**Figure 1.2**D).

Collectively the measurements of growth rates suggest that MRL8996 is better adapted in elevated temperatures and has a faster growth rate in complete medium and nutrient-limited conditions while Fol4287 tolerates osmatic and cell wall stress conditions better than MRL8996. This observation may reflect the unique adaption of the human pathogen to the host defense associated with mammalian endothermy and homeothermy and the plant pathogen to diverse environmental stresses.

1.3.2 Genome Assemblies

1.3.2.1 Improved assembly of Fol4287

Newly created genome assembly with improved base quality and assembly continuity. The reference genome Fol4287 was initially assembled to 6× read depth using Sanger

sequencing (hereafter will be called 2010 assembly) [15]. Low sequence coverage had resulted in a high error rate at the nucleotide level. Coupled with the intrinsic dynamicity of the *F. oxysporum* genome, it was necessary that an improved assembly had to be generated combining Illumina and PacBio technologies (GenBank accession: QESU00000000) [35]. The final assembly is 53.9 Mb, with 499 contigs and an N50 value of 1.3 Mb. The largest contig size is 5.7 Mb and the GC content is 47.7% (**Table 1.1**). Overall, the new assembly decreased the number of contigs, substantially increased the N50, and improved the base quality with high short read coverage.

Table 1.1 The assembly statistics for Fol4287 and MRL8996 genome assemblies.

	Fol4287	*Fol4287*	*MRL8996*
Reference	Ma et al., 2010	Ayhan et al., 2018	Zhang et al., 2019
Sequencing Platform	Sanger	Illumina HiSeq 2500 & PacBio RS II	Illumina NextSeq 500 & PacBio RS II
Median read depth	6	66 + 10	171
Assembly size	60,011,252 bp	53,912,367 bp	50,113,741 bp
Number of contigs	1,371	499	251
Largest contig size	439,130 bp	5,733,288 bp	4,647,901 bp
Smallest contig size	706 bp	515 bp	1,007 bp
N50	95,416 bp	1,338,693 bp	1,727,337 bp
N90	22,978 bp	49,310 bp	129,585 bp
L50	184	11	10
%GC	48.4	47.7	47.99
GenBank Accession	AAXH01000000	QESU00000000	VLJC00000000
Estimated Genome Size	59.9 Mb	59,980,258	53,123,701

By comparing the new assembly to the 2010 assembly of *F. oxysporum* (GenBank accession: AAXH01000000), the contigs that belong to each chromosome were identified, ordered, and oriented within each chromosome. The contigs were divided into three categories, including 11 core chromosomes (CC), 4 accessory chromosomes (AC), and some unmapped contigs (U). For this study and the following chapters, the duplicate contigs were disregarded. Except for the AC sequences of chromosomes 1 and 2, CCs were placed in scaffolds. The resulting genome sequence contains 406 sequences of a total 52.5 Mb length.

The mitochondrial chromosome was assembled as a single contig of 52,424 bp. The circularity has been confirmed by the read pair orientation of LR that maps at the edges of the contig. This contig had 1800× read coverage which is 27-fold of the nuclear chromosomes.

The copies of rDNA were attached between the core and AC regions of chromosome 2. Here, the ribosomal RNA sequences were collapsed into a single contig of 7,875 bp size with about 10000× read coverage. It can be estimated that there are about 150 copies of this sequence in the genome. The LR pair orientation suggests that the sequences are consecutive.

The telomeric repeats ([CCCTAA]×n) were also collapsed in a single sequence of 50 bp strips with coverage of about 5000×. This sequence was attached to a subtelomeric sequence of about 700 bp with 2400× read coverage (36 copies). The linkages to this contig have been found at the right end of chromosome 7.

Another feature captured by this sequencing data was the candidate centromeric sequences in CCs (**Table 1.2** & **Figure 1.3**). Also seen as small peaks in the read coverage, the centromeric sequences are AT-rich regions with no annotated genes. The average length of the centromeric sequences with no gaps in the assembly is 50 kb, and the average GC ratio is 21.6%. Because of highly fragmented sequences of the LS regions and many copy number variations, LS chromosomes centromeric sequences are not detected.

Table 1.2 Candidate centromeric sequences of Fol4287. * If there are gaps in the chromosome sequences, the last column is 'Yes'.

Chromosome	*Start*	*End*	*Length*	*GC%*	*Gaps**
Chr1	1,605,845	1,651,984	46,139	20.19	No
Chr2	2,599,425	2,649,237	49,812	20.29	No
Chr4	2,356,316	2,400,955	44,639	20.47	Yes
Chr5	878,411	927,246	48,835	22.83	No
Chr7	2,683,990	2,718,030	34,040	20.83	Yes
Chr8	3,309,827	3,325,198	15,371	19.77	Yes
Chr9	1,991,880	2,030,636	38,756	20.76	Yes
Chr10	2,893,899	2,921,279	27,380	29.70	Yes
Chr11	462,415	499,782	37,367	18.49	No
Chr12	1,303,134	1,338,290	35,156	21.13	Yes
Chr13	1,366,212	1,434,440	68,228	23.31	No

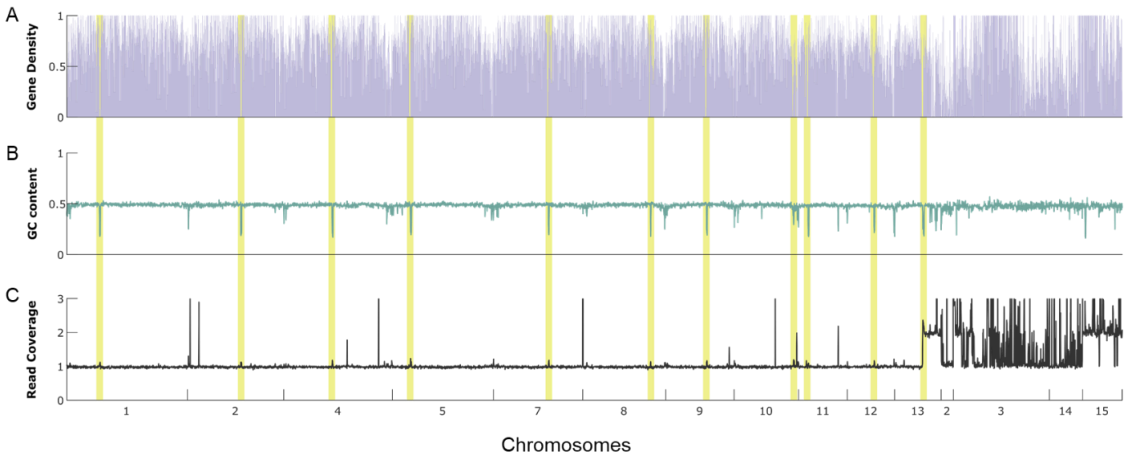


Figure 1.3 Candidate centromere sequences of Fol4287 genome sequence. The candidate centromeric sequences are noted in yellow and are detected using low gene density (A), low GC ratio (B), and slightly increased short read coverage (C). The genomic features were averaged in 10 kb windows.

1.3.2.1.1 Large segmental duplication and deletion events in Fol4287 genome

To compare the copy number variations (CNVs) between the 2010 assembly and the Fol4287 strain that was used in this study, the paired-end reads were mapped to reference genome sequence (**Figure 1.4**A-B). The CCs were unchanged except for the small arm of chromosome 13 which is duplicated in the new genome. The detailed analysis for this portion is explained below (1.3.2.1.2). The ACs, however, showed a high number of CNVs. There are small deletions in chromosomes 14 and 15 and large segmental deletions and duplications in chromosomes 3 and 6. In the 2010 genome assembly, these two chromosomes were found to be recently duplicated [15]. However, this event is not present in the genome of the strain that has been recently sequenced, instead, there is a 1 Mb region that is unique to chromosome 3 and duplicated.

When the paired-end reads are mapped to the new assembly, the read coverage distribution has one large peak that corresponds to single-copy sequences in the genome and a second small peak that represents the two-copy regions (**Figure 1.4**C). There are also three or more copy number contigs are also present in the genome (**Figure 1.4**D). When the sequence duplications are considered the estimated genome size for the Fol4287 genome increases to 59,980,258 bp which is comparable to the reported genome size in the 2010 assembly [15].

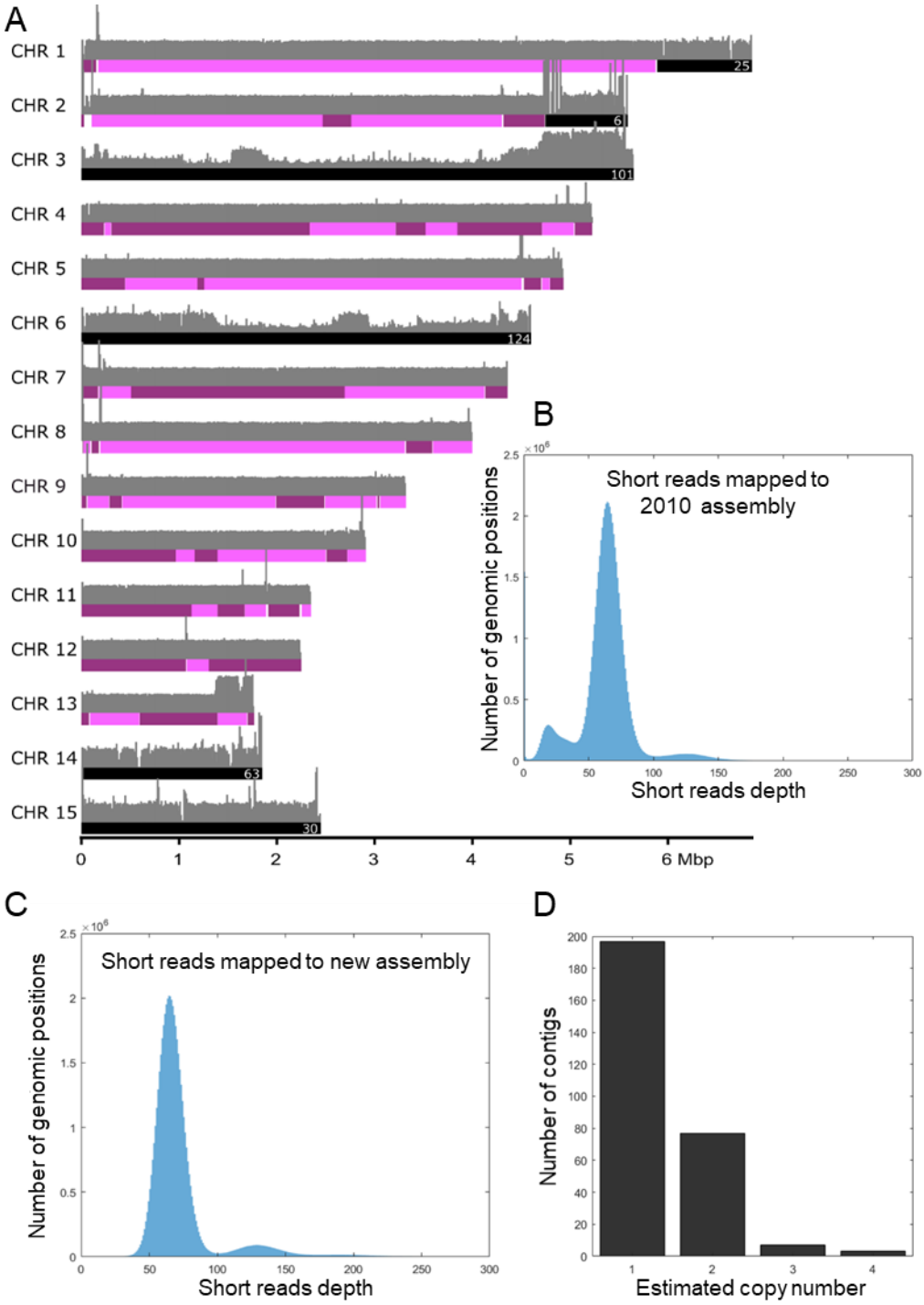


Figure 1.4 Read coverage distributions of Fol4287 genome. (A) Assembled contigs mapped to the 2010 assembly. For each chromosome, the gray plot shows the coverages of reads mapped to the 2010 genome sequence; the horizontal bars below show core (light/dark purple) or accessory (black) contigs of the new assembly of Fol4287 genome mapped to the reference genome. The numbers of contigs included in accessory regions are printed on the bars. (B) Histogram of the read coverages in A. (C) Distribution of the read coverages mapped to the new assembly. (D) The number of contigs with estimated copy numbers.

1.3.2.1.2 Recent chromosome duplication confirmed by single chromosome sequencing.

Coverage of the resequencing reads indicated a duplication of the small arm of chromosome 13 including the centromeric sequence in Fol4287 (**Figure 1.4**A). To test if this duplication was attached to a chromosome or not, our collaborators Di Pietro Lab in the University of Córdoba, Spain performed pulsed-field gel electrophoresis and confirmed the presence of a minichromosome with a size of 1 Mb (**Figure 1.5**A). The 1 Mb band was excised, DNA was isolated and sequenced via Illumina NovaSeq 6000. The sequencing produced 23,446,092 total paired reads with 150 bp length. The reads were mapped to the WT genome and had 10.5× median coverage but with enrichment in the small arm of chromosome 13, as well as the unmapped contigs U_2, and U_3 (**Figure 1.5**B).

The minichromosome was assembled separately using the PacBio reads of the Fol4287 whole-genome sequencing mapped to the sequences selected in **Figure 1.5**B. The assembly has 13 contigs with a total size of 975,569 bases and 12 of them have expected read coverage. The assembled 6 contigs align to the small arm of chromosome 13 (Chr13p), 4 contigs to U_2 and U_3; two contigs have subtelomeric sequences; and one has rRNA sequence (**Figure 1.5**C). Using the orientation shown in **Figure 1.5**C, the minichromosome sequence was constructed. There were no variations observed in the non-repeat sequences of minichromosome and Chr13p, making the two copies indifferentiable.

There are 318 annotated genes in minichromosome with enrichment in lipid catabolic processes (4 genes, *p*-value: 8.2×10^{-4}), amino acid transport (2 genes, *p*-value: 2.1×10^{-3}), transmembrane transport (25 genes, *p*-value: 6.1×10^{-3}), and anion transport (4 genes, *p*-value: 6.7×10^{-3}). In addition, retrotransposons Skippy are found in the minichromosome 5 times more than they are found in other places in the genome.

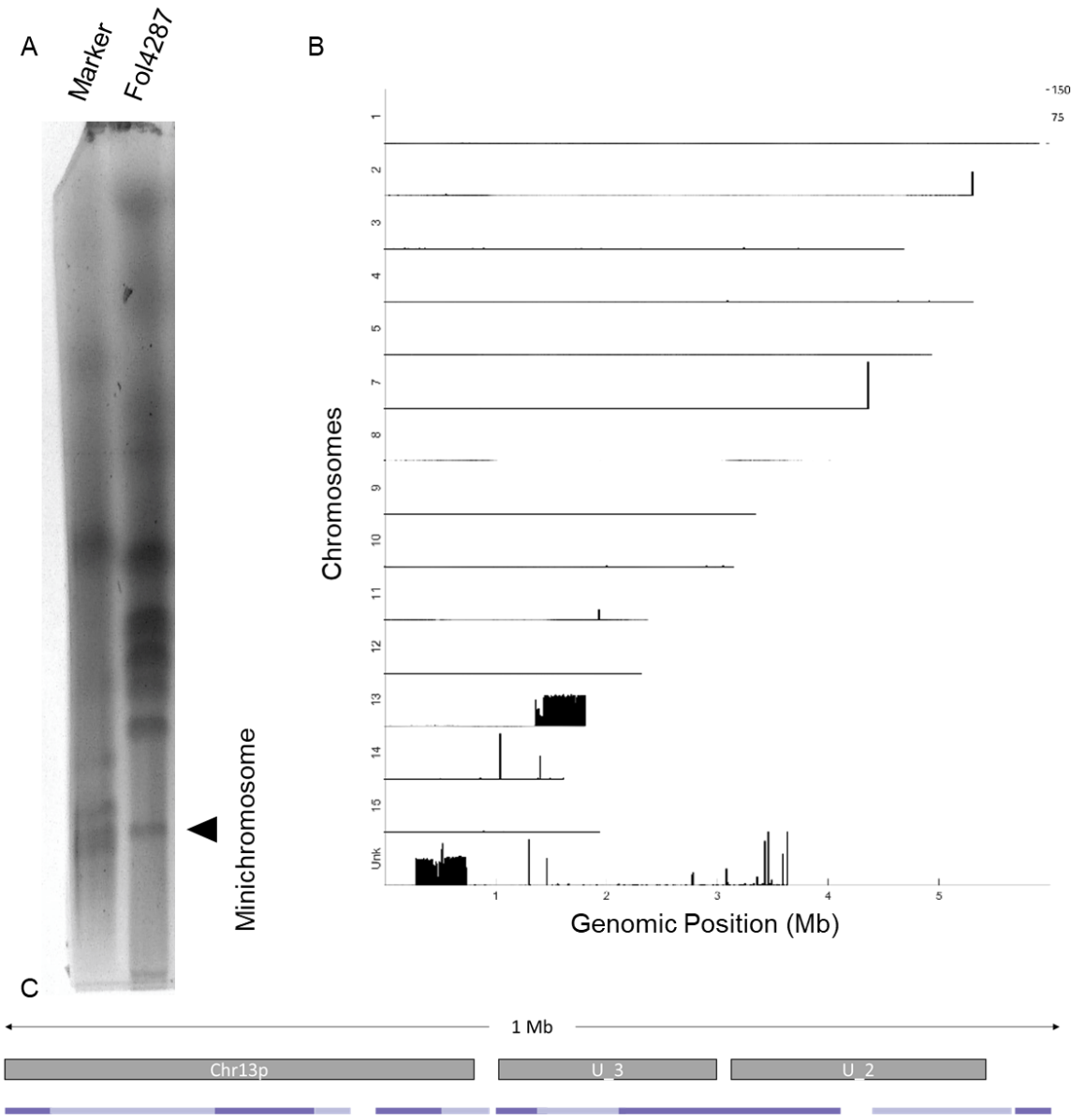


Figure 1.5 Verification of the minichromosome. (A) Pulsed field gel electrophoresis image for Fol4287 performed by Cristina López Díaz. The arrowhead points to the minichromosome which was cut from the gel and the re-isolated DNA was sequenced. (B) The normalized coverage of the sequencing reads mapped to the Fol4287 assembly. Y-axis labels are the chromosome numbers. The enrichment in the sequences corresponds to, the small arm of chromosome 13 (Chr13p), unmapped U_2 and U_3 contigs, rRNA sequence in chromosome 2; subtelomeric sequence in chromosome 7, Skippy transposon in chromosome 11, and subtelomeric sequence and Skippy transposon in chromosome 14. (C) The representation of the contig arrangements for the minichromosome assembly. Grey boxes are the scaffolds from Fol4287 genome assembly, and the alternating dark and light purple lines show the contigs of the assembled enriched PacBio reads (see main text for details).

1.3.2.2 de novo genome assembly of MRL8996

The MRL8996 (or NRRL 47514) assembly was generated using PacBio and Illumina reads. PacBio RSII platform was used to generate 113,678 raw sequences between 62,168 to 35 bp in length. Total of 84,766,484 paired-end reads with length 151 bp generated by Illumina NextSeq 500. The assembled genome size is 50.11 Mb with 252 contigs and an N50 of 1.73 Mb, 2 Mb smaller than Fol4287 (**Table 1.1**) [32].

Unlike Fol4287, the coverage distribution of MRL8996 has a single peak, even though there are many contigs with more than 1× coverage, primarily in short AC sequences (**Figure 1.6**). The largest sequence with 2× coverage has a size of 113 kb. Clearly, the MRL8996 genome lacks large segmental duplications which are observed in the Fol4287 genome. Once calibrated with sequence coverage, the estimated MRL8996 genome size is 53.1 Mb, 6.9 Mb smaller than the Fol4287 genome.

The mitochondrial chromosome of MRL8996 is assembled in a single contig of 48.6 kb size with 88× coverage. Pseudo chromosomes were assigned based on the alignments to the Fol4287 assembly (**Figure 1.6**C). The centromeric sequences were predicted by the drop in GC%, increase in transposon density, and the length of the sequence (**Table 1.3**). The average centromeric sequence length of the MRL8996 genome is 36.4 kb. A fragment of it with 6.5 kb size is detected in the pseudo chromosome 2 (mtDNA: 42156–48644 in C02.2: 42156–48644). The end of the same contig also includes about 8.4 kb ribosomal DNA repeats with 100× coverage (C02.2: 4575509–4583909).

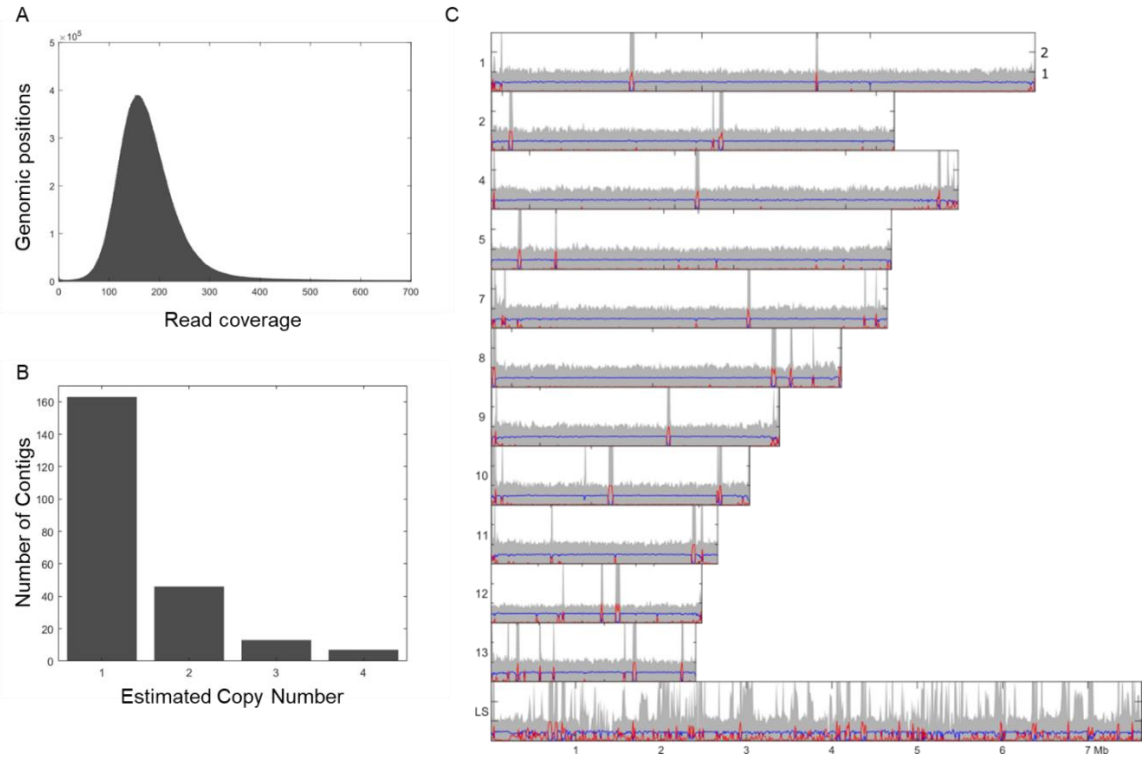


Figure 1.6 The genome assembly of MRL8996. (A) The read coverage distribution. (B) Histogram of estimated their copy numbers of the contigs. (C) Read coverage (gray histograms), GC content (blue lines), and repeat distribution (red lines) of the assembly. The chromosomes are constructed according to Fol4287 chromosomes. LS: lineage-specific or accessory chromosome sequences.

Table 1.3 The predicted centromere sequences of MRL8996 genome.

Chromosome	*Position [sequence: start–end]*	*Length (bp)*
Pseudo Chr1	C01.1: 1,626,826–1,675,309	48,484
Pseudo Chr2	C02.2: 2,531,568–2,576,358	44,791
Pseudo Chr4	C04.4: 1–37,222	37,222
Pseudo Chr5	C05.1: 752,297–769,654	17,358
Pseudo Chr7	C07.1: 3,004,341–3,033,730	29,390
Pseudo Chr8	C08.3: 1,386,996–1,435,613	48,618
Pseudo Chr9	C09.2: 1,483,529–1,519,045	35,517
Pseudo Chr10	C10.2: 194,437–244,467	50,031
Pseudo Chr11	C11.2: 1,643,162–1,680,227	37,066
Pseudo Chr12	C12.3: 388,940–435,982	47,043
Pseudo Chr13	C13.4: 330,064–365,623	35,560

1.3.3 Genome comparisons

1.3.3.1 Structural comparison

The chromosomes of MRL8996 and Fol4287 were separated by contour-clamped homogeneous electric field (CHEF) gel and at least 12 and 10 distinct chromosomes were identified, respectively (**Figure 1.7**A). The number of chromosomes and their sizes differ in two strains as seen in the genome assemblies (**Table 1.4**).

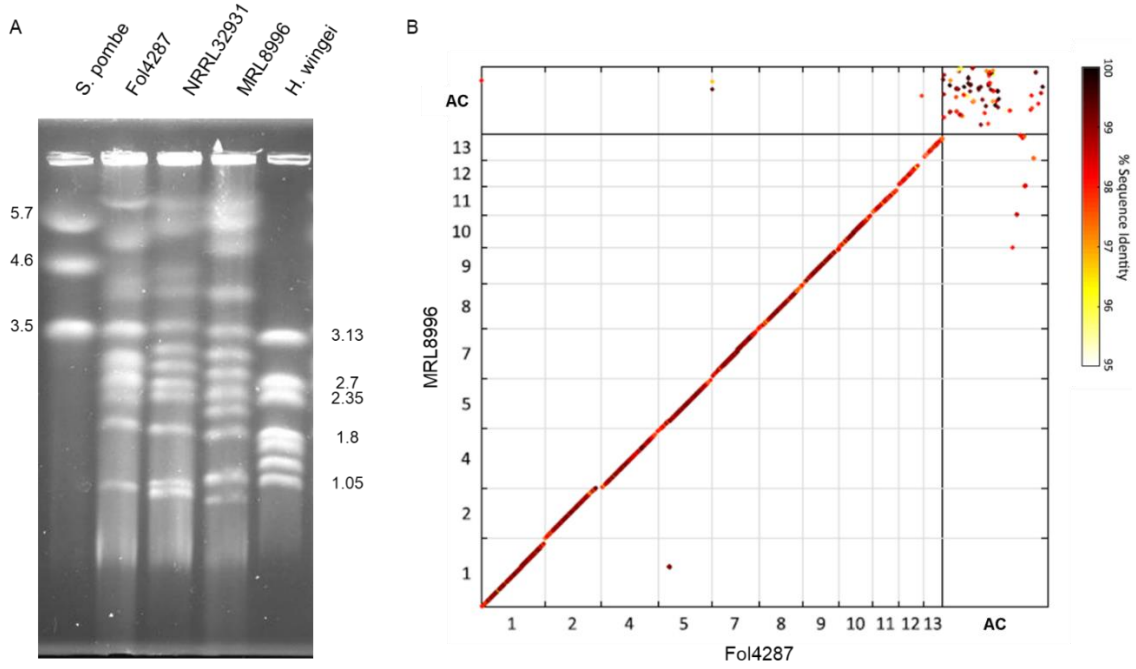


Figure 1.7 Comparison of Fol4287 and MRL8996 genomes. (A) Chromosomes of Fol4287, MRL8996, and NRRL32931, another *F. oxysporum* clinical isolate, are separated by CHEF gel. The band sizes of the two markers are labeled in Mb. The experiment was performed in collaboration with Shunsuke Kotera at Tokyo University of Agriculture and Technology, Japan. (B) Pairwise genome comparison of MRL8996 and Fol4287. The accessory chromosome (AC) sequences are combined in one section.

Even though there are differences in the chromosome sizes, pairwise sequence analysis showed that core chromosomes were shared between two strains. AC sequences, however, are highly diverged (**Figure 1.7**B). The sequence identity of the shared LS sequences is also slightly lower than the shared core sequences (**Table 1.5**). Similarly, the telomeric sequences of core chromosomes and the smallest 3 core chromosomes

(chromosomes 11, 12, and 13) have lower sequence identity than non-subtelomeric potions of the other core chromosomes.

Table 1.4 The chromosome sizes (bp) in the Fol4287 and MRL8996 genome assemblies. Minichromosome is the size of the constructed minichromosome as mentioned in 1.3.2.1.2. 'Chr 2 AC' is the accessory region attached to chromosome 2 which also includes rDNA contig. Since the accessory chromosomes of MRL8996 are highly fragmented, the total size of the accessory sequences is given.

		Fol4287	*MRL8996*
Core	Chr 1	5,906,624	6,346,428
	Chr 2	4,706,971	4,721,865
	Chr 4	5,308,927	5,329,564
	Chr 5	4,935,765	4,671,957
	Chr 7	4,366,282	4,647,901
	Chr 8	4,044,692	4,092,967
	Chr 9	3,342,190	3,376,480
	Chr 10	3,145,249	3,019,420
	Chr 11	2,375,198	2,653,038
	Chr 12	2,312,736	2,434,747
	Chr 13	2,253,417	2,380,455
	Minichromosome	975,569	
Accessory	Chr 2 AC	556,119	6,396,764
	Chr 3/6	3,953,562	
	Chr 14	1,275,783	
	Chr 15	1,614,409	
	Unknown	2,437,670	
Mitochondrial	mtDNA	52,424	48,644

Table 1.5 Alignment metrics of the pair-wise comparisons in Figure 1.7B. Size: the alignment size of the masked genomes in bp. Seq ID%: the weighted sequence identity percentage of the alignments. CC: core chromosomes. AC: accessory chromosomes.

	Size (bp)	*Seq ID %*
All	23,766,897	98.78
CC	23,410,155	98.79
AC	356,742	98.01

Another clinical isolate, NRRL32931, was included in this analysis to emphasize the differences between plant pathogenic and human pathogenic *F. oxysporum* (Appendix B2). Two clinical isolates share 17% of MRL8996 AC sequences while only 7% are shared between Fol4287 and MRL8996. Likewise, while 37% of AC sequences of NRRL 32931 genomes have homologous sequences in the MRL8996 genome, there is only 2.3% in

the Fol4287 genome. Furthermore, the sequence identity between the clinical isolates is higher in core chromosomes than between a clinical isolate and Fol4287 reflecting the evolutionary closeness of MRL8996 and NRRL32931 [32]. Most interestingly, the two human pathogenic strains share almost identical fragments in AC sequences (Appendix B2).

1.3.3.2 Comparison based on GO functional annotation

The genome annotations of Fol4287 were lifted from the 2010 assembly while the MRL8996 genome was annotated *de novo* by Joint Genome Institute [87,88]. The number of AC genes in Fol4287 and MRL8996 genomes are 4969 and 2017, respectively.The genes in AC sequences were identified and the enriched terms in their gene ontology (GO) were analyzed (**Figure 1.8** &**Table 1.6**). Through this document, for Fol4287 genes, the locus tag IDs starting with 'FOXG_' are used, while for MRL8996, 'protein_id' numbers assigned by JGI were used.

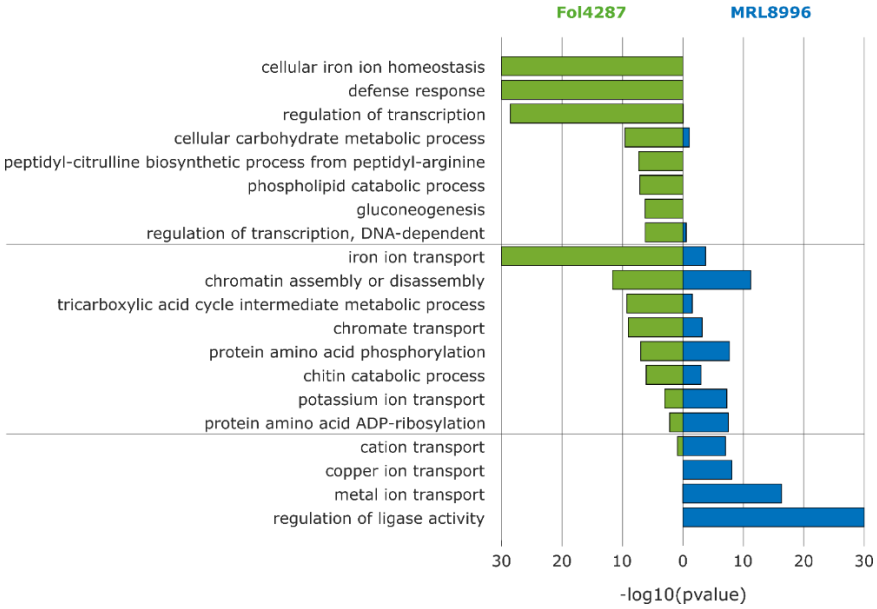


Figure 1.8 Enriched gene ontology terms of Fol4287 and MRL8996 accessory chromosomes (AC). Top panel: terms enriched only in Fol4287 ACs, middle panel: terms enriched in ACs of both genomes, bottom panel: terms enriched only in MRL8996 ACs. For *p*-values = 0, 'log10(p-values)' is set to 30. Only GO-terms with gene counts > 1 are considered.

Table 1.6 Enriched gene ontology terms of Fol4287 and MRL8996 accessory chromosomes. The yellow-colored background indicates significant enrichment.

		Fol4287			***MRL8996***		
	GO term description	*p-value*	*LS*	*All*	*p-value*	*LS*	*All*
gene expression regulation	regulation of transcription	2.79E-29	83	186	8.06E-01	1	53
	regulation of transcription, DNA-dependent	6.00E-07	221	1342	2.83E-01	43	718
	chromatin assembly or disassembly	2.27E-12	28	58	6.03E-12	12	21
post-translational regulation	protein amino acid ADP-ribosylation	6.62E-03	2	4	2.99E-08	5	6
	protein amino acid phosphorylation	1.02E-07	95	464	1.95E-08	31	192
	peptidyl-citrulline biosynthetic process from peptidyl-arginine	4.91E-08	7	8	N/A	0	2
metabolism	cellular carbohydrate metabolic process	2.60E-10	19	35	8.68E-02	1	9
	gluconeogenesis	5.39E-07	10	17	N/A	0	4
	phospholipid catabolic process	7.52E-08	10	15	N/A	0	6
	chitin catabolic process	8.43E-07	14	31	1.03E-03	5	22
ion transporters	cation transport	1.31E-01	14	90	8.66E-08	14	52
	metal ion transport	9.25E-01	2	45	4.78E-17	24	57
	iron ion transport	0	2	2	1.75E-04	2	3
	potassium ion transport	9.77E-04	10	32	5.39E-08	6	9
	copper ion transport	N/A	0	13	8.00E-09	9	18
response to environment	regulation of pH	2.22E-01	3	20	1.56E-03	3	10
	chromate transport	9.42E-10	10	12	6.70E-04	2	4
	defense response	0	4	4	N/A	0	0
	response to oxidative stress	2.71E-05	14	39	N/A	0	22

1.3.3.2.1 Regulation of gene expression

One of GO terms enriched in Fol4287 ACs is regulation of transcription (regulation of transcription and regulation of transcription, DNA-dependent). Collectively, 20% of all annotated genes with these GO terms (1528 all, 304 AC) are found in AC sequences. However, this ratio is only 5% in MRL8996.

On the other hand, chromatin assembly and disassembly genes are enriched in both strains. And 11 of 12 AC genes in MRL8996 are also present in Fol4287 with 10 of them being AC genes.

1.3.3.2.2 Post-translational regulation

Protein amino acid phosphorylation genes are significantly enriched in both strains' ACs and some of these genes are orthologous. For example, out of 95 AC genes under this term in Fol4287, 38 are present in the MRL8996 genome (26 unique genes, 21 are in ACs). 31 AC genes in MRL8996, 24 are present in Fol4287 genome (17 unique genes 14 are in ACs).

Protein amino acid ADP-ribosylation genes are significantly enriched in both strains but while 2 copies in Fol4287 (FOXG_12427 and FOXG_14110) have an ortholog in MRL8996 AC sequences, no other MRL8996 AC genes have orthologs in Fol4287 AC sequences (1 ortholog is found in a CC).

On the other hand, Fol4287 AC genes are enriched in genes involved in peptidyl-citrulline biosynthetic process from peptidyl-arginine while there are no genes with this annotation in MRL8996 AC sequences and none of the Fol4287 AC genes have orthologs in MRL8996.

1.3.3.2.3 Metabolism

Fol4287 ACs are enriched in cellular carbohydrate metabolic process (specifically tricarboxylic acid cycle intermediates), gluconeogenesis, phospholipid catabolic process, and chitin catabolic process, while MRL8996 ACs are only slightly enriched in the chitin catabolic process. Three of 7 chitinases in Fol4287 ACs are also present in MRL8996 AC sequences.

1.3.3.2.4 Ion transporters

AC genes of MRL8996 are highly enriched in metal and cation transporters specifically, iron, copper, and potassium. On the other hand, Fol4287 AC genes are slightly enriched in potassium ion transporters, and the 2 iron ion transporters in the Fol4287 genome are found in AC chromosomes.

MRL8996 AC potassium ion transporters are mostly present in Fol4287 ACs (5 out of 6), while the majority of Fol4287 AC potassium ion transporters are not in the MRL8996 genome (6 out of 10).

While the orthologs of the iron ion transporters in Fol4287 (Ferritin-1 homologs FOXG_16665 and FOXG_16728) are present in MRL8996 CCs, there are no orthologs of the iron ion transporters found in MRL8996 AC sequences (ferroportin 1 and 2 homologs) in the Fol4287 genome.

Interestingly, the majority of the iron, copper, and metal ions in MRL8996 AC sequences are absent in the Fol4287 genome (10 of 33 have homologs in Fol4287 CC and ACs), while most of these genes are also present in the genome of the other clinical isolate NRRL32931 (25 of 33 genes).

1.3.3.2.5 Genes related to environmental stresses

Fol4287 and MRL8996 ACs show distinct GO term enrichments in genes related to the environmental response.

Both strains have chromate transport genes enrichment in AC regions.

While Fol4287 ACs are significantly enriched in response to oxidative stress genes, none of the genes annotated in this category in the MRL8996 genome are in ACs. Additionally, all 4 genes annotated as the defense response in Fol4287 (2 chitinases, 1 TPR_REGION domain-containing protein, and 1 phospholipase family protein) are in ACs and none of them are present in the MRL8996 genome.

In contrast, genes related to pH are enriched in MRL8996. Out of all three Na^+/H^+ antiporter genes in this category, only one has an ortholog in the Fol4287 genome.

In addition, *pacC*, the transcription factor mediating alkaline pH while having only one copy in Fol4287 (FOXG_02222 in chromosome 5), has additional truncated copies in the MRL8996 genome. The full-length gene with protein id 139670 is located in C05.5 while truncated paralogs are in C13.3 (306577) and AC sequences (499403, 498041, and 422193).

1.3.3.2.6 Unique genes

In addition to the mentioned GO terms, there are a few unique genes found in both strains. MRL8996 AC sequences have a homolog of Adenine phosphoribosyltransferase (APRTase, 530363) and an alkylmercury lyase (481305); while Fol4287 ACs has a gene coding for Dimer_Tnp_hAT domain-containing protein which involves in pyrimidine nucleoside metabolism (FOXG_21841), a membrane-bound Ankyrin repeat gene involved in peptidoglycan biosynthesis (FOXG_17522), a transmembrane electron transporter

involved in photosynthesis (FOXG_17114), and a terpene cyclase/mutase family member protein-coding gene in hopanoid biosynthesis (FOXG_17453).

1.3.3.3 Distinct transposon profiles in these two fungal genomes

The transposons of Fol4287 and MRL8996 were *de novo* predicted and annotated using the curated TE library [15]. Comparative repeat analysis revealed that the repeat contents of Fol4287 and MRL8996 are comparable (6.6% and 5.4%, respectively), and while the CCs are lacking in terms of TE content (2.97% and 3.11% for Fol4287 and MRL8996, respectively), ACs are enriched in TEs (27.17% and 23.83%, respectively) (**Table 1.7**). For both genomes, the retrotransposons occupy the highest number of genomic positions. While in Fol4287 the majority of transposons in ACs are DNA transposon, in MRL8996 it is the retrotransposons.

1.3.3.3.1 Class I transposons (Retrotransposons)

Different families of retrotransposons are present or have the highest abundance between Fol4287 and MRL8996 genomes.

For LTR transposons, although a Copia family transposon (labeled as CopiaActive for its activity observed in Chapter 2) is present with high abundance in both, the most copy number transposon family in MRL8996 is not present in Fol4287. (**Figure 1.9**A). LINE transposons have a higher copy number in MRL8996 and the most abundant one is not present in Fol4287 (**Figure 1.9**B). For both strains, SINEs are highly abundant. Out of the 3 SINE transposon families while the Foxy2 family is more abundant in Fol4287, For MRL8996 it's the Foxy3 family. (**Figure 1.9**C).

Table 1.7 The transposable element compositions of Fol4287 and MRL8996.

		Fol4287				*MRL8996*			
Order	Superfamily	*# of Elements*	*Length Occupied*	*% of Length Occupied in CCs*	*% of Length Occupied in ACs*	*# of Elements*	*Length Occupied*	*% of Length Occupied in CCs*	*% of Length Occupied in ACs*
Class I (retrotransposons)		**2393**	**1378862**	**1.11**	**9.03**	**2081**	**1501682**	**1.82**	**11.01**
LTR		*1972*	*1167705*	*1.07*	*7.05*	*1454*	*1087216*	*1.63*	*5.83*
	Copia	357	195736	0.11	1.47	72	48382	0.00	0.75
	Gypsy	1615	971969	0.96	5.58	1382	1038834	1.63	5.09
LINE		*345*	*192457*	*0.02*	*1.85*	*442*	*350843*	*0.12*	*4.67*
	R1	26	7325	0.02	0.01	56	55944	0.02	0.77
	Tad1	319	185132	0.01	1.85	352	277772	0.09	3.73
SINE		*76*	*18700*	*0.01*	*0.13*	*185*	*63623*	*0.07*	*0.51*
	tRNA	25	1990	0.00	0.00	43	4348	0.01	0.00
Class II (DNA transposons)		**2088**	**1198103**	**0.55**	**9.74**	**909**	**532268**	**0.18**	**7.06**
TIR		*2033*	*1115601*	*0.47*	*9.21*	*871*	*473470*	*0.18*	*6.16*
	Tc1Mar-Fot1	849	498262	0.29	3.78	193	146589	0.09	1.67
	hAT	428	199246	0.03	1.89	183	97271	0.01	1.48
	Mutator	149	121421	0.08	0.87	69	39970	0.01	0.57
	pogo	223	139586	0.01	1.37	112	99564	0.03	1.37
	PiggyBac	282	141332	0.06	1.15	199	49113	0.02	0.62
Helitron		*55*	*82502*	*0.07*	*0.52*	*38*	*58798*	*0.00*	*0.90*
Unknown		**5563**	**1403616**	**1.31**	**8.40**	**3050**	**849145**	**1.10**	**5.76**
Total		***10044***	***3980581***	***2.97***	***27.17***	***6040***	***2883095***	***3.11***	***23.83***

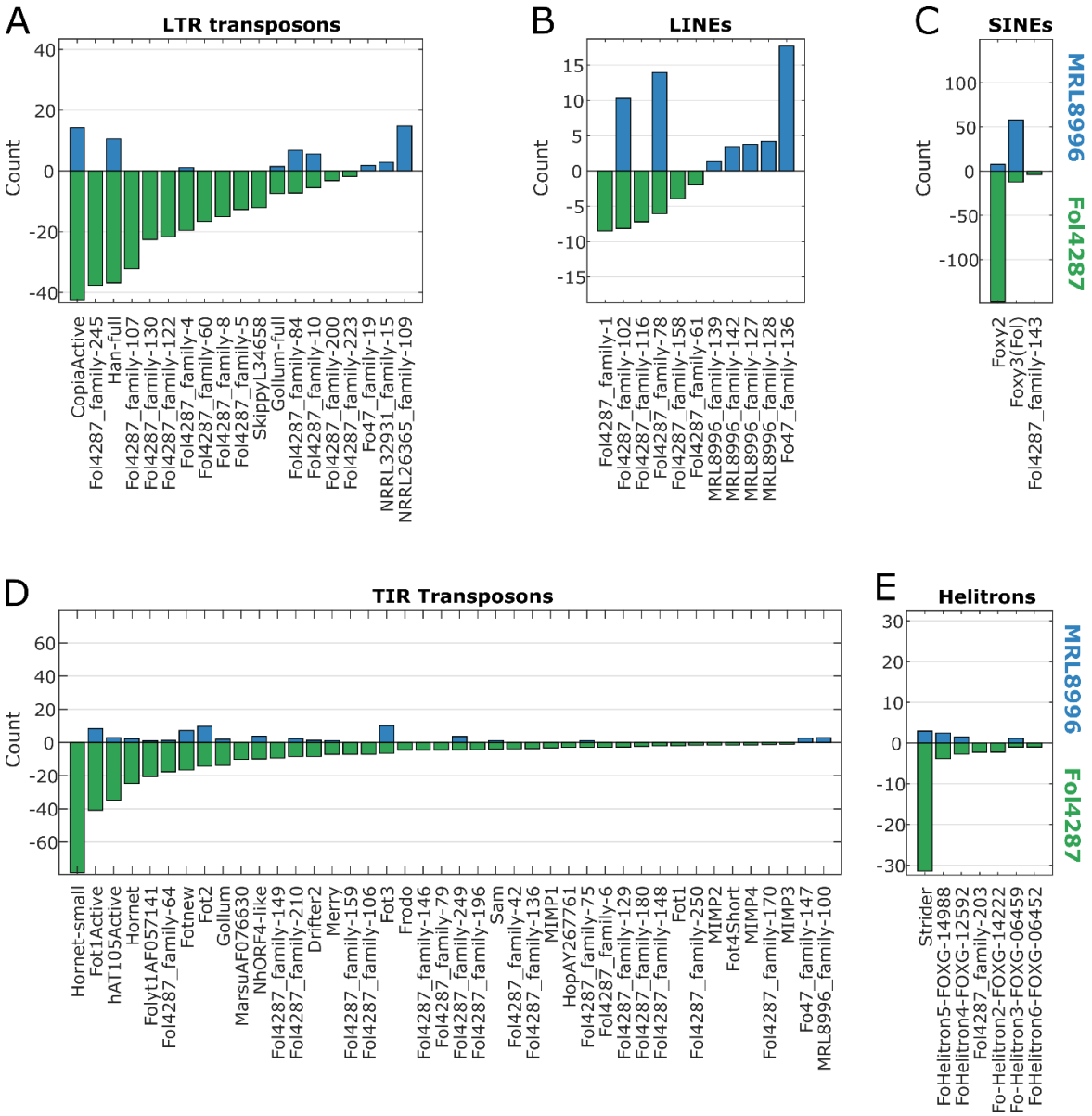


Figure 1.9 Transposon abundances in the Fol4287 and MRL8996 genomes. (A-C) Class I retrotransposons LTR (A), LINEs (B), and SINEs (C). (D-E) Class II DNA transposons TIRs (D) and Helitrons (E). Similar transposable elements are combined in one family and the element with maximum abundance is represented.

1.3.3.3.2 Class II transposons (DNA Transposons)

Overall, there are more copy numbers of DNA transposons in Fol4287, and the largest difference is seen in DNA transposons. Although most of the DNA transposons are shared, Fol4287 has a higher copy number for the majority (**Figure 1.9**D). While a hAT family miniature Hornet (or Hormin) element have the highest abundance Fol4287 genome, it's

absent in MRL8996. A Fot1 transposon (labeled as Fot1Active for its activity observed in Chapter 2) however is one of the highest copy number DNA transposons in both genomes. Miniature impala elements or MIMPs, which are known to be involved in phytopathogenic are not present in MRL8996.

Helitrons, similarly, have a higher abundance in Fol4287. Strider family while having more than 30 copies in the Fol4287 genome, have less than 10 copies in the MRL8996 genome (**Figure 1.9**E). There are also many unshared unannotated transposons between two genomes (Appendix B3).

1.4 Discussion

This chapter laid out the foundation to dissect host-specificity of the cross-kingdom fungal pathogenicity using the phytopathogen *F. oxysporum* f. sp. *lycopersici* 4287 and the human eye pathogen *F. oxysporum* MRL8996. Unique sets of accessory chromosomes established the structural and functional division within each genome. Distinct phenotypic variation indicated a genotype and phenotype correlation and that dissecting adaptation under unique environmental stresses is feasible.

1.4.1 Phenotypic characterization revealed unique niche adaptation

As the human pathogen MRL8996 is better adapted to elevated temperature and pH, has faster hyphal growth *in vitro* than Fol4287 while the plant pathogen Fol4287 is relatively more tolerant to some stress conditions.

The colony growth of MRL8996 is faster in the majority of the tested conditions which suggests an intrinsic hyphal growth fitness advantage over its counterpart. But the main advantage of MRL8996 has over Fol4287 is its growth rate in elevated temperature. The human eye which MRL8996 was isolated from has a temperature of 30°C to 37°C with the cornea being the coldest part [91]. For the fungus to be able to infect and colonize the eye, it needs to overcome temperatures up to 37°C. In the experiments, 34°C standard temperature was used. In addition to growing faster than Fol4287, MRL8996 doesn't have a strong inhibition when compared to its growth in 28°C, the standard incubation temperature.

Interestingly pH change did not result in any growth difference for neither strain. Transcriptionally, pH change results in major gene regulation rewiring in both Fol4287 and NRRL 32931, a clinical isolate, albeit very differently [92]. On the other hand, change in

temperature does not affect the transcriptions to that extent but results in high growth inhibition (Appendix B4).

On the other hand, Fol4287 seems to be more tolerant to other stress conditions such as osmatic and cell wall stress, and both strains are equally tolerant to oxidative stress. Plant host provides different stress conditions than human eye conditions. For example, rapid oxidative burst is used as the early plant defense against microorganisms, and neutralization of reactive oxygen species (ROS) is required for host colonization [93,94]. Plant immunity also activates chitinases and glucanases to disrupt the fungal cell wall [95,96]. A similar effect can be achieved by Congo Red which inhibits the β-glucan synthesis and causes weakened cell wall [97].

1.4.2 The genotypic comparison revealed distinct sets of accessory chromosomes enriched for genes that could be linked to the functional adaptation

High-quality genome assemblies brought genetic variations under the light. In addition to the conventional wisdom that subtelomeric regions are more diverse with low sequence identity, the plant pathogenic strain Fol4287 and the keratitis strain MRL8996 each have a unique set of accessory chromosomes.

Both sets of ACs are highly repetitive and highly fragmented in both genome assemblies. It is without a question that the repetitive nature contributes to the highly dynamic behaviors of these ACs. However, their transposon profiles are very distinct. In Fol4287, DNA transposons are very diverse and have a high number of copies. In MRL8995 however, while the copy numbers of DNA transposons are very low, retrotransposons especially LINE transposons are highly abundant.

Distinct sets of genes encoded in the ACs probably contribute directly to the phenotypic differences we observed using our phenotypic assays.

Notably, the ACs of Fol4287 were enriched transcription factors, as reported before [15], the MRL8996 genome did not show the same phenomenon. Instead, chromatin assembly/disassembly genes were enriched in both strains. This signals a major difference in the gene regulation mechanism between the two strains.

For the post-translational regulation categories, Fol4287 ACs were enriched in the protein citrullination process which is a post-transcriptional modification of histones and involves in autoimmune diseases in humans and found in plants as well [98,99].

On the other hand, the post-transcriptional modification which is enriched in MRL8996 was ADP-ribosylation which involves in DNA damage repair, development, metabolism, and stress response in humans and, and immune response by plants [100,101]. In addition, glucose metabolism and chitinases were enriched in ACs.

Ion transporters are a distinct category between Fol4287, and MRL8996. While potassium ion transporters are enriched in Fol4287, in MRL8996 metal ion channels are enriched. Iron transporters and iron homeostasis genes are found in many human pathogens [102]. This GO-term is also enriched in NRRL32931, another *F. oxysporum* clinical isolate [32]. The copper-binding oxidoreductase, ceruloplasmin, which is found in NRLL32931 ACs is also present in MRL8996.

Similarly, pH regulation genes were also enriched in MRL8996. The expansion in the PacC family, the transcription factors that regulate the response to alkaline pH were also observed in NRL32931 [32]. Fol4287, on the other hand, has enrichment in oxidative stress response while MRL8996 LS chromosomes do not have any genes in this category.

1.4.3 Strategies to correlate genotypic variations with phenotypic differences

In this study, the foundation of the genotype-phenotype map for *F. oxysporum* strains had been laid out. This is the first step in establishing the fitness landscape of this fungus and understanding its adaptation [103]. However, at the comparative genomics level, only the static differences are captured. The next chapters will attempt to observe the evolution in real-time.

CHAPTER 2 EXPERIMENTAL EVOLUTION OF *F. OXYSPORUM* F. SP. *LYCOPERSICI* 4287

2.1 Introduction

The idea of evolution by natural selection proposed by Charles Darwin in 1859 established the foundation of evolutionary biology. Following biological processes such as mutation, selection, adaptation, and speciation biologists predict patterns that have given the rise of the diversity of organisms and how they change over time. This chapter describes an effort to understand the evolvability and the genetic architecture enabling the effective adaptation of the asexual fungal species complex *Fusarium oxysporum* applying experimental evolution, one of the most powerful tools to understand the adaptation mechanisms of the populations and observing the fitness landscapes, under controlled environments [104].

Studying evolution in a laboratory environment with controlled parameters such as population size, selection pressure, and mutation rates helps to test hypotheses with precision and replicates. Working with microbes shortens the evolutionary time scale. Storing the intermediate populations as frozen gives the possibility to inspect the fossil records when needed; the evolution of the population can even be tracked in real-time using fluorescence markers [105,106]. In addition, forcing the culture into a challenging condition to adapt to might improve the fitness of the industrially important microbes [107]

Experimental evolution typically starts with a clonal microbial culture in a controlled laboratory environment, and it is let grow until it reaches the highest population density or predetermined population size. From the grown culture a subpopulation with fixed size collected and transferred to a new environment with fresh sources. This growth and dilution cycle, or passages, continues until it reaches several generations, an expected phenotypic change is observed, a number of passages achieved, or indefinitely [108].

While the process is underway, the evolving populations would adapt to the laboratory condition through a set of genotypic and phenotypic changes [109].

Several approaches have been used to culture growth and dilution depending on the purpose of the experiment [108,109]. One of the most popular methods is the serial dilution of a batch culture where the culture is grown until it reaches the highest population density it can get by depleting the resources and after a fixed time interval, a randomly or nonrandomly selected subpopulation is transferred to the new environment with fresh resources. During the transfers, the population is forced through a bottleneck. The most famous experiment of this type is the long-term evolution experiment by Lenski et al. which had started from a single *Escherichia coli* colony and grown as 12 evolving populations since 1988 and reached 11,000 passages and 73,500 generations over 32 years [110–112].

Another type is the continuous cultures such as chemostats [113] or microfluidics [114] where the population density is not allowed to change by automatically diluting the sample regularly using fresh media. This experiment differs from the serial dilution since the growth phase of the organism, nor the population size does not change and there is no bottleneck effect.

Finally, experimental evolution can be used to study host-microbe interaction by inoculation or infection of either plant or animal plant hosts. Surviving and propagated individuals can be used to passage a new host. Since *in vivo* environment would be much complex than the well-defined laboratory conditions, these type of experiments helps to understand the more complex relationships between host and pathogen [115].

Adaptive evolution occurs when a selection acts on genetic mutations that confer adaptation advantage to the organism under the selective condition. Most common

mutations include single nucleotide variations (SNVs), insertions and deletions (INDELs), transposon insertions and excisions variations, large segmental deletion, duplication, and translocations. Most of the mutations are neutral or deleterious, but some mutations can be beneficial and drive the adaptation [116]. The effect of a mutation on adaptation depends on its fitness effect over background genotype in the selective environment and the size of the population. Population bottlenecks occur when a population's size is reduced as genetic drift acts more quickly to a drastic reduction of genetic variation in small populations. The bottleneck effect can be as extreme as a single cell if the selection of propagation is through single colony selection in a petri dish. This type of experiment is used to accumulate mutations and they are used to estimate intrinsic mutation rates [117].

The effect of a mutation on adaptation depends on its fitness effect over background genotype in the selective environment and the size of the population. For any beneficial mutation to escape the random drift and get established in a population, the population size of the mutant, n, needs to be equal or greater than c/s; where s is the fitness effect of the mutation over the ancestor's fitness, and c is the half of the offspring number variance [105,116,118]. Once established, the mutant subpopulation grows exponentially with a rate of s, until the population acquires a new beneficial mutation, and they compete to be the dominant mutation.

In small asexual populations with rare beneficial mutations, and under a strong selection, successive selective sweeps where an established beneficial mutation gets fixed in the population before the appearance of the next beneficial mutation, the population dynamic is described as "strong-selection weak-mutation" or "successional mutations regime [116]. If the population size is large enough and the beneficial mutations may appear on different lineages simultaneously (before getting fixed), under a strong selection, similar selective sweeps can be observed ("strong-selection strong-mutation" or "concurrent-mutations

regime” where the beneficial mutation with higher *s* is fixed but lower *s* goes extinct. Not often, more than one lineages with different beneficial mutations form a stable coexistence where neither goes extinct nor gets fixed [119]. And the “clonal interference” can be observed when the dominant subpopulation is outcompeted by an emerging smaller subpopulation [120,121]

All experimental evolution designs carry diverse selection pressures, including media type with fixed nutrient source [110], lack of an important nutrient[122], media with growth-limiting-drug [123], or a host organism [115]. The rate of beneficial mutations might change over time with the occurrence of mutators [124]. Simple environments can select for loss-of-function mutations [125,126]. For example, in the long-term evolution experiment with *E. coli*, after 2,000 generations the *rbs* operon which encodes for ribose utilization genes were inactivated by mutations [125]. Similarly, when a yeast population was forced to propagate asexually, the resulting population accumulate mutations in mating-related pathways [121].

One phenomenon observed in experimental evolution experiments is genetic parallelism that describes independently acquired mutations on the same genes [127]. Parallelism or the predictability of the evolving population occurs when limited options are available for an organism to act on a given fitness landscape [128]. Such phenomenon is extremely powerful in associate genotypic variations observed in the evolving population and the environmental condition the organism is selected through.

The other phenomenon observed in experimental evolution experiments is epistasis. A stepwise optimization epistasis model describes a mutation that is dependent on the genetic background with the presence or absence of mutations in one or more other genes. In this case, epistatic mutations usually have different effects on their own than when they occur together. Beneficial mutations might occur and fix quickly in the population in the

short term. However, the longer the experiments last, the adaptation decreases [109,129,130]. The reason for that is the diminishing returns epistasis where each beneficial mutation increases the fitness of its ancestor by an amount lower than it would increase the fitness of the original ancestor [109]. In contrast to this stepwise optimization model, all-or-none epistasis might also happen. In this model, the fitness increase or innovation does not happen until all required mutations occur in the same background, therefore it happens, but rarely [109,131].

In the sexually propagating organisms, the evolution experiments showed that the rate of adaptation is faster than asexually propagating organism because (1) the recombination and combine the beneficial mutations from different lineages into the same context and (2) it can eliminate the deleterious mutations that might have hijacked a beneficial mutation from an otherwise fit background [132]. *F. oxysporum*, unlike most of the other Ascomycetes including other *Fusarium* species, does not have a known sexual stage in its life cycle, therefore lacks this important tool of evolution. Although there is evidence of recombination [133–135] and different mating types [34], the lack of sexual reproduction should have come with a cost. However, *F. oxysporum* is still a very successful pathogen that can adapt to environments and hosts across kingdoms from animals to plants. To investigate the mechanisms of its adaptability, evolution experiments were carried out using *F. oxysporum*.

The experimental evolution study of *F. oxysporum* described in this chapter is based on a collaboration between our lab and Dr. López Díaz and Dr. Di Pietro at the University of Córdoba, Spain. All bench-side experiments were conducted in Di Pietro Lab. In these experiments, tomato-plant pathogenic *F. oxysporum* f. sp. *lycopersici* 4287 (Fol4287) underwent serial dilutions or passages in vitro and through host inoculation (**Figure 2.1**A & Methods 2.2.1). Briefly, the starting conidia of clonal Fol4287 (10^7 conidia/ml) was dip-

inoculated on the roots of *Solanum lycopersicum* (tomato) plants and after 2 weeks the fungus was reisolated from the stems of the infected plants. After a short growth stage in the liquid medium, a new plant of the same age was infected with collected conidia from the previous passage. This process was repeated for 10 passages in 5 independent replicates. Similarly, the same starting isolate of Fol4287 (10^6 conidia) was spot-inoculated on plates of complete media (yeast extract peptone dextrose agar) or minimal medium. After one week of incubation, the conidia collected from the colony were inoculated on a new plate and this process was repeated for 10 passages for 5 independent replicates for both medium conditions.

At the end of the 10^{th} passage, the virulence of the evolved populations on tomato plants was tested. While none of the plant-passaged populations didn't show any difference from the ancestor, four complete medium-passaged populations and two of the minimal media-passaged populations showed reduced virulence (**Figure 2.1**B-D).

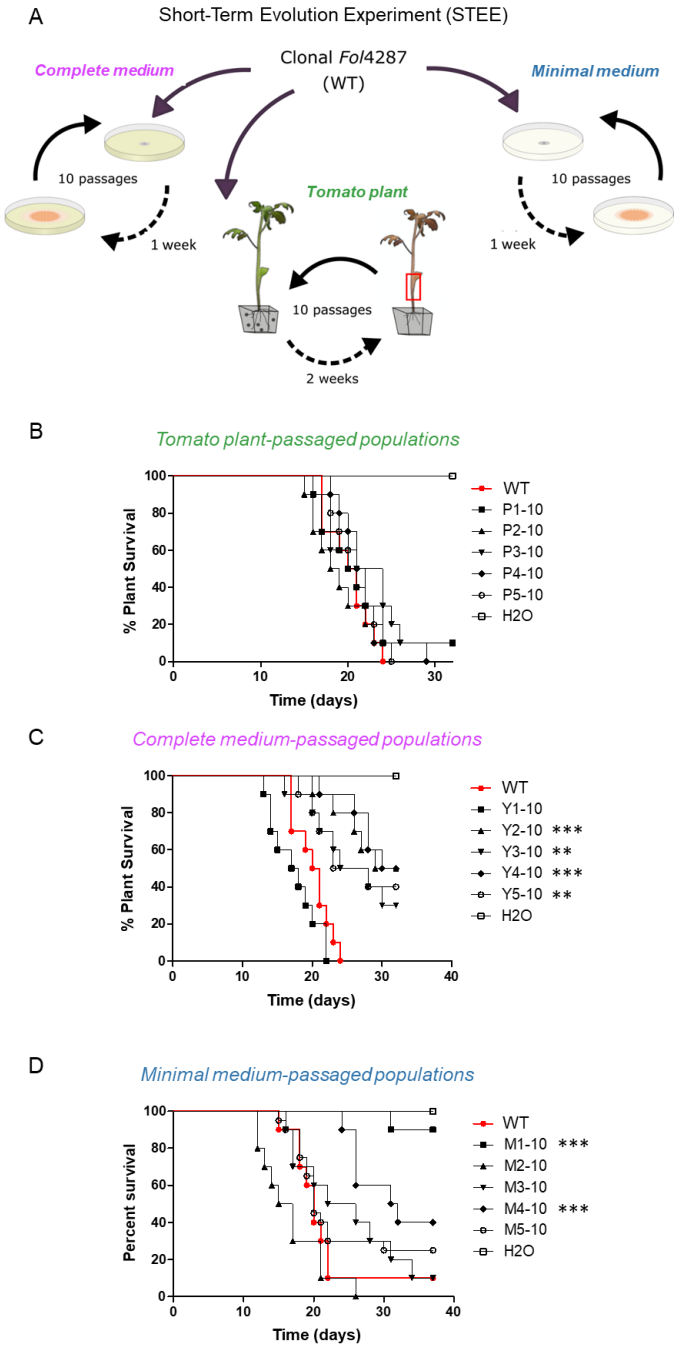


Figure 2.1. The short-term experimental evolution of Fol4287 (conducted by Cristina López Díaz at the University of Cordoba, Spain)**.** (A) Experimental design. Microconidia of the clonal isolate Fol4287 were inoculated on roots of tomato plants and plants were kept growing for 2 weeks. The fungus was then re-isolated from the stem (xylem) and used to inoculate a new healthy plant. Similarly, conidia from the same starting isolate were used to inoculate complete or minimal media plates and incubated at 28°C for 7 days. Conidia collected from the colony were then used to inoculate a fresh plate. All three processes were repeated 10 times with 5 biological replicates. In each passage, 10^7 (plants) or 10^6 conidia (plates) were inoculated. (B-D) Survival curves from plant infection assays using conidia of the ancestral clone (WT), plant-passaged populations (B), complete media-passaged populations (C), or minimal media populations (D). Tomato plants were inoculated with WT, evolved populations, or water (negative control) and the percent survival of plants was plotted. "*" annotates the level of significant changes from the WT killing profile using the Logrank test.

2.2 Methods

2.2.1 Short-term evolution experiment in vivo and in vitro (performed by Cristina López Díaz)

From a clonal Fol4287 culture, 10^7 conidia/ml were dip-inoculated on roots of 2-week-old *Solanum lycopersicum* (tomato) plants. After 12 to 14 days, the stems were surface sterilized in 1% NaOCl for 2 min, rinsed twice for 2 minutes in distilled water. Three pieces of the stem of approximately 1 cm length were cut with a sterile blade and placed on Potato Dextrose Agar (PDA) with chloramphenicol. After 1.5 days, the mycelia emerging from the xylem vessels were transferred to Potato Dextrose Broth for 1.5 days. The produced microconidia were collected by filtration through a nylon filter (Monodur, pore diameter 10–15 µm) and centrifugation, counted, and used to inoculate another healthy 2-week-old plant (10^7 conidia/ml). This process was repeated for 10 consecutive passages, with 5 independent replicate lines. Between each passage, a fraction of the collected microconidia was stored in 30% of glycerol at -80°C for later use.

Similarly, the same starting clone was used to spot-inoculate plates of complete medium (Yeast Extract Peptone Dextrose Agar: 0.3 % w/v yeast extract, 1% w/v Bactpeptone, 2% w/v dextrose, 1.5% w/v agar) or minimal medium (0.1% w/v KH_2PO_4, 0.05% w/v $MgSO_4 \cdot 7H_2O$, 0.05% w/v KCl, 0.2% w/v $NaNO_3$, 3% w/v sucrose, and 1,5% w/v Agar) and plates were incubated at 28°C for 7 days. The microconidia were recollected by adding sterile water to the colony surface and filtering the conidial suspension through a sterile nylon filter. From the obtained conidial suspension, 10^6 conidia were used to spot-inoculate fresh medium plates. This process was repeated for 10 consecutive passages, with 5 independent replicate lines for both medium types. Between each passage, a fraction of the collected microconidia were stored in 30% of glycerol at -80°C for later use [68].

The 15 populations obtained from the 10th passage, the initial ancestor culture, all intermediate populations of plant-passaged P2, and the 1st, 5th, and 9th passage intermediate populations of complete medium-passaged line Y3 and minimal medium-passaged line M4 were grown in PDB for 4 days and genomic DNA was extracted using the CTAB DNA extraction method.

2.2.2 Whole-genome sequencing of ancestor and passaged populations

The library preparation and the whole genome shotgun sequencing of the ancestor and the final plant-passaged and complete medium-passaged population samples were performed by the Tufts University Core Facilities with Illumina HiSeq 2500 Platform (2×71 cycles). Final minimal medium-passaged populations were sequenced with Illumina NextSeq 500 platform (2×75 cycles) at Genomics Resource Laboratory at the University of Massachusetts Amherst. To confirm if the more error-prone NextSeq 500 platform introduced any bias to our analysis, minimal medium-passaged M4-10 population was re-sequenced using Illumina MiSeq (2×75 cycles) platform at Genomics Resource Laboratory at the University of Massachusetts Amherst, MA, USA. The intermediate populations of P2, Y3, and M4 were sequenced by Novogene Company Limited, UK via Illumina NovaSeq 6000 platform. FastQC version 0.11.5 was used to check raw read qualities [136].

2.2.3 Mapping and variant calling

2.2.3.1 Mapping

For all analyses, the previously assembled and annotated *F. oxysporum* f. sp. *lycopersici* 4287 (WT) genome was used as reference (accession QESU00000000, Methods 1.2.3) [35].

The raw reads from sequenced populations and the ancestor (WT) paired-end Illumina reads of Fol4287 mapped to the reference using BWA mem (version 0.7.15) with the following options `'-t 8 -M -a -v 1'` [72]. The alignments then were processed using Picard (version 2.17.8) CleanSam, Samtools (version 1.3) view, Picard FixMateInformation, Picard MarkDuplicates, and Samtools index with default options [76,77]. Samtools flagstat and Picard CollectRawWgsMetrics were used to assess the mapping quality.

For WT and the final populations, the base quality scores were recalibrated via GATK BaseRecalibrator (version 3.57) using initial variant calling with FreeBayes (version v1.0.2-33-gdbb6160) for `'--knownSites'` option [73,136].

2.2.3.2 Variant calling

2.2.3.2.1 SNV and INDEL calling

The single nucleotide variants (SNVs) and small insertions deletions (INDELs) were called via GATK Mutect2 (version 4.1.4.1) using WT sample as "normal" and otherwise default options. The one command line per sample group was run (e.g., for plant-passaged final populations: WT, P1-10, P2-10, P3-10, P4-10, and P5-10 samples were analyzed together; for Y3 intermediate populations: WT, Y3-1, Y3-5, Y3-9 samples were analyzed together). The calls were filtered via GATK FilterMutectCalls using default options.

All detected variations were manually inspected and further filtered using Integrative Genomics Viewer (IGV) [137]. The allele frequency (AF) values for high coverage regions were adjusted to reflect real frequencies. For each detected variation in the samples that the intermediates sequenced, allele frequency for each population was reported even if they were not detected by the tools. AFs lower than 0.1 were filtered out for most of the analyses.

2.2.3.2.2 TIV calling

To be able to detect transposable element (TE) insertions, new software was developed: TEfinder [138]. TE annotations detected previously (Methods 1.2.6) were used.

TEfinder is a bash pipeline for detecting TE insertions using paired-end sequencing data. For this, an assembled genome and pair-end sequencing of the sample are required. The pipeline inputs four files: read alignments, reference fasta genome, the TEs of the reference genome, and a list of the TEs to search (**Figure 2.2**A).

TEfinder relies on paired-end sequencing and uses information on discordant reads, which are reads that do not match the expected orientation or insert size. The pipeline goes through 4 main steps:

i. Identification of discordant reads: BEDtools intersect is used to extract the primary alignments to the known TEs, then Picard FilterSamRecords is used to filter those alignments and their pairs. The pairs which are not mapped to the same sequence or have an insert size larger than excepted are filtered as discordant reads.
ii. Grouping of discordant reads: the clusters of reads mapped to non-TE regions are identified (**Figure 2.2**B) using BEDtools merge and grouped by their read orientation.
iii. Definition of the TE insertion sites: The plus and minus strand groups of the read clusters are filtered by their orientation.
iv. Filtering of new insertion sites: The insertion sites in repeat regions, with plus and minus strand read number bias, and with a low number of read evidence are also filtered out.

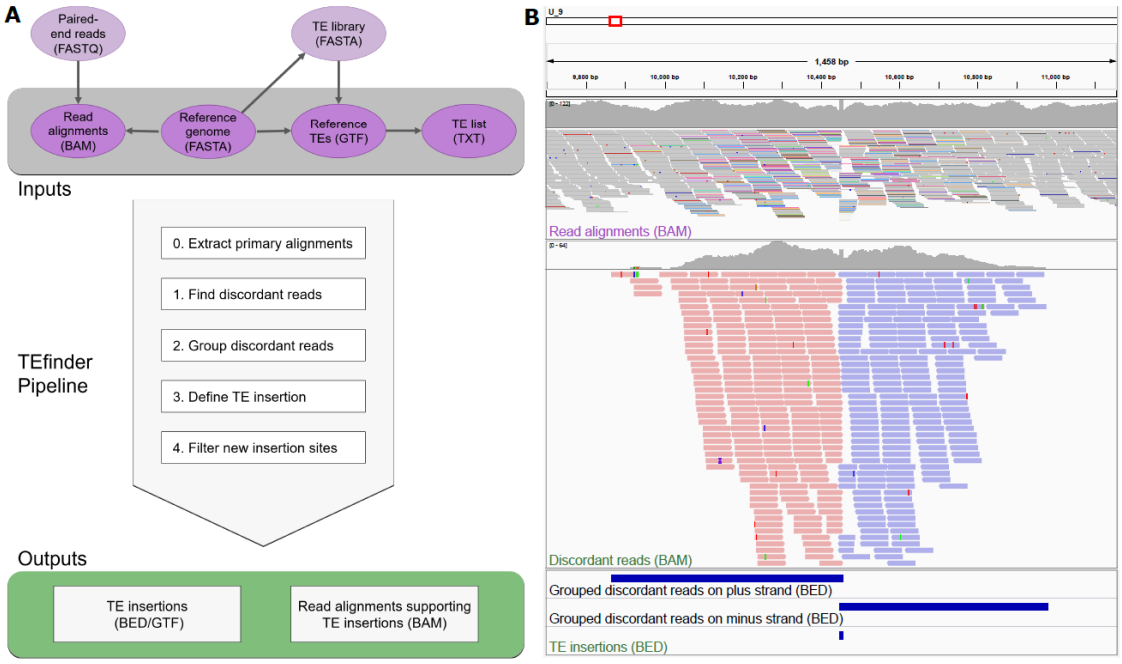


Figure 2.2 TEfinder: a pipeline to detect TE insertion events. (A) The basic workflow of the TEfinder pipeline. The required input files are in dark purple, while other files that are used to create the input files are in light purple. Arrows show the dependencies of the files. The steps of the pipeline are in the light gray box and the outputs are in the green box. (B) Visualization of a new small Hornet (TIR/hAT) insertion event using Integrative Genomics Viewer (IGV) [137]. The data were derived from the Y3-10 population. (Top) input read alignments file, (middle) output discordant reads alignment file, and (bottom) intermediate and output BED files are displayed. The aligned reads in the top panel are shown in squished mode with reads mapped to different sequences colored non-gray. The aligned reads in the middle panel are shown in collapsed mode and forward reads are in red while reverse reads are in blue. The histograms above the alignments in the top and middle panels show the aligned read counts. The position of the insertion event reported by TEfinder coincides with the duplicated target site (TSD).

The transposon insertion variations (TIVs) detected by TEfinder were further filtered out by their presence in WT using BEDtools intersect with the option '-v' and manually inspected on IGV. The AFs were estimated manually.

2.2.3.2.3 CNV calling

BEDtools genomecov with the option '-d' was used to output the read depth at a single nucleotide resolution for the final populations and WT. Then, a custom MatLab code was used to calculate median coverages of 10 kb windows of the genome. The read coverages were normalized to the median read coverages of the samples. Finally, the final population samples were normalized by the WT.

Hierarchical clustering was performed with the normalized copy numbers of the loci which shows more than 20% change in any sample and the AC contigs and some unmapped contigs were ordered accordingly. This ordering was used for both read coverage and genomic feature maps.

The coverages of the small arm of chromosome 13, duplicated region of chromosomes 3, and chromosome 15 were multiplied by 2 to reflect the real copy numbers. The regions with 10% or larger changes were noted as copy number variations. For the circular map, normalized read depths in 10 kb windows of all final populations were averaged.

The genome sizes were estimated from the raw 10 kb windows normalized by sample median coverage (Estimated genome size: $\sum_{i=1}^{w} d_i . 10000$, where *i* is the 10 kb windows, *w* is the total number of 10 kb windows, and d_i is the median read depth of the 10 kb window normalized to overall median read coverage). One-way ANOVA was performed in three passaging groups on MatLab.

2.2.4 Genomic Features

For TE content, the simple repeats were removed from the previously identified repeats (Methods 1.2.6). BEDtools genomecov with the option ‘-d’ and a custom MatLab code were used to generate median TE content over 10 kb windows. The mean GC content of the genome was calculated over 10 kb windows via a MatLab code.

ChIP-seq raw reads for H3K4me2 and H3K27me3 in Fol4287 were retrieved from NCBI GEO accession GSE1212839. The reads were mapped to the Fol4287 genome via BWA mem and the alignments were processed as previously described. The read coverage was detected via BEDtools genomecov with the option ‘-d’. The reads from the two experimental replicates were summed and the genome coverage was calculated as the median values over 1 kb windows.

One-tail Fisher's exact test was used to test the null hypothesis that there are no nonrandom associations between the variations (TIVs or SNVs & INDELs) and sites with H3K27me3.

Circos version 0.69-6 was used to plot the circular genome map [139].

2.2.5 RNA sequencing

100 ml Potato Dextrose Broth (PDB, BD Difco, USA) was inoculated with 20 µl frozen stocks of Fol4287 and incubated at 28°C, 140 rpm for 4 days. The mycelium was filtered by filter paper, rinsed with sterile water, and dried using sterile paper towels. 200 mg mycelium was weighed and mixed with 20 ml of yeast extract peptone dextrose broth (YPD: 0.3% w/v yeast extract, 1% w/v Bactpeptone, 2% w/v dextrose) medium with pH adjusted to 5.0 or 7.4 using citric acid-phosphate buffer (for pH 5.0: 53.25% v/v 0.1 M citric acid and 46.75% v/v 0.2 M Na_2HPO_4; for pH 7.4: 13.05% v/v 0.1 M citric acid and 86.95% v/v 0.2 M Na_2HPO_4; the prepared buffers were mixed with 2× YPD medium in 1:1 ratio) at 28°C and 34°C, respectively, in 50 ml tubes with three replicates, then it was incubated for 1 h at specified temperature. At the end of the treatment, the mycelium was filtered and dried as described above. For 34°C treatments, the filtration process was performed in the incubator.

100 mg mycelium was weighed and homogenized in 1 ml of Invitrogen TRIzol Reagent, USA in a 1.5 ml Next Advance Pink RINO RNA-free lysis kits, USA. Then, it was homogenized by Next Advance Bullet Blender Tissue Homogenizer at speed 10 for min. The samples were incubated in ice for 1 min and the bullet blender step was repeated, followed by 5 min room temperature incubation. 0.2 ml chloroform was added to samples and mixed. After 2–3 min, they were centrifuged for 18 min at 10,000×g, 4°C. The top phase was transferred to a new 1.5 ml RNase-free tube. 500 ml of ethanol was added.

Samples were loaded to QIAGEN RNeasy columns, and QIAGEN RNeasy Mini Kit protocol was followed. The RNA was eluted in 30 µl of RNase-free water and stored in -20°C.

The RNA was quantified with Qubit RNA BR Assay and quality was checked by Agilent Bioanalyzer. The samples were sequenced by Genomic Resource Laboratory at the University of Massachusetts Amherst, USA via Illumina NextSeq 500 platform with 2×76 cycles.

The reads were mapped to the Fol4287 genome sequence using Hisat2 (version 2.0.5) with options `'-p 15 -X 500 --dta'` [140]. Alignments were sorted with Samtools (version 1.3) sort, duplicates were marked with Picard (version 2.0.1) MarkDuplicates and indexed by Samtool index. TEcount (version 2.2.1) was used with default options to generate the read counts mapped to TEs [141]. Finally, the read counts were normalized across samples by calculating the geometric means across sample sets and normalizing each sample by their median in a custom MatLab code.

2.2.6 TE sequence alignment

The sequence alignment of Hornet and Hormin was performed by NCBI Blast [142]. The domains were annotated by aligning to a Restless-like transposase (Uniprot accession: A0A395M6G3) by NCBI Blast and Conserved Domain-search [143].

2.3 Results

2.3.1 Surveying genotypic variation over short-term experimental evolution via whole-genome sequencing

2.3.1.1 Data generation and quality

The extracted batch genomic DNA samples from the final populations passaged through the plants, complete medium, or minimal medium were sequenced along with the ancestor clone (WT) via short-read paired-end sequencing using either Illumina HiSeq 2500 or NextSeq 500 platforms. To capture enough population variation in the batch samples, an average of 98× read coverage was generated for the final populations (**Table 2.1** & Appendix A1). This read depth may provide information about the allele frequencies of the mutations such as single nucleotide variations (SNVs) as low as 5% in the population. However, to ensure high confidence in variant calling, a 10% cut-off was used for the variation analyses below.

As all evolved populations start from the same WT clone, all variations were compared to the WT Fol4287 reference genome sequence [35]. More than 99% of all reads from each experiment could be mapped to the reference. This further supports the high quality of all data generated for this study.

The data quality of NextSeq 500 is generally lower than HiSeq 2500. To confirm that there was no sequencing technology-caused difference between samples, the M4-10 population was additionally sequenced via the MiSeq platform which uses the same four-channel sequencing by synthesis technology as HiSeq 2500 [144]. After the analyses were completed, there were no detected differences between two M4-10 sequencings (data not shown) therefore only NextSeq 500 results are reported hereafter. P3-10 and P5-10 samples were observed to be sharing the variants with a certain AF ratio; therefore,

the P3-10 sample which was suspected to be contaminated with the P5-10 sample was removed from our analysis.

Table 2.1. The resequencing of ancestor (WT) and sequencing of the final evolved populations of Fol4287. * Read counts and GC% were calculated by FastQC [136]; † median coverage was estimated by Picard CollectRawWgsMetrics [76]; ‡ Mapped read % was calculated by Samtools flagstat [77].

Sample	*Platform*	*Read count* *	*Read length*	*GC%* *	*Median Coverage* †	*Mapped read %* ‡
WT	HiSeq 2500	28960270x2	71	46	66	99.31
P1-10	HiSeq 2500	36174997x2	71	47	82	99.41
P2-10	HiSeq 2500	42747800x2	71	47	94	99.36
P4-10	HiSeq 2500	37683197x2	71	47	85	99.17
P5-10	HiSeq 2500	39602682x2	71	47	92	99.34
Y1-10	HiSeq 2500	29798233x2	71	47	69	99.41
Y2-10	HiSeq 2500	41312962x2	71	47	97	99.38
Y3-10	HiSeq 2500	39952004x2	71	47	94	99.44
Y4-10	HiSeq 2500	33929161x2	71	47	80	99.42
Y5-10	HiSeq 2500	36560273x2	71	47	85	99.37
M1-10	NextSeq 500	43783567x2	75	45	96	98.99
M2-10	NextSeq 500	50188420x2	75	45	108	98.88
M3-10	NextSeq 500	54029793x2	75	46	122	98.79
M4-10	NextSeq 500	60497167x2	75	47	139	99.02
M4-10	MiSeq	27840501X2	75	47	67	99.53
M5-10	NextSeq 500	59633905x2	75	46	132	99.05

2.3.1.2 Variant Calling

Comparing to the WT genome assembly, four types of variations were detected among these evolved populations: single nucleotide variations (SNVs), small insertions deletions (INDELs), transposon insertion variations (TIVs), and copy number variations (CNVs) (summarized in **Figure 2.3**).

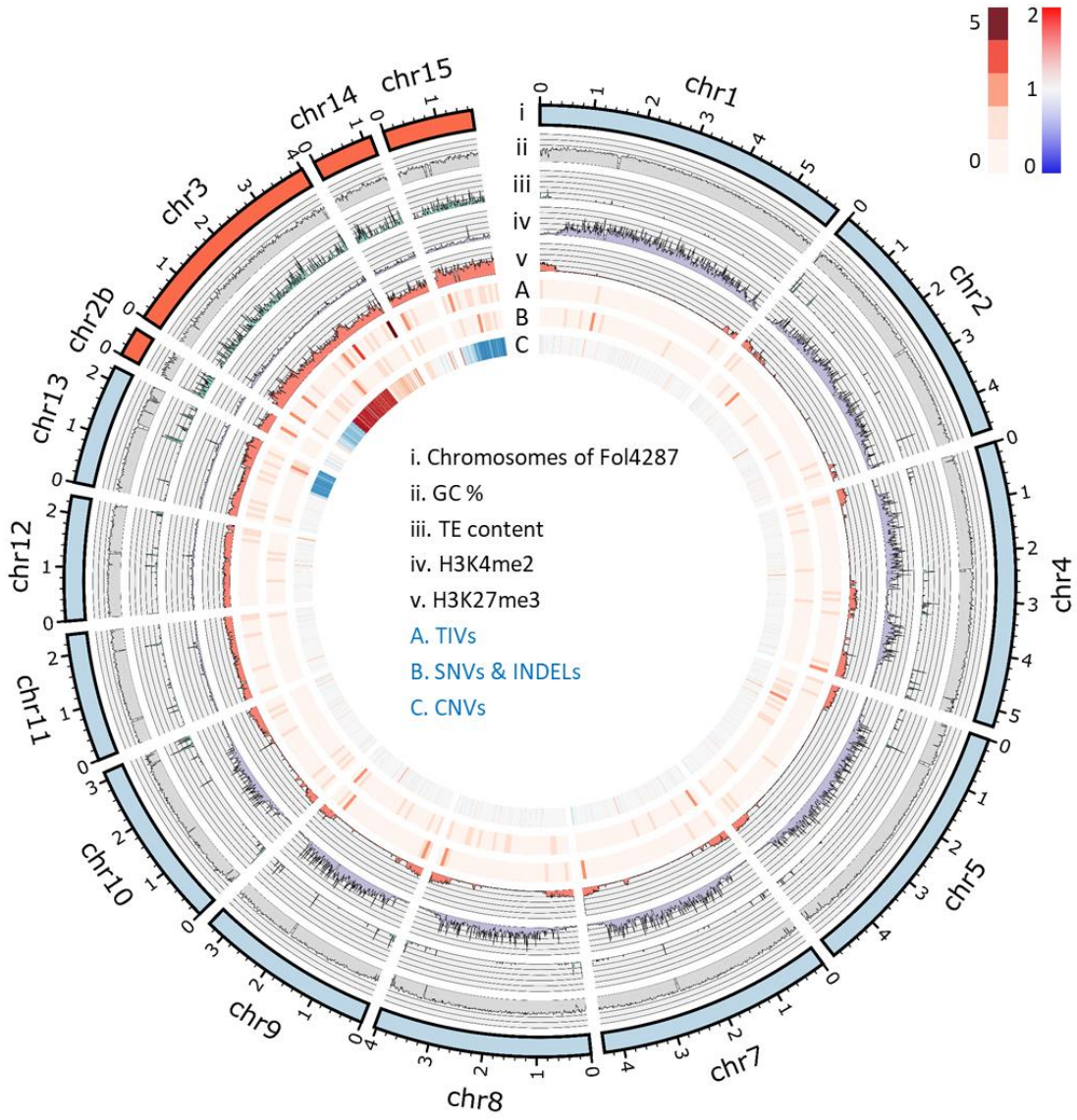


Figure 2.3 Chromosomal distribution of the mutation events. (i) The chromosomes of Fol4287. Accessory chromosomes are shown in orange, while the core chromosomes are in blue. (ii) GC % content and (iii) TE content calculated in 10 kb windows. The coverages of chromatin immunoprecipitation sequencing mapping of histone marker associated with euchromatin, H3 lysine 4 dimethylation (H3K4me2) (iv), and the histone marker associated with heterochromatin, H3 lysine 27 trimethylation (H3K27me3) (v) (GEO accession: GSE121283) [145]. (A) TIVs, (B) SNVs and INDELs distribution calculated by 100 kb windows. Light pink to red color bar is in the upper right. (C) the averaged CNV changes calculated in the 10 kb window. Blue to red color bar is in the upper right corner.

2.3.1.2.1 CNVs

To detect and map copy number variations (CNVs) including large segmental duplications and deletions, the single-base resolution read depth was calculated for the whole genome

before plotting the median values for 10 kb windows (**Figure 2.4**). Overall the detected CNVs were more likely to occur in accessory chromosomes (ACs) (*p*-value with Fisher's exact test for the nonrandom associations between the ACs and regions with > 10% CNV change in any sample: 7.08×10^{-253}) (**Figure 2.3**C) and the CNV hotspots (regions showing CNVs in more than one sample) were independent of the selection environment.

Five recurrent large segmental events were detected:

i. Loss of minichromosome 13 was observed in all complete medium-passaged populations and the M1-10 and M4-10 populations with varying frequencies in the populations.
ii. Deletion of 1 Mb region of chromosome 3 was observed in P2-10, Y1-10, M1-10, and M3-10.
iii. Duplication of a 2.1 Mb region of chromosome 3 seems to be linked to duplication of a 0.5 Mb region of chromosome 15. These sequences could be located in the same chromosome. This event appears to be linked with event ii and was observed in P2-10, Y1-10, M1-10, M3-10, and M5-10 populations.
iv. Duplication of an additional 1.2 Mb sequence of chromosome 3 was observed in two minimal medium-passaged populations (M1-10 and M5-10).
v. Deletion of the second copy of chromosome 15. This event was observed in all complete medium passaged populations, as well as in the P5-10, M3-10, and M5-10 populations.

Additional sample-specific small (< 500 kb) CNV events were detected in core chromosomes (CCs) 4, 9, and 10; and ACs 2, 3, and 14 were detected in CCs are all located in the subtelomeric regions.

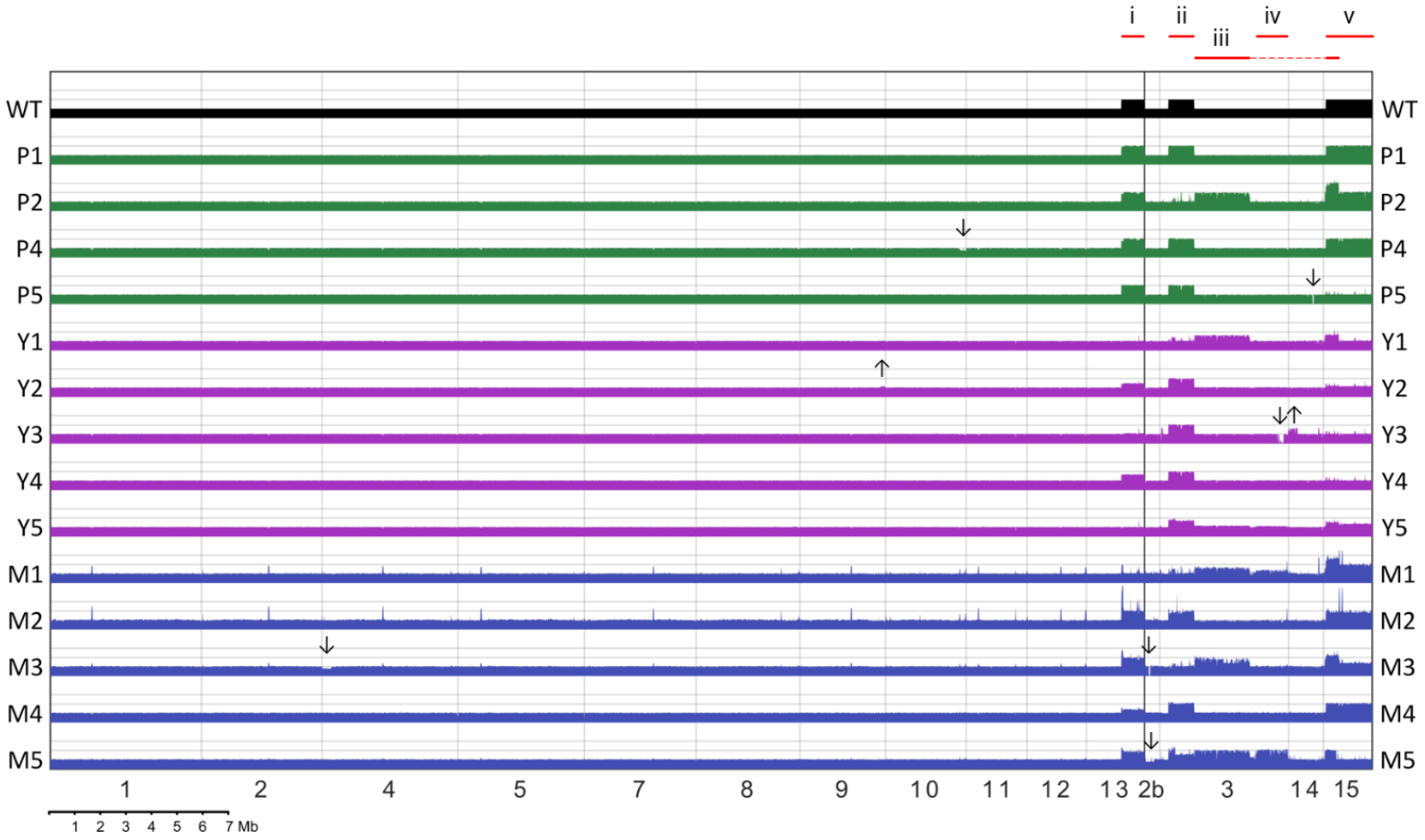


Figure 2.4 The copy number variations (CNVs) in the final populations are presented as read depth over the whole genome of Fol4287. The median read depths calculated over 10 kb-window regions were normalized first to the median read coverage of the sample, and then to WT. Note that some regions in WT are duplicated and thus represented as 2× coverage, such as the small arm of chromosome 13 (corresponding to the minichromosome), a large duplication in chromosome 3, and a whole chromosomal duplication of chromosome 15. The contig order of the accessory chromosomes is explained in Methods (2.2.3.2.3). CNV events that are observed in more than one sample are annotated as roman numerals with the regions indicated with red lines. The dotted red line represents linked events. Unique CNVs are shown as arrows pointing down for decreased copy number or up for increased copy number.

Most of the CNVs were confirmed experimentally by the Di Pietro Lab at the University of Cordoba using qPCR with specific primers for the CNVs in chromosomes 3/6, 1/15, 13/minichromosome, or 14 as well as by pulsed-field gel electrophoresis and Southern blot [146].

Large segmental duplications and deletions events resulted in genome size variation. In the populations passaged through complete medium, the genome size was reduced by 1.4 Mb on average while in plant and minimal medium-passaged populations, it was increased by 0.6 and 2.3 Mb, respectively (**Figure 2.5**).

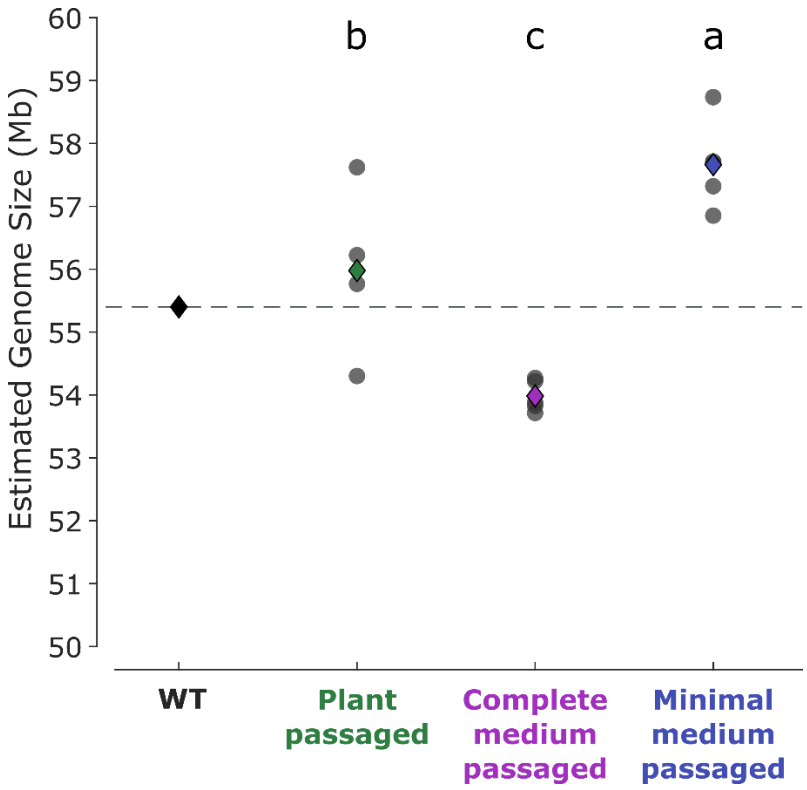


Figure 2.5 Genome size changes of final evolved populations. Gray dots: estimated genome sizes of the final populations. Colored diamonds: estimated mean genome sizes of the final evolving populations. Black diamond and gray dotted line: estimated genome size of the ancestor. One-way ANOVA was performed for final populations which returned the *p*-value of 9.8×10^{-5}. The different letters above the plot indicate the significantly different groups.

2.3.1.2.2 SNVs and INDELs

The single nucleotide variations (SNVs) and small insertions deletions (INDELs) were analyzed using the GATK best practices pipeline [86]. A total number of 71 SNVs and INDELs were identified among the 14 final populations. Among those, 58 passed the allele frequency cut off (**Table 2.2** & **Figure 2.6**). On a global level, these SNPs are randomly distributed across the whole genome (**Figure 2.3**B).

Table 2.2. The number of filtered mutations in final populations. SNV: Single nucleotide variation, INDEL: small insertion or deletion, TIV: transposon insertion variation.

	Samples	*SNV*	*INDEL*	*TIV*	*Total*
Plant-passaged populations	P1-10	4	1	6	11
	P2-10	4	1	4	9
	P4-10	1	0	8	9
	P5-10	9	2	10	21
	All	*18*	*4*	*28*	*50*
Complete medium-passaged populations	Y1-10	2	1	4	7
	Y2-10	1	2	6	9
	Y3-10	2	2	4	8
	Y4-10	2	0	5	7
	Y5-10	4	0	7	11
	All	*11*	*5*	*26*	*42*
Minimal medium-passaged populations	M1-10	4	0	2	6
	M2-10	1	0	0	1
	M3-10	4	0	3	7
	M4-10	7	1	4	12
	M5-10	2	1	1	4
	All	*18*	*2*	*10*	*30*
Total		**47**	**11**	**64**	**122**

SNV and INDEL numbers are consistent between evolved populations with an average of 4 SNPs and 1 INDELs per sample. On average, plant-passaged populations have 5.5, complete medium-passaged populations have 3.2, and minimal medium-passaged populations have 4 SNPs and INDELs. One outlier is the plant-passaged P5-10 population with twice as many mutations as the rest. The other is a minimal medium-passaged M2-10 population with only one SNV.

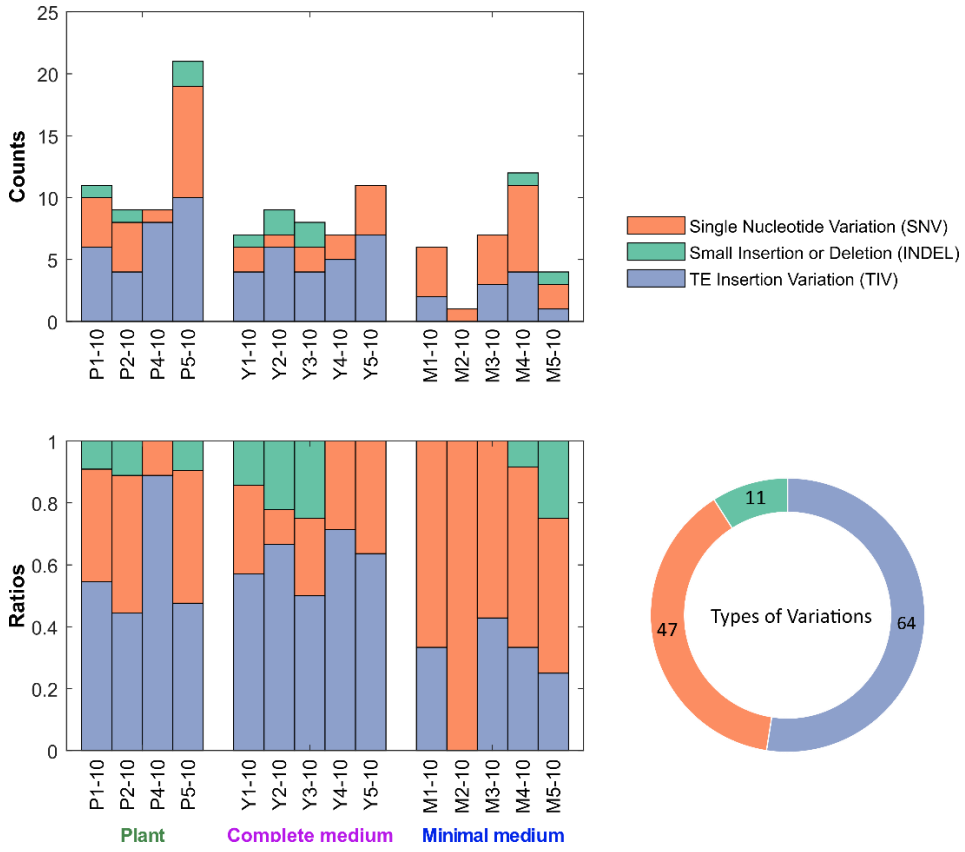


Figure 2.6 Mutation summary of final populations. (A) The number of filtered mutations in final populations including single nucleotide variations (SNVs), small insertions and deletions (INDELs), and TE insertion variations (TIVs). (B) The ratio of observed mutations in A. (C) A pie chart of the total number of mutations in all final populations.

Only one SNV in an intergenic region was observed exactly at the same position in two independent populations, P4-10 and M3-10. Only one gene, FOXG_00016, was mutated in multiple final populations (Y2-10, Y3-10, and Y4-10) (more details in 2.3.2.2).

For plant-passaged populations, 10 SNVs and INDELs reached fixation, while in both complete and minimal medium-passaged populations 4 events reached fixation (**Figure 2.7**A-C).

To ensure the accuracy of variant callings, some events were confirmed with PCR by and Sanger sequencing Cristina López Díaz in Di Pietro Lab [146].

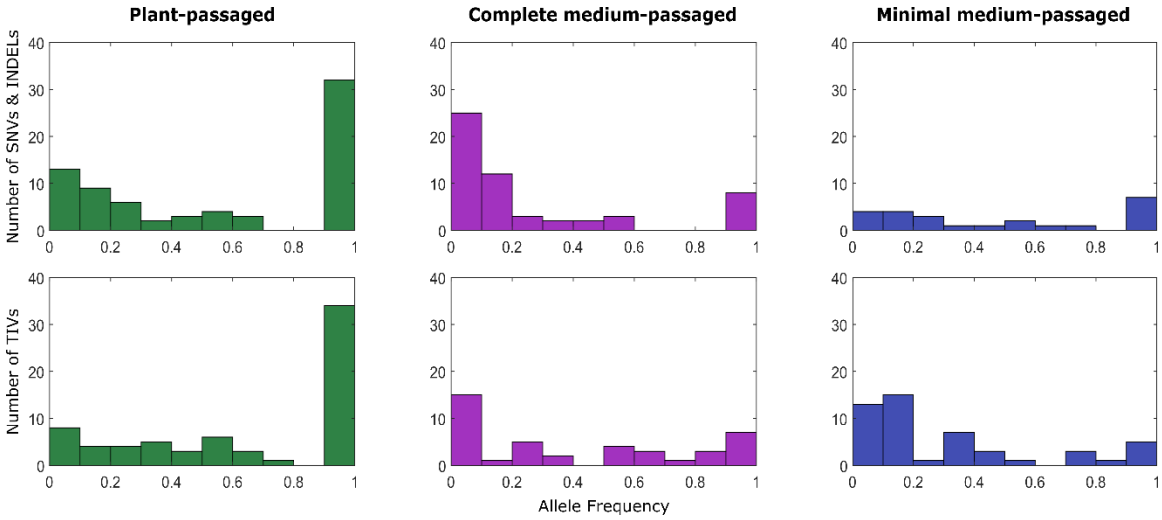


Figure 2.7 Allele frequency (AF) distributions of the final evolved populations. AF distributions single nucleotide variations and small indels (SNVs & INDELs) (A-C) and transposon insertion variations (TIVs) (D-F). Green: plant-passaged, purple: complete medium-passaged, blue: minimal medium-passaged populations.

2.3.1.2.3 TIVs

TE insertions in sequenced final populations were detected by TEfinder, a tool we developed for this study [138]. Briefly, it utilizes the pair information of the discordantly mapped reads in Next-Generation sequencing to detect read pairs whose one read mapping a known TE whereas the other mapping a different location. The read clusters of the pairs to known TE mapping reads are identified as a new insertion site. After the filtering of the possible insertion sites present in the WT sample, AFs were estimated manually by inspecting the alignments in Integrative Genomics Viewer (IGV) [137]. Of 82 detected TIVs, 64 passed AF filter of 10%, accounting for more than 50% of all variants called among all final populations. On average, plant-passaged populations had 7, complete medium-passaged populations 5.2, and minimal medium-passaged populations 2 TIVs per sample. Overall, minimal medium-passaged populations had lower TIV counts than plant and complete medium-passaged populations (**Table 2.2** & **Figure 2.6**).

TE insertions are biased for genomic sites with H3 lysine 27 trimethylation.

In contrast to the random distribution observed for SNVs and INDELs, TIVs had a biased distribution toward accessory chromosomes (ACs). In core chromosomes (CCs) the frequency of novel TE insertion was 1.03 events/Mb. While in ACs, it was 3.86 events/Mb, which is 3.75 times higher than CCs (**Figure 2.3**A). Moreover, TE insertions were accumulated at the ends of CCs. Both ACs and ends of CCs of *F. oxysporum* f. sp. *lycopersici* 4287 are known to be H3 lysine 27 trimethylated [145].

To test the hypothesis that TE insertions are more likely to occur at genomic loci with H3K27me3, a set of publicly available chromatin immunoprecipitation sequencing (ChIP-seq) data of histone markers H3K4me2 and H3K27me3 with Fol4287 generated by Fokkens et al. [145] was used (NCBI GEO accession: GSE121283). H3K4me2, a marker of euchromatin, was mainly observed in the CCs with exception of chromosomes 11, 12, and 13, which are also known as “fast-core” [145], and was absent in the subtelomeric and centromeric regions confirming the analysis performed in the original study (**Figure 2.3**iv). On the other hand, H2K27me3, a marker for heterochromatin, was mainly observed in ACs, fast core chromosomes 11, 12, and 13, and subtelomeric and centromeric regions of CCs (**Figure 2.3**v).

While the 1 kb genomic loci with an observed SNV or INDEL followed the background distribution of the whole Fol4287 genome histone marker profile (**Figure 2.8**C-D), TE insertions are accumulated in locations with H3K27me3 (**Figure 2.8**A-B). This bias towards a heterochromatin-associated histone marker was further supported via one-tailed Fisher’s Exact Test for the hypothesis that this association is random which returned *p*-values of 0.28 and 6.12×10^{-16} for SNVs & INDELs and TIVs, respectively.

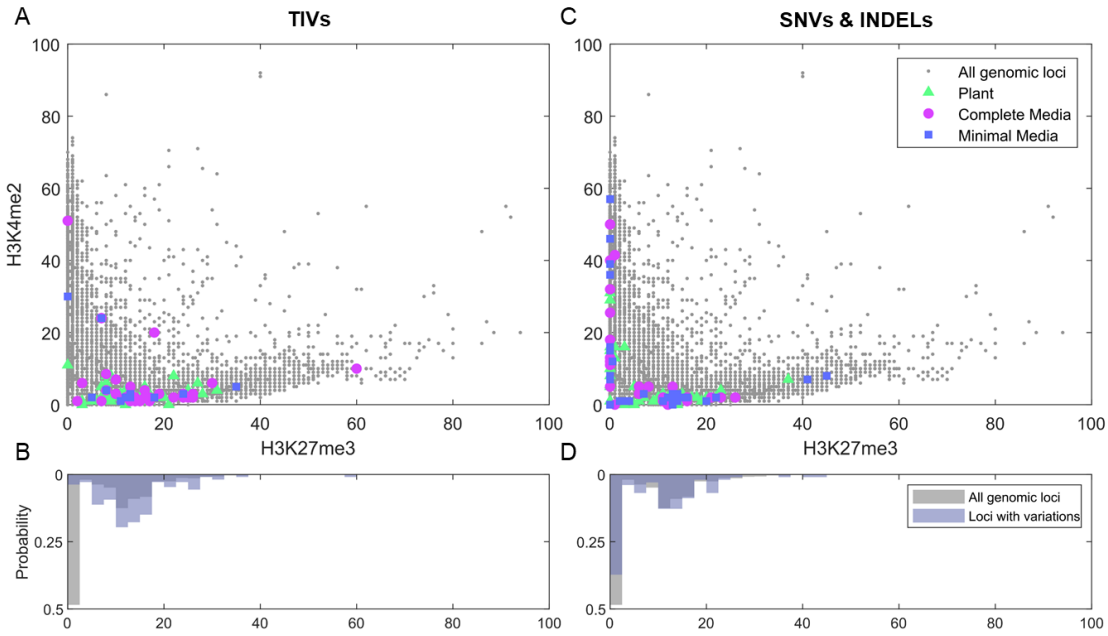


Figure 2.8 Histone marker levels estimated of genomic loci with variations. (A, C) The scatter plot of H3K27me3 and H3K4me2 read depths of 10 kb genomic windows from ChIP-seq mapping, (B, D) the histogram of H3K27me3 read depth values. The loci with TIVs (A, B) and SNVs & INDELs (C, D) are emphasized.

The activity of DNA Transposons is the major cause of variation in plant and complete medium passaged populations.

In this short-term experimental evolution study, more than 90% of the TIVs events were caused by Class II DNA transposons (**Table 2.3**). A hAT superfamily TIR element "miniature Hornet" or Hormin (also called Hornet-small) constituted half of the total insertion events. While Hormin was the most active TE in plant and complete medium-passaged populations, only one novel Hormin insertion was detected in minimal medium-passaged populations. The second most active transposon was a TcMariner superfamily TIR transposon Fot1Active which constituted 23% of all TIV events. By contrast, retrotransposons (LTR, LINE, SINE) have very limited activity overall and are not active in complete medium-passaged populations. A single insertion of Helitron was detected in a plant-passaged population.

Table 2.3 The total number of transposon insertion variations by the active transposable elements in the final populations.

Class	*Order*	*Superfamily*	*TE*	*Plant*	*Complete medium*	*Minimal medium*	*Total*
DNA	TIR	hAT	Hormin	13	18	1	32
DNA	TIR	TcMariner	Fot1Active	9	3	3	15
DNA	TIR	hAT	hAT105Active	1	2	2	5
DNA	TIR	Mutator	HopAy267761	0	2	0	2
DNA	TIR	hAT	Merry	1	0	0	1
DNA	TIR	hAT	hAT210-short-Active	0	0	1	1
DNA	TIR	hAT	hAT210Active	0	1	0	1
DNA	Helitron	Helitron	Strider	1	0	0	1
Retro	LTR	Copia	CopiaActive	2	0	0	2
Retro	LTR	Copia	Gollum-full	0	0	2	2
Retro	SINE	Foxy	FoxyAJ250814	1	0	0	1
	Unknown		UnknownActive	0	0	1	1

More than half of the variations detected in the final populations are TIVs and more than 70% of these are caused by Hormin and Fot1Active, the two most abundant TIR transposons in the Fol4287 genome (**Figure 1.9**D). Hormin (759 bp size) is a miniature element of its full-length counterpart Hornet that is about 3 kb long and encodes a functional transposase (**Figure 2.9**A). Therefore, the transposition of Hormin is likely to depend on the transposase produced by Hornet. Using the RNA sequencing data generated in complete medium, expression of different TEs was detected and the transposons active in the evolution experiment were found to be all highly expressed. Hornet was the highest expressed DNA transposon and one of the highest expressed TEs overall and certainly could be responsible for the transposition of the Hormin (**Figure 2.9**B).

In plant-passaged populations, 12 TIVs reached fixation, while 6 and 3 TIVs in complete medium and minimal medium-passaged populations, respectively. While no same site insertion was observed, 5 evolved populations had TIVs in the same gene (see 2.3.2.2). Out of 64 total filtered TIVs, 28 were in a gene (23 in the coding region) and 36 were in intergenic regions.

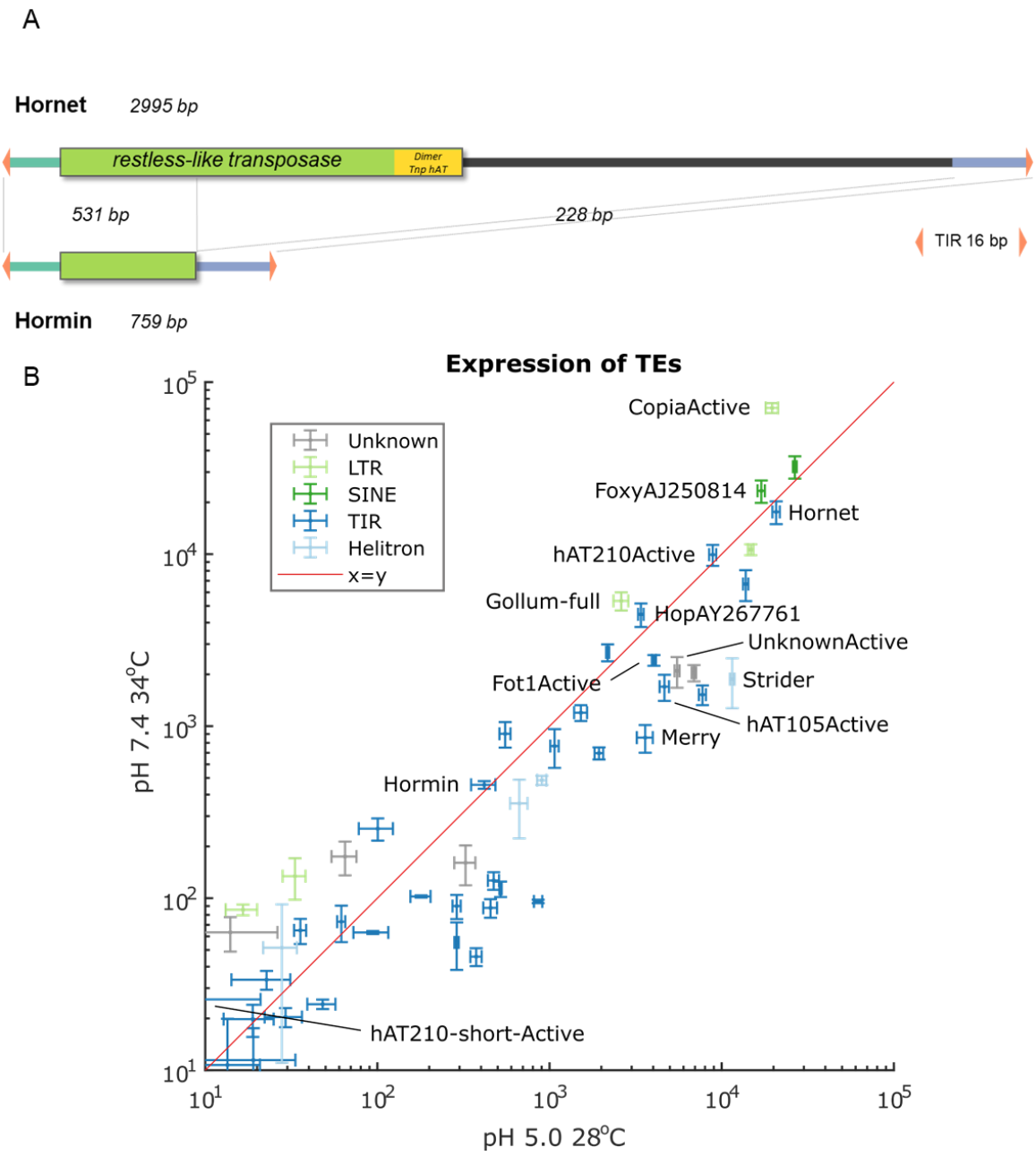


Figure 2.9 Active transposable elements in short-term evolution experiment. (A) Sequence annotations of Hornet and Hormin hAT/TIR transposons. The homologous parts are indicated with dotted lines. The sequence of restless-like transposase with hAT family C-terminal dimerization region (Dimer_Tnp_hAT) is shown in green and yellow boxes. The nonhomologous part is shown in gray while 16 bp terminal inverted repeats (TIR) are shown as orange arrowheads. (B) Normalized expression of different Fol4287 TEs in YPD medium pH 5.0 at 28°C and in YPD medium pH 7.4 at 34°C. The active TEs in the short-term evolution experiment are labeled. The error bars represent the standard deviation of three biological replicates.

Thirteen Hormin insertions in complete medium-passaged populations were selected for experimental validation by PCR, which confirmed the predictions in 12 out of 13 cases.

The one TIV event that failed in PCR amplification only had 22% allele frequency in the population [138].

2.3.2 Signatures of selection observed in evolving populations

To further understand the population dynamics in the evolving lines, all intermediate passages from population P2 as well as the 1st, 5th, and 9th passages from Y3 and M4 populations were sequenced (**Table 2.4**).

Table 2.4 The resequencing of intermediate populations of P2, Y3, and M4 populations. * Read counts and GC% were calculated by FastQC [136]; † median coverage was estimated by Picard CollectRawWgsMetrics [76]; ‡ Mapped read % was calculated by Samtools flagstat [77].

Sample	*Platform*	*Read count* *	*Read length*	*GC%* *	*Median Coverage* †	*Mapped read %* ‡
P2-1	NovaSeq 6000	33740585x2	150	47	129	99.84
P2-2	NovaSeq 6000	35350164x2	150	47	133	99.52
P2-3	NovaSeq 6000	30808735x2	150	47	149	99.46
P2-4	NovaSeq 6000	40135158x2	150	47	117	99.83
P2-5	NovaSeq 6000	25129836x2	150	47	97	99.84
P2-6	NovaSeq 6000	29422555x2	150	47	110	99.83
P2-7	NovaSeq 6000	29334468x2	150	47	112	99.81
P2-8	NovaSeq 6000	34525828x2	150	47	130	99.81
P2-9	NovaSeq 6000	30816598x2	150	47	118	99.83
Y3-1	NovaSeq 6000	36650345x2	150	47	144	99.80
Y3-5	NovaSeq 6000	33840605x2	150	47	135	99.79
Y3-9	NovaSeq 6000	32487798x2	150	47	129	99.83
M4-1	NovaSeq 6000	31458985x2	150	47	124	99.79
M4-5	NovaSeq 6000	28297638x2	150	47	111	99.82
M4-9	NovaSeq 6000	26822474x2	150	47	108	99.81

P2 was selected because the final population showed an average number of SNVs, INDELs, and TIVs. Complete medium-passaged Y3 and minimal medium-passaged M4 populations were selected for their dramatic pathogenicity phenotype and the presence of mutations in genes of the Velvet complex (see 2.3.2.2). The 1st and 9th passages of the Y3 and M4 lines were sequenced to estimate how many mutations emerged in a single passage, while the 5th passage was selected to monitor the populations at the midpoint of the experiment. This set of data was generated using Illumina NovaSeq 6000 platform

with an average of 123× read coverage and had over 99% mapping ratio to the reference. The same analysis methods as in the final populations were applied. In addition to the previously discovered mutations, 33 new SNVs & INDELs (P2: 7, Y3: 12, M4: 14) and 26 new TIVs (P2: 13, Y3: 10, M4: 3) were identified.

2.3.2.1 Evidence for strong selection in the plant-passaged P2 population

Sequencing of intermediate populations enabled higher confidence in the variant calls with lower frequencies when these variants were detected in multiple intermediate populations. Therefore, many background mutations were confirmed in the intermediate populations.

Sequence data from all populations of the P2 line captured emergence, growth, and in some cases extinction of evolutionary events (**Figure 2.10**). A total of 9 SNVs, 3 INDELs, and 17 TIVs were detected among all populations of the P2 sample; and 4 SNVs, one INDEL, and 4 TIVs were fixed in the final population. Out of 29 detected mutations, 23 passed the AF > 0.1 filter at least one of the passages.

The first mutations fixed in the evolving P2 line were a Hormin insertion in a Zn(2)-C6 fungal-type transcription factor gene (T068 in FOXG_20646) and a Fot1Active insertion downstream of an FA desaturase domain-containing protein-encoding gene (T047 near FOXG_15821). These mutations were fixed as soon as they emerged in a single passage. In the second passage, two new mutations emerged. After their fixation, 4 new mutations emerged at the end of the first passage: two TIVs, one SNV, and one INDEL. By the second passage, only the INDEL in an intergenic region had survived and was fixed (I019) while others went extinct. Four new mutations emerged (2 SNVs, 1 INDEL, and 1 TIV) in the next passage, and again only one of them (S001, a nonsense SNV in FOXG_11096, Amb_all domain-containing protein-encoding gene). This pattern, with an average of 2.9

new mutations emerging, and 0.9 of them getting fixed in 2.2 passages, continued for all 10 passages (**Figure 2.10**).

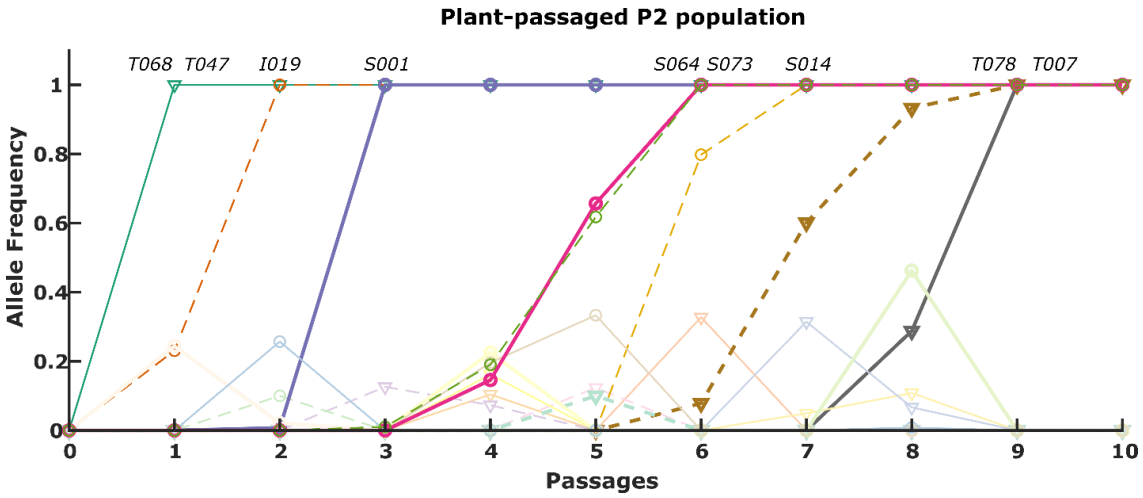


Figure 2.10 The mutation dynamics of the plant-passaged P2 population. TIVs are shown as triangles while SNVs and INDELs are circles. Dotted lines are mutations in accessory chromosomes, solid lines are in core chromosomes. Mutations in coding sequences are plotted with tick lines, in non-coding regions are with thin lines. The fixed mutations are in dark colors and the extinct mutations are lighter colored. The names of the fixed mutations are indicated above the fixation point. Only the mutations that passed AF > 0.1 filter are shown.

Reaching fixation within a short evolutionary time indicates successive selective sweeps. The mutation pattern observed along the P2 line is characteristic of a population evolving with strong mutation under strong selection [116].

The strong selection pattern was uniquely observed in plant-passaged populations. This is also supported by the distribution of allele frequencies of all variations among the final populations. In three evolving lines, the AF distributions follow a bimodal distribution for both TIVs and SNVs & INDELs (**Figure 2.7**) with peaks at AF ranging between 0–0.1 and 0.9–1. However, only in plant-passaged populations, the peak is the greatest at AF 0.9–1 which suggests a different type of selection acting on *in vitro*-passaged populations.

A total of 29 and 24 mutations were observed in Y3 and M4 populations, respectively (**Figure 2.11**). While these numbers are comparable to the P2 population, the numbers of

fixed mutations were only 5 and 2, respectively, further supporting the different selection strengths between *in vivo* and *in vitro* passages.

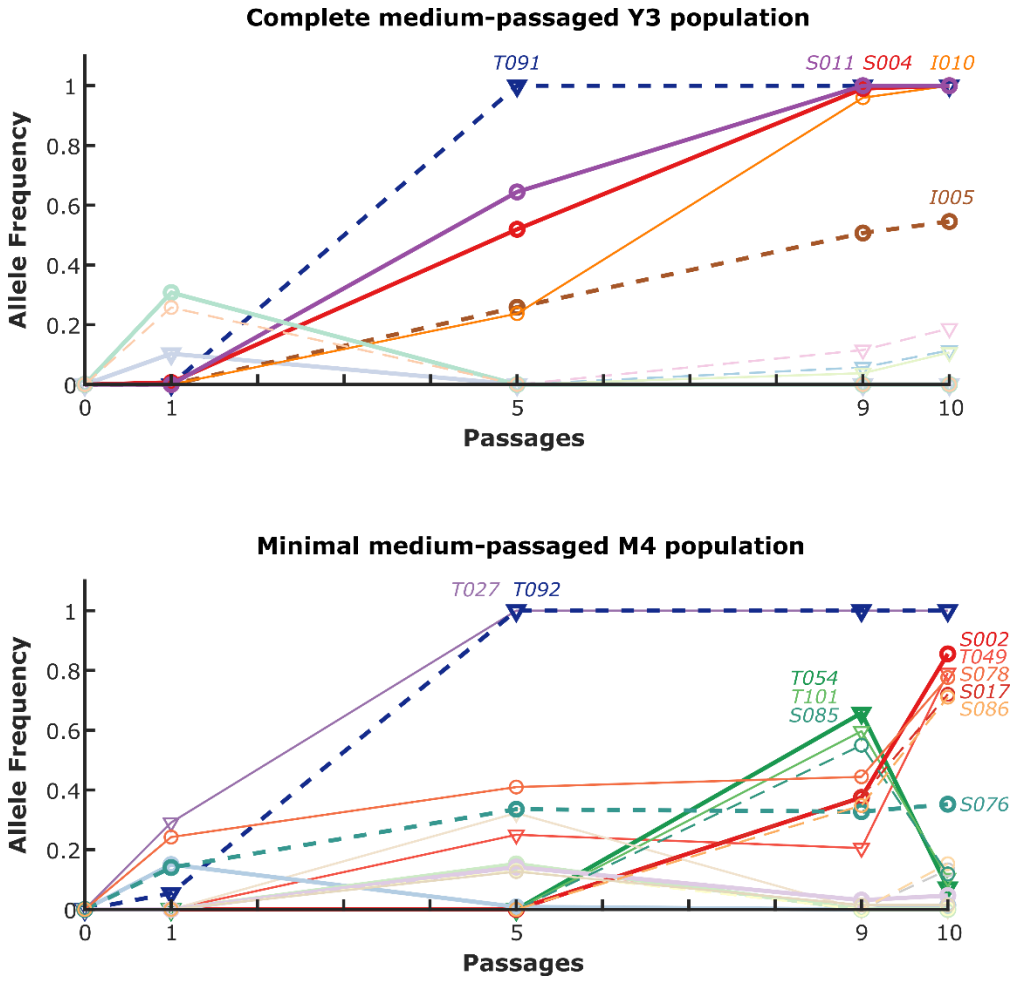


Figure 2.11 The mutation dynamics of the plate-passaged Y3 (A) and M4 (B) populations. Mutations in *velB* (S004) and *veA* (S002) genes are in red, while the transposon insertions in FOXG_21009 (T091 and T092) are in dark blue. For B, the mutations belonging to two competing subpopulations are shown in red gradient colors (S002, T049, S078, S017, and S086) and green gradient colors (T054, T101, and T085). TIVs are shown as triangles while SNVs and INDELs are circles. Dotted lines are mutations in accessory chromosomes, solid lines are in core chromosomes. Mutations in coding sequences are plotted with tick lines, in non-coding regions are with thin lines. The high allele frequency (AF) mutations are in dark colors and the extinct or low AF mutations are lighter colored. The names of the fixed or high AF mutations are indicated by the same color as the line.

When only the successive passages were considered (0^{th}–1^{st} and 9^{th}–10^{th}), the average mutation rates per passage for Y3 and M4 populations are 6 and 3.5, respectively. However, 95% of the emerging mutations are observed in the 1^{st} passage and only one mutation is observed in the 10^{th} passage of M4 and none in the 10^{th} passage of Y3.

Similarly, the number of new mutations in P2 populations decreases in the subsequent passages (**Figure 2.12**).

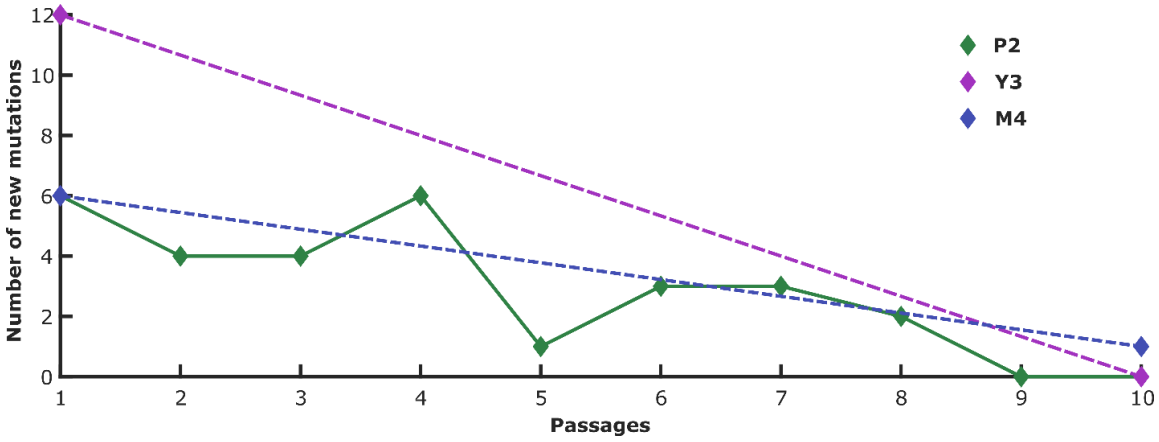


Figure 2.12 The number of new mutations emerged in single passages in P2, Y3, and M4 populations. Dashed lines indicate missing data.

2.3.2.2 Sequential mutation pattern observed involving Velvet Complex genes

Among 64 mutated genes in this study (including mutations in 1 kb up and downstream of the gene), two carried had mutations in multiple evolving populations: FOXG_21009 and FOXG_00016, which were selected in 5 and 4 independent populations, respectively.

Strikingly, four out of five complete medium-passaged (Y2, Y3, Y4, and Y5) populations carried TIVs in an uncharacterized gene FOXG_21009 (**Figure 2.13**A). These four events occurred at different locations of the coding regions of the gene. Interestingly, the same mutation event is also observed in one minimal medium-passaged population (M3). All the TIVs in complete medium-passaged populations were caused by Hormin and mutations in Y2-10, Y3-10, and Y4-10 populations were fixed, while in the Y5-10 population it reached 0.6 AF (**Figure 2.13**A). The mutation in the M4-10, minimal medium-passaged population, was caused by another hAT superfamily transposon (labeled “hAT105Active” for this study) and was also fixed in the population. While fixed mutations in plate-passaged populations were rare (**Figure 2.7**E-F), and this TE insertion event was selected

in independent passages, no such mutation was observed among plant-passaged populations. Most strikingly all populations with a TIV in this gene showed reduced virulence on tomato plants (**Figure 2.1**C-D).

Gene FOXG_21009 is located on accessory chromosome 3, encodes an uncharacterized protein of 255 aa, and is expressed in complete medium conditions (data not shown).

The other recurring mutations appeared in the Velvet complex. Among 16 SNVs and INDELs in complete medium-passaged final populations, 3 occurred in the *velB* gene, which encodes a subunit of the Velvet complex. In Y3-10 and Y4-10, nonsense mutations resulted in truncated proteins while in Y2-10, there was a 14 amino acid deletion (**Figure 2.13**A).

The Velvet complex is a well-characterized major regulator in filamentous fungi containing three subunits, VelB, VeA, and LaeA (**Figure 2.13**B). The activated Velvet complex promotes growth and pathogenicity-related functions such as hyphal growth, mycotoxin production, and nitrogen utilization while inhibiting sporulation [21,22,147]. In addition to the mutations in *velB*, a nonsense mutation in the subunit *veA* was also observed in the M4 line.

There was clear epistatic relation between the recurring TIVs in the gene FOXG_21009 and the SNVs/INDELs in the Velvet complex genes. All populations with fixed mutations in FOXG_21009 had an additional mutation in the coding sequence of either *velB* or *veA* genes. To investigate the dynamics of the mutations in FOXG_21009 and the Velvet complex genes, the intermediate passages were examined (**Figure 2.11**). Although not all passages were sequenced, it can be noted that mutations in the Velvet complex genes (red: S004 and S002) are emerging after fixation of the TE insertion events in FOXG_21009 (dark blue: T091 and T092), both in Y3 and M4 populations. For Y2 and Y4

populations, the same sequential order of mutation events was confirmed by PCR in the Di Pietro Lab [146]. This further supports an epistatic relationship between the two genes, one located in CC and the other in AC.

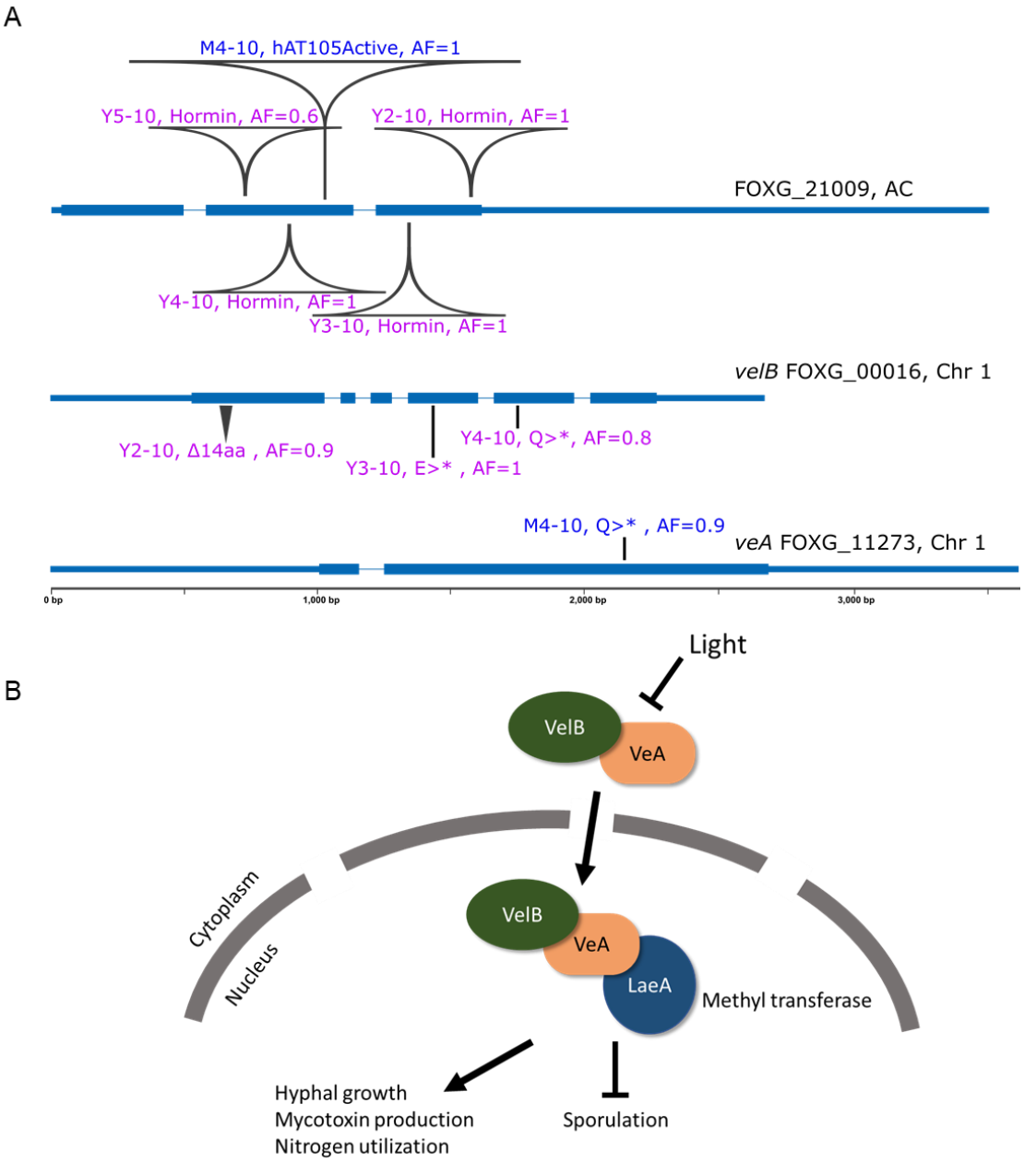


Figure 2.13 Mutations related to the Velvet complex. (A) TE insertions in FOXG_21009, an accessory chromosome (AC) gene, SNVs and an INDEL in *velB*, and an SNV in *veA* genes, top to bottom. The gene models and the transposon sizes are relative to the coordinates. The amino acid changes caused by SNVs (black lines) and INDEL (black triangle) are noted. AF: allele frequency, *: stop codon. (B) The schema of the Velvet complex mechanism [21].

Knockout mutants of *velB* and *veA* showed decreased virulence in tomato plants [22,147]. All the mutations in the Velvet complex genes have severe consequences which might explain the virulence decrease in these populations. What are potential evolutionary advantages associated with these changes? In our plate evolution experiments, the selection pressure consists of competition for faster sporulation, one of the two growth phases of *F. oxysporum* the other being hyphal growth. A growth competition assay between the initial population (WT) and the final evolved population as well as a sporulation assay for all confirmed that this short-term evolution assay had already been selected for faster colony growth and higher sporulating individuals [146].

Another interesting observation in the M4 line was that the dominant subpopulation with T054, T101, and S085 mutations in the 9th passage was replaced by a different subpopulation with mutations S002 (*veA*), S086, and S017 in the 10th passage (**Figure 2.11**B). This finding suggests clonal interference between these subpopulations. In both Y3 and M4 evolving lines, many small frequency events and few fixed events were observed, unlike in the plant-passaged population P2 suggesting a weaker selection in the plate condition (**Figure 2.7**).

2.4 Discussion

F. oxysporum is a devastating pathogen that reproduces primarily through asexual reproduction. Understanding genetic mechanisms underpinning its evolvability has been an intriguing research topic. This study demonstrates the feasibility of short-term experimental evolution (STEE) for understanding genetic mechanisms that underpin organism adaptation in the tomato pathogen *F. oxysporum* f. sp. lycopersici. Results collected after 10 passages already illustrate the power in detecting genetic variations, including fixed mutations, and correlating these genotypes with tractable phenotypes such as virulence, fungal growth, and conidiation. Even after a short period of time, experimental evolution revealed several interesting genomic dynamics of this organism and some interesting evolutionary events including genetic parallelism, epistasis, and clonal interference.

2.4.1 The role of DNA transposons

In Fol4287, while the retrotransposons are ancient and scattered through the whole genome DNA transposons have a recent family expansion and accumulated in ACs (see 1.3.3.3) [15]. This activity was also observed in the STEE with Fol4287 as the most striking result was the activity of DNA transposons. Novel transposon insertions displayed a significant bias toward genomic sites with H3 lysine 27 trimethylation, including accessory chromosomes and ends of core chromosomes of *F. oxysporum* f. sp. *lycopersici* 4287 which are known to be H3 lysine 27 trimethylated [145]. Hormin is the most active transposon in STEE, both in planta, and complete medium. Hormin as well as other TIVs were not as active in minimal medium-passaged populations which could be attributed to either the effect of mutations not being beneficial enough to be selected or a general low transposition activity in minimal medium conditions. Since Hormin is a non-autonomous transposon that lacks a transposase-encoding gene, it depends on the expression of the

transposases from other TEs. Our RNA-Seq data generated in complete medium confirms the high level of expression of the transposase of Hornet, the full-length autonomous counterpart of Hormin. Collectively, active transposable elements targeting the flexible accessory genomic regions may result in a high rate of mutagenesis that enables a rapid adaptation of the *F. oxysporum* genome.

2.4.2 Genetic parallelism and predictability of evolution

This short-term evolution revealed parallelism and possible epistasis between the loss-of-function mutations in gene FOXG_21009 and subunits of the Velvet complex, supporting potential predictability of fungal adaptation [103,127]. The Velvet complex is known to be important in plant virulence [21,22,147], however, no previous studies were conducted on the accessory chromosome gene FOXG_21009. While no sequence homology can be predicted, FOXG_21009 is expressed in complete medium conditions. The sequentiality of the mutations where FOXG_21009 acquires a TIV first and the mutation gets fixed in the population before an SNV or INDEL occurs in Velvet complex genes, signals an epistatic relationship between these two genes. The possibility of FOXG_21009 mutation increasing the evolvability of Fol4287 and changing the fitness landscape to prepare the organism to acquire mutations that could otherwise be non-beneficial is worth investigating. This result emphasizes the power of experimental evolution in studying genotype-phenotype correlation, particularly for genes of unknown function.

The mutation in gene FOXG_21009 gets rapidly fixed in the population signaling a very strong beneficial mutation effect. While in the Y3 line, the fixation of the mutation on *velB* is fast, the fixation of *veA* in the M4 line is slower. From the competition and conidiation experiments, mutations in Velvet complex genes are shown to have a high fitness effect due to the selection pressure for faster conidiation. The reason for the delay in the fixation in the M4 line could be due to the clonal interference between two competing

subpopulations. By the 10th passage, the subpopulation with the *veA* mutation wins and dominates the population. If M4 populations were allowed to propagate more, the mutation might also become fixed within one or a few passages.

While all populations with the FOXG_21009 mutation lose virulence on tomato plants, the M1-10 population which lacks this mutation or a mutation in Velvet complex genes has the same phenotype. M1-10 has three fixed or nearly fixed mutations: a nonsynonymous SNV in the gene coding for Guanine nucleotide-binding protein α-subunit, an SNV downstream of a small uncharacterized gene, and an “UnknownActive” element insertion in the coding sequence of the gene encoding for Gentisate 1,2-dioxygenase. Either of these three might be responsible for the observed loss of virulence on the plant host.

2.4.3 Other events commonly observed in evolution experiments

CNV was the most expected type of variation in the evolution experiments given the dynamic nature of the *F. oxysporum* genome. It is also frequently detected in other evolution experiments with fungi [122,148]. These events caused the genome size of complete medium-passaged populations to decrease and minimal-passaged populations to increase in independent populations. A reduction in genome size was observed in previous experimental evolution studies and represents a way to lose unused costly function in a simple environment by decreasing the genetic load [109,111,149]. In minimal medium-passaged populations, however, the genome size increased even more than in plant-passaged populations, which is a more complex environment than the defined minimal medium.

Through the STEE, in the 3 lines in which we sequenced the intermediate populations, a decrease in the novel mutations was observed in the later passages. A decrease in beneficial mutation rate was observed in previous experimental evolution studies and may

be caused by diminishing-returns epistasis, in which each beneficial mutation the fitness effect decreases due to the presence of other beneficial mutations [109,129].

2.4.4 Perspectives and future directions

F. oxysporum has two growth modes: hyphal and conidial. A germ tube forms from a single conidium and starts developing into a hypha. Hyphae continue to extend and conidiation occurs from specialized phialide cells in the hyphae [150]. This two-mode growth cycle makes the estimation of the number of DNA replications over a certain time period, and hence a generation time, impossible with the current knowledge on this fungus. This and our experimental method of focusing on the final populations lead us to disregard the mutation dynamics per generation time.

A concern in the plant passages was the bottleneck effect. If the number of hyphae growing through the xylem is very low, collecting the fungus from the xylem would create a very narrow bottleneck for the population, hence would change the adaptation dynamics. However, we were able to detect varying allele frequencies for the mutations, which suggests that the bottleneck effect was not extreme and the population size was not too small [116].

For the final populations, we set the cut-off limit for the allele frequency to 10%. Since our population size at the bottleneck was 10^6–10^7, the mutant linage subpopulation size needs to have 10^5 individuals to be able to be captured by our detection limit. For any beneficial mutation to escape the random drift and establish the population size of the mutant, n, needs to be at least c/s where s is the fitness effect of the mutation over ancestor's fitness, and c is the half of the offspring number variance [116]. If s is around 2% and c is around 1, then the random drift threshold becomes only 50 individuals after that point the beneficial mutations are established [105]. Since our detection limit is way above the

possible random drift of the neutral and deleterious mutation frequencies, any mutation detected here, provided it is not a sequencing error, should thus be beneficial or a hitchhiking neutral/deleterious mutation. The sequencing errors are mostly taken into account and removed from the analysis by the variant calling software and our high allele frequency cut-off.

Although evolution experiments offer many advantages, they cannot replicate the evolution in real life where many stress conditions exist, and other organisms are sharing the niche. For example, the acquisition of novel genes through horizontal gene transfer cannot be observed under single organism evolution conditions.

The results of this study demonstrate the high evolvability of *F. oxysporum* f. sp. *lycopersici* 4287 in a relatively short time. It also raises a question for the research community who studies this fungus and propagates it under lab conditions where it may get passaged many times. When Fol4287 encounters a new environment, many beneficial mutations occur that increase the fitness in that environment and might bear a cost for its fitness in the host or the isolation environment.

CHAPTER 3 EXPERIMENTAL EVOLUTION OF *F. OXYSPORUM* MRL8996

3.1 Introduction

Fusarium oxysporum has been isolated from soil, air, plants, and animals, and even from International Space Station [151]. Even with such widespread geographic diversity, a single clade was shown to be responsible for the recent emergence of *F. oxysporum* as an opportunistic pathogen in humans [13,34].

A phylogeny study with 88 clinical (isolated in hospital, human, or animal) and 18 environmental (isolated plant or soil) isolates of *F. oxysporum*, placed 90% of all clinical isolates including NRRL 32931, isolated from a leukemia patient with disseminated fusariosis (see Chapter 1) into same clade (Clade 3), and 70% of all clinical isolates single clonal linage inside Clade 3 [34]. Apart from NRRL 32931, Clade 3 also includes Fol4287 and MRL8996 isolates, as well as some plant pathogens such as *F. oxysporum* f. sp. *radicis-lycopersici* NRRL 26381 and *F. oxysporum* f. sp. *melonis* NRRL 26406 [34], the biocontrol *F. oxysporum* 47 isolated from soil [152], and two isolates from International Space Station [151].

As described in Chapter 1, *F. oxysporum* MRL8996 (NRRL 47514), the *Fusarium* keratitis strain, was selected as a representative of human pathogens. The strain was originally isolated in 2006 from the cornea of a patient with contact lens-associated multistate outbreak fungal keratitis at Cleveland Clinic Foundation in Ohio, USA [48].

Even though the plant pathogenic *F. oxysporum* f. sp. *lycopersici* 4287 (Fol4287) and MRL8996 share the same phylogenetic clade, as shown in Chapter 1, they are very distinct in terms of morphology, growth rate, accessory chromosome (AC) gene content, and repeat content.

While the AC genes of Fol4287 are enriched in transcription factors, chromatin assembly disassembly proteins, defense response genes, peptidoglycan metabolism, and biosynthesis, etc., the AC genes enriched in MRL8996 are related to the cholesterol biosynthesis process, organomercury catabolic process, and metal ion transport. Their transposable element (TE) contents are also different: Helitrons and Miniature impala elements which are known to be related to the pathogenicity of the Fol4287 are absent in MRL8996 [29,31].

Some transposons could be inactive, while some are expected to be mobile, and their transposition could contribute to genetic variation both at the individual and the population level. As a matter of fact, in Chapter 2, the activity of the mobile Fol4287 transposons is observed. Most transpositions in Fol4287 are caused by TIR transposons with the highest activity observed in Hormin, a hAT superfamily miniature Hornet element. Hormin however is not present in MRL8996 and therefore no such activity would be expected. In minimal medium passages populations, the activity of Hormin was very limited, and instead, a TcMariner element Fot1Active and other hAT superfamily DNA transposons were active. As shown in Chapter 1, although Fot elements and some hAT elements are present in MRL8996 they have very low abundance in the genome. Instead in MRL8996, an expansion in long and short interspersed nuclear elements (LINEs and SINEs) are observed.

TEs play important roles in many biological processes such as cancer biology [153], stem cell [154], or host-pathogen interactions [155]. They can be utilized by their host as promoters or transcription factor binding sites [156]. They can change the genome by introducing structural variations either through their transposition activity or by providing sequence homology for homologous recombination; disrupt protein sequence and causing loss of function by inserting into genes; affect transcriptional and translational regulation

by insertion in or excision out of the promoter, intronic, or regulatory sequences; or altering gene expression by inserting into an expressed region and causing epigenetic silencing of nearby genes [156–159].

Apart from the differences both strains have, their hosts also differ in many aspects. The plant and animal hosts present different biotic and abiotic stresses for the fungi. The biotic stresses are caused by the host immunity against the fungal pathogen. While the plants have PAMP-triggered and effector-triggered immunity (see 1.1), macrophages and neutrophils are responsible for the immune response against the fungus by damaging the hyphae [64,160,161].

The abiotic stresses include but are not limited to temperature, pH, and nutrient limitation. The pH of the tomato plant xylem is around pH 5-6 while the pH of the eye is around pH 7.4 [162,163]. The human eye has an internal temperature of around 30-35°C with the anterior the lowest temperature [91]. While pH change does not affect the growth rate in either strain, MRL8996 is shown to have a higher tolerance to the temperature (**Figure 1.1**).

In this study, the adaptation of the eye infecting strain of *F oxysporum*, MRL8996 compared to the plant pathogenic *F. oxysporum* f. sp. *lycopersici* 4287. Considering that plant pathogenic and clinical strains differ in terms of genome and phenotype, the hypothesis was that the adaptation mechanisms to human eye conditions are different than the adaptation to the plant host. Similar to Chapter 2, an evolution experiment was conducted using MRL8996 (**Figure 3.1**A). In these experiments, the main focus was pH and temperature adaptation as these aspects are the two abiotic differences between human eye conditions and the plant environment.

3.2 Methods

3.2.1 Short-term evolution experiment

Following medium plates were prepared using 10 mm-diameter petri dishes:

(A) Modified Czapek-Dox (minimal medium, MM) Agar: 0.1% w/v KH_2PO_4, 0.05% w/v $MgSO_4 \cdot 7H_2O$, 0.05% w/v KCl, 0.02% v/v Trace Elements Solution, 0.2% w/v $NaNO_3$, 3% w/v sucrose, and 2% w/v Agar; Trace Elements Solution: 5% w/v citric acid, 5% w/v $ZnSO_4 \cdot 6H_2O$, 1% w/v $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$, 0.25% w/v $CuSO_4 \cdot 5H_2O$, 0.05% w/v $MnSO_4$, 0.05% w/v H_3BO_3, 0.05% w/v $Na_2MoO_4 \cdot 2H_2O$)

(B) Yeast Extract Peptone Dextrose (YPD) Agar: 0.3 % w/v yeast extract, 1% w/v Bactpeptone, 2% w/v dextrose, 1.5% w/v agar)

(C) MM agar with 5.0% v/v 0.1 citric acid and 40.0% v/v 0.2 Na_2HPO_4 for pH 7.4

(D) YPD agar with 6.5% v/v 0.1 M citric acid and 43.5% v/v 0.2 M Na_2HPO_4 for pH 7.4

(E) MM agar with 5.0% v/v 0.1 citric acid and 40.0% v/v 0.2 Na_2HPO_4 for pH 7.4

(F) YPD agar with 6.5% v/v 0.1 M citric acid and 43.5% v/v 0.2 M Na_2HPO_4 for pH 7.4

The same frozen stock of MRL8996, two cultures were started in Potato Dextrose Broth medium (PDB, BD Difco, USA), incubated at 28°C, 140 rpm for 7 days, and the conidia were collected. A and B plates were inoculated with 10^6 conidia and incubated at 28°C for 7 days. With the second starting cultures, C and D plates were inoculated with 10^6 conidia and incubated at 34°C for 7 days. Similarly, the conidia were collected from Fol4287 culture which was incubated in PDB at 28°C 140 rpm for 7 days was inoculated in E and F plates (10^6 conidia each) and incubated at 34°C for 7 days.

After 7 days, the conidia were recollected by adding 4 ml sterile water and transferring the mycelium-conidia suspension through a sterile column with cotton (a piece of cotton placed in a 5 ml Eppendorf epTIPs which is placed in a 15 ml centrifuge tube and positive

pressure is applied via Eppendorf pipette). From the recollected conidia suspension, 10^6 conidia were used to inoculate fresh medium plates of the same kind and the plates were incubated at the same temperature as the previous cultures. This process was repeated 10 times with 5 independent replicates for each medium type. Two samples of evolving D populations were lost due to persistent contamination problems. A portion of collected spores was stored at -80°C after each cycle.

3.2.2 Phenotyping

The phenotyping was performed the same way as the ancestor strains (details in 1.2.2.1). The growth rates of each replicate and dilution were calculated as the slope of the growth curve for 3 days.

The mean growth rates (*R*) of each evolving population (*s*) were normalized to the mean growth rates of its ancestor (*w*) in the same condition ($R_s/R_w = N_s$) and the two-sample *t*-test was performed between *s* and *w* replicates. For nonsignificant values, N_s is set to 1.

Tolerance to high temperature was defined as previously described (1.2.2.2). Each replicate of samples at 34°C (3 replicates) was divided by each replicate of the same samples at 28°C (3 replicates) in the otherwise same condition (total 9 values generated) and mean values are calculated. N_s is calculated for each sample and medium condition, and the two-sample *t*-test was performed between the values for evolving samples and the ancestor samples. For nonsignificant changes, the mean values are set to 1.

3.2.3 Whole genome sequencing

The 10^{th} passages of 28 populations, and their initial ancestor cultures (A0 for MRL8996, and E0 for Fol4287), were grown from the frozen stock in the PDB medium at 28°C 140 rpm for 4 days. The mycelium was collected using 2-layer EMD Millipore Miracloth. The excess liquid was removed using sterile paper towels on the outside of the Miracloth. The

mycelium was portioned into around 200 mg pieces using a sterile spatula, placed in 1.5 ml tubes, and shocked froze in liquid nitrogen followed by storage at -80°C for DNA extraction.

The tissue samples were unfrozen and 500 µl CTAB extraction buffer (3% w/v Cetyltrimethyl ammonium bromide, 1.4 M NaCl, 20 mM Ethylenediaminetetraacetic acid pH 8, 0.1 M Tris-HCl pH 8, 3% w/v polyvinylpyrrolidone MW: 40 kDa, 0.2% v/v β-mercaptoethanol) and 10 µl β-mercaptoethanol are added to the tissue samples and homogenized using a pestle pellet motor. Then, QIAGEN DNeasy Plant Mini Kit was used with the suggested protocol with the following changes: all material was combined into a single column for each sample, an additional centrifugation step was added to remove excess liquid from the tubes after all the wash steps, and the genomic DNA was eluted into 30 µl nuclease-free water.

The extracted DNA was quantified using Qubit dsDNA BR Assay Kit and sequenced by Novogene Company Limited, UK via Illumina NovaSeq 6000 platform with 2×150 cycles.

3.2.4 Variant Calling

The mapping and the variant calling were performed as described previously (2.2.3) with exception of the following changes:

BEDtools genomecov with the option ‘-d’ was used to output the read depth at a single nucleotide resolution for the final populations and the ancestors [78]. Then, a custom MatLab code was used to calculate median coverages of 10 kb windows of the genome. The read coverages were normalized to the median read coverages of the samples. For accessory chromosome sequences (AC) the final population samples were normalized by the ancestor. For core chromosomes (CCs) normalization to ancestor was not applied because of the extreme changes around centromeric sequences caused by very low

coverage in the A0 sample. Hierarchical clustering was performed with the normalized copy numbers for AC contigs were ordered accordingly. This ordering was used for **Figure 3.5**C and **Figure 3.6**.

For Fol4287 samples (E and F populations) the coverages of the small arm of chromosome 13 (2×), the end of the portion of the same chromosome 13 small arm (4×), increased copy number region of chromosomes 3 (3×, 6×, and 4×), and chromosome 15 (3×) were multiplied by indicated numbers to reflect the reality.

For the dynamic AC regions, the number 10 kb bins of AC sequences with more than 10% copy number (CN) change with respect to the ancestor (CN > 1.1 or CN < 0.9) were divided by the total 10 kb bins in ACs for all final populations. The two-way t-test was applied between the values of the evolving populations of MRL8996 and Fol4287.

The passed transposon insertion variation (TIV) events detected by TEfinder were filtered and the events which are also detected in ancestors were removed. The mean coverage of the reads mapped to the insertion site which is determined by TEfinder (R_a) was extracted from the output of BEDtools genomecov which is used for the copy number variations. The number of reads that are evidence for the insertion detected by TEfinder (R_t) was extracted. 14 TIVs are arbitrarily selected and the allele frequencies (AFs) are manually estimated by inspecting them in Integrative Genomics Viewer (IGV) [137]. A linear fit was applied to the AFs with respect to R_t/R_a values which resulted in the function $AF = (R_t/R_a) + 0.01812/1.526$ with R^2 = 0.93. This function was applied to all TIVs to estimate AFs (Appendix B5).

For the boxplots with beeswarm plots, plotSpread function in MatLab was used [164]. One-way ANOVA in MatLab was applied for the number of mutations and the allele frequencies of the mutations in sample sets.

3.2.5 Gene ontology (GO) enrichment analysis

Gene annotations in 1.2.10 with GO_term_enrichment_analysis.m script (available at https://github.com/d-ayhan/tools) which utilizes the hypergeometric cumulative distribution function was used to analyze the GO term enrichments in chromosome 12.

3.2.6 Annotation of Foxy5

The transposable element (TE) sequences are generated previously (Methods 1.2.6). The MRL8996_family-5 element was blasted against annotated TE sequences and had matches with Foxy/SINE family elements [142]. MEGA X was used to generate multi-sequence alignments between available annotates Foxy elements (FoxyAJ250814, Foxy2, and Foxy3) [165]. MRL8996_family-5 is renamed Foxy5.

The motif enrichment in the insertion sites of Foxy5 was analyzed by MEME version 5.0.5 using 500 bp up and downstream sequences of insertion sites (462 sequences with 1 kb sizes) [166].

RNA sequencing was performed as described before (2.2.5) using MRL8996.

3.2.7 Other statistics

Two-sample t-test was applied for the ratio of mutations that are fixed or almost fixed to all detected mutations between MRL8996 and Fol4287 final populations evolving in the same condition.

The 1 kb up and downstream regions of the genes were extracted from the gene annotation file and any genes intersecting with those sequences were removed. The positions which intersect with the mutations were counted and one-tailed Fisher's exact test was applied for the null hypothesis that the association with the mutations and the 1 kb up or downstream regions of the genes are random. Similarly, the coding sequences

(CDS) of the genes were extracted and the mutations on those regions were counted. One-tailed Fisher's exact test was applied.

The differences in allele frequency distributions in Fol4287 and MRL8996 evolved populations were analyzed by Kolmogorov-Smirnov test using the unfiltered numbers of TIVs, SNVs, and INDELs in each set of evolving conditions.

All calculations were performed in MatLab [67].

3.3 Results

3.3.1 Short-term evolution experiment set-up

Similar to the short-term evolution experiments (STEE) in Chapter 2, the ancestor spores of MRL8996 (A0) was passaged in four conditions: minimal medium (MM) agar plates at 28°C (A populations), complete medium (yeast extract peptone dextrose, CM) agar plates at 28°C (B populations), MM with pH adjusted to 7.4 plates at 34°C (C populations), and CM with pH adjusted to 7.4 at 34°C (D populations). For comparison Fol4287 plate-evolved populations from 2.2.1, MM-passaged (M populations), and CM-passaged (Y populations), as well as new sets of populations passaged in MM pH 7.4 and CM pH 7.4 plates at 34°C (E and F populations, respectively) using ancestor Fol4287 spores (E0), were added to the analysis (**Figure 3.1**A).

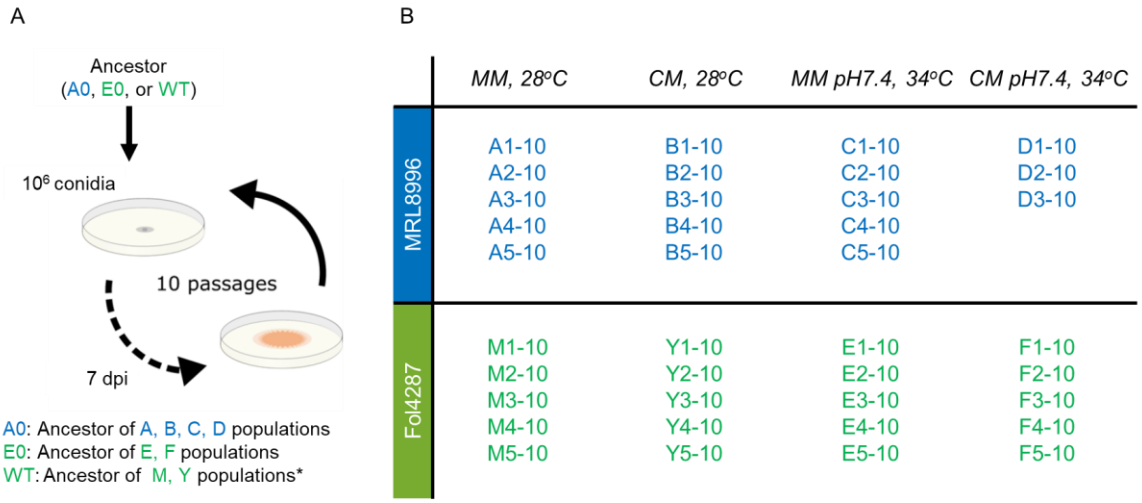


Figure 3.1 Short-term evolution experiments with the clinical isolate MRL8996. (A) The design for the evolution experiments. For comparison between plant pathogen Fol4287, the samples analyzed in Chapter 2 (*) and two more Fol4287 samples were added. The initial 10^6 ancestor conidia (A0, E0, or WT) were inoculated in one of the following *in vitro* conditions: minimal medium (MM, modified Czapex-Dox), yeast extract peptone dextrose (CM), or MM with pH adjusted to 7.4 with citric acid-sodium phosphate buffer (MM pH7.4), or CM with pH adjusted to 7.4 with citric acid-sodium phosphate buffer (CM pH7.4). MM and CM plates were incubated at 28°C (A/M and B/Y populations, respectively), while MM pH7.4 and CM pH7.4 plates were incubated at 34°C (C/E and D/F populations, respectively) for 7 days. Then the conidia were collected, and a fresh plate of the same kind was inoculated. This was repeated for 10 passages. (B) Overview of the final populations. For each condition, there were 3–5 independently evolving populations.

Each sample set had five biological replicates except D populations of which 2 populations were lost due to persistent contamination problems (**Figure 3.1**B). Hereafter the final populations will be called only by the sample name without "-10" since in this study the intermediate populations are not investigated.

3.3.2 Changed phenotype over the short-term evolution experiment

To investigate the phenotypic changes of the evolved population large-scale phenotyping experiment was performed using 7 medium conditions that were described in 1.2.2.1 at 28°C and 34°C.

3.3.2.1 Growth rates in evolving conditions

The hyphal growth rates of final populations were tested in their evolving condition (**Figure 3.2**). Even though the fitness is not directly tested under these conditions, some significant changes can be observed. In evolved MRL8996 populations, two A populations, two B populations, one C population, and two D populations showed significant hyphal growth rate increase when compared to the ancestor (A0), while one A population had a significant growth rate decrease. Overall, these changes were very small in evolved MRL8996 populations regardless of the condition.

Most drastic changes happened in evolved Fol4287 populations passaged at the high temperature (

Figure 3.2). All E and F populations had a significant growth rate increase in their evolving conditions (in MM pH 7.4 at 34°C and CM pH 7.4 at 34°C, respectively) within some cases more than a 2-fold growth rate increase was observed.

There is not much difference observed for the growth rates at pH 5 or pH 7.4 in the same nutrient condition (**Figure 3.3**). However, in different medium conditions at the same temperature as their evolving condition, (e.g., A populations in CM 28°C or B populations

in MM 28°C), while 28°C passaged populations display reduced growth rates when compared to ancestor and their own evolving conditions, 34°C passaged populations showed similar or sometimes higher growth rates (C populations in CM, pH 7.4 at 34°C and D populations in MM pH 7.4 at 34°C).

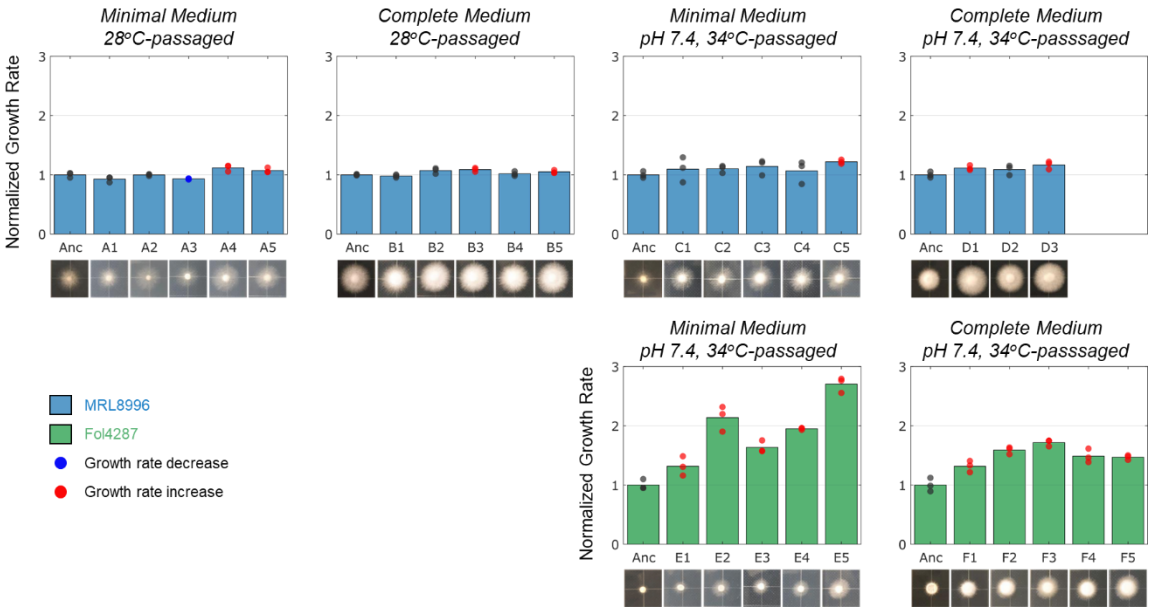


Figure 3.2 The growth rates of the final populations and the ancestors in their evolving environments. The plots represent the growth rates of each population in stated medium condition, normalized to the ancestors (Anc). The bar plot shows the average values of three replicates while data points show their values. Gray data points show the ancestor values or the populations with no significant change when compared to the ancestor. Red and blue data points represent the significant growth rate increase or decrease, respectively (p-value < 0.05 with two-sample t-test). The pictures under the plots are representative of 3 replicates at 2-day post-inoculation. The initial inoculation amount was 10^3 conidia.

3.3.2.2 Growth in stress conditions

If an organism adapts to one type of selection condition, there might be fitness effect reduction in another condition because of loss of function mutations (e.g., virulence decrease of Y-populations mentioned in Chapter 2) [127,167]. This effect was tested using three stress conditions which were described in Chapter 1: 0.6 M NaCl for osmatic stress, 1 mM hydrogen peroxide for oxidative stress, and 1 mg/ml Congo Red for cell wall stress.

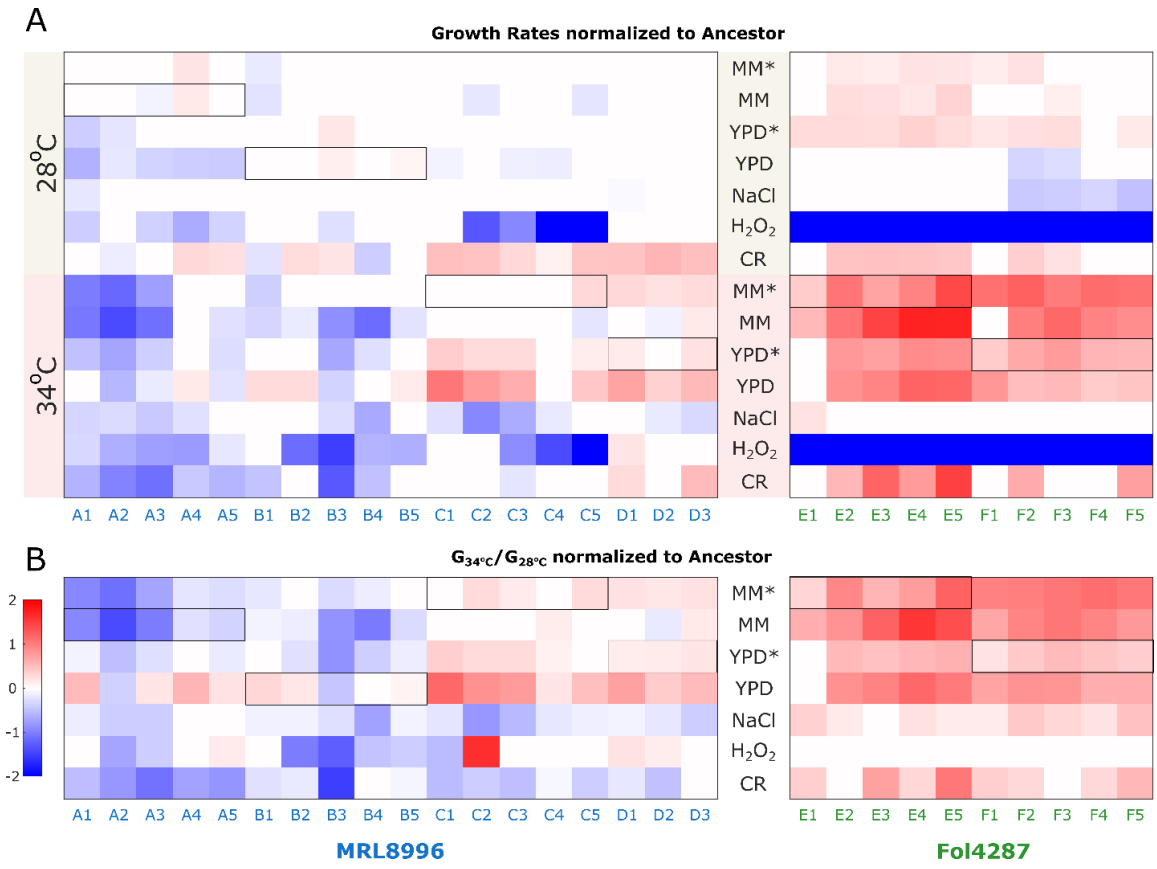


Figure 3.3 Phenotypes of the evolved populations in various conditions. (A) The growth of evolved MRL8996 and Fol4287 populations with initial conidia count of 10^3 incubated in minimal medium (MM), minimal medium pH adjusted to 7.4 (MM*), yeast extract dextrose peptone (YPD), YPD pH adjusted to 7.4 (YPD*), 0.6 M sodium chloride for osmotic stress (NaCl), 1 mM hydrogen peroxide for oxidative stress (H_2O_2), 1 mg/mL Congo Red for cell wall stress (CR) agar plates for 3 days at 28°C or 34°C. The growth rates are calculated as mm per day and normalized to Ancestor (A0 or E0) growth rates, averaged, and $\log_2$-transformed. The conditions where the samples are adapted are shown in black rectangles. (B) The temperature adaptation is shown as the averaged growth rates at 34°C ($GR_{34°C}$) divided by rates at 28°C ($GR_{28°C}$). The values are then normalized to the ancestor (A0 or E0), and $\log_2$-transformed. The values with non-significant changes (*p*-value > 0.05 with two-sample *t*-test) are set to 0.

At 28°C, most of the evolved samples showed increased fitness to the cell wall stress (**Figure 3.3**A & **Figure 3.4**). But for oxidative stress or osmotic stress, the evolving population either showed no significant change or decreased growth rate. In oxidative stress condition, all E and F populations, as well as two of C populations had an all-or-none growth effect that was dependent on the inoculated conidia count and therefore had dramatic growth rate reduction. The three stress conditions when coupled with the temperature stress cause a decreased fitness effect in most of the evolved populations.

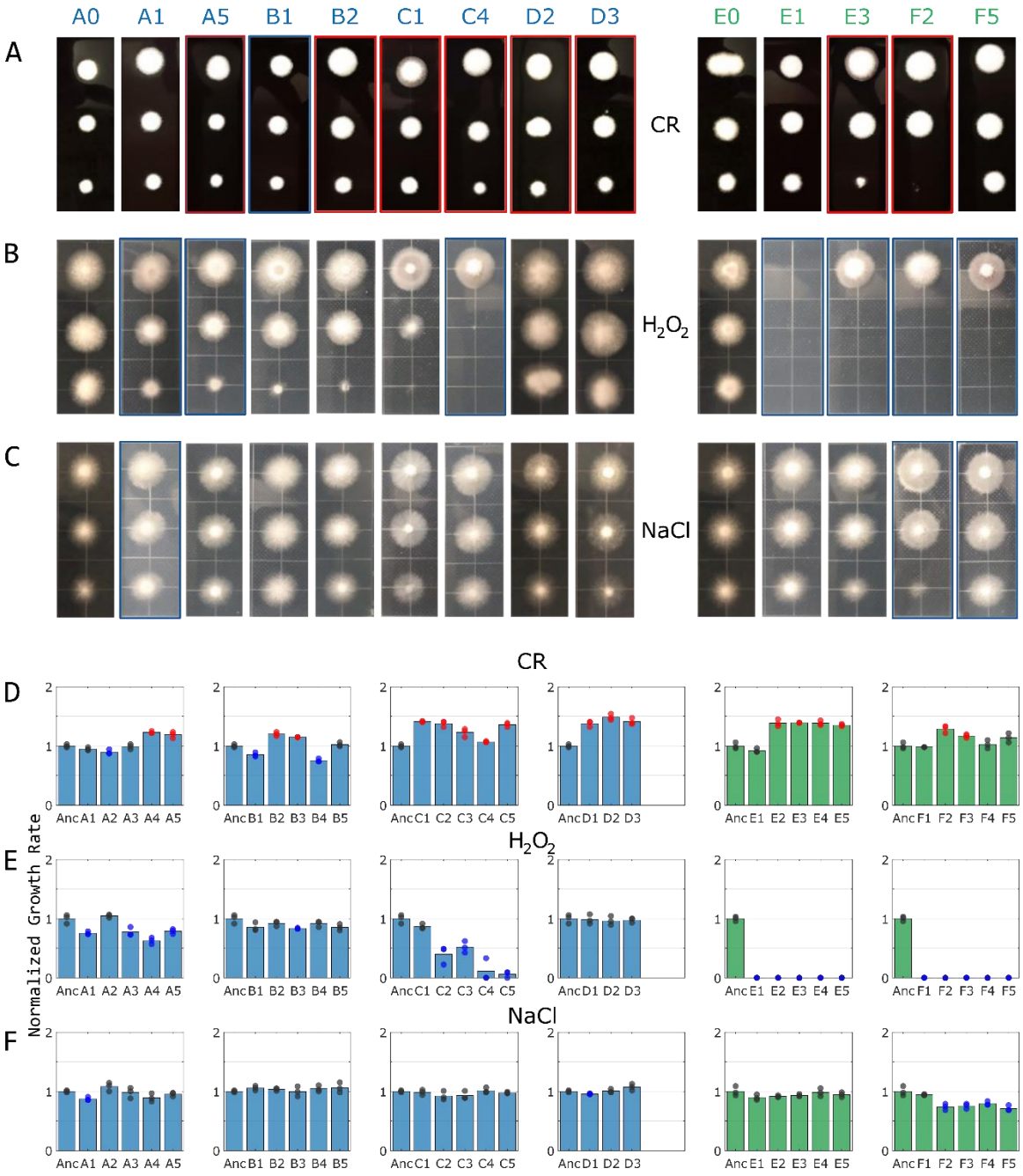


Figure 3.4 The growth of the selected final populations and their ancestors in stress conditions. (A-B) CR: 1 mg/mL Congo Red for cell wall stress; H_2O_2: 1 mM hydrogen peroxide for oxidative stress; NaCl: 0.6 M sodium chloride for osmotic stress agar plates, incubated at 28°C. Each picture is representative of 3 replicates at 2-day post-inoculation with 10^4, 10^3, and 10^2 initial inoculated conidia (top to bottom). The red and blue outlines show a significant increase or decrease, respectively, from the ancestor (A0 or E0) growth considering only 10^3 conidia inoculation (*p*-value < 0.05 with two-sample *t*-test). (D-F) The growth rates of each population (10^3 initial conidia) CR, H_2O_2, and NaCl medium plates, normalized to the ancestors (Anc). The bar plot shows the average values of three replicates while data points show their values. Gray data points show the ancestor values or the populations with no significant change when compared to the ancestor. Red and blue data points represent the significant growth rate increase or decrease, respectively (p-value < 0.05 with two-sample t-test).

3.3.2.3 Temperature adaptation

Overall temperature tolerance (the growth at 34°C relative to the growth at 28°C) is increased in high temperature passaged populations when compared to the ancestor (**Figure 3.3**B). Fol4287 samples showed the highest increase of the fitness effect for temperature tolerance. For A populations temperature tolerance is decreased while for B populations this effect is mixed.

When the growth rate ratios between 34°C and 28°C (not normalized to the ancestors) are considered, C and D populations are observed to have a higher temperature tolerance than E and F populations, and some of A and B populations have slower growth than E and F populations.

3.3.3 Surveying genotypic variation in STEE by whole genome sequencing

3.3.3.1 Data generation and quality

The final populations as well as the ancestors were sequenced using Illumina NovaSeq 6000 platform with 2×150 cycles which produced 38-55 million reads (**Table 3.1** & Appendix A). The raw reads were mapped to the reference genomes of MRL8996 and Fol4287 (1.3.2) and resulted in > 99.6% mapping percentage with 71 to 119 median read depth.

3.3.3.2 Variant calling

Comparing to the ancestor genome assemblies, four types of variations were detected among these evolving populations: single nucleotide variations (SNVs), small insertions deletions (INDELs), transposon insertion variations (TIVs), and copy number variations (CNVs) (summarized in **Figure 3.5**).

Table 3.1 Whole-genome shotgun sequencing and mapping metrics for short-term evolution experiment final populations and the ancestors. MAD: median absolute deviation.

Sample	*Total Reads*	*Mapped %*	*Median of read depth*	*MAD of read depth*
A0	42,891,444	99.92	90	8
A1	42,915,083	99.83	97	9
A2	43,399,174	99.72	97	9
A3	43,260,387	99.91	88	8
A4	44,147,997	99.90	94	9
A5	54,035,923	99.82	113	10
B1	43,382,967	99.91	91	8
B2	37,958,493	99.93	78	8
B3	43,575,971	99.91	95	10
B4	41,319,802	99.91	86	8
B5	41,257,829	99.88	87	8
C1	44,858,844	99.77	96	12
C2	51,745,511	99.66	119	11
C3	40,488,352	99.60	87	9
C4	44,148,299	99.90	99	9
C5	44,978,757	99.82	101	10
D1	40,342,126	99.83	95	9
D2	49,496,334	99.89	109	10
D3	50,064,004	99.78	113	9
E0	45,045,811	99.80	81	8
E1	44,413,103	99.82	84	8
E2	41,070,844	99.82	74	8
E3	47,816,031	99.82	76	8
E4	47,704,258	99.82	89	9
E5	43,330,965	99.78	79	8
F1	47,167,467	99.84	86	10
F2	38,270,486	99.84	71	12
F3	44,203,804	99.86	87	8
F4	52,653,023	99.82	95	11
F5	55,505,123	99.86	103	9

3.3.3.2.1 CNVs

The single-base resolution read depths were calculated for the whole genome, then the median values for 10 kb windows were plotted and the evolved samples were compared to their ancestor. (**Figure 3.5**C & **Figure 3.6**). The major observed copy number variation event is the loss of chromosome 12. The loss of chromosome 12 events was seen in 17 of the 18 evolved populations at various frequencies in the population. Chromosome 12 is lost completely in the 34°C passaged populations and the event is even observed in the ancestor. It is mostly conserved in B populations.

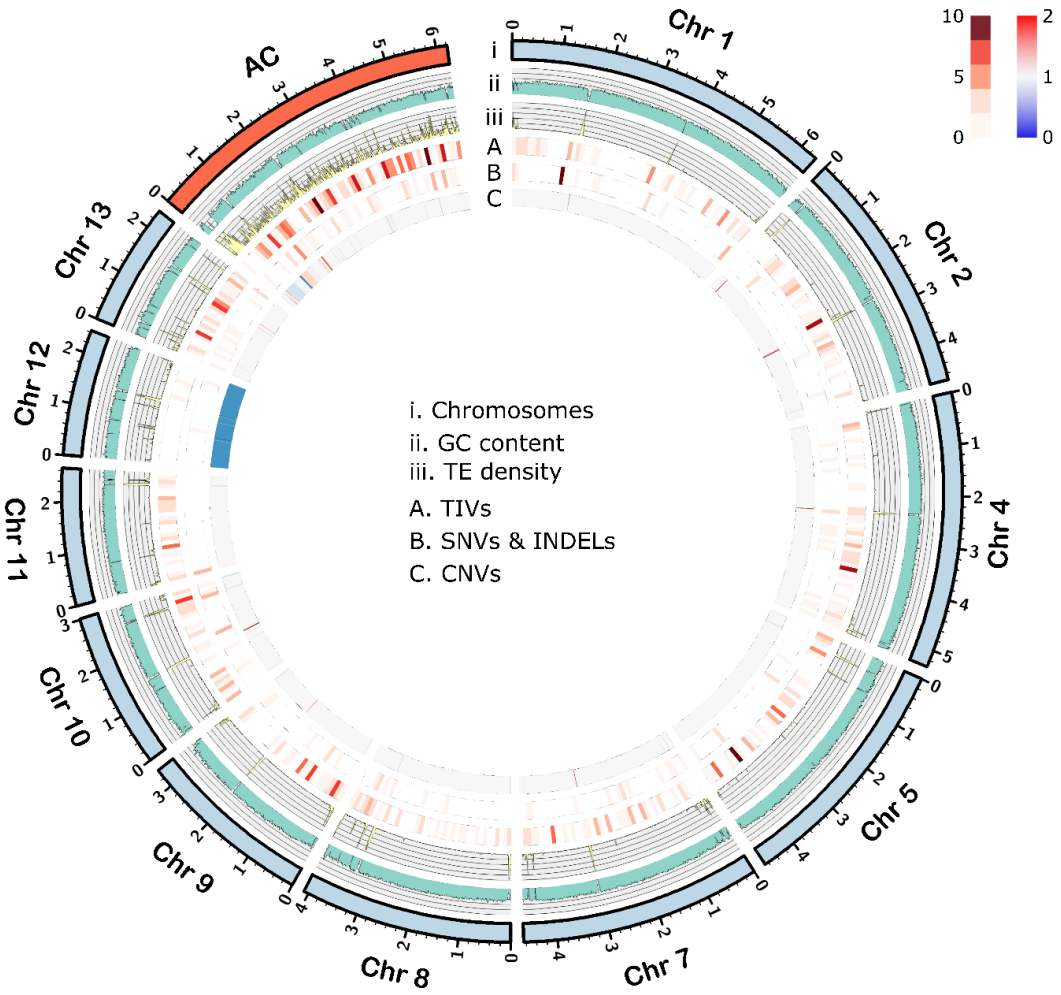


Figure 3.5 Chromosomal distribution of the mutation events. (i) The pseudochromosomes of MRL8996. Accessory chromosome (AC) sequences are shown in orange, while the core chromosomes are in blue. (ii) GC % content and (iii) TE content calculated in 10 kb windows. (A) transposon inversion variation (TIV), (E) single nucleotide variation (SNV) and small insertion/deletion (INDEL) distributions calculated in 100 kb windows. Light orange to red color bar is in the upper right. (C) The averaged copy number variation (CNV) changes are calculated in 10 kb windows. Blue to red color bar is in the upper right corner.

In M, Y, and F populations of Fol4287, no copy number variations in chromosome 12 were observed. The only copy number variation of chromosome 12 which was observed in E populations is found in the E5 population where the large arm of chromosome 12 from centromere has 23% reduced coverage (**Figure 3.7**).

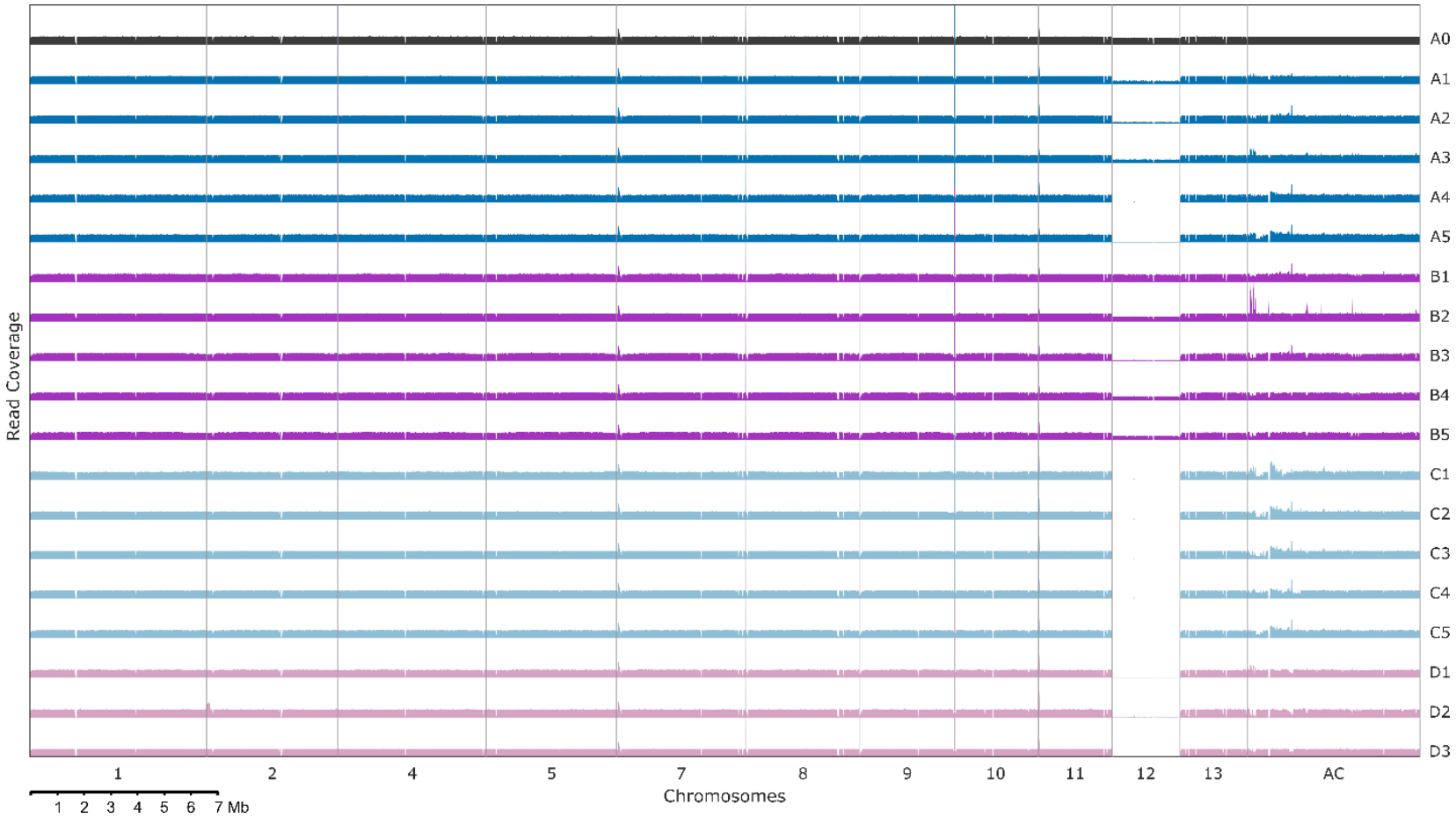


Figure 3.6 The read coverage distribution of MRL8996 chromosomes. The read coverages of the core chromosomes are normalized to the median read coverage of the sample. The read coverages of the accessory chromosome (AC) sequences are normalized to both the median read coverages in and the median read coverages of A0. For both 10 kb windows are used. The core chromosome contigs are merged into pseudochromosomes (see **Figure 1.6**C). AC sequences are ordered by clustering (details in Methods 3.2.4).

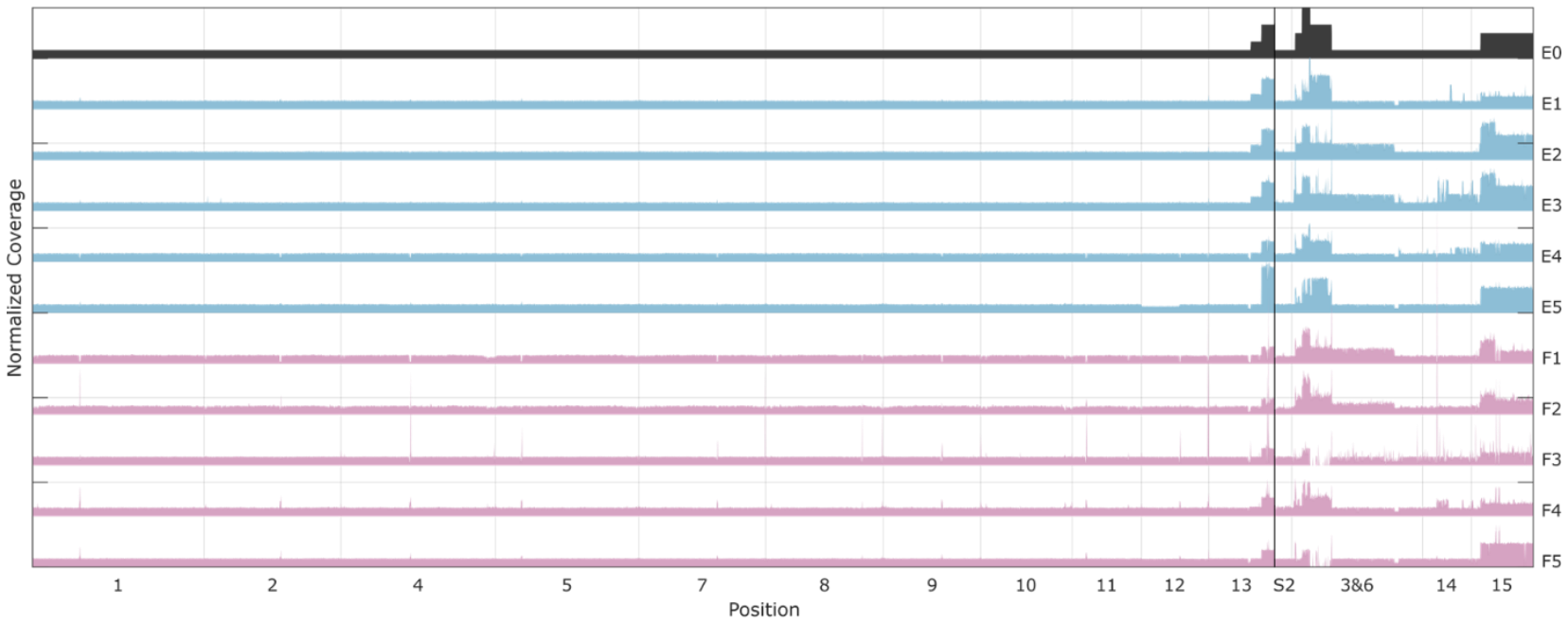


Figure 3.7 The read coverage distribution of Fol4287 chromosomes. The median read depths calculated over 10 kb-window regions are normalized to first median read coverage of the sample, then to E0. Some regions in E0 are duplicated and thus represented as 2–6× coverage (see Methods 3.2.4 for more details).

Chromosome 12 of MRL8996 and Fol4287 are more diverged than the rest of the core chromosomes. While the overall genome alignment between two genome is 56.96% (with sequence identity cut off 95%), it's 22.32% for chromosome 12 (**Figure 1.7**B). There are 890 genes in MRL8996 chromosome 12 and enriched in sphingolipid metabolic process (*p*-value = 0), lysosome organization and biogenesis (*p*-value = 0), pathogenesis (*p*-value = 0), carbohydrate metabolic process (*p*-value = 4.7×10^{-6}), regulation of endo-1,4-beta-xylanase activity (*p*-value = 2.9×10^{-5}), response to toxin (*p*-value = 3.2×10^{-4}), electron transport (*p*-value = 5.0×10^{-3}), and sulfate transport (*p*-value = 7.1×10^{-3}) gene ontology (GO) terms. Same GO-terms are also enriched in Fol4287 chromosome 12 except for xylanase activity genes which was caused by missing annotation (*p*-values of the above GO-terms: 1.2×10^{-3}, 1.2×10^{-3}, 0.017, 0.021, N/A, 4.1×10^{-4}, 1.2×10^{-3}, 0.007, respectively).

Therefore, the gene content difference cannot be the reason for this conditional disposability of this chromosome in MRL8996. However, since some of GO-terms such as sphingolipid metabolic process, lysosome organization, and biogenesis, carbohydrate metabolic process, or electron transport are also enriched in Fol4287 accessory chromosome (AC) genes (*p*-values: 0.015, 0.015, 9.5×10^{-5}, 4.0×10^{-4}, respectively), those gene groups might be involved in Fol4287 host colonization and not necessary for MRL8996 survival *in vitro* at 34°C.

In previous studies, the loss of chromosome 12 in Fol4287 was observed at a very low frequency (5.2×10^{-6} per conidia) *in vitro* and Fol4287 stains without chromosome 12 showed similar pathogenicity in tomato seedlings and a similar phenotype as wildtype Fol4287 in various growth media except on medium with ribose as the carbon source [16]. Observing this event happening at a high frequency, and even observing in the ancestor at 19% signals a different function of chromosome 12 in MRL8996 and Fol4287.

Chromosome 12 of Fol4287, along with chromosome 11 and chromosome 13 are methylated with H3K27me3 (**Figure 2.3**C) [145]. As shown in Chapter 2, chromosome 13 of Fol4287 shows copy number variations in evolving populations. Similarly, in this study, E and F populations also showed the same loss of duplication at the small arm of chromosome 13 (**Figure 3.7**) and an additional duplication at the subtelomeric region in the small arm of chromosome 13. These associations of CNVs with chromosomes 12 and 13 support that chromosome destabilization might be connected to the histone methylation markers [168,169].

AC sequences also show some duplications and deletion events. However, when compared to the average of 49.3% of AC regions which are dynamic in Fol4287 (**Figure 2.4** & **Figure 3.7**), MRL8996 with an average dynamic AC ratio of 19.2% are much more stable (regions considered with more than 10% copy number change; *p*-value with two-way *t*-test: 8.5×10^{-7}).

Connected to the chromosome 12 loss, all MRL8996 populations showed decreased genome size up to 5% while MM-passaged Fol4287 populations had increase genome size (**Figure 3.8**). CM pH 7.4 34°C-passaged Fol4287 (F) populations had the highest genome size reduction (up to 10%).

The increases or decreases in read depth on centromeric sequences are also observed in MRL8996 similar to Fol4287.

3.3.3.2.2 SNVs, INDELs, and TIVs

The variant calling was performed as described in 2.3.1.2.2 and 2.3.1.2.3. A total of 960 single nucleotide variations (SNVs), small insertions and deletions (INDELs), and transposon insertion variations (TIVs) were detected by the GATK Best practices pipeline

and TEfinder [86,138]. Among them, 699 variants with allele frequency (AF) larger than 10% were filtered (**Table 3.2**, **Figure 3.5**A-B).

The number of mutations in MRL8996 samples was on average 1.8 times the number of mutations in Fol4287 samples and higher in both SNVs & INDELs and TIVs. While the numbers of SNVs & INDELs are consistent across samples (*p*-value: 0.067 with one-way ANOVA), numbers of TIVs showed significant variations (*p*-value: 2.7×10^{-7} with one-way ANOVA) (**Figure 3.9**). The average number of TIVs in 34°C-passaged populations was higher than 28°C-passaged populations in both strains. The number of events was especially higher for TIVs in MM pH7.4 34°C passaged MRL8996 populations (C populations: 37.2 events per sample on average), when compared to MM 28°C-passaged MRL8996 populations (A populations: 9.4 events per sample) and MM pH7.4 34°C-passaged Fol4287 populations (E populations: 11 events per sample). MM-passaged Fol4287 populations (M: 2 events per sample) had the lowest number of TIV events in all sample sets.

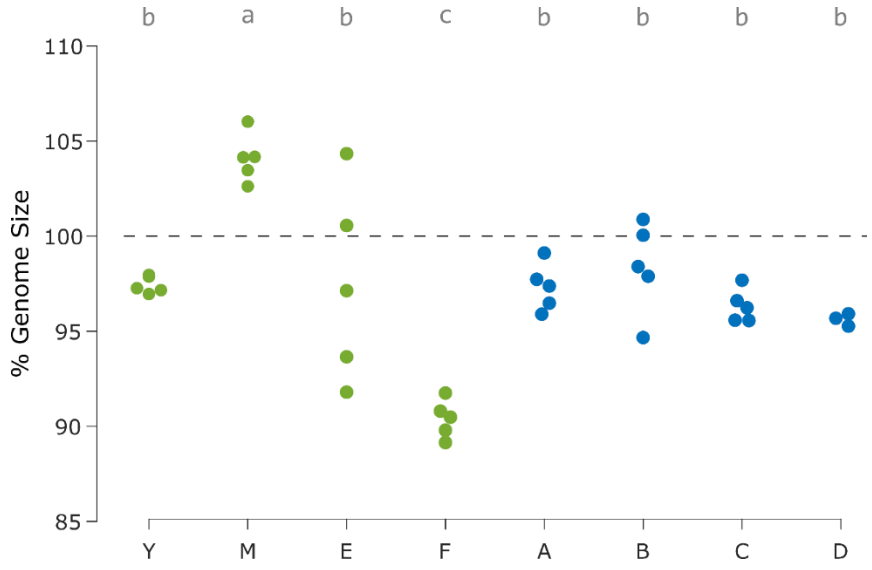


Figure 3.8 The genome size change of the evolved populations with respect to their ancestors. Green: Fol4287, Blue: MRL8996 evolved populations. Letters above the boxes indicate the significant variation differences in the ANOVA 1-way test where the sample sets with the same letters do not have significantly different variance (p-value: 2.0×10^{-8}).

Table 3.2 The number of filtered mutations in the final populations. SNV: Single nucleotide variation, INDEL: small insertion or deletion, TIV: transposon insertion variation.

	MRL8996				Fol4287			
	Sample	*SNVs & INDELs*	*TIVs*	*Total*	*Sample*	*SNVs & INDELs*	*TIVs*	*Total*
MM 28°C passaged populations	*A1*	3	12	15	*M1*	4	2	6
	A2	7	9	16	*M2*	1	0	1
	A3	6	6	12	*M3*	4	3	7
	A4	3	14	17	*M4*	8	4	12
	A5	3	6	9	*M5*	3	1	4
	All	*22*	*47*	*69*	*All*	*20*	*10*	*30*
CM 28°C passaged populations	*B1*	2	2	4	*Y1*	3	4	7
	B2	4	8	12	*Y2*	3	6	9
	B3	4	9	13	*Y3*	4	4	8
	B4	11	5	16	*Y4*	2	5	7
	B5	3	10	13	*Y5*	4	7	11
	All	*24*	*34*	*58*	*All*	*16*	*26*	*42*
MM pH 7.4 34°C passaged populations	*C1*	3	19	22	*E1*	1	8	9
	C2	3	44	47	*E2*	2	9	11
	C3	8	53	61	*E3*	3	3	6
	C4	2	51	53	*E4*	3	24	27
	C5	2	19	21	*E5*	5	11	16
	All	*18*	*186*	*204*	*All*	*14*	*55*	*69*
CM pH 7.4 34°C passaged populations	*D1*	11	28	39	*F1*	3	27	30
	D2	5	19	24	*F2*	4	20	24
	D3	10	25	35	*F3*	6	14	20
					F4	3	28	31
					F5	4	20	24
	All	*26*	*72*	*98*	*All*	*20*	*109*	*129*
	Total	*90*	*339*	*429*	*Total*	*70*	*200*	*270*

E and F populations of Fol4287 samples show comparable AF distribution pattern for both SNVs & INDELs and TIVs with the previous study (Chapter 2) in which the distributions have two peaks around 0-0.1 and 0.9-1 AFs (**Figure 3.10**), signaling that the response to the selections in samples from Chapter 2 and newly generated Fol4287 samples are comparable.

For both Fol4287 and MRL8996, TIVs consist of the majority of the events (74.0% and 79.0%, respectively).

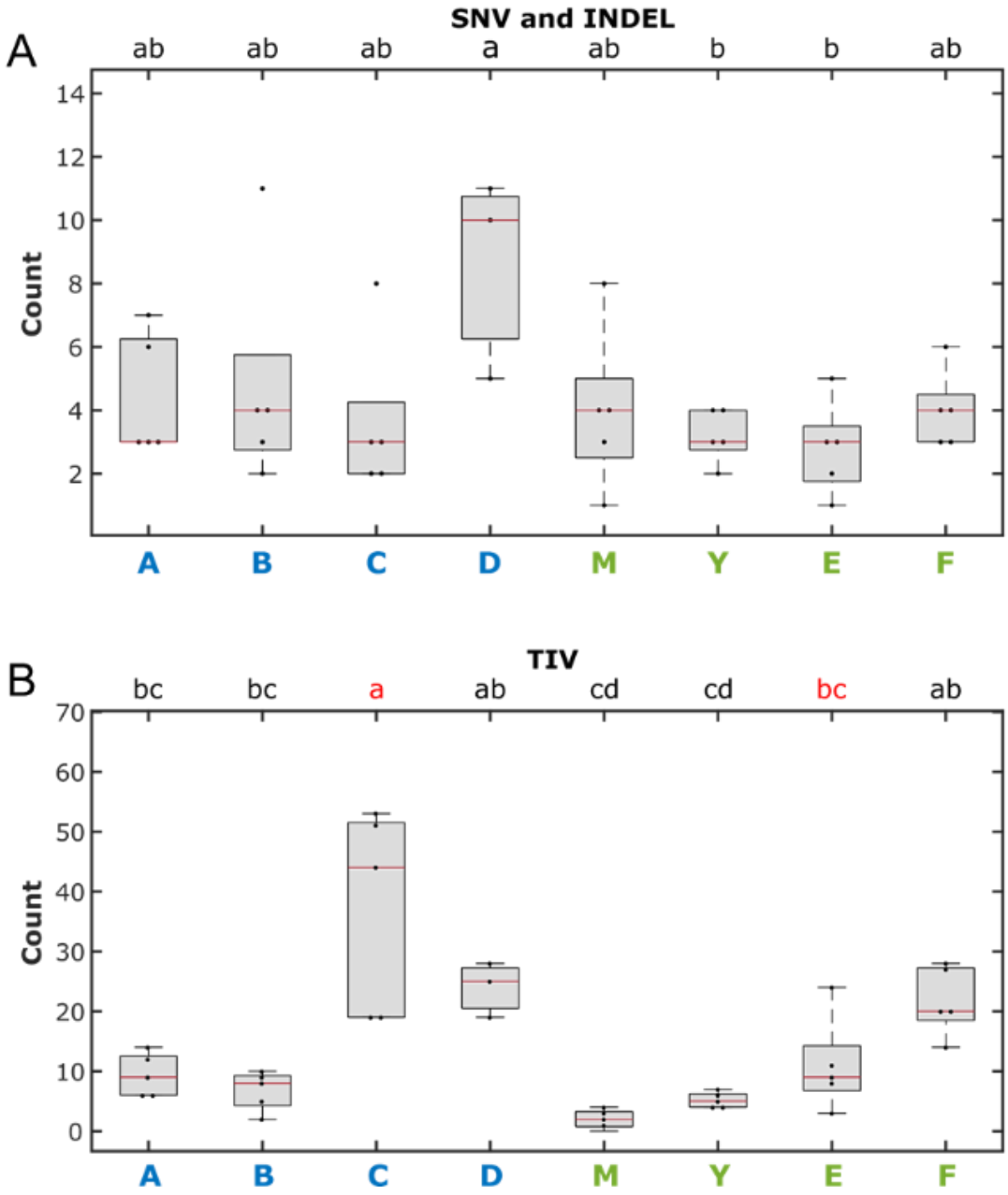


Figure 3.9 Number of SNVs and INDELs (A) and TIVs (B) in final populations. The number of filtered mutations is shown as box plots. Letters above the boxes indicate the significant variation differences in the ANOVA 1-way test where the sample sets with the same letters do not have significantly different variance (*p*-values for SNVs and INDELs (A): 0.067; for TIVs (B): 2.7×10^{-7}). Red lines indicate the median, the gray boxes cover from 25th to 75th percentiles, and the whiskers indicate the extreme non-outlier data points. All data points are shown as black dots.

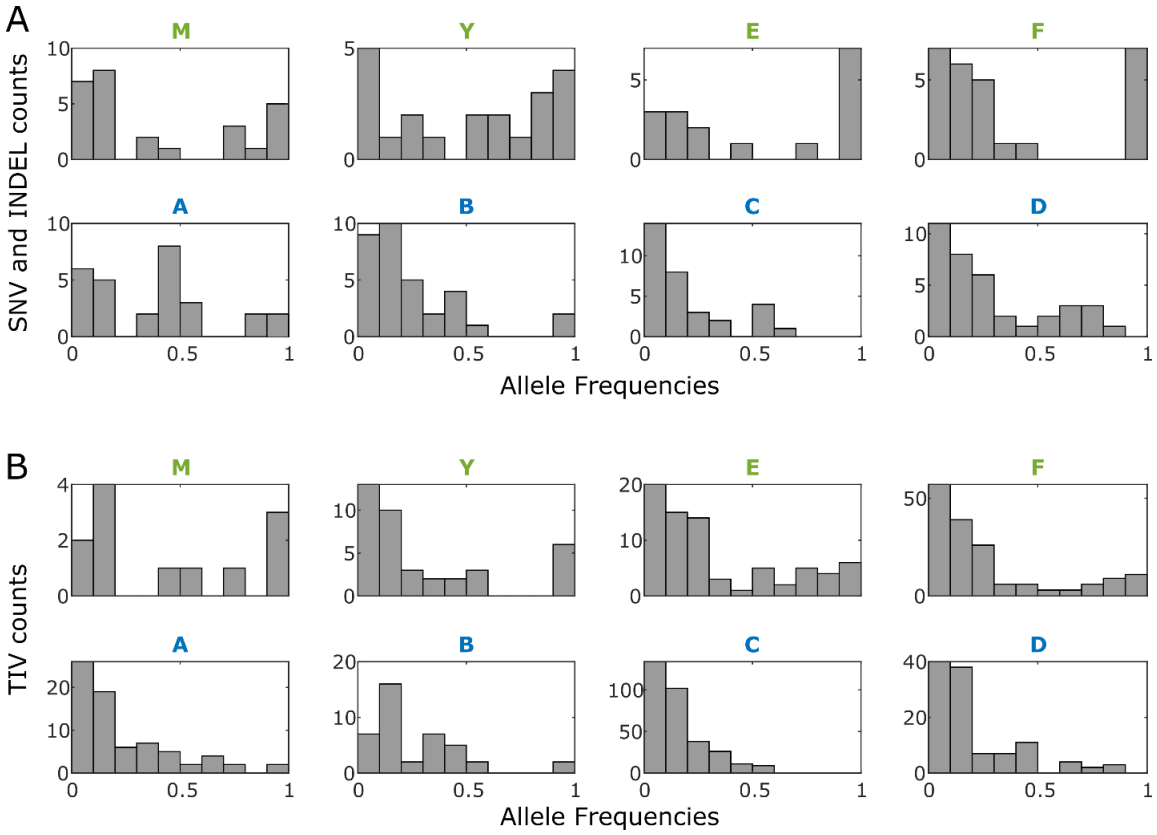


Figure 3.10 Distribution of allele frequencies (AF) of the mutations in the evolution experiments. (A) AF distributions of single nucleotide variation (SNV) and small insertions deletion (INDEL). (B) AF distributions of transposon insertion variation (TIV). No filtration is applied. M, Y, E, and F are Fol4287 populations; A, B, C, and D are MRL8996 populations.

3.3.4 Transposon activity in STEE

Like it was observed in Chapter 2, the TE insertions were more frequent in ACs than they were in CCs in MRL8996 (11.26 insertions/Mb and 6.11 insertions/ Mb, respectively).

A more detailed analysis revealed that the TE insertion sites were not random in MRL8996. TIVs detected in Fol4287 followed the gene density ratio of the Fol4287 genome which 64% is compromised by genes. Whereas in MRL8996, while the gene density is 59%, the TIVs were found to be in intergenic regions about 75% of the time (**Figure 3.11**A). In addition, TIVs detected in MRL8996 were disproportionally on 1 kb upstream or downstream of the genes (*p*-values: 1.78×10^{-8}, 3.82×10^{-6}, 9.41×10^{-41}, and 1.70×10^{-18} for A, B, C, and D populations, respectively). This might mean that transposons responsible

for TIVs in the MRL8996 genome have an affinity to the 1 kb up or downstream regions of the genes or that the evolution selects for the TIVs inserted in these regions presumably for their effect on the regulation of gene expressions. On the other hand, there is no bias for the up or downstream of the genes observed for SNVs and INDELs in Fol4287 (**Figure 3.11**B).

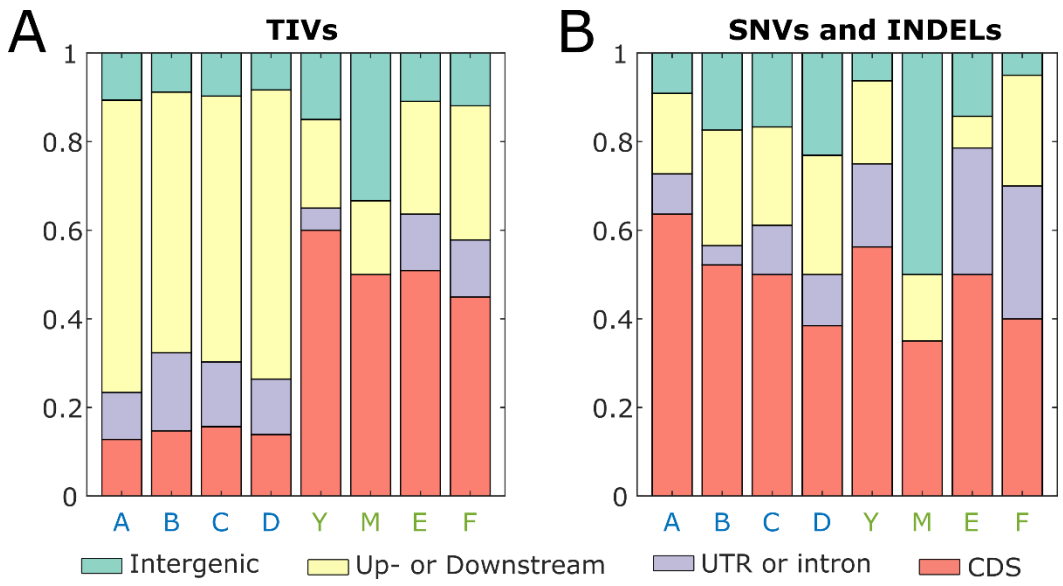


Figure 3.11 The ratios of the mutations in coding sequences (CDS), UTR or intronic regions, 1 kb up or downstream of the genes, or more than 1 kb distance to the genes in filtered mutations. The variants numbers from game groups are averaged and normalized to the total numbers. The association of TIVs in up- and downstream of the genes for MRL8996 samples is not random (*p*-values for one-tailed Fisher's exact test for A, B, C, and D are 1.78×10^{-8}, 3.82×10^{-6}, 9.41×10^{-41}, and 1.70×10^{-18}, respectively). The association of SNVs/INDELs and CDS for MRL8996 A, B, and D samples are not random (*p*-values for one-tailed Fisher's exact test for are 0.0027, 0.0217, and 0.0060, respectively).

Transpositionally active transposons in Fol4287 and MRL8996 evolving populations are distinct. The only TE which is shared between ML8996 and Fol4287 evolved samples is a LTR/Copia family transposon, CopiaActive (**Figure 3.12** & **Table 3.3**). In MRL8996, SINEs and LINEs are active while in Fol4287 TIR transposons are active. The number of detected events in MRL8996 is higher than in Fol4287 even with the higher sample number in Fol4287.

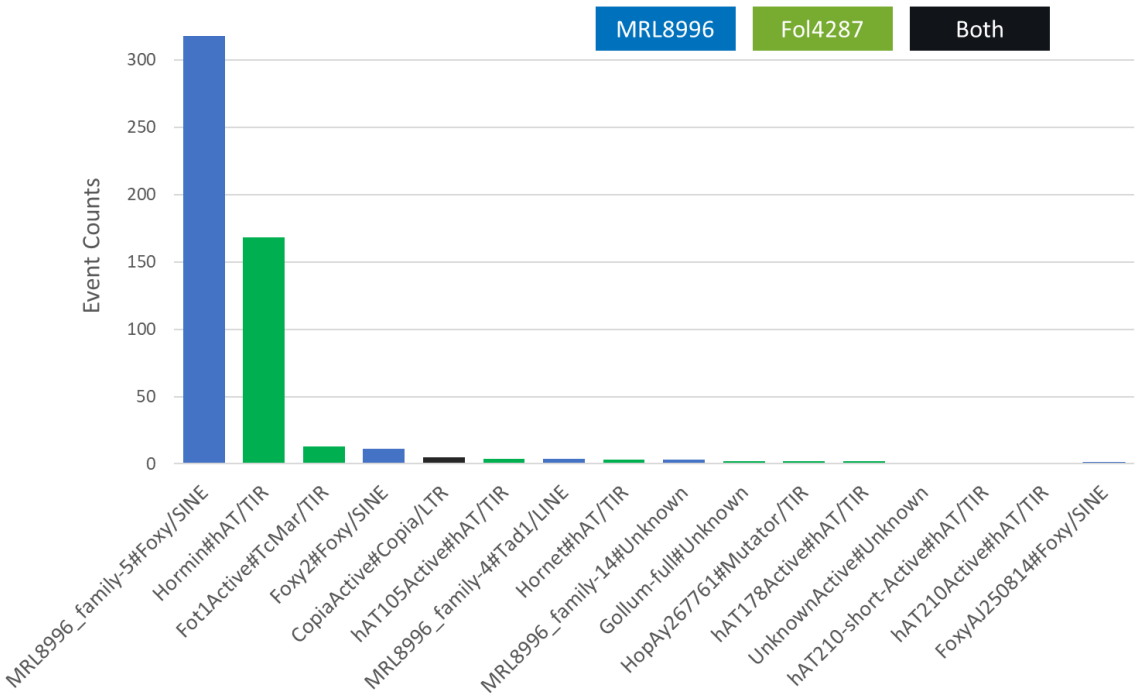


Figure 3.12 Total filtered number of transposon insertion events by transposable element (TE) across the final populations. Blue: TEs which are only active in MRL8996, green: TEs which are only active in Fol4287, black: TE that is active in both strains.

Table 3.3 The frequency of mutations observed in the final populations. The frequencies are calculated as the total number of filtered insertion events/number of samples.

Order	*Superfamily*	*Family*	***A***	***B***	***C***	***D***	***M***	***Y***	***E***	***F***
SINE	*Foxy*	*Foxy5*	8.4	6.4	35.8	21.7				
SINE	*Foxy*	*Foxy2*	0.8	0.4	0.4	1				
LINE	*Tad1*	*MRL8996_family-4*			0.4	0.7				
SINE	*Foxy*	*FoxyAJ250814*			0.2					
LTR	*Copia*	*CopiaActive*	0.2			0.3			0.6	
LTR	*Copia*	*Gollum-full*					0.4			
TIR	*hAT*	*Hormin*					0.2	3.6	9.6	20.2
TIR	*TcMar*	*Fot1Active*					0.6	0.6	0.4	1
TIR	*hAT*	*hAT105Active*					0.4	0.4		
TIR	*hAT*	*Hornet*							0.2	0.4
TIR	*Mutator*	*HopAy267761*						0.4		
TIR	*hAT*	*hAT178Active*							0.2	0.2
TIR	*hAT*	*hAT210-short-Active*					0.2			
TIR	*hAT*	*hAT210Active*						0.2		
Unknown		*MRL8996_family-14*			0.4	0.3				
Unknown		*UnknownActive*					0.2			

The highest activity transposon in MRL8996 evolved populations is ‘MRL8996_family-5’. Although RepeatModeller [79] was not able to classify it, ‘MRL8996_family-5’ was annotated as Foxy Superfamily short interspaced nuclear element (SINE) transposon manually because of its homology to other Foxy elements, repeats at 5’ end, adenine rich 3’ end, and it’s variable target sequence duplication site length [170] and will be called Foxy5 hereafter. Foxy5 clusters with Foxy3 element, but it is about 600 bp longer than Foxy3 and has an open reading frame on its non-homologous region (**Figure 3.13**). Enrichment analysis with the insertion sites of Foxy5 revealed that the motif ‘TACCCTRNA’ is enriched near the insertion site (E-value: 3.3×10^{-15}).

Figure 3.13 Comparison of Foxy superfamily transposable elements Foxy3 and Foxy5. While the gray section shows the homologous sequence, the orange section represents the nonhomologous open reading frame (ORF) including the 3’ end sequence of Foxy5. The blue boxes represent the repeats at 5’ ends, and the purple ticks represent the Box A and B sequences that recruit RNA polymerase III.

As discussed in Chapter 1, Foxy5 is the most abundant TE family in MRL8996 (labeled as Foxy3(fol) in **Figure 1.9**C). It also has the highest expression at the CM broth medium (**Figure 3.14**). Although Foxy3-like family transposons are present in Fol4287 (a small number when compared to Foxy2), Foxy5 is unique to MRL8996. While in MRL8996 Foxy2 is also present, albeit not as abundant as Foxy5, and has the second-highest expression level.

Foxy5, like other SINEs, lacks reverse transcriptase and endonuclease genes and therefore probably depended on LINEs for its transposition. The Tad1 superfamily LINE ‘MRL8996_family-4’ which is an active LINE in C and D populations, the highest

expressed LINE in RNAseq experiments (**Figure 3.14**), and the third most abundant TE in MRL8996 (**Figure 1.9**, clustered with 'Fol4287_family-102') or the CRE-Cn11 superfamily LINE 'MRL8996_family-43' (clustered with 'Fo47_family-136') which is the highest abundant LINE in MRL8996 genome are the top two candidates for the LINEs that are responsible for Foxy5's transposition.

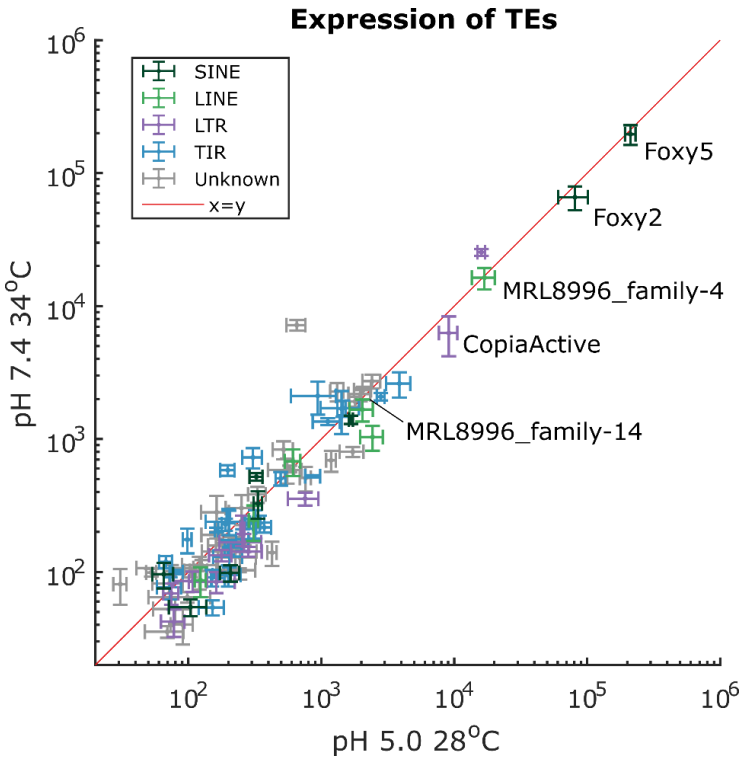


Figure 3.14 The normalized expressions of MRL8996 transposable elements (TEs) in CM medium pH 5.0 at 28°C and in CM medium pH 7.4 at 34°C. Only the active TEs in the short-term evolution experiment are labeled. The error bars represent the standard deviation of the three biological replicates.

3.3.5 Weak selection in MRL8996 evolving populations

The number of fixed mutations in evolved MRL8996 populations was very low while in evolved Fol4287 populations many fixed mutations were detected even in plate conditions which lead to a weak selection regime. The ratio of fixed or almost fixed mutations ($0.9 < AF \leq 1$) to all detected mutations in the populations were higher in Fol4287 populations (*p*-values: 0.02 for M vs A; 0.08 for Y vs B; 4×10^{-4} for E vs C; and 0.02 for F vs D with two-

sample *t*-test) (**Figure 3.10**) and the overall distributions were significantly different in 28°C-passaged and MM pH 7.4 34°C-passaged populations (*p*-values: 0.03 for M vs A, 0.02 for Y vs B, 1.5×10^{-6}, for E vs C, and 0.1 for F vs D with two-sample Kolmogorov-Smirnov test) (**Figure 3.15**). This signals that the selection pressure for MRL8996 was lower than Fol4287 in the same conditions.

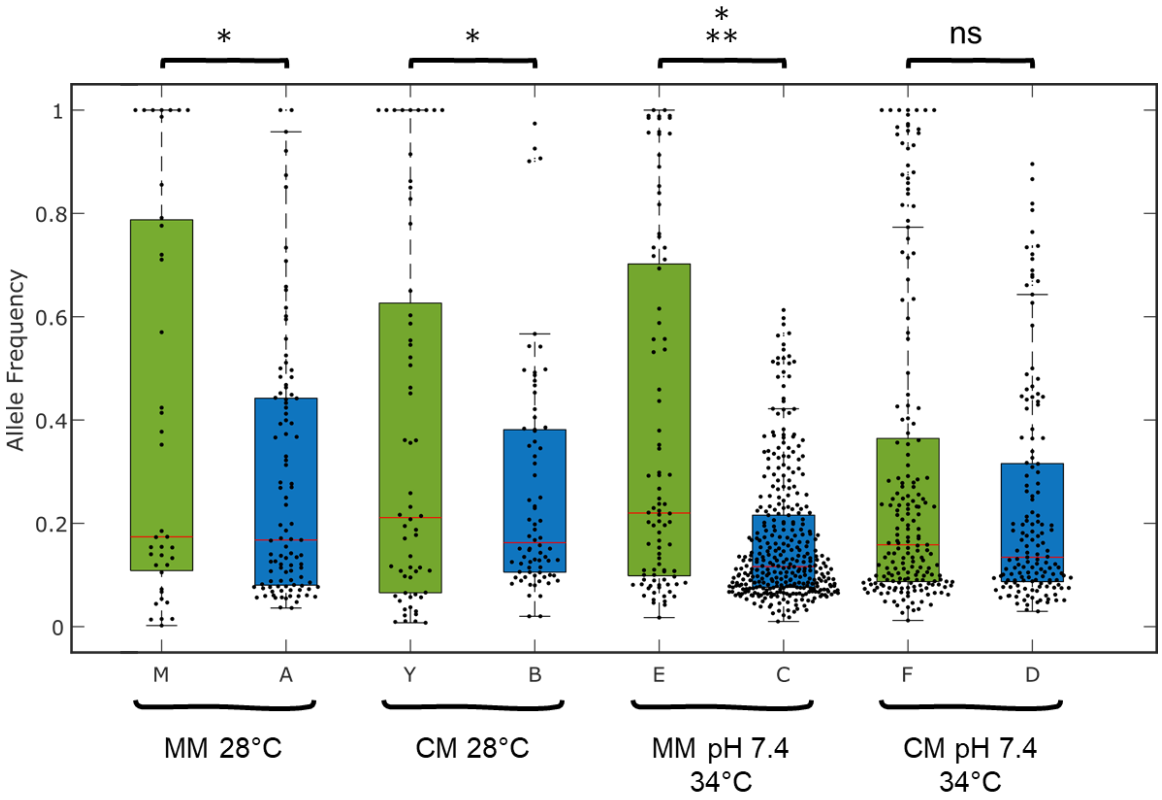


Figure 3.15 The allele frequency distributions of the detected variations. The distribution of allele frequencies of Fol4287 evolved samples (green) and MRL8996 evolved samples (blue) are shown as box plots. Red lines indicate the median, the green, and blue boxes cover from 25th to 75th percentiles, and the whiskers indicate the extreme non-outlier data points. All data points are shown as black dots. Pairwise comparisons were performed by two-sample Kolmogorov-Smirnov test. (*p*-values: 0.03, 0.02, 1.5×10^{-6}, and 0.1 for M-A, Y-B, E-C, and F-D comparisons, respectively).

MM pH 7.4 34°C-passaged MRL8996 (C) populations, had the highest number of TIVs but there were no detected fixed mutation events. While pH 7.4 at 34°C condition might be a cause for the increased activity of TIVs, the small frequency events with no fixation signal a very weak selection model in this condition for the MRL8996 population [116].

Additionally, the SNVs and INDELs are more likely to be located in CDS regions of the genes in MRL8996 evolved samples with the exception of C populations (*p*-values: 0.0027, 0.0217, and 0.0060, for A, B, and D populations respectively). Again, this supports that different selection pressure for C populations applies (**Figure 3.11**B).

3.3.6 Mutation patterns

There were 20 genes that were mutated in multiple evolved populations (**Figure 3.16**). The most striking ones are the Velvet complex genes which were also discussed in Chapter 2.

While *velB* is mostly mutated in Fol4287 evolved populations (four Y populations, one C population, and one D population and), *veA* is mostly targeted in MRL8996 evolved populations (five A populations, one B population, one C population, one M populations, and one E population). In total, 14 independent populations had mutations on a subunit of the Velvet complex. The mutations in *velB* in Y and M populations were shown to occur after the mutations on the FOXG_21009 gene. There is one homolog of FOXG_21009 with 41.2% sequence identity in MRL8996: protein ID 90294 which doesn't have any mutations. There are no alternative genes that seem to have an epistatic relation with the Velvet complex genes in MRL8996.

Other genes mutated in MRL8996 samples include a histidine kinase (431163), a glyoxylase (429645), 60S ribosomal protein P2 (201076), transcription factors (153193, 449065, 339274), and other unknown genes.

For Fol4287 evolving populations, an asparaginase (FOXG_04868), a nucleotidase (FOXG_03639), and a defense mechanism gene (FOXG_15014) are mutated, among other unknown genes.

gene	Fol4287	MRL8996	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	M1	M2	M3	M4	M5	Y1	Y2	Y3	Y4	Y5	E1	E2	E3	E4	E5	F1	F2	F3	F4	F5
"D:SnoaL-like"	FOXG_12017	362521	U		U				U							I																								
"alpha-L-rhamnosidase"	FOXG_10845	343264		U											U																									
"Transcription initiation factor TFIID"	FOXG_05243	153193		C				C																																
"Glyoxylase"		429645			I													I																						
"TF"	FOXG_11440	449065			C													C																						
"60S acidic ribosomal protein P2"	FOXG_09282	201076				DS	DS							DS			DS																						
"Adenylate cyclase-associated protei	FOXG_04210	400564				US	US									US			US																					
"TF"		339274													DS				DS																					
"Sensory transduction histidine kinas	FOXG_01684	431163												C			C												C										
veA	FOXG_11273	17519	C	C	C	C	C				C		C											C						C										
velB	FOXG_00016	20006															C			C							C	C	C											
	FOXG_21009	90294																						C			C	C	C	C										
	FOXG_21002	479978																												C	C		C			C			
"Asparaginase"	FOXG_04868	406006																															C		DS	DS				
	FOXG_21511																															U	U						
	FOXG_15875																														DS								DS
"IMP-specific 5'-nucleotidase 1"	FOXG_03639	411523																														U							U
"Defense mechanisms"	FOXG_13532																														C			C			C		
	FOXG_15014	479844																													C	C	C	C		C	C	C	C

Figure 3.16 List of Genes mutated across more than one evolved sample. The protein IDs are provided for MRL8996 genes, Color scale is 0 to 1 AF for white to dark red. Locations with respect to the genes are C: CDS, U: UTR, I: intron, DS: downstream, and US: upstream.

The elevated temperature tolerance increase is the highest for evolved Fol4287 populations, even though the growth rates are not yet higher than evolved MRL8996 samples. The tolerance to elevated temperature can be caused by one of some of the mutations shared by E and F populations such as FOXG_15014 (unknown) in E2, E3, E4, E5, F2, F3, F4, and F5; FOXG_04868 (an asparaginase) in E4 and F1; FOXG_21511 (unknown) in E3 and E4; FOXG_15875 (unknown) in E2 and F5; FOXG_03639 (a nucleotidase) in E3 and F5; and FOXG_13532 (a defense-related gene) in E2, E5, and F3. All these genes but most importantly the gene FOXG_15014 which encodes for an unknown protein of 600 aa and is mutated in 8 out 10 populations are candidates for future research of elevated temperature adaptation of Fol4287.

3.4 Discussion

This study investigated the evolution mechanism of the *F. oxysporum* keratitis isolate MRL8996 in comparison to the tomato pathogen Fol4287 using experimental evolution. Even in a short time, experimental evolution resulted in many differences in terms of genotype and phenotype for both *F. oxysporum* strains. These experiments demonstrated that experimental evolution can be applied to filamentous fungi and can be used to study the pathogenicity and adaptability of cross-kingdom pathogens.

F. oxysporum, as it lacks a known sexual cycle, is a surprisingly efficient pathogen as it can adapt to a wide host range from plants to animals. In this study, we discovered that its rapid adaptation is likely due to the activity of transposable elements. In Fol4287, the activity of transposon suggested its involvement in the evolvability of the organism, as the TE insertion in an accessory chromosome gene gained (FOXG_21009) prepared a major growth regulator (Velvet complex) to be able to gain loss-of-function mutations. In MRL8996, Velvet complex genes were also mutated however without this interaction as the true ortholog of FOXG_21009 is not present in MRL8996.

While the mechanisms of evolutions were shared in both strains such as the dynamic chromosomes and the activity of transposons, the actors that use those mechanisms were different.

3.4.1 Loss of chromosome 12

In a short time of 10 weeks, all 34°C adapted MRL8996 populations and some 28°C adapted MRL8996 populations lost the core chromosome 12. The only copy number variation observed in Fol4287 chromosome 12 was the deletion of the large arm in the E5 population which happened in a very small frequency. Fol4287 chromosome 12 can be disposable without any significant fitness cost as but the frequency of that event in lab

conditions was found to be very low [16]. Here, it is shown that MRL8996 chromosome 12 is disposable, and the loss of chromosome events is very frequent.

Although the sequence homology of chromosome 12 in both strains is lower than what is expected for a core chromosome, there is a significant gene content difference between the two. There might be a difference in terms of epigenetic markers which are known to be stabilizing the chromosomes such as H3K9me3 or H3K27me3 [168,169]. As shown in Chapter 1, the chromatin assembly/disassembly genes are enriched in both genomes' ACs suggesting an important gene regulation and adaptation related to epigenetics. Although H3 methylation regulation genes such as kmt1 or kmt6 do not have mutations in the final populations, H3 methylation and its effect on the adaptation should be investigated.

3.4.2 Transposon activity

In both genomes, the most active TEs in short-term evolution experiments (STEEs) were also the most abundant TEs in their genome. However, Hormin and Foxy5, the most active transposons in Fol4287 and MRL8996, respectively, are not present in the other genome. Foxy5 constitutes more than 60% of all detected events in this study and 94% of the MRL8996 transposition events. Foxy5 is shown to have a bias for 1 kb up or downstream regions of genes and specifically binds to the motif 'TACCCTRNA', unlike Hormin or other TIRs which had a bias for regions with H3K27me3 and no specific binding motif. This difference might affect the consequence of the mutations. While the genes are knocked out by Hormin/TIR element insertions, Foxy5 might affect the gene expression [171].

When compared to the 28°C passaged populations, 34°C passaged populations displayed higher TIV counts for evolved populations of both MRL8996 and Fol4287. This was not the case for SNVs and INDELs. This can be the result of either (i) higher transposases

expression and therefore transposition activity at elevated temperatures or (ii) that at elevated temperatures the adaption selects for many TIVs. From the TE transcription analysis, it can be concluded that the former might not be the case, because the transcription levels of LINEs or TIRs are not higher at 34°C in complete medium conditions. Therefore, it needs to be investigated whether the selection for TIVs causing this effect.

3.4.3 Selection pressure

Even though having the highest number of TIVs or highest number of SNVs, there is no observed mutation fixation event observed in 34°C-passaged MRL8996 populations and the mutations were accumulated at lower frequencies. Additionally, unlike other MRL8996 evolved populations, in MM pH 7.4 34°C-passaged MRL8996 populations, the mutation events on coding sequences of the genes are not significantly enriched, unlike others. C1, C5, and D3 populations have mutations in *veA* or *velB* genes of the Velvet complex, which was correlated with a strong mutation effect for conidiation in Chapter 2. Therefore, the selection observed in 34°C-passaged MRL8996 populations is weaker than what is observed in evolved Fol4287 populations which also had a weaker selection than in plant-passaged populations. The effect of this phenomenon in phenotype can be observed as detectable but not drastic adaption.

3.4.4 Genotype-phenotype map

In this study, the evolved populations were also tested under various conditions including their evolving environment, and compared against their ancestor. Although the methods used in this study cannot be translated into fitness effect over the ancestors as the final populations and the ancestor were tested separately and only hyphal growth rate, a feature that was not selected for in STEE, some change in phenotypes could be captured. Most of the evolved populations either grows faster or not significantly different than their ancestor under their evolving conditions. The most drastic difference was observed for

Fol4287 high temperature evolved populations with growth rate differences sometimes 2-fold or higher. This might be related to diminishing returns epistasis where each beneficial mutation increases the fitness to some extent but as the mutations accumulate the effect decreases [109]. While the ancestor Fol4287 starts with a very low fitness to elevated temperatures, the ancestor MRL8996 is already has a high fitness. The effect of the beneficial mutations that accumulate during STEE in Fol4287 evolved populations, therefore, is higher over the ancestor's fitness.

Additionally, 28°C evolved MRL8996 populations were lost their adaptation to 34°C. While no clear connection between genotype and phenotype can be drawn, the genes which are mutated multiple times, as well as the loss of chromosome 12, should be studied for their effect on the temperature adaptation of MRL8996 and Fol4287.

CONCLUSION

Fusarium oxysporum species complex is a group of filamentous fungi that can cause vascular wilt disease in many plant species and local or disseminated infections in humans. One of the ways *F. oxysporum* can adapt to such a wide range of hosts is through accessory chromosomes (ACs) which are enriched in host-specific genes and repeat content. This dissertation investigated the mechanisms that drive the adaptive evolution in the cross-kingdom pathogen *F. oxysporum* using comparative genomics and an experimental evolution approach.

Using two strains, *F. oxysporum* f. sp. *lycopersici* 4287 a plant pathogenic isolate, and *F. oxysporum* MRL8996 a keratitis isolate, as model organisms, first the genome assemblies were generated, and genomic and phenotypic comparisons were performed. Different ACs with different gene and transposon contents are found in the plant pathogenic and keratitis strains. In the plant pathogenic strain ACs, transcription factors, carbohydrate metabolism, defense response, and oxidative stress response genes and DNA transposons were enriched. In the keratitis strain metal ion transporters and pH regulation genes, LINE and SINE transposons were enriched. In both chromatin assembly disassembly genes and protein kinases were enriched. While the keratitis strain was better adapted to the elevated temperatures, the plant pathogenic strain had more tolerance to some stress conditions.

Short-term evolution experiments were carried out under 9 conditions with two strains and 42 populations were generated. In a short time of 10 passages, many changes were observed. Genotyping and phenotyping were performed in comparison to the ancestors. In general, when the populations were passaged in a non-host-specific condition, the loss of adaptation was observed in both strains.

In the evolved populations of plant pathogenic strain, copy number changes were observed in ACs and chromosome 13 while in the evolved populations of keratitis isolate chromosome 12 was lost frequently.

For both strains, the activity of transposons was the major cause of variations while different TEs were active. In the evolved populations of plant pathogenic strain, a DNA transposon constitutes the majority of the transposition events while in the evolved populations of the keratitis strain, a SINE transposon has the highest activity. In both cases, the most active TE was nonautonomous and had already the highest copy number in their class in the genome of the ancestor.

Through the short-term evolution experiments, diverse selection pressures were captured under different conditions such as strong selection under plant-passaged populations and weak selection in plate passaged populations. Genetic parallelism was observed with many genes that were mutated in multiple independent populations. Velvet complex subunits were mutated in 14 independent populations in both strains passaged in plate conditions revealing the importance of this protein complex in *F. oxysporum* adaptation. In the plate-passaged populations of the plant pathogenic isolate, the mutations on these genes were correlated with virulence loss and found to have an epistatic relation with an uncharacterized AC gene.

Overall, in this dissertation it was discovered that (I) *F. oxysporum* pathogenic strains to plant and human host adapt rapidly in short term evolution experiments; (II) the genome of *F. oxysporum* is plastic but two strains have different behavior regarding copy number variations; (III) transposable element activities are the driving forces, even though different genomes are using a different type of transposons; (IV) active transposable elements are encoded in ACs, and novel insertions are enriched among ACs. In conclusion, the accessory chromosomes provide evolutionary hotspots for *F. oxysporum*.

APPENDIX A

Extended Table A1

The list of the detected mutations that passed the AF ≥ 0.1 filter in all Fol4287 final populations in Chapter 2 and Chapter 3. CHR: chromosome; POS: position; AF: Allele frequency; TYPE: mutation types (TIV: transposon insertion variation; SNV: single nucleotide variation; INDEL: small insertion or deletion); MUTATION: if TYPE is TIV – the TE name, if TYPE is SNV or INDEL – nucleotide change; GENE: the locus tag ID of the gene if the mutation is in a gene or 1 kb region of a gene (starts with "FOXG_" followed by 5-digit numbers); LOC: location of the mutation with respect to gene; SAM: sample name (all samples are the 10th population); AA CHANGE: if the mutation is SNV or INDEL in CDS, the change in the amino acid ({FS}: frameshift, *: stop codon).

CHR	*POS*	*AF*	*TYPE*	*MUTATION*	*GENE*	*LOC*	*SAM*	*AA CHANGE*
Chr1	40235	1.00	TIV	Fot1Active		intergenic	F4	
Chr1	99759	0.18	TIV	hAT105Active	10848	CDS	Y4	
Chr1	808419	1.00	SNV	G→A	11096	CDS	P2	W>*
Chr1	1377646	0.86	SNV	G→A	11273	CDS	M4	Q>*
Chr1	1378251	0.99	SNV	C→T	11273	CDS	E1	G>D
Chr1	1549218	0.73	SNV	A→G	11316	CDS	E4	D>D
Chr1	1704182	0.86	INDEL	AGTCAACATCAC CACTCGTCTCAC CCTCCTCTACCT CCTCCGCC→A	16	CDS	Y2	SQHHHS SHPPLP PPP>S
Chr1	1704921	1.00	SNV	G→T	16	CDS	Y3	E>*
Chr1	1705310	0.83	SNV	C→T	16	CDS	Y4	Q>*
Chr1	1705397	0.12	TIV	hAT105Active	16	CDS	Y4	
Chr1	2007148	1.00	INDEL	AGAAT→A	110	CDS	P5	{FS}
Chr1	2688697	0.75	TIV	Fot1Active	17859	upstream	E2	
Chr1	4297750	0.25	INDEL	CTT→C	836	upstream	E4	
Chr1	5548691	0.14	TIV	Hormin	1244	CDS	E4	
Chr1	5861915	0.98	SNV	T→A	1365	UTR 3'	E3	
Chr10	168552	0.36	SNV	G→C	17339	upstream	Y1	
Chr10	199289	0.21	TIV	Hormin	13577	UTR 3'	F2	
Chr10	263593	0.11	TIV	Hormin	21213	downstream	E3	
Chr10	331457	0.40	SNV	C→A	13626	CDS	P5	S>*
Chr10	560381	0.15	TIV	Hormin	13710	upstrem	M1	
Chr10	720467	0.24	SNV	G→T	13761	UTR 5'	F2	
Chr10	2076351	0.19	TIV	Hormin	11575	CDS	Y5	
Chr10	2335215	0.28	TIV	Hormin	20594	downstream	F4	
Chr10	2510312	0.16	TIV	Hormin	11740	UTR 3'	F4	
Chr10	2537857	0.89	TIV	Hormin	11753	UTR 3'	E5	
Chr10	2852261	0.12	TIV	Hormin	20622	CDS	F3	
Chr10	2879863	0.43	TIV	Fot1Active	11880	downstream	P1	
Chr11	136385	0.15	TIV	Hormin	16521	upstream	E4	
Chr11	357450	0.97	TIV	Hormin	16612	CDS	F1	

Chr11	448484	0.27	TIV	Hormin	22490	CDS	F4	
Chr11	620599	0.82	TIV	Hormin	9576	CDS	F4	
Chr11	736737	1.00	TIV	UnknownActive	9621	CDS	M1	
Chr11	1427171	0.26	TIV	Hormin	9881	upstream	F4	
Chr11	1700410	0.59	TIV	Hormin	9993	CDS	Y1	
Chr11	1736981	0.17	TIV	Hormin	20200	upstream	F5	
Chr11	1993741	0.63	TIV	Hormin	20221	CDS	F4	
Chr11	2095436	0.11	TIV	Hormin	10139	CDS	E4	
Chr11	2121728	0.29	TIV	Hormin	10151	CDS	F1	
Chr11	2205347	0.11	TIV	Fot1Active	10187	intron	Y1	
Chr11	2228025	1.00	SNV	C→A	10196	CDS	P2	D>E
Chr12	611833	0.11	TIV	Hormin	13300	downstream	E4	
Chr12	637191	0.19	TIV	Hormin	21134	downstream	F2	
Chr12	748889	0.37	SNV	T→A	21146	CDS	P1	N>Y
Chr12	1196545	0.76	TIV	Hormin	13532	CDS	E2	
Chr12	1196837	0.87	TIV	Hormin	13532	CDS	F3	
Chr12	1197771	0.82	TIV	Hormin	13532	CDS	E5	
Chr12	1277101	0.42	INDEL	T→TG	13564	CDS	F5	{FS}
Chr12	1411207	0.24	TIV	Hormin	14506	downstream	F1	
Chr12	1497480	0.68	SNV	C→A	14537	downstream	P1	
Chr12	1518717	0.49	TIV	Hormin	21578	CDS	F4	
Chr12	1740415	0.56	TIV	Hormin	21593	CDS	E1	
Chr12	1873554	0.31	TIV	Hormin	14683	upstream	F5	
Chr12	1965602	1.00	TIV	Hormin	14710	downstream	P4	
Chr12	2184428	0.57	TIV	Gollum-full	14777	upstrem	M3	
Chr12	2278042	0.11	TIV	Fot1Active		intergenic	Y3	
Chr13	124647	0.15	TIV	Gollum-full	11926	CDS	M3	
Chr13	205320	0.10	TIV	Hormin	11955	CDS	F4	
Chr13	234005	1.00	SNV	G→T	11967	CDS	P5	C>F
Chr13	331677	1.00	TIV	Hormin	20646	UTR 5'	P2	
Chr13	396343	0.23	TIV	Hormin	20653	UTR 5'	F5	
Chr13	436287	0.20	TIV	Hormin	12035	CDS	F2	
Chr13	479346	0.19	TIV	Hormin	12055	CDS	F2	
Chr13	551645	0.29	TIV	Hormin	12083	UTR 5'	F4	
Chr13	596363	1.00	TIV	Hormin	12102	intron	P5	
Chr13	950540	0.43	TIV	Hormin	12245	CDS	F1	
Chr13	961964	0.11	TIV	hAT178Active	20697	downstream	E4	
Chr13	963521	0.63	TIV	Hormin	12250	CDS	F1	
Chr13	998251	1.00	TIV	Hormin	12262	CDS	F1	
Chr13	1126729	0.20	TIV	Hormin	12313	CDS	E5	
Chr13	1353694	0.85	TIV	Hormin	12396	downstream	F4	
Chr13	1454865	0.45	TIV	FoxyAJ250814	22594	upstrem	P4	
Chr13	1560820	0.11	TIV	Hormin	16841	CDS	F3	

Chr13	1720296	0.20	TIV	Hormin	22627	CDS	F2	
Chr13	1795228	0.11	TIV	Hormin	16932	CDS	E4	
Chr2	38691	0.75	TIV	Hormin		intergenic	F1	
Chr2	60812	0.64	TIV	Fot1Active	5596	upstrem	P1	
Chr2	215731	0.26	TIV	Hormin	5653	CDS	F5	
Chr2	341746	0.22	TIV	Hormin	5698	CDS	F5	
Chr2	438257	0.37	TIV	Hormin	5733	CDS	F1	
Chr2	495068	1.00	TIV	Hormin	5759	CDS	E5	
Chr2	496534	0.28	TIV	Hornet	5759	CDS	F4	
Chr2	981843	1.00	SNV	C→T		intergenic	M3	
Chr2	1811573	1.00	SNV	G→T	6212	CDS	P5	N>K
Chr2	1855356	0.16	SNV	G→A	6225	CDS	F3	L>L
Chr2	3297552	1.00	SNV	C→T	8386	CDS	Y3	K>K
Chr2	4060694	0.53	TIV	Strider	8643	downstream	P1	
Chr2	4237634	0.71	TIV	Hormin		intergenic	F1	
Chr2	4262023	0.67	TIV	Hormin	8721	CDS	F1	
Chr2	4303357	1.00	TIV	Hormin	8732	CDS	P2	
Chr2	4460997	0.20	TIV	Hormin	8789	upstream	E5	
Chr2	4480886	0.10	SNV	C→T	8795	upstream	F3	
Chr2	4593917	1.00	TIV	Hormin	19902	intron	E2	
Chr4	113163	0.96	TIV	Hormin	7396	CDS	E1	
Chr4	180626	0.28	TIV	Hormin	7423	downstream	F4	
Chr4	187966	0.10	TIV	Fot1Active		intergenic	M4	
Chr4	225472	1.00	TIV	Fot1Active	19485	upstrem	M4	
Chr4	257159	0.25	TIV	Hormin	7450	CDS	F5	
Chr4	296707	0.25	TIV	Hormin	7465	CDS	F2	
Chr4	964195	0.97	SNV	C→T	19530	CDS	F3	F>F
Chr4	969486	0.14	SNV	G→A	7715	CDS	M3	K>K
Chr4	1764528	0.93	SNV	C→A	7978	CDS	F1	G>G
Chr4	2228498	0.60	SNV	T→C	19573	UTR 5'	Y5	
Chr4	2554114	0.97	INDEL	TGG→T	3744	UTR 5'	F3	
Chr4	2584710	0.24	TIV	Hormin	18643	CDS	E4	
Chr4	2815212	0.17	TIV	Hormin	3837	CDS	E4	
Chr4	2856410	0.33	TIV	Hormin	3854	CDS	F5	
Chr4	2919290	1.00	TIV	Hormin	3878	upstrem	Y2	
Chr4	3194484	0.46	TIV	HopAy267761	18706	upstrem	Y1	
Chr4	3406468	0.25	TIV	Hormin	18726	downstream	F2	
Chr4	4011320	0.36	SNV	T→C	4264	UTR 3'	F5	
Chr4	4422652	0.16	SNV	G→A	4392	downstream	F3	
Chr4	4868544	0.14	TIV	Hormin	18852	CDS	F4	
Chr4	4976932	0.87	TIV	Hormin	4606	CDS	F3	
Chr4	5001593	0.25	TIV	Hormin	4615	upstream	F2	
Chr4	5116450	0.10	TIV	Hormin	4648	UTR 5'	F5	

Chr4	5224675	0.45	TIV	Hormin	4682	intron	F4	
Chr4	5234420	1.00	TIV	Hormin	4683	CDS	P5	
Chr4	5294483	0.17	TIV	hAT210-short-Active	22915	upstrem	M3	
Chr5	237643	0.35	TIV	Hormin	15258	downstream	F5	
Chr5	246356	0.10	TIV	Hormin	15263	CDS	F2	
Chr5	328733	0.15	SNV	G→A	15290	CDS	F4	D>N
Chr5	471500	0.19	SNV	C→A		intergenic	M5	
Chr5	478389	0.11	TIV	Hormin	21890	downstream	Y2	
Chr5	553215	0.18	TIV	Hormin	115370	UTR 3'	F2	
Chr5	906106	0.65	SNV	C→T		intergenic	P1	
Chr5	984888	0.78	SNV	G→A	1429	CDS	Y4	W>*
Chr5	1041335	0.96	SNV	C→T	1446	UTR 5'	F1	
Chr5	1782816	0.91	SNV	G→A	1671	UTR 3'	P5	
Chr5	1830302	0.21	SNV	C→T	1684	CDS	Y5	H>Y
Chr5	2643702	0.29	SNV	G→A	18210	CDS	E5	R>R
Chr5	4415008	0.50	TIV	Fot1Active	18320	downstream	P5	
Chr5	4573749	1.00	TIV	Hormin	18341	CDS	F3	
Chr5	4621181	1.00	TIV	Fot1Active	2547	upstrem	P5	
Chr5	4814697	0.84	TIV	Hormin	2619	CDS	F1	
Chr5	4883393	1.00	SNV	C→T	18368	downstream	M1	
Chr7	136758	0.26	TIV	Hormin	4737	CDS	Y5	
Chr7	151972	0.95	TIV	Hormin	4742	CDS	E1	
Chr7	222245	0.14	SNV	T→C		intergenic	M1	
Chr7	361199	0.14	TIV	Hormin	4816	intron	F5	
Chr7	504780	0.72	TIV	Hormin	4868	CDS	E4	
Chr7	505424	0.24	TIV	Hormin	4868	downstream	F1	
Chr7	773685	0.99	SNV	G→A	4946	UTR 3'	E5	
Chr7	819893	0.51	SNV	A→G	4961	CDS	Y2	S>P
Chr7	1784774	0.12	SNV	C→T		intergenic	M3	
Chr7	3922566	0.22	TIV	Hormin	10657	CDS	E1	
Chr7	4105265	0.99	SNV	G→A	10726	CDS	E3	G>R
Chr7	4252890	0.58	TIV	Hormin	10785	CDS	P1	
Chr7	4310263	0.59	TIV	Hormin	10805	intron	E1	
Chr8	56893	0.24	SNV	C→T	2672	downstream	F2	
Chr8	144179	0.18	TIV	Hormin	2695	downstream	E4	
Chr8	297947	0.12	TIV	Hormin	2741	UTR 5'	F2	
Chr8	317212	0.81	TIV	Hormin		intergenic	F1	
Chr8	400085	0.11	SNV	G→A	2782	UTR 3'	Y5	
Chr8	521943	0.17	SNV	G→T	18413	CDS	M5	N>K
Chr8	796746	1.00	TIV	Hormin	2937	downstream	P4	
Chr8	806994	0.24	TIV	Hormin	2942	CDS	F1	
Chr8	1827271	0.12	SNV	G→A	3235	CDS	E5	G>S
Chr8	1942407	0.42	SNV	G→A	3265	CDS	M2	A>T

Chr8	2001140	1.00	INDEL	TAC→T	3284	upstream	Y3	
Chr8	2759036	0.13	TIV	Hormin	3526	CDS	F1	
Chr8	2822545	0.93	TIV	Hormin	3547	CDS	F1	
Chr8	3098850	0.99	SNV	G→T	3639	UTR 3'	E2	
Chr8	3098850	0.99	SNV	G→T	3639	UTR 3'	F5	
Chr8	3476706	0.96	SNV	C→T	15669	CDS	F3	A>T
Chr8	3478403	0.27	INDEL	G→GCA	15669	upstream	F4	
Chr8	3501891	0.20	TIV	Hormin	15677	upstream	E4	
Chr8	3740423	0.17	TIV	Hormin	15766	upstrem	Y5	
Chr8	3744711	0.16	TIV	Hormin	15768	UTR 3'	E4	
Chr8	3859630	0.23	TIV	Hormin	22141	downstream	E4	
Chr8	3882739	1.00	TIV	Fot1Active	15821	upstrem	P2	
Chr8	3898537	0.52	TIV	Hormin	15825	upstrem	Y5	
Chr8	3961001	1.00	SNV	T→C	15845	upstream	P5	
Chr8	4007950	0.22	TIV	Hormin		intergenic	F4	
Chr9	77914	0.79	TIV	Fot1Active		intergenic	M4	
Chr9	149354	0.24	TIV	Hormin	8890	UTR 5'	F4	
Chr9	199716	1.00	TIV	Fot1Active	8909	UTR 3'	P5	
Chr9	268292	0.89	TIV	Hormin	8940	upstream	F3	
Chr9	366756	0.29	TIV	Hormin	8984	CDS	E1	
Chr9	386898	1.00	SNV	C→A	8994	CDS	P5	D>Y
Chr9	663217	0.15	TIV	Hormin	9093	CDS	F3	
Chr9	1439606	0.99	SNV	C→T	9359	CDS	M1	G>S
Chr9	2079315	0.95	INDEL	TCTC→T	12649	CDS	F1	SP>S
Chr9	2819547	0.11	TIV	CopiaActive	12879	CDS	E2	
Chr9	3149623	0.91	TIV	Fot1Active	13004	downstream	E5	
Chr9	3189780	0.77	TIV	Fot1Active	13018	downstream	F1	
Chr9	3189780	0.23	TIV	Fot1Active	13018	downstream	F2	
Chr9	3193363	0.38	INDEL	TTACCTAAAACC CTTAATATTATT TTATAACTTAAT ATTTAATTATAT TAATAATTATA→ T	13018	upstream	P1	
S_14.11	18339	1.00	TIV	Hormin	21552	upstream	F5	
S_14.12	3192	0.91	INDEL	C→CG	16428	UTR 3'	Y2	
S_14.17	21706	0.17	TIV	Hormin		intergenic	F1	
S_14.2	9356	0.11	TIV	Hormin	17266	upstream	F2	
S_14.2	56529	0.11	TIV	Hormin	22711	downstream	F2	
S_14.2	66739	0.63	TIV	hAT105Active	17251	CDS	P4	
S_14.2	96874	0.27	TIV	Hormin	22703	CDS	E4	
S_14.21	14783	0.40	TIV	Fot1Active	16460	upstream	F1	
S_14.21	16127	0.62	TIV	Hornet		intergenic	E4	
S_14.25	869	1.00	TIV	Hormin		intergenic	Y2	
S_14.25	16791	0.14	TIV	Hormin	22501	downstream	Y4	

S_14.30	6571	0.15	SNV	C→G		intergenic	M4	
S_14.46	5904	0.69	TIV	Hormin		intergenic	P1	
S_14.5	36692	0.95	TIV	Hormin	21511	UTR 3'	E3	
S_14.5	37013	0.84	TIV	Hormin	21511	UTR 3'	E4	
S_15.13	10571	0.94	TIV	Hormin	22180	upstream	F3	
S_15.15	1954	0.88	TIV	Hormin		intergenic	F1	
S_15.16	13755	0.13	TIV	Hormin	22093	upstream	F5	
S_15.18	9672	0.35	SNV	C→T	15600	CDS	M4	G>D
S_15.4	47811	0.36	TIV	Hormin	15552	downstream	Y4	
S_15.5	21812	0.14	TIV	Hormin	22586	CDS	P4	
S_15.5	56189	0.11	TIV	Hormin		intergenic	F3	
S_15.6	17237	0.24	TIV	Hormin	16731	CDS	E4	
S_15.6	29360	0.85	INDEL	AGCTGTTGATAA AG→A	16733	upstream	Y1	
S_15.6	46534	0.38	TIV	Hormin	22577	intron	E5	
S_15.7	12392	0.11	TIV	Hormin	22538	downstream	F4	
S_15.8	1113	0.94	TIV	CopiaActive	22059	intron	P5	
S_1n15.11	9390	0.19	TIV	hAT210Active	15093	upstream	Y3	
S_1n15.2	62481	0.60	TIV	Hormin	15002	upstream	F5	
S_1n15.4	8664	0.12	TIV	Hormin	14349	CDS	F5	
S_1n15.4	51267	0.51	SNV	C→T	21995	downstream	P2	
S_1n15.6	33360	0.24	TIV	Hormin	18091	CDS	P2	
S_1n15.7	4856	0.18	TIV	Hormin	15027	CDS	F5	
S_1n15.7	32374	0.18	SNV	C→T	21951	downstream	P5	
S_1n15.8	6395	0.41	TIV	hAT105Active	14462	downstream	M5	
S_1n15.8	23597	0.56	TIV	Hormin	15014	CDS	E4	
S_1n15.8	23604	0.56	TIV	Hormin	15014	CDS	F5	
S_1n15.8	23605	0.44	TIV	Hormin	15014	CDS	E2	
S_1n15.8	23607	0.54	TIV	Hormin	15014	CDS	E3	
S_1n15.8	23618	0.29	TIV	Hormin	15014	CDS	F2	
S_1n15.8	23618	1.00	TIV	Hormin	15014	CDS	F3	
S_1n15.8	23619	0.86	TIV	Hormin	15014	CDS	F4	
S_1n15.8	23628	0.53	TIV	Hormin	15014	CDS	E5	
S_1n15.9	3355	0.57	TIV	Hormin	15114	CDS	F4	
S_2.1	37193	0.19	TIV	Hormin	15857	CDS	F4	
S_2.2	2415	0.85	TIV	Hormin	15875	downstream	E2	
S_2.2	2415	0.96	TIV	Hormin	15875	downstream	F5	
S_2.4	2231	0.34	TIV	Hormin		intergenic	E4	
S_2.4	138829	1.00	TIV	CopiaActive	15919	upstrem	P5	
S_2.4	139696	0.23	TIV	Hormin	22265	CDS	F4	
S_2.4	143021	1.00	TIV	Hormin	22266	CDS	P5	
S_2.4	227274	0.20	TIV	Hormin	15935	downstream	E4	

S_2.6	11302	0.79	TIV	Hormin		intergenic	F3	
S_3.1	3933	0.10	SNV	T→C	14820	UTR 5'	E2	
S_3.1	39040	0.10	SNV	A→T	6924	CDS	E4	R>W
S_3.1	57021	0.15	TIV	Hormin	6928	intron	F2	
S_3.1	57048	0.72	TIV	Hormin	6928	intron	F1	
S_3.1	66929	0.13	TIV	Hormin	16080	upstream	F2	
S_3.14	16040	0.30	TIV	Hormin		intergenic	E4	
S_3.18	861	0.11	TIV	Hornet	21844	CDS	F4	
S_3.2	100966	0.13	TIV	Hormin	14890	CDS	F4	
S_3.3	44672	0.47	TIV	Hormin	21846	CDS	P5	
S_3.8	28723	0.55	INDEL	T→TGAGCCGGAGCCG	14947	CDS	Y3	G>GSGS
S_3.9	9510	0.22	TIV	Hormin	14995	downstream	F2	
S_3n6.1	17468	1.00	TIV	Hormin	12502	upstream	F3	
S_3n6.1	26500	1.00	INDEL	AAAAGGTTGCTGATGTTGAACGTACCTGCTTCAGTAAGATGATCGCCATTTTG→A	20948	CDS	M5	{FS}
S_3n6.1	143239	0.40	TIV	Hormin	12524	UTR 5'	F4	
S_3n6.12	23039	0.39	TIV	Hormin	22333	downstream	F1	
S_3n6.14	23814	0.21	TIV	HopAy267761	16226	upstrem	Y2	
S_3n6.17	13280	0.35	TIV	Hormin	13998	CDS	E1	
S_3n6.18	7929	0.15	TIV	Hormin		intergenic	F1	
S_3n6.29	9935	0.72	SNV	T→C	6486	downstream	M4	
S_3n6.3	29721	1.00	SNV	G→A	12431	downstream	P2	
S_3n6.32	18431	0.50	TIV	Merry		intergenic	P4	
S_3n6.4	2065	0.69	TIV	CopiaActive		intergenic	E2	
S_3n6.5	37911	0.43	TIV	Hormin	21002	CDS	F2	
S_3n6.5	38123	0.22	TIV	Hormin	21002	CDS	E4	
S_3n6.5	38167	0.36	TIV	Hormin	21002	CDS	Y5	
S_3n6.5	38178	0.71	TIV	Hormin	21002	CDS	E2	
S_3n6.5	38640	0.99	SNV	T→C	12611	CDS	E5	K>T
S_3n6.52	1585	0.10	TIV	hAT178Active	14222	upstream	F1	
S_3n6.9	2604	0.99	INDEL	G→GA	22360	upstream	P5	
S_6.1	111521	0.27	TIV	Fot1Active	22790	upstrem	P4	
S_6.11	15731	0.20	TIV	CopiaActive	12405	downstream	E2	
S_6.16	18272	0.22	TIV	Hormin	7102	CDS	E4	
S_6.17	11479	0.22	TIV	Hormin	19603	CDS	E5	
S_6.19	15464	0.14	TIV	Hormin	7043	CDS	F1	
S_6.2	47771	0.13	TIV	Hormin	19594	intron	F4	
S_6.20	16370	0.22	TIV	Hormin		intergenic	Y5	
S_6.21	10816	0.11	TIV	Hormin		intergenic	P1	
S_6.23	10730	0.11	TIV	Fot1Active	17123	downstream	Y3	
S_6.23	11847	0.13	INDEL	CTCTA→C	17122	upstream	M4	

S_6.3	54530	0.46	SNV	T→C		intergenic	E3	
S_6.6	35810	0.38	SNV	C→T		intergenic	M1	
S_6.7	33599	0.13	TIV	Hormin		intergenic	E1	
S_6.9	27752	0.57	SNV	T→C	7021	upstream	P5	
U_1	35953	0.36	TIV	Hormin	17029	CDS	F5	
U_1	84803	0.45	TIV	Hormin	17010	CDS	Y1	
U_1	157464	0.13	TIV	Hormin	16977	CDS	F3	
U_11	28557	0.13	TIV	Hormin	17608	downstream	F5	
U_12	54982	0.12	TIV	Hormin	22894	upstream	F5	
U_14	67706	0.73	TIV	Hormin	17560	CDS	E4	
U_15	38846	0.12	SNV	G→T		intergenic	M4	
U_16	31615	0.16	TIV	Hormin		intergenic	E5	
U_17	8140	0.25	TIV	Fot1Active	17527	downstream	P4	
U_19	24807	0.16	SNV	C→A	17327	UTR 5'	F2	
U_2	40924	0.23	SNV	G→T	17596	CDS	Y5	C>F
U_2	118452	0.23	SNV	A→G	17616	CDS	F4	M>T
U_2	163549	0.78	SNV	C→T		intergenic	M4	
U_2	169262	0.32	SNV	T→G		intergenic	P1	
U_23	10080	0.17	INDEL	C→CCAATCAAAT CAATCAAAT	21687	upstream	F2	
U_24	80	0.96	INDEL	TTATTTATAGAT AATTAATAAAGG AGA→T		intergenic	E5	
U_25	496	0.23	TIV	Hormin		intergenic	F2	
U_25	12980	0.71	SNV	A→C		intergenic	M4	
U_3	60462	0.11	TIV	Hormin	22773	downstream	F4	
U_3	99058	0.49	SNV	C→A	22768	downstream	P5	
U_3	142447	0.29	TIV	Hormin	22762	upstream	E5	
U_3	185060	0.14	TIV	Hormin	17137	CDS	F4	
U_32	15081	1.00	TIV	Fot1Active		intergenic	P5	
U_37	2876	0.16	TIV	Hormin		intergenic	F1	
U_39	774	1.00	TIV	Hormin	17680	CDS	Y2	
U_41	5485	0.28	SNV	T→A		intergenic	F5	
U_45	53	1.00	INDEL	AAGATATTATCC TAGACTAT→A		intergenic	P2	
U_54	983	0.65	SNV	A→T		intergenic	Y1	
U_6	34365	0.72	TIV	Hormin		intergenic	F1	
U_6	62983	1.00	TIV	Fot1Active	14853	downstream	F3	
U_9	10330	1.00	TIV	Hormin	21009	CDS	Y2	
U_9	10450	1.00	TIV	Hormin	21009	CDS	Y3	
U_9	10704	1.00	TIV	hAT105Active	21009	CDS	M4	
U_9	10778	1.00	TIV	Hormin	21009	CDS	Y4	
U_9	10811	0.55	TIV	Hormin	21009	CDS	Y5	

U_9	57854	1.00	SNV	T→C	intergenic	M3
U_9	57854	1.00	SNV	T→C	intergenic	P4

Extended Table A2

The list of the detected mutations that passed the AF ≥ 0.1 filter in all MRL8996 final populations in Chapter 3. CHR: chromosome; POS: position; AF: Allele frequency; TYPE: mutation types (TIV: transposon insertion variation; SNV: single nucleotide variation; INDEL: small insertion or deletion); MUTATION: if TYPE is TIV – the TE name, if TYPE is SNV or INDEL – nucleotide change; GENE: the protein ID of the gene if the mutation is in a gene or 1 kb region of a gene; LOC: location of the mutation with respect to gene; SAM: sample name; AA CHANGE: if the mutation is SNV or INDEL in CDS, the change in the amino acid ({FS}: frameshift, *: stop codon).

CHR	*POS*	*AF*	*TYPE*	*MUTATION*	*GENE*	*LOCATION*	*SAM*	*AA CHANGE*
C_10101	265931	0.91	TIV	Foxy5	414961	upstream	B1	
C_10101	303929	0.33	TIV	CopiaActive	358238	CDS	D3	
C_10101	319132	0.60	TIV	MRL8996_family-14	358240	upstream	C1	
C_10101	415666	0.90	TIV	Foxy5	865	upstream	D2	
C_10101	529204	0.45	SNV	C→T	476075	CDS	D1	T>I
C_10101	573039	0.48	TIV	Foxy5	481665	downstream	C4	
C_10101	573040	0.10	TIV	Foxy5	481665	downstream	C3	
C_10101	783598	0.14	TIV	Foxy5	481741	downstream	C3	
C_10101	791773	0.14	TIV	Foxy5	448857	downstream	C2	
C_10101	797760	0.35	TIV	Foxy5	476166	CDS	C4	
C_10101	1101468	0.25	INDEL	GCATCACATCACATCA→G		intergenic	B4	
C_10102	389635	0.22	TIV	Foxy5	7282	downstream	C3	
C_10102	492070	0.19	SNV	A→C	436466	CDS	C3	P>P
C_10102	515462	0.74	SNV	C→G	449065	CDS	D1	E>Q
C_10102	515462	0.51	SNV	C→G	449065	CDS	A3	E>Q
C_10102	678096	0.34	TIV	Foxy5	8673	downstream	C4	
C_10102	712074	0.13	TIV	Foxy5	334137	intron	C4	
C_10102	910981	0.55	TIV	Foxy5	10053	upstream	C5	
C_10102	1199959	0.22	TIV	Foxy5	11360	downstream	C4	
C_10102	1229697	0.36	TIV	Foxy5	11439	UTR 5'	D1	
C_10102	1315060	0.11	TIV	Foxy5	416660	upstream	D1	
C_10102	1343249	0.19	TIV	Foxy5	449258	upstream	C2	
C_10102	1367556	0.44	TIV	Foxy5	476730	downstream	D1	
C_10102	1494511	0.37	SNV	T→C		intergenic	C3	
C_10102	1494511	0.30	SNV	T→C		intergenic	C4	
C_10102	1555694	0.10	TIV	Foxy5	359497	upstream	C4	
C_10102	1581254	0.37	TIV	Foxy5	416846	upstream	C1	
C_10102	1586393	0.10	TIV	Foxy5	436829	CDS	D2	
C_10102	1619941	0.42	INDEL	AGAG→A	449337	CDS	A4	KR>K
C_10102	1622309	0.11	TIV	Foxy5	416869	intron	C3	
C_10102	1701619	0.37	TIV	Foxy5	12496	upstream	C4	
C_10102	1731104	0.25	SNV	G→A	482558	CDS	C3	G>E

C_10102	1734382	0.11	TIV	Foxy2	436885	upstream	C1	
C_1011	38946	0.10	SNV	G→T	509620	CDS	D1	I>M
C_1011	41383	0.21	SNV	C→T		intergenic	B2	
C_1011	80760	0.27	TIV	Foxy5	388334	intron	C4	
C_1011	113574	0.54	SNV	G→C		intergenic	B5	
C_1011	140704	0.56	TIV	Foxy2	343264	UTR 5'	A2	
C_1011	140705	0.15	TIV	Foxy5	343264	UTR 5'	C3	
C_1011	293016	0.50	TIV	Foxy5	441374	upstream	A1	
C_1011	347158	0.15	TIV	Foxy5	388474	UTR 5'	C1	
C_1011	597247	0.35	TIV	Foxy5	482744	downstream	C2	
C_1011	667063	0.67	TIV	Foxy5	509783	UTR 3'	D3	
C_1011	667063	0.10	TIV	Foxy5	509783	UTR 3'	D2	
C_1011	735071	0.40	TIV	Foxy5	14352	downstream	A3	
C_1011	1398022	0.87	SNV	G→A	17519	CDS	A5	Q>I
C_1011	1398153	0.60	INDEL	CGA→C	17519	CDS	A1	{FS}
C_1011	1398402	0.45	SNV	G→A	17519	CDS	A3	T>I
C_1011	1398424	0.43	SNV	A→C	17519	CDS	A2	F>V
C_1011	1398556	0.32	INDEL	A→AG	17519	CDS	A2	{FS}
C_1011	1398588	0.92	SNV	G→A	17519	CDS	A4	S>F
C_1011	1398783	0.57	SNV	C→T	17519	CDS	C1	G>D
C_1011	1398831	0.13	SNV	G→T	17519	CDS	B4	P>H
C_1011	1451387	0.48	TIV	Foxy5	315242	downstream	A4	
C_1011	1484971	0.17	TIV	Foxy5		intergenic	C5	
C_1011	1691921	0.36	TIV	Foxy5	465621	downstream	C5	
C_1011	1695656	0.35	TIV	Foxy5	441710	intron	B2	
C_1011	1729245	0.26	SNV	G→A	20006	CDS	D3	G>S
C_1011	1729595	0.11	TIV	Foxy5	20006	CDS	C5	
C_10111	40016	0.18	SNV	G→A		intergenic	D1	
C_10111	68767	0.31	TIV	Foxy5	417003	upstream	C2	
C_10111	99835	0.37	TIV	Foxy5	382226	CDS	A3	
C_10111	107063	0.28	TIV	Foxy5	528675	upstream	C2	
C_10112	99284	0.12	TIV	Foxy5	477209	downstream	C2	
C_10112	161528	0.13	TIV	Foxy5	417557	upstream	B1	
C_10112	166824	0.14	TIV	Foxy2	462342	CDS	D2	
C_10112	311132	0.12	TIV	Foxy5	483577	upstream	D2	
C_10112	423440	0.23	TIV	Foxy5		intergenic	C2	
C_10112	764597	0.45	TIV	Foxy5	25671	upstream	D1	
C_10112	784673	0.26	TIV	Foxy5	437485	CDS	D3	
C_10112	835458	0.15	TIV	Foxy5	462536	UTR 5'	C4	
C_10112	1013802	0.12	TIV	Foxy5	477568	downstream	C3	
C_10112	1084188	0.20	TIV	Foxy5	26564	upstream	C2	
C_10112	1134010	0.38	TIV	Foxy5	477617	downstream	B2	
C_10112	1212376	0.12	TIV	MRL8996_family-4	529141	UTR 5'	C3	

C_10112	1290111	0.17	TIV	Foxy5	437681	upstream	B5	
C_10112	1336486	0.12	TIV	Foxy5	418299	upstream	C5	
C_10112	1370590	0.35	TIV	Foxy5	360808	downstream	C2	
C_10112	1385131	0.56	TIV	Foxy5	383369	intron	C5	
C_10112	1609086	0.37	TIV	Foxy5	335841	downstream	A4	
C_10112	1720511	0.13	TIV	Foxy5	437846	CDS	C3	
C_1012	221462	0.32	TIV	Foxy5	30191	downstream	D3	
C_1012	305586	0.14	SNV	C→A	465785	CDS	D1	E>D
C_1012	461363	0.12	SNV	G→A	32222	UTR 5'	C5	
C_10122	394651	0.96	SNV	C→G	450524	CDS	A3	S>S
C_10124	16483	0.53	TIV	Foxy5	450773	downstream	A1	
C_10125	125700	0.93	TIV	Foxy5	419324	upstream	B5	
C_10125	326460	0.12	SNV	G→T	520380	CDS	B4	T>K
C_1013	1073210	0.13	TIV	Foxy5	426424	CDS	C5	
C_1013	1107405	0.10	TIV	Foxy5	43481	upstream	D2	
C_10131	478	0.72	SNV	C→T		intergenic	D3	
C_10131	253744	0.11	TIV	Foxy5	384862	upstream	B3	
C_10132	87641	0.25	INDEL	TG→T	485393	UTR 5'	D1	
C_10132	213523	0.14	SNV	C→A	46067	CDS	B4	S>I
C_10132	221205	0.21	TIV	Foxy5	451129	downstream	C3	
C_10133	62049	0.32	TIV	Foxy5	479026	upstream	C4	
C_10133	62050	0.17	TIV	Foxy5	479026	downstream	C2	
C_10133	78073	0.41	TIV	Foxy5	362521	UTR 5'	A3	
C_10133	78073	0.17	TIV	Foxy5	362521	UTR 5'	B2	
C_10133	78074	0.20	TIV	Foxy5	362521	UTR 5'	A1	
C_10133	335971	0.12	TIV	Foxy5	419941	CDS	C3	
C_10133	494852	0.31	TIV	Foxy5	420054	UTR 3'	C1	
C_10133	520339	0.37	INDEL	T→TCTA	385351	UTR 3'	D1	
C_10133	751350	0.13	TIV	Foxy5		intergenic	C1	
C_10134	133835	0.33	TIV	Foxy5	363046	upstream	C2	
C_10134	178878	0.39	TIV	Foxy5	339587	upstream	A3	
C_10134	279077	0.18	TIV	Foxy5	339706	downstream	C4	
C_10134	304614	0.18	TIV	MRL8996_family-4	48421	upstream	C4	
C_10134	433117	0.33	TIV	Foxy5	451528	CDS	C1	
C_10134	535626	0.11	TIV	Foxy5		intergenic	C2	
C_10134	716870	0.90	SNV	C→A	385960	upstream	B1	
C_10134	872830	0.14	TIV	Foxy5	386060	UTR 3'	C4	
C_10134	885806	0.13	TIV	Foxy5	386064	upstream	C3	
C_10134	885815	0.15	TIV	Foxy5	386064	upstream	C5	
C_10134	1016962	0.76	SNV	G→T	339274	downstream	D2	
C_10134	1016962	0.19	SNV	G→T	339274	downstream	C3	
C_10134	1040467	0.23	TIV	Foxy5	451735	upstream	C1	
C_1014	224076	0.42	TIV	Foxy5	51795	upstream	C2	

C_1014	224077	0.11	TIV	Foxy5	51795	upstream	C4
C_1014	492215	0.23	TIV	Foxy5		intergenic	B4
C_1015	184782	0.41	TIV	Foxy2	55526	upstream	B3
C_1015	749725	0.22	TIV	Foxy5	466674	upstream	C4
C_1015	1066629	0.12	TIV	Foxy5	427015	upstream	C2
C_1015	1182145	0.47	TIV	Foxy5	393213	upstream	D2
C_1015	1193716	0.11	TIV	Foxy5	454313	UTR 3'	D3
C_1015	1257136	0.17	TIV	Foxy5	466864	upstream	D1
C_1015	1545098	0.26	TIV	Foxy5	442608	intron	C4
C_1015	1835280	0.44	TIV	Foxy5	467084	CDS	A2
C_1015	1856174	0.48	INDEL	C→CA	393636	upstream	A2
C_1015	1856174	0.15	INDEL	C→CA	393636	upstream	C3
C_1015	1867645	0.48	SNV	C→T	442680	inron	B3
C_1015	1901330	0.10	TIV	Foxy5		intergenic	C4
C_1022	504847	0.23	TIV	Foxy5	454648	upstream	C4
C_1022	623884	0.17	SNV	A→G	64862	upstream	C3
C_1022	643463	0.13	TIV	Foxy2	64977	downstream	D2
C_1022	715885	0.30	TIV	Foxy5	511465	upstream	C4
C_1022	715885	0.20	TIV	Foxy5	511465	upstream	D2
C_1022	802467	0.13	TIV	Foxy5	394333	downstream	C3
C_1022	1214633	0.68	TIV	Foxy5	523488	upstream	D1
C_1022	1221733	0.24	TIV	Foxy5	67649	UTR 3'	C2
C_1022	1333540	0.20	TIV	Foxy5		intergenic	D1
C_1022	1356884	0.13	TIV	Foxy2		intergenic	B5
C_1022	1446880	0.17	TIV	Foxy5	454912	u	C3
C_1022	1446881	0.42	TIV	Foxy5	454912	upstream	B5
C_1022	1447325	0.44	TIV	Foxy5	454912	upstream	C1
C_1022	1506974	0.13	TIV	Foxy5	69742	upstream	D2
C_1022	1746433	0.33	TIV	Foxy5	71108	downstream	B3
C_1022	1769036	0.49	TIV	Foxy5	71208	upstream	C1
C_1022	2295134	0.48	TIV	Foxy5		intergenic	D1
C_1022	2441930	0.18	TIV	Foxy5	75922	downstream	D3
C_1022	2445192	0.13	TIV	Foxy5	428168	CDS	C3
C_1022	2445193	0.28	TIV	Foxy5	428168	CDS	D3
C_1022	2445193	0.20	TIV	Foxy5	428168	CDS	C4
C_1022	2445194	0.22	TIV	Foxy5	428168	CDS	D1
C_1022	2445195	0.12	TIV	Foxy5	428168	CDS	C2
C_1022	2485735	0.16	TIV	Foxy5	76274	UTR 5'	C2
C_1022	2718911	0.12	TIV	Foxy5	395900	downstream	C2
C_1022	2819602	0.27	TIV	Foxy5	468065	upstream	A1
C_1022	2974501	0.18	SNV	C→T	79211	upstream	D3
C_1022	2994475	0.45	TIV	Foxy5	455297	downstream	D1
C_1022	3002318	0.12	TIV	Foxy5	79359	upstream	D1

C_1022	3065600	0.20	TIV	Foxy5		intergenic	D1	
C_1022	3813399	0.54	TIV	Foxy5	443480	upstream	B5	
C_1022	4133679	0.21	SNV	C→T	524023	CDS	D1	L>F
C_1041	202021	0.44	TIV	Foxy5	370923	upstream	A1	
C_1041	276163	0.31	TIV	Foxy5	90690	downstream	C1	
C_1041	276163	0.16	TIV	Foxy5	90690	upstream	C4	
C_1041	330230	0.14	TIV	Foxy5	443803	upstream	D3	
C_1041	353660	0.49	TIV	Foxy5	348374	intron	B5	
C_1042	163250	0.62	TIV	Foxy5	91528	upstream	A5	
C_1042	228732	0.24	SNV	G→A	443907	CDS	C3	D>N
C_1042	612280	0.64	SNV	C→T		intergenic	D3	
C_1042	773126	0.20	TIV	Foxy5		intergenic	C4	
C_1042	862791	0.19	SNV	G→A	321269	CDS	B4	P>S
C_1042	1259219	0.10	TIV	Foxy5	469251	upstream	B4	
C_1043	220531	0.19	TIV	Foxy5	104854	upstream	D3	
C_1044	69443	0.68	SNV	G→A	429645	intron	D1	
C_1044	69443	0.50	SNV	G→A	429645	intron	A3	
C_1044	89120	0.13	TIV	Foxy5	469492	upstream	D2	
C_1044	260181	0.81	TIV	Foxy5	106421	CDS	D3	
C_1044	463438	0.12	TIV	Foxy5	524646	UTR 3'	C4	
C_1044	463587	0.11	INDEL	G→GA	524646	UTR 3'	C4	
C_1044	523148	0.11	TIV	Foxy5	107190	downstream	A2	
C_1044	523148	0.14	TIV	Foxy5	107190	downstream	C4	
C_1044	573087	0.14	SNV	C→A	322115	CDS	B4	N>K
C_1044	715982	0.25	SNV	C→G	321638	CDS	B2	V>L
C_1044	858931	0.49	TIV	Foxy5	400000	upstream	D1	
C_1044	858933	0.29	TIV	Foxy5	400000	upstream	D3	
C_1044	913857	0.11	TIV	Foxy5	513449	upstream	C3	
C_1044	1085792	0.49	SNV	C→T	430019	CDS	B2	S>S
C_1044	1150578	0.44	TIV	Foxy5	513531	UTR 3'	D1	
C_1044	1238523	0.15	TIV	Foxy5	108975	downstream	C3	
C_1044	1238523	0.11	TIV	Foxy5	108975	upstream	B4	
C_1044	1456933	0.32	TIV	Foxy5	400564	upstream	D2	
C_1044	1456933	0.28	TIV	Foxy5	400564	upstream	C4	
C_1044	1456933	0.71	TIV	Foxy5	400564	upstream	A5	
C_1044	1456934	1.00	TIV	Foxy5	400564	upstream	A4	
C_1044	1479658	0.16	TIV	Foxy5	469997	CDS	B2	
C_1044	1479659	0.32	TIV	Foxy5	469997	CDS	B3	
C_1044	1509731	0.18	TIV	Foxy5	444589	downstream	C2	
C_1044	1633882	0.27	TIV	Foxy5		intergenic	A2	
C_1045	119772	0.44	TIV	Foxy5	430266	downstream	D1	
C_1045	586794	0.61	SNV	A→G	470291	CDS	C1	I>M
C_1045	759797	0.29	TIV	Foxy5	457071	CDS	C2	

C_1045	803867	0.15	SNV	G→T	115842	downstream	B2	
C_1045	830234	0.10	TIV	Foxy5	401371	downstream	C3	
C_1045	878767	0.54	TIV	Foxy5	457099	CDS	C5	
C_1045	896867	0.14	SNV	C→A	350620	CDS	C2	G>V
C_1045	983906	0.20	TIV	Foxy5	457119	CDS	A1	
C_1045	1141496	0.11	TIV	MRL8996_family-4	444990	upstream	D3	
C_1051	91516	0.46	TIV	Foxy5	514022	downstream	D1	
C_1051	91517	0.13	TIV	Foxy5	514022	downstream	C3	
C_1051	176523	0.21	TIV	Foxy5	116623	downstream	B5	
C_1051	190750	0.73	TIV	Foxy5		intergenic	D3	
C_1051	434314	0.33	TIV	Foxy5	401769	UTR 5'	C3	
C_1051	434513	0.20	TIV	Foxy5	401769	upstream	D1	
C_1051	434513	0.28	TIV	Foxy5	401769	upstream	A1	
C_1051	487097	0.11	TIV	Foxy5	117104	CDS	C3	
C_1052	446382	0.11	TIV	Foxy5	120987	downstream	D3	
C_1052	611405	0.10	TIV	Foxy5	470959	upstream	D1	
C_1052	837489	0.52	SNV	A→C	431163	CDS	C5	T>P
C_1052	837828	0.51	SNV	A→G	431163	CDS	C2	T>A
C_1052	878389	0.11	TIV	Foxy5	124592	UTR 5'	C5	
C_1052	1127707	0.25	TIV	Foxy5	445442	downstream	C4	
C_1052	1203564	0.29	TIV	Foxy5	127290	UTR 5'	C4	
C_1052	1203564	0.37	TIV	Foxy5	127290	UTR 5'	A3	
C_1053	222259	0.43	TIV	Foxy5	431347	intron	D1	
C_1053	222259	0.27	TIV	Foxy5	431347	intron	C3	
C_1053	222259	0.11	TIV	Foxy5	431347	intron	B3	
C_1053	222259	0.11	TIV	Foxy5	431347	intron	D2	
C_1054	23533	0.59	TIV	Foxy5	431405	downstream	C5	
C_1054	23533	0.11	TIV	Foxy5	431405	upstream	D2	
C_1054	100555	0.71	TIV	Foxy5	471287	upstream	D3	
C_1055	27195	0.33	TIV	Foxy5		intergenic	A2	
C_1055	203423	0.15	TIV	Foxy5	404305	downstream	D3	
C_1055	398077	0.17	TIV	Foxy5	404542	downstream	A1	
C_1055	538145	0.47	TIV	Foxy5	404707	downstream	C4	
C_1055	569529	0.13	TIV	Foxy2	139565	upstream	A4	
C_1055	569531	0.12	TIV	Foxy5	139565	upstream	D2	
C_1055	828299	0.87	TIV	Foxy5		intergenic	D3	
C_1055	894940	0.66	SNV	C→A	141735	CDS	D3	E>D
C_1055	932905	0.37	SNV	A→C		intergenic	C3	
C_1055	971632	0.20	TIV	Foxy5	142116	upstream	D2	
C_1055	971632	0.11	TIV	Foxy5	142116	upstream	A2	
C_1055	971634	0.52	TIV	Foxy5	142116	upstream	C2	
C_1055	971634	0.51	TIV	Foxy5	142116	upstream	C5	
C_1055	971634	0.37	TIV	Foxy5	142116	upstream	C1	

C_1055	971634	0.19	TIV	Foxy5	142116	upstream	D3	
C_1055	971634	0.39	TIV	Foxy5	142116	upstream	B3	
C_1055	971635	0.48	TIV	Foxy5	142116	upstream	C3	
C_1055	971635	0.34	TIV	Foxy5	142116	upstream	C4	
C_1055	1511056	0.20	TIV	Foxy5	492185	upstream	C4	
C_1055	1695859	0.10	SNV	G→A	325641	CDS	C1	L>L
C_1071	49175	0.22	INDEL	AG→A		intergenic	D1	
C_1071	362552	0.13	SNV	G→T	492403	CDS	D3	E>D
C_1071	384545	0.10	TIV	Foxy5	492408	upstream	C5	
C_1071	745922	0.44	SNV	T→C	458519	CDS	A4	K>K
C_1071	1019404	0.51	TIV	MRL8996_family-14	472310	CDS	C1	
C_1071	1235962	0.18	TIV	Foxy5	472392	downstream	C4	
C_1071	1509110	0.17	TIV	Foxy5	327206	upstream	C4	
C_1071	1509111	0.15	TIV	Foxy5	327206	upstream	C3	
C_1071	1595408	0.21	TIV	Foxy5	151417	downstream	C2	
C_1071	1805071	0.12	TIV	Foxy2	153046	downstream	A1	
C_1071	1805072	0.36	TIV	Foxy5	153046	upstream	C3	
C_1071	1805072	0.12	TIV	Foxy5	153046	upstream	C4	
C_1071	1829670	0.14	SNV	C→T	153193	CDS	A2	Q>*
C_1071	1829670	0.97	SNV	C→T	153193	CDS	B1	Q>*
C_1071	1947901	0.17	SNV	T→C	154318	downstream	A1	
C_1071	2135346	0.11	TIV	Foxy5	407223	UTR 3'	C3	
C_1071	2225196	0.47	SNV	G→A	493011	CDS	B5	A>V
C_1071	2423080	0.18	TIV	Foxy5	157919	UTR 5'	C4	
C_1071	2481108	0.31	TIV	Foxy5	158484	CDS	C4	
C_1071	2481645	0.14	TIV	Foxy5	158484	CDS	C3	
C_1071	2651754	0.17	TIV	Foxy5	327531	upstream	A4	
C_1071	2709363	0.13	TIV	Foxy5	432977	upstream	C3	
C_1071	2758423	0.13	TIV	Foxy5	472873	CDS	A5	
C_1071	2789022	0.17	TIV	Foxy5	161360	upstream	D1	
C_1071	3098404	0.14	TIV	Foxy5	165204	upstream	A2	
C_1071	3399886	0.25	TIV	Foxy5	408183	UTR 3'	C3	
C_1071	3545875	0.17	TIV	Foxy5		intergenic	C4	
C_1071	3800345	0.42	TIV	Foxy5	433315	upstream	C4	
C_1071	3835483	0.15	TIV	Foxy5	169328	upstream	C3	
C_1071	4277770	0.10	TIV	Foxy5	433469	downstream	C2	
C_1071	4307229	0.15	TIV	Foxy5	408867	downstream	C2	
C_1071	4466436	0.25	TIV	Foxy5	526707	upstream	D3	
C_1071	4471549	0.12	TIV	Foxy5	526709	UTR 3'	A4	
C_1081	3237	0.18	TIV	Foxy5	354739	downstream	D1	
C_1082	61584	0.13	TIV	Foxy5	173098	downstream	C4	
C_1082	61649	0.12	TIV	Foxy5	173098	UTR 5'	D2	
C_1082	396139	0.25	TIV	Foxy5	473734	downstream	A1	

C_1082	874584	0.13	TIV	Foxy5		intergenic	C2	
C_1082	1227989	0.31	TIV	Foxy5		intergenic	C2	
C_1082	1447017	0.18	TIV	Foxy5	410279	upstream	D1	
C_1082	1478940	0.21	SNV	A→G	459920	downstream	D3	
C_1082	1576694	0.13	TIV	Foxy5	181501	UTR 5'	B3	
C_1083	508912	0.11	TIV	Foxy5	186160	UTR 5'	C2	
C_1083	1003690	0.18	TIV	Foxy5	411367	CDS	B5	
C_1083	1241194	0.10	TIV	Foxy5	191172	intron	D1	
C_1083	1396927	0.31	SNV	A→G		intergenic	D1	
C_1083	1457476	0.23	TIV	Foxy5	356074	intron	C3	
C_1083	1594624	0.51	SNV	G→C	411808	upstream	C2	
C_1083	1677266	0.32	TIV	Foxy5	517314	CDS	C3	
C_1083	1682454	0.10	TIV	Foxy5	495088	CDS	D2	
C_1083	1881482	0.13	TIV	Foxy5	447803	upstream	C3	
C_1083	1898513	0.34	TIV	Foxy5	527392	downstream	C4	
C_1083	1902050	0.27	SNV	C→A	378927	CDS	D3	W>L
C_1083	2002595	0.27	TIV	Foxy5	356400	CDS	C3	
C_1083	2052976	0.63	TIV	Foxy5	460476	upstream	D3	
C_1083	2137217	0.11	TIV	Foxy5	379050	CDS	A3	
C_1091	113635	0.34	TIV	MRL8996_family-14	330536	downstream	D3	
C_1091	169532	0.39	TIV	Foxy5	412248	downstream	A4	
C_1091	283358	0.19	TIV	Foxy5	527486	downstream	C3	
C_1091	283358	0.11	TIV	Foxy5	527486	upstream	A2	
C_1091	384644	0.24	TIV	Foxy5	412369	downstream	C2	
C_1091	497719	0.23	TIV	Foxy5	448269	upstream	C2	
C_1091	506015	0.11	TIV	Foxy2	460596	upstream	C3	
C_1091	506880	0.15	SNV	G→A		intergenic	B5	
C_1092	300709	0.12	INDEL	C→CA	198496	CDS	D2	{FS}
C_1092	305069	0.20	SNV	G→A	435173	CDS	B3	A>V
C_1092	323807	0.38	TIV	Foxy2	475114	downstream	D3	
C_1092	368589	0.22	TIV	Foxy5	331863	downstream	C3	
C_1092	368590	0.41	TIV	Foxy5	331863	downstream	C5	
C_1092	368590	0.37	TIV	Foxy5	331863	downstream	C4	
C_1092	424807	0.16	TIV	Foxy5	475143	upstream	D3	
C_1092	495337	0.29	TIV	Foxy5	412885	downstream	C3	
C_1092	626535	0.16	TIV	Foxy5	200443	downstream	C4	
C_1092	702467	0.19	TIV	Foxy5	201076	downstream	C2	
C_1092	702531	0.13	TIV	Foxy5	201076	downstream	C5	
C_1092	702531	0.65	TIV	Foxy5	201076	downstream	A4	
C_1092	702533	0.66	TIV	Foxy5	201076	downstream	A5	
C_1092	721591	0.20	TIV	Foxy5	495702	downstream	C2	
C_1092	907830	0.14	SNV	G→A	202647	CDS	B4	T>M
C_1092	1023572	0.23	TIV	Foxy5	448269	upstream	C2	

C_1092	1111632	0.16	TIV	Foxy5	495825	CDS	C3
C_1092	1410659	0.73	TIV	Foxy5	495947	upstream	A5
C_1092	2387085	0.38	SNV	C→A	414568	UTR 3'	B3
C_1092	2444947	0.19	SNV	G→A	475786	upstream	A2
C_1092	2491983	0.14	TIV	Foxy5	414611	upstream	A2
C_1092	2551164	0.16	TIV	Foxy5	461276	CDS	D1
C_1092	2583948	0.38	TIV	Foxy5	448591	downstream	B2
S_1	184315	0.50	SNV	G→A		intergenic	D3
S_10	53547	0.42	TIV	Foxy5	421853	downstream	D1
S_10	53547	0.13	TIV	Foxy5	421853	upstream	C2
S_10	53548	0.14	TIV	Foxy5	421853	upstream	C3
S_102	74	0.31	INDEL	G→GAT	424872	UTR 3'	A2
S_115	12548	0.10	TIV	Foxy5	452823	downstream	B2
S_116	10924	0.50	TIV	Foxy5		intergenic	B5
S_12	100096	0.33	TIV	Foxy5		intergenic	C4
S_13	59210	0.15	INDEL	A→AT	521378	downstream	D2
S_14	18955	0.22	TIV	Foxy5	521392	downstream	C5
S_14	30666	0.13	TIV	Foxy5	452109	intron	B2
S_14	30667	0.16	TIV	Foxy5	452109	CDS	C1
S_15	28107	0.33	TIV	Foxy5	497167	CDS	C3
S_15	31017	0.18	TIV	Foxy5	440212	downstream	C3
S_152	4832	0.11	TIV	Foxy5	425234	CDS	C2
S_16	45701	0.17	TIV	Foxy5	422242	upstream	C1
S_17	43998	0.16	TIV	Foxy5	497384	UTR 3'	C3
S_19	21411	0.57	TIV	Foxy5	341207	upstream	B4
S_2	39077	0.45	INDEL	GA→G	439762	upstream	B4
S_2	39139	0.35	SNV	T→G	439762	upstream	B4
S_2	39168	0.29	SNV	A→T	439762	upstream	B4
S_2	162928	0.24	TIV	Foxy5	451828	upstream	A4
S_2	164677	0.26	TIV	Foxy5	421250	CDS	C3
S_2	164678	0.14	TIV	Foxy5	421250	CDS	C2
S_2	164678	0.14	TIV	Foxy5	421250	CDS	C4
S_20	81007	0.12	TIV	Foxy5	497722	downstream	A4
S_20	81008	0.52	TIV	Foxy5	497722	downstream	C2
S_21	69563	0.20	TIV	Foxy5		intergenic	C2
S_21	69564	0.11	TIV	Foxy5		intergenic	C1
S_21	79546	0.22	TIV	Foxy5	497778	upstream	C5
S_22	68193	0.42	TIV	Foxy5	497813	downstream	C2
S_22	68193	0.17	TIV	Foxy5	497813	downstream	C4
S_24	24753	0.16	TIV	Foxy5	452268	upstream	C2
S_24	40900	0.14	INDEL	G→GCA		intergenic	A1
S_25	46854	0.18	TIV	Foxy5	222217	upstream	C2
S_25	52736	0.13	TIV	Foxy5	341466	CDS	C3

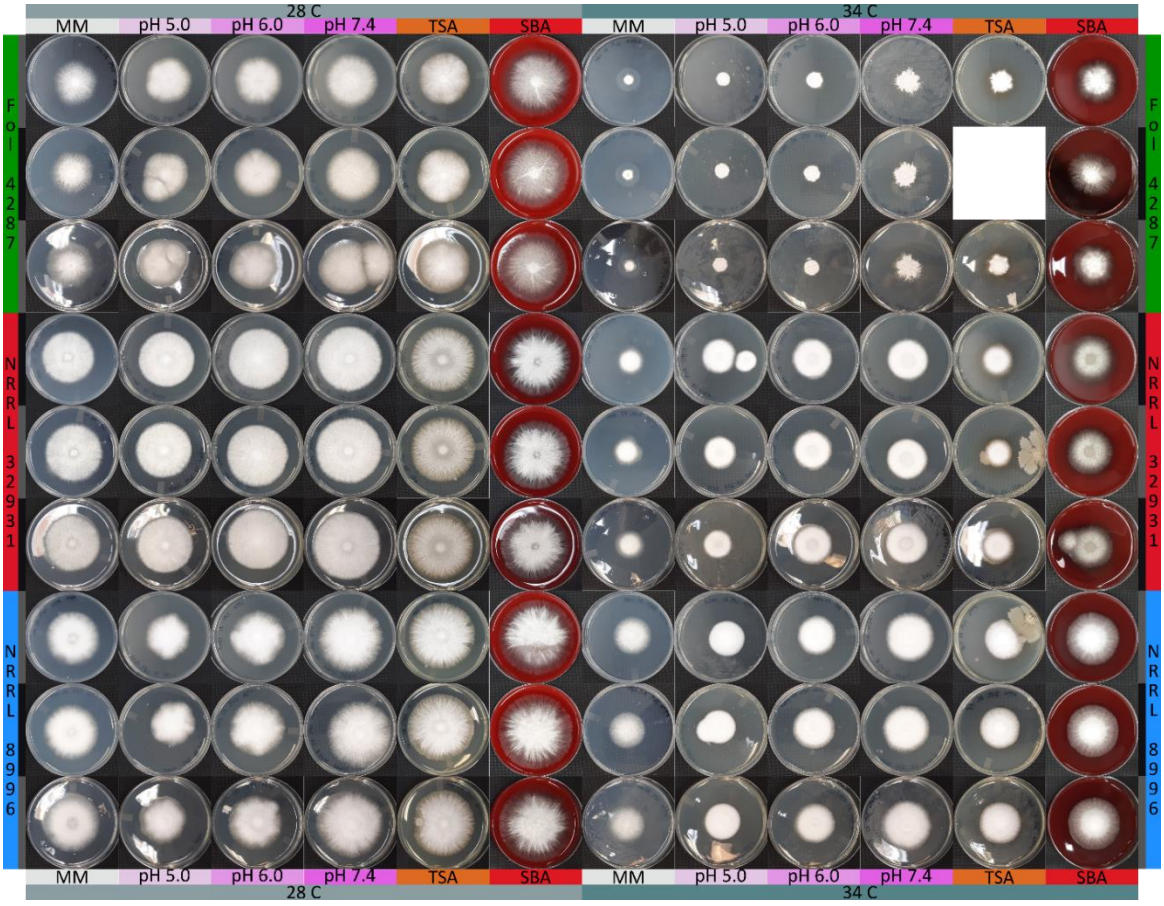
S_25	52736	0.11	TIV	Foxy5	341466	CDS	C2	
S_25	60100	0.13	TIV	Foxy5	222419	upstream	A1	
S_26	68245	0.23	TIV	Foxy5		intergenic	C4	
S_26	68246	0.60	TIV	Foxy5		intergenic	A5	
S_26	68247	1.00	TIV	Foxy5		intergenic	A4	
S_26	68251	0.11	TIV	FoxyAJ250814		intergenic	C3	
S_28	4694	0.38	TIV	Foxy5	341581	upstream	C1	
S_28	24487	0.10	TIV	Foxy5	521613	upstream	C4	
S_29	9668	0.11	TIV	Foxy5	341614	downstream	C5	
S_3	48061	0.46	SNV	C→G	340084	CDS	A3	G>R
S_3	127671	0.47	SNV	G→A	464124	CDS	A3	I>I
S_3	154946	0.13	TIV	Foxy5	451874	upstream	C4	
S_30	9787	0.41	TIV	Foxy5	423108	upstream	C3	
S_32	4667	0.22	TIV	Foxy5	313056	CDS	C4	
S_32	4668	0.27	TIV	Foxy5	313056	CDS	C3	
S_32	49810	0.11	TIV	Foxy5	423252	upstream	D1	
S_32	62285	0.13	TIV	CopiaActive		intergenic	A4	
S_33	32783	0.36	TIV	Foxy5	387382	CDS	B3	
S_33	32784	0.69	TIV	Foxy5	387382	CDS	D3	
S_34	40500	0.45	SNV	C→T	423291	CDS	A5	T>T
S_38	25540	0.16	TIV	Foxy5	521715	downstream	D1	
S_39	40539	0.10	TIV	Foxy5	464664	upstream	C4	
S_4	57258	0.23	SNV	G→T	306907	downstream	B3	
S_4	102580	0.20	TIV	Foxy5		intergenic	C1	
S_41	992	0.14	TIV	Foxy5	296096	upstream	C3	
S_41	7857	0.28	TIV	Foxy5	225518	upstream	A1	
S_41	17204	0.30	TIV	Foxy5		intergenic	D2	
S_44	8372	0.13	SNV	G→C	440716	upstream	D3	
S_49	10701	0.19	TIV	Foxy5	387651	intron	C3	
S_49	10702	0.13	TIV	Foxy5	387651	intron	C2	
S_49	10703	0.27	TIV	Foxy5	387651	intron	C5	
S_49	16375	0.17	TIV	Foxy5	365194	upstream	A4	
S_49	16376	0.17	TIV	Foxy5	365194	upstream	C3	
S_49	16376	0.14	TIV	Foxy5	365194	upstream	C4	
S_5	145419	0.38	TIV	Foxy5		intergenic	C1	
S_57	34790	0.15	TIV	Foxy5	424035	downstream	B5	
S_58	14706	0.23	TIV	Foxy5	480899	upstream	C3	
S_59	6912	0.50	TIV	Foxy5	424071	CDS	B3	
S_59	22747	0.25	TIV	Foxy5		intergenic	C2	
S_6	32505	0.11	TIV	Foxy5	340335	CDS	C2	
S_6	99062	0.11	TIV	Foxy2	340335	CDS	A4	
S_6	106566	0.12	TIV	Foxy5	340311	downstream	C3	
S_6	121615	0.44	TIV	Foxy5	340370	downstream	C3	

S_62	19391	0.85	SNV	A→T		intergenic	A5	
S_7	18789	0.37	TIV	Foxy5	479979	upstream	D1	
S_7	90007	0.15	TIV	Foxy5	421605	downstream	B2	
S_7	98332	0.15	TIV	Foxy5		intergenic	C4	
S_8	109224	0.16	SNV	C→A	300415	upstream	A2	
S_82	10645	0.13	SNV	G→C	342716	CDS	B4	R>T
S_87	6813	0.20	TIV	MRL8996_family-4	441051	upstream	D3	
S_87	8566	0.11	TIV	Foxy5	424591	downstream	B4	
S_89	2028	0.58	SNV	T→C	441061	downstream	D2	
S_89	3071	0.82	INDEL	TACATA→T	441061	upstream	D2	

APPENDIX B

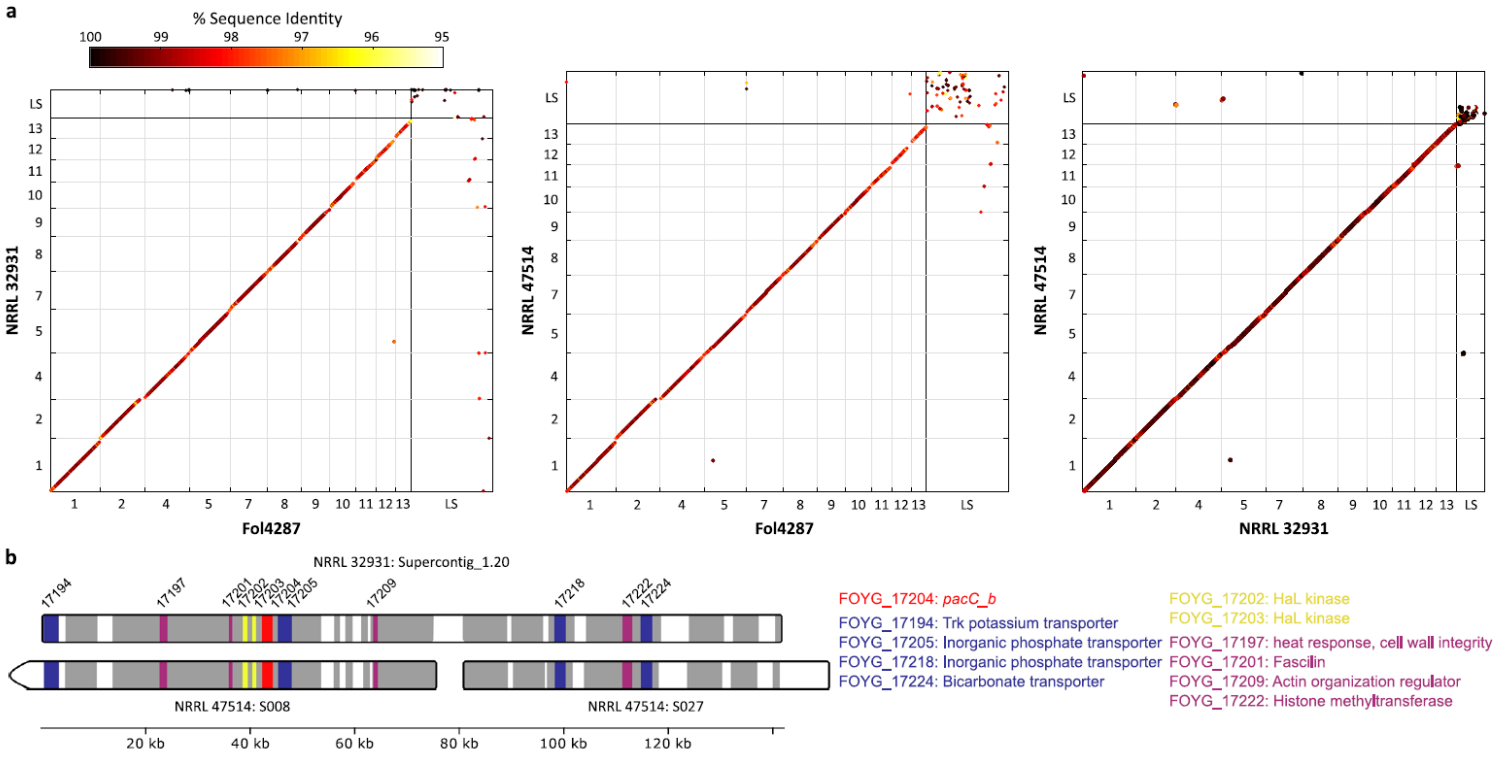
Extended Figure B1

The growth of Fol4287, NRRL 32931, or MRL8996 in following medium types: MM agar, modified Czapek-Dox; YPD agar with pH adjusted to 5.0, 6.0, and 7.4; TSA, Tryptic Soy Agar (BD Difco, USA); and SBA, TSA with 5% sheep blood (Remel, USA). The YPD agar medium pH was adjusted using citric acid-sodium phosphate buffer. 0.1 M citric acid was mixed with 0.2 M Na_2HPO_4 with following volumes: For pH 5.0: 53.25 ml and 46.75 ml; for pH 7.4: 39.55 ml and 60.45 ml; for pH 7.4: 13.05 ml and 86.95 ml, respectively. The buffer was mixed with 100 ml 2× YPD agar. 5 µl of conidia solution with 5×10^6 conidia/ml concentration was inoculated in small petri dishes with agar medium. The cultures were incubated at either 28°C or 34°C and photographed on the third day. There were three replicates per sample and there was contamination observed in some TSA plates at 34°C.



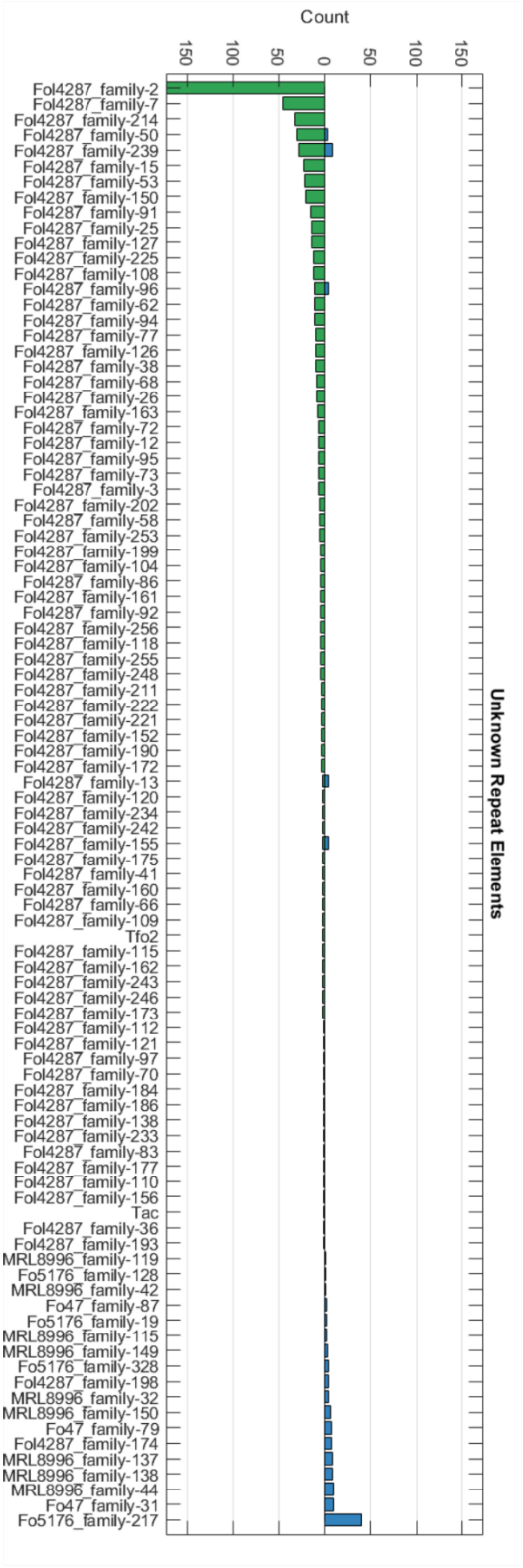
Extended Figure B2

(a) Pairwise genome comparisons of NRRL 32931, NRRL 47514 (MRL8996), and Fol4287 (LS: lineage-specific sequences or accessory chromosome (AC) sequences). 37% NRRL 32931 masked AC sequences have homologous sequences in the MRL8996 genome, whereas there is only 2.3% in the Fol4287 genome. With a larger AC region in the MRL8996 genome, two human pathogens shared AC sequences and represent 17% of MRL8996 masked AC sequences; about 7% of MRL8996 masked AC sequences also have homologous sequences in the genome of Fol4287. (b) Alignment of Supercontig 20 of chromosome 14 of NRRL 32931 using ClustalW with default parameters in MEGA X [165]. The supercontig includes *pacC* paralog *pacC_b* to MRL8996 AC contigs S008 and S027. Syntenic regions are shown in gray and colored bars. Annotated genes are colored, the gene locus tag numbers are labeled, and more information about the genes is next to the diagram. Although whole sequences of Supercontig 20 and S027 are shown, the S008 contig is partial. After removing repetitive sequences, the aligned sequences between Supercontig 20 and two NRRL 47514 contigs (S008 and S027) are almost identical, with 99.16% and 98.98% identity across the entire region [32].



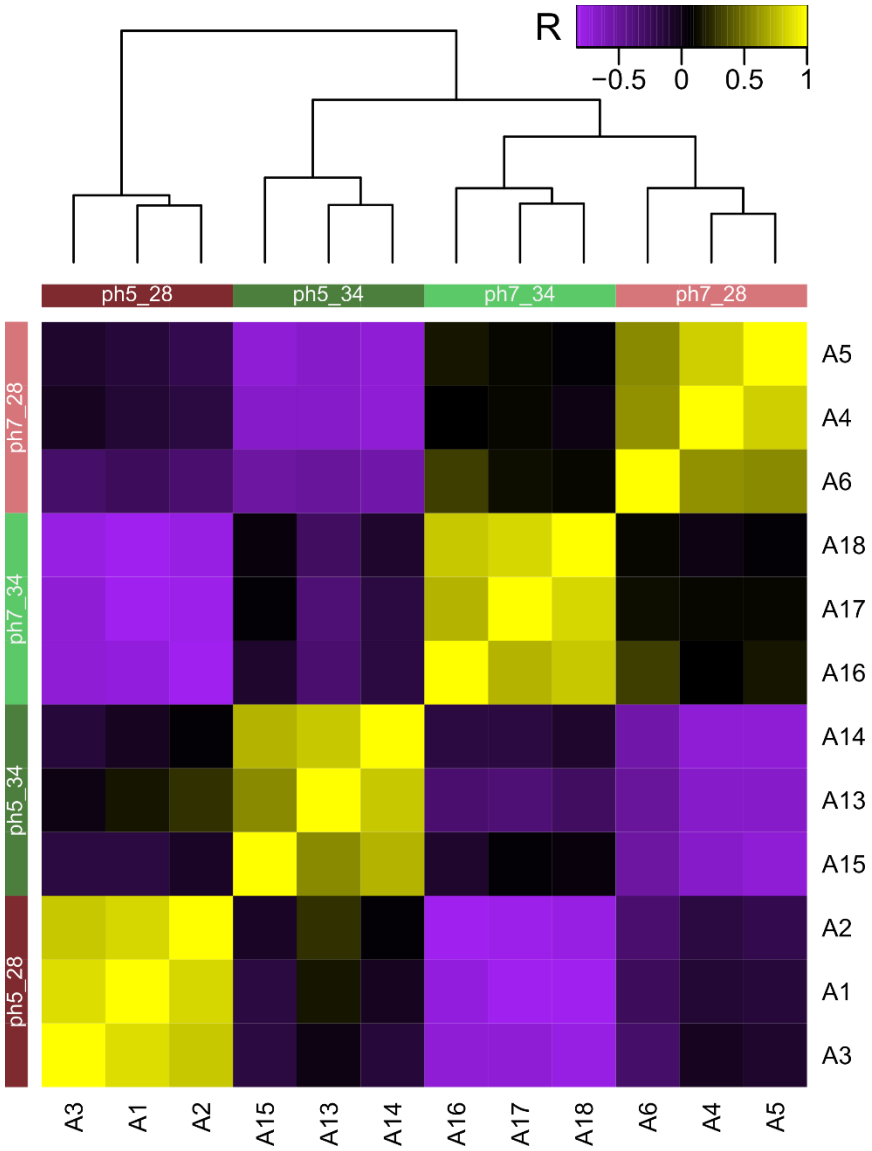
Extended Figure B3

The unclassified repeat element abundances in Fol4287 (green) and MRL8996 (blue).



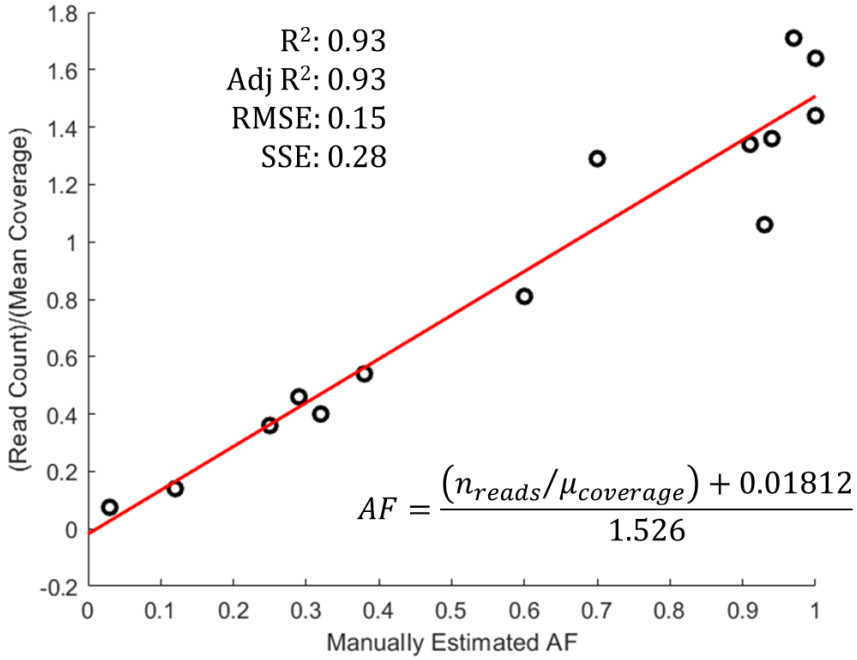
Extended Figure B4

RNA sequencing experiment was carried as described in Methods 2.2.5 with MRL8996 in YPD pH 5.0 at 28°C, YPD pH 7.4 at 28°C, YPD pH 5.0 at 34°C, and YPD pH 7.4 at 34°C. The mapping was performed as described. Transcripts per million (TPM) values were generated by Stringtie version 1.3.0 with options '`-e -B -p 8`' and raw read counts were converted [172]. DESeq2 version 1.30.0 was used to perform differential expression analysis and the hierarchical clustering with heatmap was generated using 8048 differentially expressed genes [173].



Extended Figure B5

The allele frequency estimation function for transposon insertion calling. For more information see the main text (3.2.4). Linear regression model statistics: R^2: R-squared; Adj R^2: adjusted R-squared; RMSE: Root Mean Square Error; SSE: Sum of squared errors. n_{reads}: the number of evidence reads for the insertion reported by TEfinder; $\mu_{coverage}$: averaged coverage in the InsRegion reported by TEfinder.



BIBLIOGRAPHY

1. Gordon TR. Fusarium oxysporum and the Fusarium Wilt Syndrome. Annual review of phytopathology. 2017;55: 23–39. doi:10.1146/annurev-phyto-080615-095919

2. Ma L-J, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, et al. Fusarium Pathogenomics. Annual Review of Microbiology. 2013;67: 399–416. doi:10.1146/annurev-micro-092412-155650

3. Michielse CB, Rep M. Pathogen profile update: Fusarium oxysporum. Molecular plant pathology. 2009;10: 311–24. doi:10.1111/j.1364-3703.2009.00538.x

4. Di Pietro A, Madrid MP, Caracuel Z, Delgado-Jarana J, Roncero MIG. Fusarium oxysporum: exploring the molecular arsenal of a vascular wilt fungus. Molecular Plant Pathology. 2003;4: 315–325. doi:10.1046/j.1364-3703.2003.00180.x

5. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. Molecular Plant Pathology. 2012;13: 414–430. doi:10.1111/j.1364-3703.2011.00783.x

6. Edel-Hermann V, Lecomte C. Current Status of Fusarium oxysporum Formae Speciales and Races. Phytopathology. 2019;109: 512–530. doi:10.1094/PHYTO-08-18-0320-RVW

7. Schäfer K, Bain JM, Pietro AD, Gow NAR, Erwig LP. Hyphal growth of phagocytosed Fusarium oxysporum causes cell lysis and death of murine macrophages. PLoS ONE. 2014;9: 1–8. doi:10.1371/journal.pone.0101999

8. Mukherjee PK, Chandra J, Yu C, Sun Y, Pearlman E, Ghannoum MA. Characterization of Fusarium Keratitis Outbreak Isolates: Contribution of Biofilms to Antimicrobial Resistance and Pathogenesis. Investigative Opthalmology & Visual Science. 2012;53: 4450. doi:10.1167/iovs.12-9848

9. Sun Y, Chandra J, Mukherjee P, Szczotka-Flynn L, Ghannoum MA, Pearlman E. A Murine Model of Contact Lens–Associated Fusarium Keratitis. Investigative Opthalmology & Visual Science. 2010;51: 1511. doi:10.1167/iovs.09-4237

10. Walther G, Stasch S, Kaerger K, Hamprecht A, Roth M, Cornely OA, et al. Fusarium Keratitis in Germany. Warnock DW, editor. Journal of Clinical Microbiology. 2017;55: 2983–2995. doi:10.1128/JCM.00649-17

11. Esnakula AK, Summers I, Naab TJ. Fatal Disseminated Fusarium Infection in a Human Immunodeficiency Virus Positive Patient. Case Reports in Infectious Diseases. 2013;2013: 1–5. doi:10.1155/2013/379320

12. Nucci M, Anaissie E. Fusarium infections in immunocompromised patients. Clinical Microbiology Reviews. 2007;20: 695–704. doi:10.1128/CMR.00014-07

13. Boutati EI, Anaissie EJ. Fusarium, a significant emerging pathogen in patients with hematologic malignancy: ten years' experience at a cancer center and implications for management. Blood. 1997;90: 999–1008.

14. Low C-Y, Rotstein C. Emerging fungal infections in immunocompromised patients. F1000 Medicine Reports. 2011;3. doi:10.3410/M3-14

15. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010;464: 367–373. doi:10.1038/nature08850

16. Vlaardingerbroek I, Beerens B, Schmidt SM, Cornelissen BJC, Rep M. Dispensable chromosomes in Fusarium oxysporum f. sp. lycopersici. Molecular Plant Pathology. 2016;17: 1455–1466. doi:https://doi.org/10.1111/mpp.12440

17. Armstrong GM, Armstrong JK. Formae speciales and races of Fusarium oxysporum causing wilt diseases. In: Nelson PE, Toussoun TA, Cook RJ, editors. Fusarium: Diseases, Biology, and Taxonomy. University Park, PA: Pennsylvania State University Press; 1981. pp. 391–399.

18. Baayen RP, O'Donnell K, Bonants PJ, Cigelnik E, Kroon LP, Roebroeck EJ, et al. Gene Genealogies and AFLP Analyses in the Fusarium oxysporum Complex Identify Monophyletic and Nonmonophyletic Formae Speciales Causing Wilt and Rot Disease. Phytopathology. 2000;90: 891–900. doi:10.1094/PHYTO.2000.90.8.891

19. Di Pietro A, García-MacEira FI, Méglecz E, Roncero MI. A MAP kinase of the vascular wilt fungus Fusarium oxysporum is essential for root penetration and pathogenesis. Molecular microbiology. 2001;39: 1140–52.

20. Jain S, Akiyama K, Mae K, Ohguchi T, Takata R. Targeted disruption of a G protein alpha subunit gene results in reduced pathogenicity in Fusarium oxysporum. Current genetics. 2002;41: 407–13. doi:10.1007/s00294-002-0322-y

21. Bayram O, Braus GH. Coordination of secondary metabolism and development in fungi: the velvet family of regulatory proteins. FEMS microbiology reviews. 2012;36: 1–24. doi:10.1111/j.1574-6976.2011.00285.x

22. López-Berges MS, Hera C, Sulyok M, Schäfer K, Capilla J, Guarro J, et al. The velvet complex governs mycotoxin production and virulence of Fusarium oxysporum on plant and mammalian hosts. Molecular Microbiology. 2013;87: 49–65. doi:10.1111/mmi.12082

23. López-Berges MS, Di Pietro A, Daboussi MJ, Wahab HA, Vasnier C, Roncero MIG, et al. Identification of virulence genes in Fusarium oxysporum f. sp. lycopersici by large-scale transposon tagging. Molecular Plant Pathology. 2009;10: 95–107. doi:10.1111/j.1364-3703.2008.00512.x

24. López-Díaz C, Rahjoo V, Sulyok M, Ghionna V, Martín-Vicente A, Capilla J, et al. Fusaric acid contributes to virulence of Fusarium oxysporum on plant and mammalian hosts. Molecular Plant Pathology. 2018;19: 440–453. doi:10.1111/mpp.12536

25. Bravo Ruiz G, Di Pietro A, Roncero MIG. Combined action of the major secreted exo- and endopolygalacturonases is required for full virulence of Fusarium oxysporum. Molecular plant pathology. 2016;17: 339–53. doi:10.1111/mpp.12283

26. Jashni MK, Dols IHM, Iida Y, Boeren S, Beenen HG, Mehrabi R, et al. Synergistic Action of a Metalloprotease and a Serine Protease from Fusarium oxysporum f. sp. lycopersici Cleaves Chitin-Binding Tomato Chitinases, Reduces Their Antifungal Activity, and Enhances Fungal Virulence. Molecular plant-microbe interactions : MPMI. 2015;28: 996–1008. doi:10.1094/MPMI-04-15-0074-R

27. Houterman PM, Speijer D, Dekker HL, DE Koster CG, Cornelissen BJC, Rep M. The mixed xylem sap proteome of Fusarium oxysporum-infected tomato plants. Molecular plant pathology. 2007;8: 215–21. doi:10.1111/j.1364-3703.2007.00384.x

28. Navarro-Velasco GY, Prados-Rosales RC, Ortíz-Urquiza A, Quesada-Moraga E, Di Pietro A. Galleria mellonella as model host for the trans-kingdom pathogen Fusarium oxysporum. Fungal Genetics and Biology. 2011;48: 1124–1129. doi:10.1016/j.fgb.2011.08.004

29. Chellapan BV, Van Dam P, Rep M, Cornelissen BJC, Fokkens L. Non-canonical Helitrons in Fusarium oxysporum. Mobile DNA. 2016;7: 1–16. doi:10.1186/s13100-016-0083-7

30. van Dam P, Rep M. The Distribution of Miniature Impala Elements and SIX Genes in the Fusarium Genus is Suggestive of Horizontal Gene Transfer. Journal of Molecular Evolution. 2017;85: 14–25. doi:10.1007/s00239-017-9801-0

31. Schmidt SM, Houterman PM, Schreiver I, Ma L, Amyotte S, Chellappan B, et al. MITEs in the promoters of effector genes allow prediction of novel virulence genes in Fusarium oxysporum. BMC genomics. 2013;14. doi:10.1186/1471-2164-14-119

32. Zhang Y, Yang H, Turra D, Zhou S, Ayhan DH, DeIulio GA, et al. The genome of opportunistic fungal pathogen Fusarium oxysporum carries a unique set of lineage-specific chromosomes. Commun Biol. 2020;3: 1–12. doi:10.1038/s42003-020-0770-2

33. Di Pietro A, Roncero MI. Purification and characterization of an exo-polygalacturonase from the tomato vascular wilt pathogen Fusarium oxysporum f.sp. lycopersici. FEMS Microbiol Lett. 1996;145: 295–299. doi:10.1111/j.1574-6968.1996.tb08592.x

34. O'Donnell K, Sutton DA, Rinaldi MG, Magnon KC, Cox PA, Revankar SG, et al. Genetic Diversity of Human Pathogenic Members of the Fusarium oxysporum Complex Inferred from Multilocus DNA Sequence Data and Amplified Fragment Length Polymorphism Analyses: Evidence for the Recent Dispersion of a Geographically Widespread Clonal Lineage and Nosocomial Origin. J Clin Microbiol. 2004;42: 5109–5120. doi:10.1128/JCM.42.11.5109-5120.2004

35. Ayhan DH, López-Díaz C, Di Pietro A, Ma L-J. Improved Assembly of Reference Genome Fusarium oxysporum f. sp. lycopersici Strain Fol4287. Rokas A, editor. Microbiology Resource Announcements. 2018;7: 4–5. doi:10.1128/MRA.00910-18

36. Müntzing A. Accessory chromosomes. Annu Rev Genet. 1974;8: 243–266. doi:10.1146/annurev.ge.08.120174.001331

37. Yang H, Yu H, Ma L-J. Accessory Chromosomes in Fusarium oxysporum. Phytopathology. 2020;110: 1488–1496. doi:10.1094/PHYTO-03-20-0069-IA

38. Miao VP, Covert SF, VanEtten HD. A fungal gene for antibiotic resistance on a dispensable ("B") chromosome. Science. 1991;254: 1773–1776. doi:10.1126/science.1763326

39. Covert SF. Supernumerary chromosomes in filamentous fungi. Curr Genet. 1998;33: 311–319. doi:10.1007/s002940050342

40. Kistler HC, Rep M, Ma L-J. Structural dynamics of Fusarium genomes. In: Rown DW, Proctor RH, editors. Fusarium: genomics, molecular and cellular biology. Norwich, United Kingdom: Horizon Scientific Press; 2013. pp. 31–41.

41. van Dam P, Fokkens L, Schmidt SM, Linmans JHJ, Kistler HC, Ma L-J, et al. Effector profiles distinguish formae speciales of Fusarium oxysporum. Environmental Microbiology. 2016;18: 4087–4102. doi:10.1111/1462-2920.13445

42. Houterman PM, Ma L, van Ooijen G, de Vroomen MJ, Cornelissen BJC, Takken FLW, et al. The effector protein Avr2 of the xylem-colonizing fungus Fusarium oxysporum activates the tomato resistance protein I-2 intracellularly. Plant J. 2009;58: 970–978. doi:10.1111/j.1365-313X.2009.03838.x

43. Takken F, Rep M. The arms race between tomato and Fusarium oxysporum. Molecular Plant Pathology. 2010;11: 309–314. doi:10.1111/j.1364-3703.2009.00605.x

44. Inami K, Yoshioka-Akiyama C, Morita Y, Yamasaki M, Teraoka T, Arie T. A Genetic Mechanism for Emergence of Races in Fusarium oxysporum f. sp. lycopersici: Inactivation of Avirulence Gene AVR1 by Transposon Insertion. Zhang Z, editor. PLoS ONE. 2012;7: e44101. doi:10.1371/journal.pone.0044101

45. Ma L-J. Horizontal chromosome transfer and rational strategies to manage Fusarium vascular wilt diseases. Molecular Plant Pathology. 2014;15: 763–766. doi:10.1111/mpp.12171

46. McGovern RJ. Management of tomato diseases caused by Fusarium oxysporum. Crop Protection. 2015;73: 78–92. doi:10.1016/j.cropro.2015.02.021

47. Chang DC, Grant GB, O'Donnell K, Wannemuehler KA, Noble-Wang J, Rao CY, et al. Multistate outbreak of Fusarium keratitis associated with use of a contact lens solution. JAMA. 2006;296: 953–963. doi:10.1001/jama.296.8.953

48. Imamura Y, Chandra J, Mukherjee PK, Lattif AA, Szczotka-Flynn LB, Pearlman E, et al. Fusarium and Candida albicans Biofilms on Soft Contact Lenses: Model Development, Influence of Lens Type, and Susceptibility to Lens Care Solutions. Antimicrobial Agents and Chemotherapy. 2008;52: 171–182. doi:10.1128/AAC.00387-07

49. O'Donnell K, Sarver BAJ, Brandt M, Chang DC, Noble-Wang J, Park BJ, et al. Phylogenetic diversity and microsphere array-based genotyping of human pathogenic Fusaria, including isolates from the multistate contact lens-associated U.S. keratitis outbreaks of 2005 and 2006. J Clin Microbiol. 2007;45: 2235–2248. doi:10.1128/JCM.00533-07

50. Kredics L, Narendran V, Shobana CS, Vágvölgyi C, Manikandan P, Indo-Hungarian Fungal Keratitis Working Group. Filamentous fungal infections of the cornea: a global overview of epidemiology and drug sensitivity. Mycoses. 2015;58: 243–260. doi:10.1111/myc.12306

51. Hassan AS, Al-Hatmi AMS, Shobana CS, van Diepeningen AD, Kredics L, Vágvölgyi C, et al. Antifungal Susceptibility and Phylogeny of Opportunistic Members of the Genus Fusarium Causing Human Keratomycosis in South India. Med Mycol. 2016;54: 287–294. doi:10.1093/mmy/myv105

52. Lalitha P, Prajna NV, Manoharan G, Srinivasan M, Mascarenhas J, Das M, et al. Trends in bacterial and fungal keratitis in South India, 2002-2012. Br J Ophthalmol. 2015;99: 192–194. doi:10.1136/bjophthalmol-2014-305000

53. He D, Hao J, Zhang B, Yang Y, Song W, Zhang Y, et al. Pathogenic spectrum of fungal keratitis and specific identification of Fusarium solani. Invest Ophthalmol Vis Sci. 2011;52: 2804–2808. doi:10.1167/iovs.10-5977

54. Wang L, Sun S, Jing Y, Han L, Zhang H, Yue J. Spectrum of fungal keratitis in central China. Clin Exp Ophthalmol. 2009;37: 763–771. doi:10.1111/j.1442-9071.2009.02155.x

55. O'Sullivan J, Gilbert C, Foster A. The causes of childhood blindness in South Africa. S Afr Med J. 1997;87: 1691–1695.

56. Ibrahim MM, Vanini R, Ibrahim FM, Fioriti LS, Furlan EMR, Provinzano LMA, et al. Epidemiologic aspects and clinical outcome of fungal keratitis in southeastern Brazil. Eur J Ophthalmol. 2009;19: 355–361. doi:10.1177/112067210901900305

57. Dudley MA, Chick EW. CORNEAL LESIONS PRODUCED IN RABBITS BY AN EXTRACT OF FUSARIUM MONILIFORME. Arch Ophthalmol. 1964;72: 346–350. doi:10.1001/archopht.1964.00970020346012

58. Hemo I, Pe'er J, Polacheck I. Fusarium oxysporum keratitis. Ophthalmologica. 1989;198: 3–7. doi:10.1159/000309951

59. Epstein AB. In the aftermath of the Fusarium keratitis outbreak: What have we learned? Clinical ophthalmology (Auckland, NZ). 2007;1: 355–66.

60. Alfonso EC, Cantu-Dibildox J, Munir WM, Miller D, O'Brien TP, Karp CL, et al. Insurgence of Fusarium Keratitis Associated With Contact Lens Wear. Arch Ophthalmol. 2006;124: 941. doi:10.1001/archopht.124.7.ecs60039

61. Gaujoux T, Chatel MA, Chaumeil C, Laroche L, Borderie VM. Outbreak of contact lens-related Fusarium keratitis in France. Cornea. 2008;27: 1018–1021. doi:10.1097/ICO.0b013e318173144d

62. Khor W-B, Aung T, Saw S-M, Wong T-Y, Tambyah PA, Tan A-L, et al. An outbreak of Fusarium keratitis associated with contact lens wear in Singapore. JAMA. 2006;295: 2867–2873. doi:10.1001/jama.295.24.2867

63. Ma SE, So K, Chung P, Tsang HT, Chuang S. A multi-country outbreak of fungal keratitis associated with a brand of contact lens solution: the Hong Kong experience. Int J Infect Dis. 2009;13: 443–448. doi:10.1016/j.ijid.2007.12.018

64. Nucci M, Marr KA, Vehreschild MJGT, de Souza CA, Velasco E, Cappellano P, et al. Improvement in the outcome of invasive fusariosis in the last decade. Clin Microbiol Infect. 2014;20: 580–585. doi:10.1111/1469-0691.12409

65. Prajna NV, Krishnan T, Rajaraman R, Patel S, Srinivasan M, Das M, et al. Effect of Oral Voriconazole on Fungal Keratitis in the Mycotic Ulcer Treatment Trial II (MUTT II): A Randomized Clinical Trial. JAMA Ophthalmol. 2016;134: 1365–1372. doi:10.1001/jamaophthalmol.2016.4096

66. Leal SM, Vareechon C, Cowden S, Cobb BA, Latgé J-P, Momany M, et al. Fungal antioxidant pathways promote survival against neutrophils during infection. J Clin Invest. 2012;122: 2482–2498. doi:10.1172/JCI63239

67. MATLAB. version 9.9.0.1570001 (R2020b). Natick, Massachusetts: The MathWorks Inc.; 2020.

68. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 1980;8: 4321–4325.

69. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. FastQC. Babraham, UK; 2012.

70. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. HybridSPAdes: An algorithm for hybrid assembly of short and long reads. Bioinformatics. 2016;32: 1009–1015. doi:10.1093/bioinformatics/btv688

71. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature methods. 2013;10: 563–9. doi:10.1038/nmeth.2474

72. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

73. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint. 2012;arXiv:1207. Available: http://arxiv.org/abs/1207.3907

74. Cameron DL, Schröder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Research. 2017;27: 2050–2060. doi:10.1101/gr.222109.117

75. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nature Methods. 2018;15: 461–468. doi:10.1038/s41592-018-0001-7

76. Picard Toolkit. 2019. Available: http://broadinstitute.github.io/picard

77. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352

78. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

79. Smit A, Hubley R. RepeatModeler Open-1.0. <http://www.repeatmasker.org>. 2015.

80. Edgar R. Usearch. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States); 2010 Jun. Available: https://www.osti.gov/sciencecinema/biblio/1137186

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

82. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome biology. 2004;5: R12. doi:10.1186/gb-2004-5-2-r12

83. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 2004;14: 1394–1403. doi:10.1101/gr.2289704

84. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26: 589–595. doi:10.1093/bioinformatics/btp698

85. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27: 722–736. doi:10.1101/gr.215087.116

86. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010;20: 1297–303. doi:10.1101/gr.107524.110

87. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research. 2014;42: D699–D704. doi:10.1093/nar/gkt1183

88. Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, et al. Fueling the future with fungal genomics. null. 2011;2: 192–209. doi:10.1080/21501203.2011.584577

89. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 2019;20: 238. doi:10.1186/s13059-019-1832-y

90. Kessel L, Johnson L, Arvidsson H, Larsen M. The Relationship between Body and Ambient Temperature and Corneal Temperature. Investigative Opthalmology & Visual Science. 2010;51: 6593. doi:10.1167/iovs.10-5659

91. Mirnezami SA, Rajaei Jafarabadi M, Abrishami M. Temperature Distribution Simulation of the Human Eye Exposed to Laser Radiation. J Lasers Med Sci. 2013;4: 175–181.

92. Yang H. ACCESSORY GENES CONTRIBUTE TO REWIRING THE TRANSCRIPTIONAL NETWORK IN FUSARIUM OXYSPORUM. University of Massachusetts Amherst. 2020. Available: https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=2994&context=dissertations_2

93. Lamb C, Dixon RA. THE OXIDATIVE BURST IN PLANT DISEASE RESISTANCE. Annu Rev Plant Physiol Plant Mol Biol. 1997;48: 251–275. doi:10.1146/annurev.arplant.48.1.251

94. Segal LM, Wilson RA. Reactive oxygen species metabolism and plant-fungal interactions. Fungal Genet Biol. 2018;110: 1–9. doi:10.1016/j.fgb.2017.12.003

95. Pusztahelyi T. Chitin and chitin-related compounds in plant–fungal interactions. Mycology. 2018;9: 189–201. doi:10.1080/21501203.2018.1473299

96. Roby D, Gadelle A, Toppan A. Chitin oligosaccharides as elicitors of chitinase activity in melon plants. Biochem Biophys Res Commun. 1987;143: 885–892. doi:10.1016/0006-291x(87)90332-9

97. Ram AFJ, Klis FM. Identification of fungal cell wall mutants using susceptibility assays based on Calcofluor white and Congo red. Nature Protocols. 2006;1: 2253–2256. doi:10.1038/nprot.2006.397

98. Alghamdi M, Alasmari D, Assiri A, Mattar E, Aljaddawi AA, Alattas SG, et al. An Overview of the Intrinsic Role of Citrullination in Autoimmune Disorders. J Immunol Res. 2019;2019. doi:10.1155/2019/7592851

99. Marondedze C, Elia G, Thomas L, Wong A, Gehring C. Citrullination of Proteins as a Specific Response Mechanism in Plants. Front Plant Sci. 2021;12. doi:10.3389/fpls.2021.638392

100. Palazzo L, Mikolčević P, Mikoč A, Ahel I. ADP-ribosylation signalling and human disease. Open Biol. 2019;9: 190041. doi:10.1098/rsob.190041

101. Feng B, Liu C, Shan L, He P. Protein ADP-Ribosylation Takes Control in Plant–Bacterium Interactions. PLoS Pathog. 2016;12. doi:10.1371/journal.ppat.1005941

102. Caza M, Kronstad JW. Shared and distinct mechanisms of iron acquisition by bacterial and fungal pathogens of humans. Front Cell Infect Microbiol. 2013;3: 80. doi:10.3389/fcimb.2013.00080

103. Visser DJ, Arjan GM, Krug J. Empirical fitness landscapes and the predictability of evolution. Nature Reviews Genetics. 2014;15: 480–490. doi:10.1038/nrg3744

104. Garland T, Rose MR. Experimental Evolution. Berkeley: University Of California Press; 2009. Available: https://www.ucpress.edu/book/9780520261808/experimental-evolution

105. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature. 2015;519: 181–186. doi:10.1038/nature14279

106. Nguyen Ba AN, Cvijović I, Rojas Echenique JI, Lawrence KR, Rego-Costa A, Liu X, et al. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. Nature. 2019;575: 494–499. doi:10.1038/s41586-019-1749-3

107. Peris D, Moriarty RV, Alexander WG, Baker E, Sylvester K, Sardi M, et al. Hybridization and adaptive evolution of diverse Saccharomyces species for cellulosic biofuel production. Biotechnology for Biofuels. 2017;10: 78. doi:10.1186/s13068-017-0763-7

108. McDonald MJ. Microbial Experimental Evolution – a proving ground for evolutionary theory and a tool for discovery. EMBO Rep. 2019;20. doi:10.15252/embr.201846992

109. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nature reviews Genetics. 2013;14: 827–39. doi:10.1038/nrg3564

110. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 generations. American Naturalist. 1991;138: 1315–1341.

111. Lenski RE. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. The ISME Journal. 2017;11: 2181–2194. doi:10.1038/ismej.2017.69

112. Lenski R. We Interrupt this Nasty Virus with Some Good News about Bacteria. In: Telliamed Revisited [Internet]. 24 Feb 2020 [cited 10 Jun 2021]. Available: https://telliamedrevisited.wordpress.com/2020/02/24/we-interrupt-this-nasty-coronavirus-with-some-good-news-about-bacteria/

113. Novick A, Szilard L. Description of the chemostat. Science. 1950;112: 715–716. doi:10.1126/science.112.2920.715

114. Nagy K, Ábrahám Á, Keymer JE, Galajda P. Application of Microfluidics in Experimental Ecology: The Importance of Being Spatial. Front Microbiol. 2018;9: 496. doi:10.3389/fmicb.2018.00496

115. Lescat M, Launay A, Ghalayini M, Magnan M, Glodt J, Pintard C, et al. Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an Escherichia coli natural isolate in the streptomycin-treated mouse gut. Microbial Ecology. 2016;26: 1802–1817. doi:https://doi.org/10.1111/mec.13851

116. Desai MM, Fisher DS. Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics. 2007;176: 1759–1798. doi:10.1534/genetics.106.067678

117. Halligan DL, Keightley PD. Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. Annual Review of Ecology, Evolution, and Systematics. 2009;40: 151–172. doi:10.1146/annurev.ecolsys.39.110707.173437

118. Fisher DS. Asexual evolution waves: Fluctuations and universality. Journal of Statistical Mechanics: Theory and Experiment. 2013;2013. doi:10.1088/1742-5468/2013/01/P01011

119. Frenkel EM, McDonald MJ, Dyken JDV, Kosheleva K, Lang GI, Desai MM. Crowded growth leads to the spontaneous evolution of semistable coexistence in laboratory yeast populations. PNAS. 2015;112: 11306–11311. doi:10.1073/pnas.1506184112

120. Fogle CA, Nagle JL, Desai MM. Clonal interference, multiple mutations and adaptation in large asexual populations. Genetics. 2008;180: 2163–2173. doi:10.1534/genetics.108.090019

121. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 2013;500: 571–574. doi:10.1038/nature12344

122. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, et al. The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast. Snyder M, editor. PLoS Genetics. 2008;4: e1000303. doi:10.1371/journal.pgen.1000303

123. Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nature genetics. 2012;44: 101–5. doi:10.1038/ng.1034

124. Tenaillon O, Toupance B, Le Nagard H, Taddei F, Godelle B. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. Genetics. 1999;152: 485–93.

125. Cooper VS, Schneider D, Blot M, Lenski RE. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of Escherichia coli B. J Bacteriol. 2001;183: 2834–2841. doi:10.1128/JB.183.9.2834-2841.2001

126. Kvitek DJ, Sherlock G. Whole Genome, Whole Population Sequencing Reveals That Loss of Signaling Networks Is the Major Adaptive Strategy in a Constant Environment. Zhang J, editor. PLoS Genetics. 2013;9: e1003972. doi:10.1371/journal.pgen.1003972

127. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. The dynamics of molecular evolution over 60,000 generations. Nature. 2017;551: 45–50. doi:10.1038/nature24287

128. Kryazhimskiy S, Tkacik G, Plotkin JB. The dynamics of adaptation on correlated fitness landscapes. Proceedings of the National Academy of Sciences. 2009;106: 18638–18643. doi:10.1073/pnas.0905497106

129. Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, et al. Tempo and mode of genome evolution in a 50,000 - generation experiment. Nature. 2016;536: 1–21. doi:10.1101/036806

130. Wiser MJ, Ribeck N, Lenski RE. Long-Term Dynamics of Adaptation in Asexual Populations. Science. 2013;342: 1364–1367.

131. Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science. 2012;335: 428–432. doi:10.1126/science.1214449

132. McDonald MJ, Rice DP, Desai MM. Sex Speeds Adaptation by Altering the Dynamics of Molecular Evolution. Nature. 2016;531: 233–236. doi:10.1038/nature17143

133. Koenig RL, Ploetz RC, Kistler HC. Fusarium oxysporum f. sp. cubense Consists of a Small Number of Divergent and Globally Distributed Clonal Lineages. Phytopathology. 1997;87: 915–923. doi:10.1094/PHYTO.1997.87.9.915

134. Brankovics B, van Dam P, Rep M, de Hoog GS, van der Lee TAJ, Waalwijk C, et al. Mitochondrial genomes reveal recombination in the presumed asexual Fusarium oxysporum species complex. BMC Genomics. 2017;18: 1–14. doi:10.1186/s12864-017-4116-5

135. Laurent B, Palaiokostas C, Spataro C, Moinard M, Zehraoui E, Houston RD, et al. High-resolution mapping of the recombination landscape of the phytopathogen Fusarium graminearum suggests two-speed genome evolution. Molecular Plant Pathology. 2018;19: 341–354. doi:10.1111/mpp.12524

136. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current protocols in bioinformatics. 2013;43: 11.10.1-33. doi:10.1002/0471250953.bi1110s43

137. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011;29: 24–26. doi:10.1038/nbt.1754

138. Sohrab V, López-Díaz C, Di Pietro A, Ma L-J, Ayhan DH. TEfinder: A Bioinformatics Pipeline for Detecting New Transposable Element Insertion Events in Next-Generation Sequencing Data. Genes. 2021;12: 224. doi:10.3390/genes12020224

139. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An Information Aesthetic for Comparative Genomics. Genome research. 2009;19: 1639–45. doi:10.1101/gr.092759.109

140. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37: 907–915. doi:10.1038/s41587-019-0201-4

141. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics. 2015;31: 3593–3599. doi:10.1093/bioinformatics/btv422

142. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421

143. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. Biol Direct. 2012;7: 12. doi:10.1186/1745-6150-7-12

144. Illumina Sequencing Technology. Illumina, Inc; 2010. Available: https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

145. Fokkens L, Shahi S, Connolly LR, Stam R, Schmidt SM, Smith KM, et al. The multi-speed genome of Fusarium oxysporum reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events . bioRxiv. 2018. doi:http://dx.doi.org/10.1101/465070

146. López Díaz C. Molecular and genetic mechanisms underlying genome plasticity in Fusarium oxysporum. University of Córdoba. 2019. Available: https://helvia.uco.es/xmlui/bitstream/handle/10396/18865/2019000001959.pdf?sequence=1&isAllowed=y

147. López-Berges MS, Schäfer K, Hera C, Di Pietro A. Combinatorial function of velvet and AreA in transcriptional regulation of nitrate utilization and secondary metabolism. Fungal Genet Biol. 2014;62: 78–84. doi:10.1016/j.fgb.2013.11.002

148. Lauer S, Avecilla G, Spealman P, Sethia G, Brandt N, Levy SF, et al. Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. de Visser JAGM, editor. PLOS Biology. 2018;16: e3000069. doi:10.1371/journal.pbio.3000069

149. Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. Bacterial genome size reduction by experimental evolution. Proc Natl Acad Sci U S A. 2005;102: 12112–12116. doi:10.1073/pnas.0503654102

150. Ruiz-Roldán MC, Köhli M, Roncero MIG, Philippsen P, Di Pietro A, Espeso EA. Nuclear dynamics during germination, conidiation, and hyphal fusion of Fusarium oxysporum. Eukaryotic Cell. 2010;9: 1216–1224. doi:10.1128/EC.00040-10

151. Urbaniak C, van Dam P, Zaborin A, Zaborina O, Gilbert JA, Torok T, et al. Genomic Characterization and Virulence Potential of Two Fusarium oxysporum Isolates Cultured from the International Space Station. mSystems. 4: e00345-18. doi:10.1128/mSystems.00345-18

152. O'Donnell K, Gueidan C, Sink S, Johnston PR, Crous PW, Glenn A, et al. A two-locus DNA sequence database for typing plant and human pathogens within the Fusarium oxysporum species complex. Fungal Genet Biol. 2009;46: 936–948. doi:10.1016/j.fgb.2009.08.006

153. Burns KH. Transposable elements in cancer. Nat Rev Cancer. 2017;17: 415–424. doi:10.1038/nrc.2017.35

154. Wang J, Huang J, Shi G. Retrotransposons in pluripotent stem cells. Cell Regen. 2020;9: 4. doi:10.1186/s13619-020-00046-4

155. Seidl MF, Thomma BPHJ. Transposable Elements Direct The Coevolution between Plants and Microbes. Trends Genet. 2017;33: 842–851. doi:10.1016/j.tig.2017.07.003

156. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits Edward. Nat Rev Genet. 2016; 1–16. doi:10.1038/nrg.2016.139

157. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19: 199. doi:10.1186/s13059-018-1577-z

158. Huang CRL, Burns KH, Boeke JD. Active transposition in genomes. Annu Rev Genet. 2012;46: 651–675. doi:10.1146/annurev-genet-110711-155616

159. Choi JY, Lee YCG. Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. PLOS Genetics. 2020;16: e1008872. doi:10.1371/journal.pgen.1008872

160. Gaviria JM, van Burik JA, Dale DC, Root RK, Liles WC. Comparison of interferon-gamma, granulocyte colony-stimulating factor, and granulocyte-macrophage colony-stimulating factor for priming leukocyte-mediated hyphal damage of opportunistic fungal pathogens. J Infect Dis. 1999;179: 1038–1041. doi:10.1086/314679

161. Winn RM, Gil-Lamaignere C, Roilides E, Simitsopoulou M, Lyman CA, Maloukou A, et al. Effects of interleukin-15 on antifungal responses of human polymorphonuclear leukocytes against Fusarium spp. and Scedosporium spp. Cytokine. 2005;31: 1–8. doi:10.1016/j.cyto.2004.07.016

162. Wilkinson S, Corlett JE, Oger L, Davies WJ. Effects of Xylem pH on Transpiration from Wild-Type andflacca Tomato Leaves1: A Vital Role for Abscisic Acid in Preventing Excessive Water Loss Even from Well-Watered Plants. Plant Physiology. 1998;117: 703–709. doi:10.1104/pp.117.2.703

163. Bonanno JA, Polse KA. Effect of Rigid Contact Lens Oxygen Transmissibility on Stromal PH in the Living Human Eye. Ophthalmology. 1987;94: 1305–1309. doi:10.1016/S0161-6420(87)80016-7

164. Jonas. plot spread points (beeswarm plot). MATLAB Cental File Exchange; 2021. Available: https://www.mathworks.com/matlabcentral/fileexchange/37105-plot-spread-points-beeswarm-plot

165. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35: 1547–1549. doi:10.1093/molbev/msy096

166. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2: 28–36.

167. Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proc Natl Acad Sci U S A. 2006;103: 9107–9112. doi:10.1073/pnas.0602917103

168. Connolly LR, Smith KM, Freitag M. The Fusarium graminearum Histone H3 K27 Methyltransferase KMT6 Regulates Development and Expression of Secondary Metabolite Gene Clusters. PLOS Genetics. 2013;9: e1003916. doi:10.1371/journal.pgen.1003916

169. Möller M, Schotanus K, Soyer JL, Haueisen J, Happ K, Stralucke M, et al. Destabilization of chromosome structure by histone H3 lysine 27 methylation. PLOS Genetics. 2019;15: e1008093. doi:10.1371/journal.pgen.1008093

170. Mes JJ, Haring MA, Cornelissen BJ. Foxy: an active family of short interspersed nuclear elements from Fusarium oxysporum. Mol Gen Genet. 2000;263: 271–280. doi:10.1007/pl00008681

171. Branco MR, Chuong EB. Crossroads between transposons and gene regulation. Philosophical Transactions of the Royal Society B: Biological Sciences. 2020;375: 20190330. doi:10.1098/rstb.2019.0330

172. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11: 1650–1667. doi:10.1038/nprot.2016.095

173. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15: 550. doi:10.1186/s13059-014-0550-8