Doctoral Dissertation (Censored)

博士論文（要約）

# Comparative Genomic Analysis of Plant Pathogenic *Colletotrichum* Fungi

（植物病原糸状菌 *Colletotrichum* 属菌の比較ゲノム解析）

A Dissertation Submitted for the Degree of Doctor of Philosophy

January 2019

平成31年1月博士（理学）申請

Department of Biological Sciences, Graduate School of Science,

The University of Tokyo

東京大学大学院理学系研究科 生物科学専攻

Ayako Tsushima

津島 綾子

## Table of contents

## Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Ken Shirasu, for his insights into my work, guidance, and patience. He provided the best environment for conducting my research and continuously encouraged me to pursue my interests. Moreover, I would like to deeply thank Dr. Pamela Gan for guiding me throughout my Ph.D., especially in bioinformatic analyses and scientific writing. Without her immense knowledge, motivation, and patience, I could not have completed this thesis. I would like to thank Dr. Naoyoshi Kumakura for having numerous discussions with me at any time of the day or night.

I would like to thank my collaborators: Dr. Mari Narusaka, Dr. Yoshihiro Narusaka, and Dr. Yoshitaka Takano, for providing research materials and expert suggestions since the beginning of my study on fungal plant pathogens; Dr. Sophien Kamoun for providing me the great opportunity to study in The Sainsbury Laboratory and to deepen my interest in evolution; Dr. Joe Win and Dr. Hiroaki Adachi for teaching me how to perform immunoprecipitation; and Dr. Frank Menke and Dr. Jan Sklenar for their technical support for the LC-MS/MS analysis. In addition, I would like to thank Dr. Richard O'Connell, Dr. Yasuyuki Kubo, and Dr. Miin-Huey Lee, who kindly provided fungal strains for my research. My thesis would not have been possible without the assistance of these people.

I must thank every single member of the Plant Immunity Research Group and my family and friends for supporting my journey. I am also grateful to the two JSPS fellowships, the Research Fellowship for Doctoral Course Students (DC2), and the Overseas Challenge Program for Young Researchers, which allowed me to expand the scope of my research.

**January 2019**

## Abbreviations

| | |
|---|---|
| AP orthogroup | all protein orthogroup |
| CCE1 | *Colletotrichum* core effector |
| cfu | colony forming unit |
| DAPI | 4′,6-diamidino-2-phenylindole |
| EC orthogroup | effector candidate orthogroup |
| ER | endoplasmic reticulum |
| FIR | flanking intergenic region |
| HR | hypersensitive response |
| LC-MS/MS | liquid chromatography/tandem mass spectrometry |
| NGS | next-generation sequencing |
| *Pto* | *Pseudomonas syringae* pv. *tomato* |
| ROI | region of interest |
| TE | transposable element |
| VIGS | virus-induced gene silencing |

## Abstract

Plant pathogens secrete proteins called effectors to manipulate host cells and promote infection. Recent studies have shown examples of effectors conserved across taxa, which are required for full pathogenicity. However, in some cases, effectors are perceived by host immune receptors, resulting in strong immune responses against the invading pathogens. To avoid this recognition, effectors are generally are under diversifying selection and differ even between strains within the same species. *Colletotrichum* fungi collectively cause anthracnose disease in a broad range of plants, although individual species have specialized in infecting limited host plants. The aim of my Ph.D. research was to understand how *Colletotrichum* fungi have adapted to various niches, by characterizing their effector gene sets.

Among *Colletotrichum* species, *C. higginsianum* has been widely used for scientific studies, as it infects the model plant *Arabidopsis thaliana*. However, the first published genome assembly of this pathogen is fragmented and possibly contains missing or incorrect gene annotations. In order to overcome this problem, I generated a more contiguous assembly of the genome of *C. higginsianum* MAFF305635-RFP by sequencing it using PacBio RS II. This genome assembly comprises of 28 contigs and is estimated to include 99.0% of all coding genes. I analyzed the conservation patterns of effector candidates amongst 24 ascomycetes, including *C. higginsianum* MAFF305635-RFP. This analysis revealed that seven effector candidate orthogroups are specifically conserved in all seven *Colletotrichum* species tested, but not in other ascomycetes. This analysis also identified species-specific effector candidates of *Colletotrichum* that may contribute to host specificity.

As few sexual morphs have been described in the genus *Colletotrichum*, most members including *C. higginsianum* appear to proliferate clonally. To determine whether *Colletotrichum*

species exhibit intra-species genomic variations in the absence of a sexual cycle, I compared the two closely-related *C. higginsianum* strains, MAFF305635-RFP and IMI 349063, which were sequenced by Zampounis et al. (2016). First, I performed whole-genome alignments between the two *C. higginsianum* strains. This analysis revealed the presence of 10 large-scale rearrangements between the two strains, including six inter-chromosomal translocations and four intra-chromosomal inversions. Whole-genome alignments also indicated that the two strains have strain-specific regions (< 99% identity, < 15 kb) that are variable in the other strain. In order to identify strain-specific variations in effectors of *C. higginsianum*, effector candidates from the two strains were compared. This analysis revealed that 8 out of 582 candidates in MAFF305635-RFP and 18 out of 576 candidates in IMI 349063 were highly variable between the two strains, with ≤ 90% query coverage. Such effector candidates showed variable conservation patterns in Ascomycota, possibly reflecting differences in their evolutionary history (e.g., *de novo* evolution, loss after speciation, and horizontal gene transfer). Transposable elements (TEs) are known to often be involved in the generation of genomic variations. To examine whether TEs contribute to the generation of genomic variations in *C. higginsianum*, the association between TEs and strain-specific regions was investigated. In the genome of MAFF305635-RFP, 29.5% of strain-specific regions were found to overlap with TEs and this is significantly higher than the case if TEs were randomly distributed in the genome (Monte Carlo method, $P < 0.001$). Further, my results indicate that the genome of *C. higginsianum* is compartmentalized into regions harboring conserved genes, which are gene-dense and TE-sparse, as well as regions with more effector candidate genes, which are gene-sparse and TE-dense.

To further characterize the effector candidates identified, I conducted a functional analysis of CCE1 (*Colletotrichum* Core Effector 1). Genus-wide comparative genomic analyses revealed that this effector candidate gene is highly conserved in the genus *Colletotrichum*. In addition, transient expression assays indicate that *CCE1* homologs from three *Colletotrichum* species infecting

different hosts induce cell death in *Nicotiana benthamiana* leaves. Furthermore, by performing *in planta* co-immunoprecipitation, I identified candidate interactors of CCE1 including cytoskeleton-related proteins, such as actin and tubulin, the Golgi-targeted protein α1-COP, and the ER-targeted protein BIP2. These data suggest that CCE1 proteins may function in promoting host cell death during infection by targeting a component found in various host plants.

In conclusion, in this thesis, I show the diversity of effector candidates and a potential mechanism for generating genomic variations in *Colletotrichum*. Given that effectors play important roles in plant-microbe interactions, variations in effector complements may contribute to the fitness of this group of fungi.

## Chapter I: General Introduction

Fungal plant diseases pose significant threats to food security and ecosystems (Fisher et al. 2012; Islam et al. 2016; McMullan et al. 2018). Among fungal plant pathogens, members of the genus *Colletotrichum*, belonging to the phylum Ascomycota, are known as causal agents of anthracnose diseases. They have a wide global distribution and cause considerable economic losses of crops (Crouch et al. 2014). For example, in Japan, the financial damage to strawberries by *Colletotrichum* species is estimated at 3.5 billion yen per year (Sato & Moriwaki 2009), and in the U.S., *C. graminicola* infections of maize are estimated to cause annual losses of more than 1 billion USD (Frey et al. 2011).

Members of the genus *Colletotrichum* are diverse and can be grouped into several major species complexes consisting of closely related species that share infection lifestyles (Figure 1-1). Members of *Colletotrichum* infect a wide variety of host plants (Cannon et al. 2012). This diversity of hosts led some researchers to speculate that virtually every crop could potentially be infected by one or more *Colletotrichum* species (Dean et al. 2012). Although members in this genus have a wide host range collectively, individual species infect limited host plants. For example, *C. higginsianum* from the *C. destructivum* species complex infects Brassicaceae plants including *Arabidopsis thaliana* (O'Connell et al. 2004), while *C. graminicola* in the *C. graminicola* species complex is restricted to infecting maize (*Zea mays*) (Crouch & Beirn 2009).

Most *Colletotrichum* species employ a hemibiotrophic lifestyle, characterized by the sequential development of a series of special cell types (Münch et al. 2008). After conidia germinate on the host, they differentiate into melanized appressoria, which function in penetration of leaf epidermal cells (Figure 1-2A, B). Following penetration, the fungi develop bulbous primary hyphae enveloped with intact host plasma membranes inside living cells (Figure 1-2C). This stage is called

the biotrophic phase because the fungi do not kill host cells at this stage. Later on during the infection, the fungi develop thin secondary hyphae and begin to destroy the host tissue (Figure 1-2D, E). This stage is referred to as the necrotrophic stage. In contrast to obligate biotrophic pathogens including the *Blumeria graminis*, which causes powdery mildew, and *Puccinia striiformis*, which causes stripe rust, *Colletotrichum* species can be cultured axenically without living hosts and are amenable to genetic manipulation, making them ideal model organisms for molecular studies (Figure 1-2F) (O'Connell et al. 2004; Spanu & Panstruga 2017; Kumakura et al. 2018).

In order to establish infections, plant pathogens have evolved secreted proteins to manipulate plant calls, called effectors (Chisholm et al. 2006). Recent studies indicate that plant pathogens orchestrate the expression of a range of effectors to attenuate plant immunity and promote colonization (Deslandes & Rivas 2012; Giraldo & Valent 2013). To date, a number of effectors have been shown to contribute to the virulence of plant pathogens during infection (He et al. 2006; Sohn et al. 2009) or to target components in plant cells to suppress host immunity (Xiang et al. 2008; Göhre et al. 2008; Sarris et al. 2015). However, some effectors can be perceived by plant-encoded resistance (R) proteins. This perception triggers strong immune responses, resulting in rapid programmed cell death, called hypersensitive response cell death (Hogenhout et al. 2009). Initial studies on effectors of plant pathogens, including those of fungi, were limited to individual effector genes, which were identified based on genetic screening of multiple strains within a single species, displaying different infection phenotypes (Flor 1971). More recently, the development of next-generation sequencing (NGS) technologies has enabled researchers to predict effector gene sets from multiple strains and species at the whole genome level and to analyze the conservation or specificity of effectors in different plant pathogens (Dong et al. 2014; Hemetsberger et al. 2015; Sanz-Martín et al. 2015). However, accurate prediction of effector gene sets is still sometimes limited because short reads generated using NGS technologies are not geared toward assembling

repeat-rich genomic regions, which often harbor effector genes (Treangen & Salzberg 2012; Möller & Stukenbrock 2017).

Due to agricultural and scientific interests in *Colletotrichum* species, genome sequencing of these fungi has been performed (O'Connell et al. 2012; Gan et al. 2013, 2016; Hacquard et al. 2016; Baroncelli et al. 2016). This has made comparative genomic analysis of *Colletotrichum* species feasible. The purpose of my Ph.D. research is to understand how this group of fungi adapt to their niches in terms of their effector gene sets. In my thesis, I conducted comparative genomic analysis to identify conserved or specialized effector candidates of *Colletotrichum* (Chapter II). To elucidate the mechanisms that generate genomic diversity, I compared the genomes of two closely-related strains belonging to *C. higginsianum* in detail (Chapter III). Furthermore, I performed functional analyses on a highly conserved effector candidate that may play a role in infection of multiple *Colletotrichum* members in different host plants (Chapter IV).

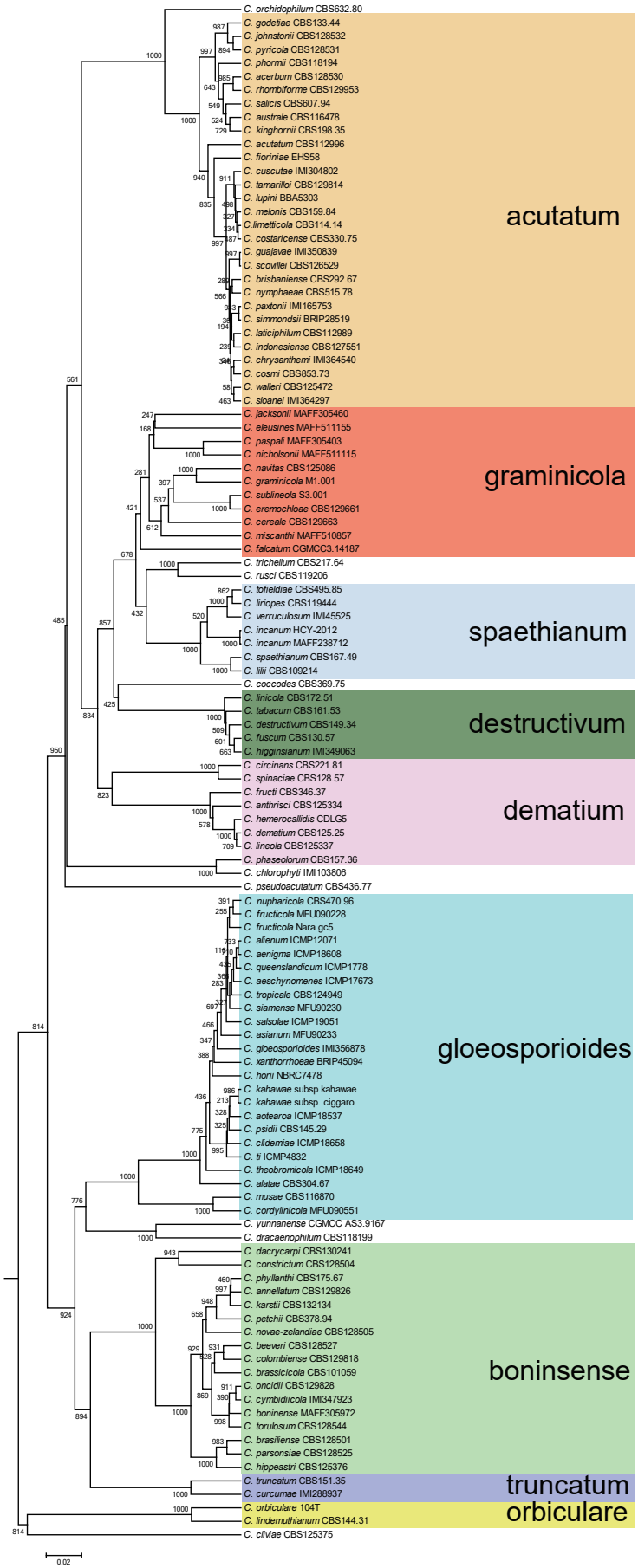



Fig. 1-1. Phylogenetic tree showing relationship of *Colletotrichum* species based on a concatenated alignment of *chitin synthase*, *actin*, internal transcribed spacer, and *tubulin* sequences. Values at the branch points represent bootstrap support values out of 1,000 replicates. Colored boxes in the tree represent each species complex. This figure is quoted from Gan, Narusaka, Kumakura, Tsushima et al. (2016).

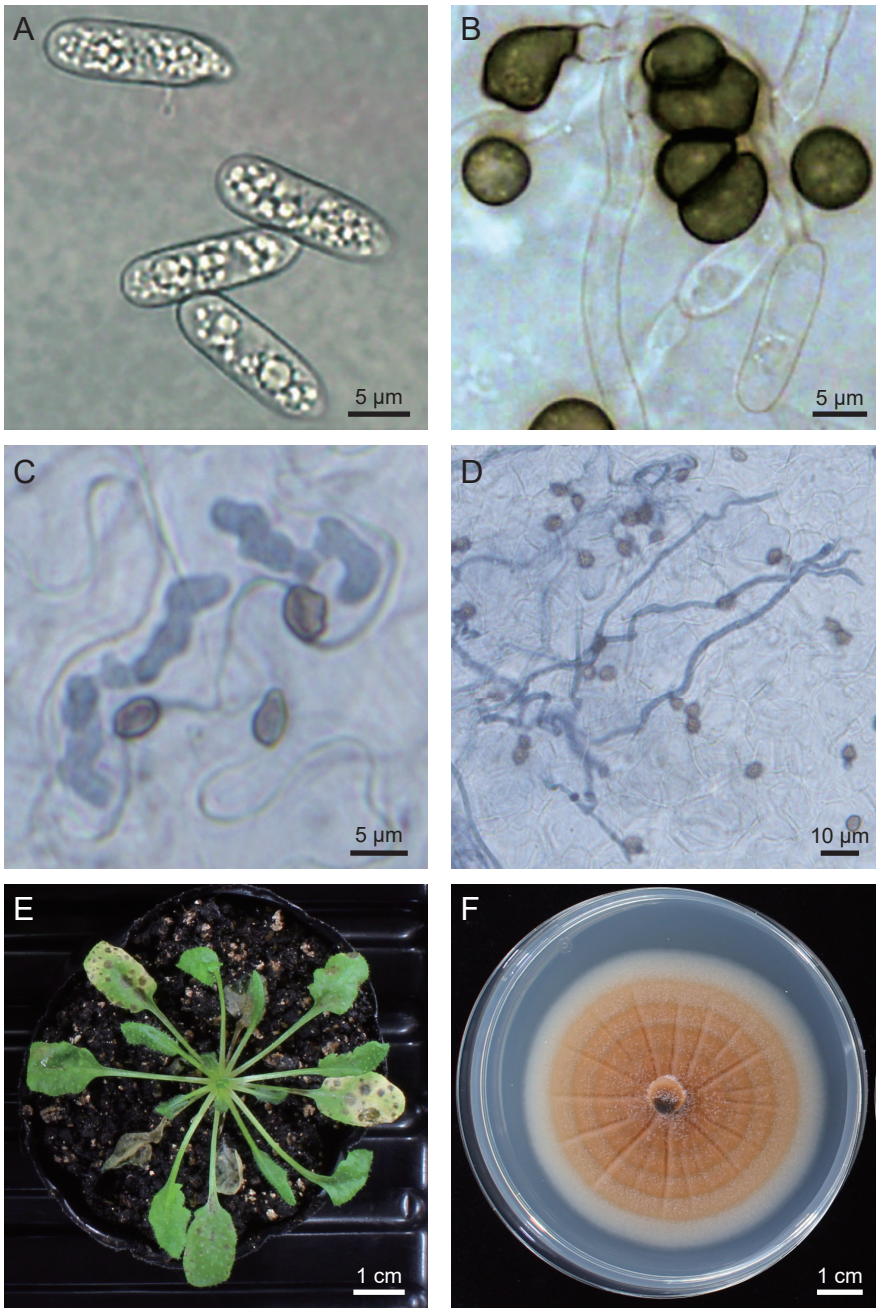



Fig. 1-2. Morphology of *Colletotrichum* species. (A) Conidia of *C. shisoi*. (B) Melanized appressoria of *C. shisoi*. (C) Trypan blue stained primary hyphae of *C. higginsianum* at 40 hours post-infection (hpi) in infected *A. thaliana* leaves. (D) Trypan blue stained secondary hyphae of *C. higginsianum* at 60 hpi in infected *A. thaliana* leaves. (E) Symptoms of *A. thaliana* plant infected by *C. higginsianum* at 9 days post-infection. (F) 7-day-old culture of *C. higginsianum* on potato dextrose agar.

## Chapter II: Conservation Patterns of Effector Candidates from *Colletotrichum* Fungi

*本章の一部は、５年以内に雑誌等で刊行予定のため、非公開。

Table 2-1. Genome assembly statistics of *Colletotrichum* species from previous studies.

| | ***C. higginsianum* IMI 349063** | ***C. graminicola* M1.001** | ***C. fructicola* Nara-gc5** | ***C. orbiculare* 104-T** |
|---|---|---|---|---|
| Isolated place | Trinidad and Tobago | The United States of America | Japan | Japan |
| Original host | *Brassica campestris* | *Zea mays* | *Fragaria x ananassa* | *Cucumis sativus* |
| Sequencer | Sanger, Roche454, Illumina | Sanger, Roche454 | Illumina | Illumina, Roche 454 |
| Total contig length (Mb) | 49.08 | 51.60 | 55.60 | 90.83 |
| Contig number | 10235 | 653 | 1241 | 526 |
| N50 Contig (Kb) | 6147.00 | 579.19 | 112.81 | 449.29 |
| GC-content (%) | 55.10 | 49.12 | 53.58 | 37.52 |
| Gene space coverage (%) complete / partial | 82.3 / 14.0 | 99.2 / 0.4 | 89.3 / 8.5 | 99.2 / 0.3 |
| Reference | O'Connell et al. 2012 | O'Connell et al. 2012 | Gan et al. 2013 | Gan et al. 2013 |

Table 2-3. Genome assembly statistics of *C. higginsianum* MAFF 305635-RFP.

| | ***C. higginsianum* MAFF 305635-RFP** | ***C. higginsianum* IMI 349063** |
|---|---|---|
| Sequencer | PacBio RS II | Sanger, Roche454, Illumina |
| Assenbler | HGAP | SeqMan NGen assembly tool |
| Total contig length(Mb) | 49.8 | 49.08 |
| Contig number | 28 | 10235 |
| N50 Contig(Mb) | 5.06 | 6147.00 |
| L50 Contig | 5 | 2404 |
| GC-content (%) | 54.61 | 55.10 |
| Gene space coverage (%) complete / fragmented | 99.0 / 0.3 | 82.3 / 14.0 |
| N per 100 kbp | 0 | 107.91 |
| Place of origin | Japan | Trinidad and Tobago |
| Host | *Brassica rapa* var. *perviridis* | *Brassica campestris* |
| Reference | This study | O'Connell et al. 2012 |

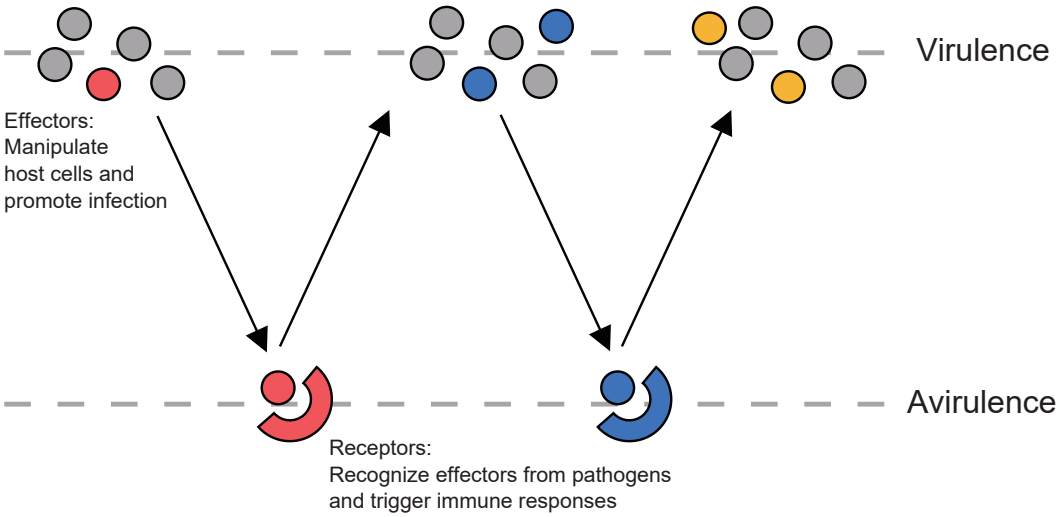



Fig. 2-1. A zigzag model illustrates co-evolution of pathogen effector complements and host immune receptors. In the first phase of this scheme, pathogens deliver effectors to manipulate host cells and to promote infection (virulence). However, a host-encoded immune receptor recognizes one effector (indicated in red), resulting in disease resistance (avirulence). In the next phase, pathogen strains are selected that have modified the recognized effector or lost it and gained new effectors (indicated in blue). Selection pressure leads new immune receptor alleles that can recognize one of effectors again. This cycle continuously turns in populations of pathogens and hosts and generates scores of effectors and immune receptors. This figure is modified from Jones & Dangl (2006).

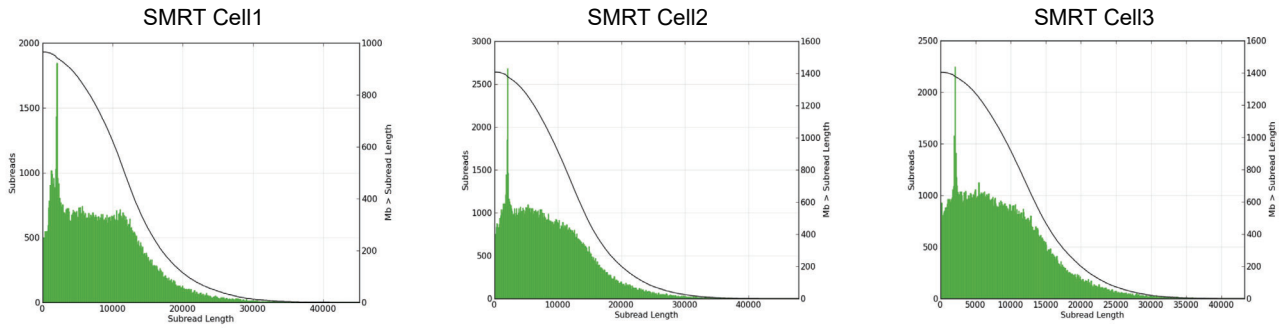



Fig. 2-2. Quality of reads derived from PacBio RSII. Graphs show the read qualities from three SMRT Cells. Green bars indicate the frequency of reads with corresponding lengths. Blacklines indicate total sequenced read lengths greater than reads of the corresponding subread length.

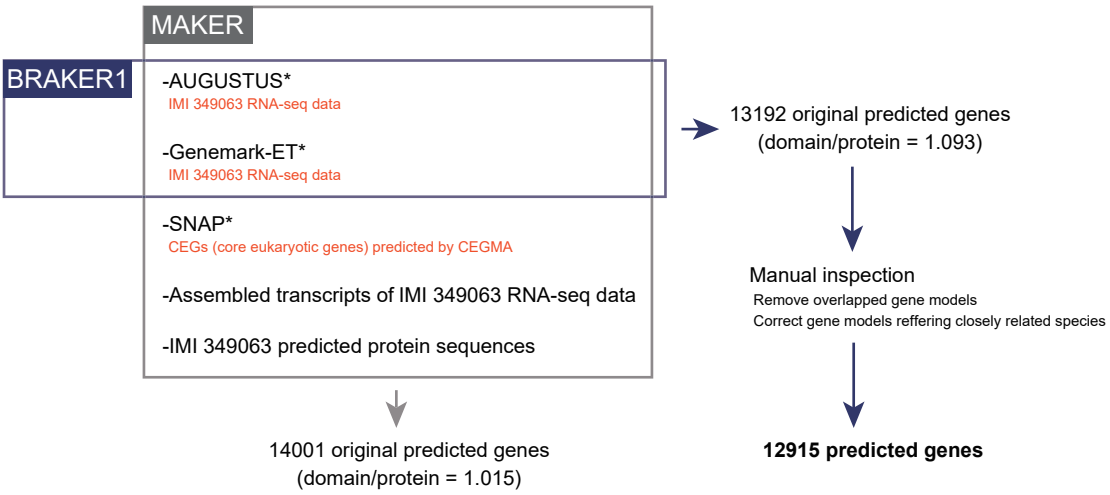



Fig. 2-3. Schemes showing the two annotation pipelines used, BRAKER1, and MAKER. Blue and gray rectangles indicate *ab initio* gene predictors and additional evidence used for BRAKER1 and MAKER, respectively. Asterisks indicate *ab initio* gene predictors. Evidence used for training of each *ab initio* gene prediction program is listed below in orange.

# Chapter III: Genomic Plasticity Mediated by Transposable Elements in *Colletotrichum higginsianum*


## Abstract

Phytopathogen genomes are under constant pressure to change, as pathogens are locked in an evolutionary arms race with their hosts, where pathogens evolve effector genes to manipulate their hosts, while the hosts evolve immune components to recognize the products of these genes. *Colletotrichum higginsianum*, a fungal pathogen with no known sexual morph, infects Brassicaceae plants including *Arabidopsis thaliana*. Previous studies revealed that *C. higginsianum* differs in its virulence towards various *A. thaliana* ecotypes, indicating the existence of coevolutionary selective pressures. However, inter-strain genomic variations in *C. higginsianum* have not been studied. Here, I compared two chromosome-level genome assemblies of *C. higginsianum* strains to identify genomic variations between them. I found that the two closely related strains vary in terms of large-scale rearrangements, the existence of strain-specific regions, and effector candidate gene sets and that these variations are frequently associated with transposable elements (TEs). *Colletotrichum higginsianum* has a compartmentalized genome consisting of gene-sparse, TE-dense regions with more effector candidate genes and gene-dense, TE-sparse regions harboring conserved genes. Additionally, analysis of the conservation patterns and syntenic regions of effector candidate genes indicated that the two strains vary in their effector candidate gene sets because of *de novo* evolution, horizontal gene transfer, or gene loss after divergence. Overall, this study shows mechanisms for generating genomic diversity in this asexual pathogen, which are important for understanding its adaption to hosts.

## Introduction

Genomic plasticity allows organisms to adapt to environmental changes and occupy novel niches. Although such adaptations can be observed in any organism, this is particularly important for pathogens that are co-evolving with their hosts (Raffaele & Kamoun 2012; Möller & Stukenbrock 2017). In these interactions, plant pathogens secrete small proteins known as effectors, which are thought to promote colonization by manipulating the host cells (Giraldo & Valent 2013). However, upon recognition by host immune receptors, effectors may also trigger strong immune responses (Dodds & Rathjen 2010; Asai & Shirasu 2015). Genes encoding effectors often have a higher degree of variation than housekeeping genes, as a signature of positive selection in coevolutionary relationships between pathogens and hosts (Dodds et al. 2006; Yoshida et al. 2009).

With the increasing availability of eukaryotic plant pathogen genomes, it has been revealed that effector genes are not distributed uniformly in pathogen genomes. Raffaele et al. (2010) found that effector genes are often associated with gene-sparse and repeat-rich genomic compartments showing higher rates of polymorphisms than other genomic regions in the genome of *Phytophthora infestans*. Such “two-speed genomes”, where genomes exhibit a bipartite genome architecture with rapidly evolving genomic regions facilitating adaptation and relatively conserved regions harboring housekeeping genes, have been widely identified in eukaryotic plant pathogens (Croll & McDonald 2012; Dong et al. 2015). Examples of these uneven patterns of genomic evolution in pathogen genomes include lineage-specific genomic regions and conditionally dispensable chromosomes that are highly variable, even within the same species, and are often required for full-pathogenicity or host specificity (Ma et al. 2010; De Jonge et al. 2013).

Using pulsed-field gel electrophoresis, PCR, or short-read sequencing technology, researchers have demonstrated that the genomes of eukaryotic plant pathogens undergo structural changes in chromosomes including chromosomal rearrangements and partial or whole chromosome

duplications or losses (Hatta et al. 2002; Chuma et al. 2003; Croll et al. 2013). The recent advent of long-read sequencers including PacBio has enabled the generation of more contiguous genome assemblies. Chromosome-level genome assemblies have allowed for more detailed analyses focusing on structural variations in the genomes of plant pathogenic fungi including *Verticillium dahliae* and *Magnaporthe oryzae* (Faino et al. 2016; Bao et al. 2017). Although such dynamic chromosomal changes can have deleterious effects on organisms, they may also play an important role in increasing genetic diversity, particularly for asexual organisms that cannot acquire genomic variations through meiotic recombination (Seidl & Thomma 2014).

*Colletotrichum* cause anthracnose disease in many plants, including important crops, and have a devastating economic impact (Crouch et al. 2014). To protect food security and understand their infection mechanisms, over 30 genomes of *Colletotrichum* have been sequenced to date (O'Connell et al. 2012; Gan et al. 2013, 2016, 2017, Baroncelli et al. 2014, 2016; Hacquard et al. 2016). Among them, *C. higginsianum* infects Brassicaceae plants, including the model plant *Arabidopsis thaliana*, as a hemibiotroph (O'Connell et al. 2004). Based on the interaction between *C. higginsianum* and *A. thaliana* as a model system, previous studies revealed that *A. thaliana* ecotypes vary in their susceptibility/resistance to *C. higginsianum* (Narusaka et al. 2004, 2009; Birker et al. 2009). This indicates that *C. higginsianum* strains have co-evolved with their hosts to promote infection and evade recognition by immune receptors. However, this pathogen appears to proliferate clonally, as its sexual cycle has never been identified (O'Connell et al. 2012). Therefore, it is important to understand whether different *C. higginsianum* strains exhibit high genetic diversity and, if so, how this pathogen achieves genomic variations in the absence of a sexual reproduction. Two different strains, MAFF 305635 isolated from *Brassica rapa* var. *perviridis* (Komatsuna) in Japan and IMI 349063 isolated from *B. campestris* subsp. *chinensis* (Pak-Choi) in Trinidad and Tobago, are frequently used in studies of *C. higginsianum* (Narusaka et al. 2004, 2009; Kleemann et al. 2012;

Takahara et al. 2016). The first version of the IMI 349063 genome assembly was released in 2012 (O'Connell et al. 2012) and a second version sequenced by PacBio is now available (Zampounis et al. 2016). However, the genome assemblies of other strains of *C. higginsianum* have not been released.

Here, I identified genomic variations between MAFF 305635-RFP, an MAFF 305635 transformant expressing monomeric red fluorescent protein (RFP) (Hiruma et al. 2010) and IMI 349063. I performed whole genome alignments using two highly contiguous assemblies of MAFF 305635-RFP and IMI 349063 to identify large-scale genomic differences between the two strains and compared the effector candidate gene complements of the two strains in detail. To determine how variations in effector candidate genes arise, I analyzed their conservation patterns in other ascomycetes and the synteny of genomic regions containing these genes.

## Material and Methods

### Fungal Strains

The details of all *C. higginsianum* strains used in this study are in Table 3-1. To extract genomic DNA, fungal strains was cultured in potato dextrose broth (BD Biosciences, Franklin Lakes, NJ, USA) at 24°C in the dark for 2 days. The genomic DNA of MAFF 305635-RFP was extracted from fungal tissue using CTAB and Qiagen Genomic-tip 500/G (Qiagen, Hilden, Germany) as described for the 1000 fungal genomes project (http://1000.fungalgenomes.org). Genomic DNA was extracted from other strains using the DNeasy Plant Mini Kit (Qiagen). Polymerase chain reaction (PCR) to detect genes on minichromosomes and highly variable effector candidate genes with presence/absence polymorphisms was performed using KOD FX Neo (Toyobo, Co., Ltd., Osaka, Japan) following the manufacturer's instructions. The primers are listed in Table 3-2. Fungal strains for infection assays were grown on potato dextrose agar (Nissui Pharmaceutical Co., Ltd., Tokyo, Japan) at 24°C for 12 h under black-light blue fluorescent bulb light/12 h dark conditions for one week.

### Genome Sequencing

SOLiD reads were generated from the MAFF 305635 wild-type strain (BioProject accession: PRJNA352900). Genomic DNA for preparing this library was obtained as described by Gan et al. (2013). Genomic sequencing was performed using a SOLiD3 sequencer (Applied Biosystems, Foster City, CA, USA) on a 600–800-base pair (bp) insert mate-paired library with 50-bp read lengths. A total of 109,156,072 paired-reads were generated.

### Whole Genome Alignments

Whole genome alignments were performed with nucmer in MUMmer 3.23 (Kurtz et al.

2004) using default settings. To remove spurious hits, the alignments were subsequently filtered by length, retaining alignments with ≥99% identity and ≥15 kb. Strain-specific regions were defined as genomic regions that were not aligned to the other genome after filtering. To confirm large-scale rearrangements, PacBio and SOLiD reads were mapped using SMRT analysis v2.3.0 and CLC Genomics Workbench8 (CLC bio, Aarhus, Denmark), respectively, using the default settings for both programs and visually examined using Integrated Genome Browser 9.0.1 (Freese et al. 2016). Mapping analysis using Illumina MiSeq reads derived from MAFF 305635 wild-type (CK7444), *vir-49*, and *vir-51* generated by Plaumann et al. (2018) (.fastq accessions: SAMN08226879, SAMN08226880, SAMN08226881) was performed using Bowtie 2 (Langmead & Salzberg 2012) with default settings. To detect genes on the minichromosomes of IMI 349063, reciprocal best hit analysis was performed. For this analysis, BLASTP analyses of the predicted proteomes of MAFF 305635-RFP and IMI 349063 were reciprocally performed with an *E*-value = $10^{-6}$ used as the threshold following Moreno-Hagelsieb & Latimer (2008).

**Prediction of Transposable Elements**

Transposable elements (TEs) were predicted as described by Castanera et al. (2016). TEs in the two *C. higginsianum* genome assemblies were predicted using RECON version 1.08 (Bao & Eddy 2002), RepeatScout version 1.0.5 (Price et al. 2005) (integrated into the RepeatModeler pipeline), and LTRharvest from GenomeTools-1.5.9 (Ellinghaus et al. 2008) with default settings for all programs. The results from LTRharvest were used as queries for BLASTN (cutoff *E*-value = $10^{-15}$) against the genome assembly and for BLASTX (cutoff *E*-value = $10^{-5}$) against the Repbase peptide database (downloaded on February 1, 2014) (Bao et al. 2015). Only sequences longer than 400 bp with more than five copies or yielding a significant hit to the described sequences in Repbase were further analyzed. The outputs from the genome assemblies of the two strains using the three

programs were merged and identical sequences were eliminated using USEARCH v9.1.13 (Edgar 2010) with the -fastx_uniques option. The obtained sequences were clustered at 80% similarity using USEARCH v9.1.13 with the -cluster_smallmem option to create a custom TE library. Consensus sequences in the library were classified using BLASTX (cutoff *E*-value = $10^{-5}$) against the Repbase peptide database, and the final libraries were used as input for RepeatMasker (http://www.repeatmasker.org). Consensus sequences without similarity to any Repbase entry were removed from the library. RepeatMasker outputs were parsed using the One_code_to_find_them_all version 1.0 (Bailly-Bechet et al. 2014) to reconstruct TE fragments into full-length copies. Among the reconstructed fragments of TEs, those longer than 400 bp were used for analysis. To perform Monte Carlo tests, 1000 trials to model TEs randomly located on the genome were generated using BEDTools version 2.25.0 (Quinlan & Hall 2010) with the shuffle -noOverlapping option. Overlap between TEs and strain-specific regions were calculated using BEDTools version 2.25.0 with the coverage option.

**Analysis of Genome Compartmentalization**

Flanking intergenic regions were calculated using R scripts as described by Saunders et al. (2014). Two-dimensional plots were created by referring to Frantzeskakis et al. (2018). Distances from genes to the nearest TEs were calculated using BEDTools version 2.27.1 (Quinlan & Hall 2010) with the closest -D a -iu -a or -D a -id -a option. Fungal universal single-copy orthologs in the two strains were identified as the best hits from BLASTP (cutoff *E*-value = $10^{-5}$) results using fungi_odb9 sequences provided in BUSCO v3.0.2 (Simão et al. 2015) as queries and *C. higginsianum* predicted proteomes as references. For synteny analysis, Easyfig 2.2.2 (Sullivan et al. 2011) was used with an identity cutoff of 70%.

**Comparative Secretome Analysis**

To predict secreted proteins, SignalP 4.1 (Petersen et al. 2011), TMHMM 2.0 (Krogh et al. 2001), and big-PI Fungal Predictor (Eisenhaber et al. 2004) were used with the default settings. In this study, effector candidate proteins were defined as predicted secreted proteins (with a signal peptide present but no transmembrane domains and glycosylphosphatidylinositol-anchors) with lengths of less than 300 amino acids. To examine whether effector candidates show similarity to known proteins, BLASTP analysis using the Swiss-Prot database (downloaded at October 22, 2016) was performed (cutoff *E*-value = $10^{-5}$). To detect variations in effector candidates, the protein sequences of effector candidates from each strain were queried against the genome assembly of the other strain using exonerate version 2.2.0 (Slater & Birney 2005) with the protein2genome option. Query coverage values of homologous sequences were calculated by reciprocally performing BLASTP between effector candidates and exonerate-predicted protein sequences. The results were inspected and manually corrected. A dendrogram based on the presence/absence patterns of highly variable effector candidate genes was drawn using the R package heatmap.2 from gplots v3.0.1.

**Conservation Patterns of Highly Variable Effector Candidate Genes**

The details of the genome and proteome sequences of 25 ascomycetes used in this analysis are shown in Table 3-3. To identify orthogroups containing highly variable effector candidates among the 25 ascomycetes, orthoMCL v2.0.9 (Li et al. 2003) was used with an *E*-value = $10^{-5}$ as the threshold and inflation value of 1.5. The alignment of CH35J_007515, CH35J_007516, and their homologs was generated using protein sequences with their predicted signal peptides removed. Sequence alignments were performed using the CLC Genomics Workbench8 (CLC bio).

**Phylogenetic Analyses**

The phylogenetic tree to classify 15 *C. higginsianum* strains was generated based on the combined alignments of *actin* (*ACT*), *chitin synthase I* (*CHS-1*), *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*), *histone H3* (*HIS3*), internal transcribed spacer (ITS), and *tubulin-2* (*TUB2*). This tree was drawn using previously identified sequences from other species in the *C. destructivum* complex (Damm et al. 2014). The DNA sequences of 15 strains were determined by direct Sanger sequencing of PCR products amplified with KOD -Plus- Neo (Toyobo) following the manufacturer's instructions. The primers are listed in Table 3-2. The phylogenetic tree of 25 ascomycetes was generated based on the combined alignments of single gene orthogroups conserved in all 25 ascomycetes obtained from the orthoMCL results. For both trees, DNA or protein sequences were aligned using MAFFT version 7.215 (Katoh et al. 2002) with the auto setting and trimmed using trimAL v1.2 (Capella-Gutiérrez et al. 2009) with the automated1 settings. The concatenated trimmed alignments were then utilized to estimate the maximum-likelihood species phylogeny with RAxML version 8.2.11 (Stamatakis 2014) with 1,000 bootstrap replicates. To generate the maximum-likelihood tree of species in the *C. destructivum* complex, the GTRCAT model was used. To generate the maximum-likelihood tree of 25 ascomycetes, PROTGAMMAAUTO was used to find the best protein substitution model and autoMRE was used to determine the appropriate number of bootstrap samples. *Saccharomyces cerevisiae* sequences were used as the outgroup in this tree. Trees were visualized using iTOL version 4.1 (Letunic & Bork 2016).

**Infection Assays**

Plants were grown at 22°C with a 10-h photoperiod for 4 weeks. Three leaves per plant were inoculated with 5 μL droplets of conidial suspensions at $5 \times 10^5$ conidia $mL^{-1}$. Plants were maintained at 22°C with a 10-h photoperiod under 100% humidity conditions after inoculation. The symptoms were observed at 6 days after inoculation and lesion areas were measured using the color

threshold function in ImageJ 1.51k (Schneider et al. 2012) using the following settings: Hue: 0-255, saturation: 110-140, and brightness: 0-255 with a square region of interest (ROI).

## Results

### Extensive Genomic Variations Between *C. higginsianum* Strains

To examine genomic differences between the two sequenced strains of *C. higginsianum*, whole-genome alignments were performed (Figure 3-1). The two strains shared 88.2% of their sequences (≥99% identity, ≥15 kb). However, 19 synteny breakpoints were detected. Remapping of the PacBio reads derived from MAFF 305635-RFP and SOLiD reads derived from MAFF 305635 wild-type to the genome assembly of IMI 349063 confirmed that at least 11 of these sites were not caused by misassembly and are present in the wild-type strain (Figure 3-2). These 19 synteny breakpoints included ten large-scale rearrangements between the two strains including six inter-chromosomal translocations and four intra-chromosomal inversions. Whole-genome alignments also revealed that the two strains contained strain-specific regions (<99% identity, <15 kb) with lower sequence similarity compared to the other strain.

Notably, chromosomes 11 and 12, which are known as minichromosomes in IMI 349063, were the two largest strain-specific regions. Indeed, only 19 of 271 genes on the minichromosomes of IMI 349063 showed reciprocal best hits in MAFF 305635-RFP (Table 3-4). Recently, Plaumann et al. (2018) reported that MAFF 305635 also contains two minichromosomes, chromosomes 11 and 12. The same study identified *vir-49* and *vir-51* as MAFF 305635-derived virulence mutants lacking chromosome 11. Mapping of the Illumina MiSeq reads derived from MAFF 305635 wild-type (CK7444), *vir-49*, and *vir-51* from their study to the genome assembly of MAFF 305635-RFP revealed that no contigs showed reduced read coverage per gene compared to the average coverage of all genes (Figure 3-3). To confirm the presence or absence of chromosomes 11 and 12 in MAFF 305635-RFP, PCR was also performed using the primer sets described by Plaumann et al. (2018) and newly designed primer sets to amplify selected genes from chromosome 12 of IMI 349063. PCR products were not detected in MAFF 305635-RFP with any primer sets designed to amplify genes on

the minichromosomes (Figure 3-4). Taken together, these results suggest that MAFF 305635-RFP lacks the minichromosomes reported in MAFF 305635 wild-type (CK7444) used by Plaumann et al. (2018).

**Association Between Genomic Variations and Transposable Elements in *C. higginsianum***

TEs are known to contribute to the generation of genomic variations. Thus, I predicted that TEs are a driving force in generating genomic variations in *C. higginsianum* in the absence of meiosis. To test this hypothesis, the TEs of *C. higginsianum* were predicted *de novo* using the pipeline described by Castanera et al. (2016). The genome coverage of TEs was estimated to be 4.6% and 5.1% in MAFF 305635-RFP and IMI 349063, respectively. Both assemblies contained several shorter contigs or chromosomes with higher TE coverage compared to the rest of the genome (Figure 3-5). Notably, in IMI 349063, the two minichromosomes showed higher TE coverage (37.3% in chromosome 11 and 21.2% in chromosome 12). Predicted TEs were classified into four types: Copia, Gypsy, and Tad1 from Class I transposable elements and TcMar-Fot1 from Class II transposable elements. Among the four types of TEs, TcMar-Fot1 showed the highest genome coverage in both strains (3.49% in MAFF 305635-RFP and 4.23% in IMI 349063) (Figure 3-6).

The association between genomic variations and TEs was assessed (Figure 3-7). The results showed that 8 of 10 large-scale rearrangements were within 10 kb of the nearest TE in MAFF 305635-RFP (Figure 3-8). Additionally, 29.5% and 29.8% of strain-specific regions were occupied by TEs in MAFF 305635-RFP and IMI 349063, respectively. This rate is significantly higher than if the TEs were randomly distributed in both genomes (highest coverage in 1,000 trials = 7.5% and 7.2% in MAFF 305635-RFP and IMI 349063, respectively, $P = 0.001$, Monte Carlo test) (Figure 3-9).

**Compartmentalization of Effector Candidate and Housekeeping Genes in the *C. higginsianum* Genome**

To determine whether the *C. higginsianum* genome contains regions enriched with effector genes, I predicted the effector candidates of *C. higginsianum*. Proteins were classified as effector candidates if they were predicted to be secreted proteins with lengths of less than 300 amino acids. This analysis revealed that both *C. higginsianum* strains have a similar number of effector candidates (582 in MAFF 305635-RFP and 576 in IMI 349063). BLASTP analysis using the Swiss-Prot database (cutoff *E*-value = $10^{-5}$) revealed that 428 (73.5%) and 427 (74.1%) effector candidates in MAFF 305635-RFP and IMI 349063, respectively, are of unknown function (Figure 3-10).

Next, I assessed the flanking intergenic regions (FIRs) of effector candidate genes and fungal universal single-copy orthologs to determine if these genes are in gene-dense or gene-poor genomic regions. Two-dimensional plots describing 5′ and 3′ FIRs indicated that fungal universal single-copy orthologs tended to be closer to flanking neighboring genes, whereas effector candidate genes were further apart from their nearest gene neighbors (Figure 3-11). These patterns were significant in both strains (*P*-value = 8.499e-15 and 8.899e-11 in MAFF 305635-RFP and IMI 349063, respectively, Wilcoxon rank sum test) (Figure 3-12A, B). The distances between effector candidate genes and their nearest TEs were also investigated. The distributions of distances between effector candidate genes and their nearest TEs are significantly different from the distances between fungal universal single-copy orthologs and their nearest TEs (*P*-value < 2.2e-16 in both strains, Wilcoxon rank sum test) (Figure 3-12C, D). In both strains, the median distances between effector candidate genes and their neighboring TEs are lower than the median distances between fungal universal single-copy orthologs and their closest TEs.

**Variations in Effector Candidate Genes Between *C. higginsianum* Strains**

To provide insight into the adaptative evolution of *C. higginsianum* effectors, variations in effector candidates were investigated (Figure 3-13A). To eliminate predicted gene variations related to differences in the annotation programs used, protein sequences of effector candidates from each strain were queried against the genome assembly of the other strain. Based on this analysis, 474 (81.4%) and 469 (80.6%) effector candidates in MAFF 305635-RFP and in IMI 349063 were identical in the other strain, indicating that the two strains generally contain a similar repertoire of effector candidates. However, 8 (1.37%) MAFF 305635-RFP and 18 (3.09%) IMI 349063 candidates were highly variable between the two strains, defined here as having ≤90% query coverage. Among them, ten candidates were detected as presence/absence polymorphisms and seven candidates showed lower query coverages because of frameshifts (Table 3-5). A total of 100 (17.2%) MAFF 305635-RFP and 89 (15.3%) IMI 349063 effector candidates were also polymorphic (containing at least one non-synonymous substitution and >90% query coverage).

To examine whether these genes varied from those in other strains of *C. higginsianum*, ten highly variable effector candidate genes with presence/absence polymorphisms were assessed by PCR in 15 *C. higginsianum* strains isolated from different geographic locations, including MAFF 305635-RFP and IMI 349063 (Table 3-1). Molecular phylogenetic analysis, using all available sequences from strains from the Destructivum species complex described by Damm et al. (2014), confirmed that all 15 strains were classified as *C. higginsianum* (Figure 3-14). Amplification of these effector candidate genes by PCR revealed that their conservation patterns varied within other strains as well (Figure 3-13B, Figure 3-15). Eight strains isolated in Japan showed different patterns for the presence/absence of these genes. In contrast, four strains from Trinidad and Tobago showed the same presence/absence patterns except IMI 349063Δ, an IMI 349063-derived strain that appears to lack chromosome 12, as no PCR bands were detected in all tested genes in chromosome 12 (Figure 3-16).

**Conservation of Highly Variable *C. higginsianum* Effector Candidate Genes**

To explore the potential mechanisms of how *C. higginsianum* acquired highly variable effector candidate genes, I investigated the conservation patterns of these genes amongst 25 ascomycetes (Figure 3-17, Table 3-3). The results showed that CH35J_003318 from MAFF 305635-RFP and CH63R_09232, CH63R_09755, CH63R_14384, CH63R_14389, CH63R_14470, and CH63R_14558 from IMI 349063 were found in either of the two strains among the tested ascomycetes. In contrast, CH35J_011924 and CH35J_002132 from MAFF 305635-RFP and CH63R_14516, CH63R_05497, CH63R_04687, and CH63R_06433 from IMI 349063 were relatively conserved in Ascomycota but absent from MAFF 305635-RFP or IMI 349063. The remaining highly variable effector candidate genes showed uneven conservation patterns that did not follow the species' phylogenetic relationships. Among them, CH35J_007515 and CH35J_007516 were found to be paralogs. Interestingly, the paralogs showed high similarities to proteins found only in *Bipolaris maydis* as shown in Figure 3-17. Querying the protein sequences of the paralogs against the NCBI nucleotide collection revealed that only *B. maydis*, *B. oryzae*, *B. victoriae*, *B. zeicola*, *B. sorokiniana*, *Aspergillus novofumigatus*, and *A. terreus* have similar sequences (TBLASTN, *E*-value $\leq 2 \times 10^{-28}$), suggesting that a potential horizontal gene transfer event had occurred (Figure 3-18).

**Synteny of Genomic Regions Encoding Highly Variable Effector Candidates**

To investigate the associations between effector candidate genes and TEs, synteny analysis of genomic regions containing highly variable effector candidate genes was performed (Figure 3-19). Syntenic regions containing 17 of 26 highly variable effector candidate genes were reconstructed. I found that eight (six in MAFF 305635-RFP and two in IMI 349063) of 17 such genes were in synteny-disrupted regions associated with TEs. However, nine highly variable effector candidate genes were not associated with synteny-disrupted regions (two in MAFF 305635-RFP and seven in

IMI 349063). Notably, seven of these effector candidate genes (two in MAFF 305635-RFP and five in IMI 349063) were highly variable because of frameshifts.

No syntenic region was identified for the remaining nine highly variable effector candidate genes. These genes in the two minichromosomes of IMI 349063 and syntenic regions could not be reconstructed because of the loss of minichromosomes in MAFF 305635-RFP. Interestingly, in IMI 349063, nine of the 18 highly variable effector candidate genes were in its two minichromosomes. Of these, five genes were completely absent from MAFF 305635-RFP. However, despite the lack of minichromosomes, related sequences were identified for four of these genes in MAFF 305635-RFP (Table 3-6).

**Both Sequenced *C. higginsianum* Strains Showed Similar Virulence Levels Towards *A. thaliana* Ecotypes Ws-2 and Ler-0**

To examine the impact of the genetic differences observed between the two sequenced isolates, I assessed the ability of both strains to cause lesions on *A. thaliana* ecotypes Ws-2 and Ler-0. A previous report showed that Ws-2 and Ler-0 were resistant and susceptible, respectively, to *C. higginsianum* IMI 349061 (Birker et al. 2009). To quantify these differences in pathogenicity, lesion areas caused by MAFF 305635-RFP and IMI 349063 were measured on Ws-2 and Ler-0. Both strains showed significantly larger lesions on Ler-0 compared to Ws-2. However, no significant differences were detected between the strains (Figure 3-20, Figure 3-21).

**Discussion**

Pathogenic microbes are closely associated with hosts, and their genomes are under selective pressure to promote effective colonization and evade host recognition. However, genomic variations in *Colletotrichum* spp. and the mechanisms underlying such genomic changes have not been widely examined. By performing comparative genomics using highly contiguous assemblies of two C. higginsianum strains, I show, for the first time in the genus *Colletotrichum*, that the genome of this plant pathogen is remarkably flexible, as represented by large-scale structural rearrangements and the presence of strain-specific regions.

Dynamic genomic changes may be beneficial for plant pathogenic fungi by allowing the rapid generation of novel genetic alleles. However, extreme alterations in genome structures also impair homologous chromosome pairing during meiosis (Kistler & Miao 1992). In the *Colletotrichum* genus, few sexual productions have been described (Vaillancourt & Hanau 1991; Rodríguez-Guerra et al. 2005; Menat et al. 2012) and most species, including *C. higginsianum*, are considered as pre-dominantly asexual. The results obtained in this study suggest that genomic plasticity in *C. higginsianum* contributes to the generation of novel genetic variations; however, this may cause difficulty in performing sexual reproduction as well.

I indicate that TEs contribute to the generation of large-scale rearrangements and strain-specific regions in *C. higginsianum*. Seidl & Thomma (2014) proposed that TEs impact genomic content not only by simple insertion or excision, but also by inducing homology-based recombination during double-strand DNA break repair. Additionally, it is possible that Class II transposable elements, which are the most abundant class of TEs in *C. higginsianum*, autonomously cause genomic rearrangements through alternative transpositions, as described for the *Ac/Ds* elements of maize (Zhang et al. 2009; Yu et al. 2011).

I found that *C. higginsianum* has features of a compartmentalized genome consisting of

gene-sparse, TE-dense regions with more effector candidate genes and gene-dense, TE-sparse regions with more conserved genes. These so-called “two-speed genomes” have also been found in other eukaryotic plant pathogens, such as *P. infestans* and *Leptosphaeria maculans* (Raffaele et al. 2010; Grandaubert et al. 2014). This suggests that having a compartmentalized genome structure is advantageous for various eukaryotic plant pathogens to allow both rapid evolution of effector genes and protection of house-keeping genes from the deleterious effects of TEs.

I identified 26 highly variable effector candidate genes between two strains of *C. higginsianum*. Further, I observed the presence/absence polymorphisms of ten highly variable effector candidate genes in 15 different *C. higginsianum* strains from different geographic locations. These genes showed various conservation patterns in Ascomycota, suggesting that *C. higginsianum* acquires differences in its effector repertoire via several different mechanisms, such as *de novo* evolution, horizontal gene transfer, or gene loss after the divergence of species.

Seven effector candidate genes were predicted to be generated through *de novo* evolution in *C. higginsianum* because the homologs of these genes were not found in other ascomycete species. In *Zymoseptoria tritici*, an effector candidate gene that is highly correlated with pathogenicity towards different wheat cultivars, *Zt_8_609*, was also suggested to have recently emerged after speciation (Hartmann et al. 2017). The mechanisms underlying *de novo* evolution of effector genes remain unclear. However, such orphan genes without homologs in other lineages may arise by duplications followed by exceeding divergence beyond the threshold of homology searches or *de novo* generation of functional open reading frames from non-coding regions (Tautz & Domazet-Lošo 2011).

The uneven conservation patterns of effector candidate genes in Ascomycota (e.g. CH35J_007515 and CH35J_007516) suggest that these genes were horizontally transferred and/or frequently gained/lost in this taxon. There are several reports of the transfer of effector genes in plant

pathogenic fungi; for example, *Avr-Pita* in *M. oryzae* is known to be horizontally transferred between individual isolates and *Ave1* in *V. dahliae* is thought to be obtained from plants (Chuma et al. 2011; de Jonge et al. 2012). Thus, in *C. higginsianum*, horizontal gene transfer events may also generate highly variable effector candidate genes displaying conservation patterns that contradict species phylogeny.

Through synteny analysis, I found that eight of the 26 highly variable effector candidate genes in synteny-disrupted regions were associated with TEs. However, the remaining 18 highly variable effector candidate genes were not detected in synteny-disrupted regions. Notably, seven and nine of these genes were found to be strain-specific because of frameshifts and the loss of minichromosomes, respectively. Therefore, TEs clearly contribute to generating variations in effector candidate genes, although other mechanisms also exist, such as DNA point mutation resulting from replication errors and entire chromosome loss.

The loss of minichromosomes in MAFF 305635-RFP and IMI 349063Δ was likely caused by the unstable nature of minichromosomes in this species, as described by Plaumann et al. (2018). Infection assays comparing the two sequenced isolates suggested that the identified genomic variations including the loss of two minichromosomes did not result in differences in the pathogenicity towards the *A. thaliana* ecotypes Ws-2 and Ler-0. However, genomic variations between the two strains may cause differences in pathogenicity when they infect other *A. thaliana* ecotypes or Brassicaceae plants. Previous reports independently showed that MAFF 305635 and IMI 349061 are avirulent on Ws-2, which harbors the dual *Resistance* (*R*) genes *RPS4* and *RRS1* (Birker et al. 2009; Narusaka et al. 2017). The direct comparison using MAFF 305635-RFP and IMI 349063 suggested that the variations in effector candidates revealed in this study did not affect recognition by RRS1 and RPS4, indicating that the effector recognized by this R protein pair remains conserved.

Overall, by comparing closely related strains of *C. higginsianum*, I identified genomic

variations in the structure and genes encoding effector candidates and potential mechanisms of altering the genome mediated by TEs in this species. The results obtained in this study improve the understanding of adaptation driven by genomic evolution in this scientifically and agriculturally important group of plant pathogenic fungi.

Table 3-1. *C. higginsianum* strains used in this study.

| No. | Strain name | Original isolated host | Country | Location | Reference |
|---|---|---|---|---|---|
| 1 | Aba1-1 | *Brassica parachinensis* | Japan | Edogawa, Tokyo | Horie et al. 1988 |
| 2 | Abc1-3 | *Brassica chinensis* | Japan | Edogawa, Tokyo | Horie et al. 1988 |
| 3 | Abcr1-2 | *Brassica cernua* | Japan | Edogawa, Tokyo | Horie et al. 1988 |
| 4 | Abj1-2 | *Brassica japonica* | Japan | Edogawa, Tokyo | Horie et al. 1988 |
| 5 | Abju1-2 | *Brassica juncea* | Japan | Edogawa, Tokyo | Horie et al. 1988 |
| 6 | Abo1-1 | *Brassica oleracea* var. *gemmifera* | Japan | Edogawa, Tokyo | Horie et al. 1988 |
| 7 | Abp1-2 | *Brassica pekinensis* | Japan | Edogawa, Tokyo | Horie et al. 1988 |
| 8 | IMI 349061 | *Brassica campestris* | Trinidad and Tobago | N.A. | |
| 9 | IMI 349063A [a] | *Brassica campestris* | Trinidad and Tobago | N.A. | |
| 10 | IMI 349063B [a] | *Brassica campestris* | Trinidad and Tobago | N.A. | |
| 11 | NBRC 6182 | *Brassica* sp. | Italy | N.A. | |
| 12 | P01 | Chinese cabbage (var. Pei-Tsai) | Taiwan | Siluo, Yulin county | |
| 13 | P02 | Chinese cabbage (var. Pei-Tsai) | Taiwan | Jhuolan, Miaoli county | |
| 14 | IMI 349063Δ [b] | *Brassica campestris* | Trinidad and Tobago | N.A. | |
| 15 | MAFF 305635-RFP [c] | *Brassica rapa* var. *perviridis* | Japan | Edogawa, Tokyo | Hiruma et al. 2010 |

a IMI 349063 was re-isolated into IMI 349063A and IMI 349063B (O'Connell et al. 2004).
IMI 349063A is the sequenced strain (Takahara et al. 2016).

b Reisolated by Yasuyuki Kubo.

c MAFF 305635 transformant expressing mRFP driven by TEF promoter (Hiruma et al. 2010)

Table 3-2. Primers used in this study.

| Primer name | 5'-3' sequence | Comment | Reference |
|---|---|---|---|
| CK5295 | CGATGGTCTTGTCTGGGTGAAC | CH63R_14381 | Plaumann et al. 2018 |
| CK5209 | GGGGACAACTTTGTATAGAAAAGTTGTTAGCCCCGCTGCCTACTTGTTCCTT | CH63R_14381 | Plaumann et al. 2018 |
| CK5230 | GGGGACAGCTTTCTTGTACAAAGTGGAATCCGAAGAACAGGCATTGTTAG | CH63R_14384 | Plaumann et al. 2018 |
| CK5231 | GGGGACTGCTTTTTTGTACAAACTTGTTGTTAGTTTGGCGTTCATTAGCC | CH63R_14384 | Plaumann et al. 2018 |
| CK5299 | GTCGGAATCGAACCAACCGTC | CH63R_14406 | Plaumann et al. 2018 |
| CK5393 | GGGGACTGCTTTTTTGTACAAACTTGTCTGGCGTTGCCTGTTCTGTC | CH63R_14406 | Plaumann et al. 2018 |
| CK5236 | GGGGACAACTTTGTATAGAAAAGTTGTTTGAGCGGAGTGGCGAGAGGAAT | CH63R_14428 | Plaumann et al. 2018 |
| CK5237 | GGGGACAACTTTGTATAATAAAGTTGTGGAAGCTAAGTCCTTGTTTGCA | CH63R_14428 | Plaumann et al. 2018 |
| CK5207 | GGGGACAGCTTTCTTGTACAAAGTGGAACACCGCAAAGGGTTCACGGCAGAT | CH63R_14500 | Plaumann et al. 2018 |
| CK5208 | GGGGACTGCTTTTTTGTACAAACTTGTCGACGCAAGAGGGAAAAGGCACGA | CH63R_14500 | Plaumann et al. 2018 |
| CK5213 | GGGGACAACTTTGTATAGAAAAGTTGTTTTCCAACCCACATTCCAGGGGCGA | CH63R_14517 | Plaumann et al. 2018 |
| CK5214 | GGGGACAACTTTGTATAATAAAGTTGTGCCGTTCTCACCCTCCTCGCATCT | CH63R_14517 | Plaumann et al. 2018 |
| CK5196 | TGCCGACAATGCTTTCGTAGA | CH63R_14534 | Plaumann et al. 2018 |
| CK5197 | TCAAGTCGTGGCAAAGTTCCT | CH63R_14534 | Plaumann et al. 2018 |
| CH63R_14520_F | ATGTTGGATATCGATGATATTCTGGCC | CH63R_14520 | This study |
| CH63R_14520_R | CTACAAGGCTCTATCCGCCAG | CH63R_14520 | This study |
| CH63R_14535_F | ATGCGGTTGTTACTGGTTTTCCTC | CH63R_14535 | This study |
| CH63R_14535_R | CTAATTTAAAATCTTCACCATGACTAAAATGAGGG | CH63R_14535 | This study |
| CH63R_14587_F | ATGGGTCTGAAGGGAAGCTG | CH63R_14587 | This study |

| | | | |
|---|---|---|---|
| CH63R_14587_R | TCAACTGAAAGCTGGCTTCTGATC | CH63R_14587 | This study |
| CH63R_14632_F | ATGAAGTTCTTCAATCTGATTTCCTTCG | CH63R_14632 | This study |
| CH63R_14632_R | TTAGAGTGTAGGCGTTGTCGC | CH63R_14632 | This study |
| CH35J_003318_F | ATGCATCTCCAAACTGTCCCCTTC | CH63J_003318 | This study |
| CH35J_003318_R | TTAGGACTCCTCGCCACCAAC | CH63J_003318 | This study |
| CH35J_007515_R | CTAGCATCCCTGGTAGACGGC | CH63J_007515 and 007516 | This study |
| CH35J_007516_R | TTAGCACCCCTGATAAACGACGAAAG | CH63J_007515 and 007516 | This study |
| CH35J_010999_F | ATGAAGGGTTCTCTCATTCTTACTCTTTC | CH63J_010999 | This study |
| CH35J_010999_R | TCAGTGGGCGCTTCTGATAAC | CH63J_010999 | This study |
| CH35J_011924_F | ATGCACGCCTCCCGCTTC | CH63J_011924 | This study |
| CH35J_011924_R | TCAGAGGTTGATGTTCAGGCCG | CH63J_011924 | This study |
| CH63R_14384_F | ATGAAGTCTCCTACTTTACTTTTTGTCTTTG | CH63R_14384 | This study |
| CH63R_14384_R | TTAAGTACTCTGATTCCAGCAGTTCC | CH63R_14384 | This study |
| CH63R_14389_F | ATGAAGCTCCTTCCTCTGTTAGC | CH63R_14389 | This study |
| CH63R_14389_R | TTAATCCACCCTGTTCTTCTAAGTGAG | CH63R_14389 | This study |
| CH63R_14470_F | ATGTGGTTCATTTTTGTCTTCGTTTTCG | CH63R_14470 | This study |
| CH63R_14470_R | CTATGCACCATCCAAGGCCC | CH63R_14470 | This study |
| CH63R_14618_F | ATGCAAGACTGGTAAGTGAAGCC | CH63R_14618 | This study |
| CH63R_14618_R | CTATGTATCAAACGGCTCGATCGC | CH63R_14618 | This study |
| CH63R_14648_F | ATGAGATTTTTATATCTCCTGTCATCCTGC | CH63R_14648 | This study |
| CH63R_14648_R | TTAAGACGTGAACTTGCCATTTCTCATG | CH63R_14648 | This study |
| ACT-512F | ATGTGCAAGGCCGGTTTCGC | *ACT* | Carbone & Kohn 1999 |
| ACT-783R | TACGAGTCCTTCTGGCCCAT | *ACT* | Carbone & Kohn 1999 |
| CHS-79F | TGGGGCAAGGATGCTTGGAAGAAG | *CHS-1* | Carbone & Kohn 1999 |
| CHS-345R | TGGAAGAACCATCTGTGAGAGTTG | *CHS-1* | Carbone & Kohn 1999 |
| GDF1 | GCCGTCAACGACCCCTTCATTGA | *GAPDH* | Templeton et al. 1992 |

| | | | |
|---|---|---|---|
| GDR1 | GGGTGGAGTCGTACTTGAGC ATGT | *GAPDH* | Templeton et al. 1992 |
| ITS-1F | CTTGGTCATTTAGAGGAAGTA A | *HIS3* | Gardes & Bruns 1993 |
| ITS-4 | TCCTCCGCTTATTGATATGC | *HIS3* | White et al. 1990 |
| CYLH3F | AGGTCCACTGGTGGCAAG | ITS | Crous et al. 2004 |
| CYLH3R | AGCTGGATGTCCTTGGACTG | ITS | Crous et al. 2004 |
| T1 | AACATGCGTGAGATTGTAAGT | *TUB2* | O'Donnell & Cigelnik 1997 |
| Bt2b | ACCCTCAGTGTAGTGACCCT TGGC | *TUB2* | Glass & Donaldson 1995 |

Table 3-3. Details of the predicted proteomes of 25 ascomycetes used in this study.

| Species | Version | Database | Reference | Code |
|---|---|---|---|---|
| *Aspergillus nidulans* fgsg a4 | verison 1 | BROAD | Galagan et al. 2005 | ANID |
| *Bipolaris maydis* C5 | AIDY00000000.1 | NCBI | Ohm et al. 2012, Condon et al. 2013 | BIMA |
| *Botrytis cinerea* B04.10 | verison 1 | BROAD | Amselem et al. 2011 | BOTC |
| *C. chlorophyti* NTL11 | MPGH00000000.1 | NCBI | Gan et al. 2017 | CCHL |
| *C. fioriniae* PJ7 | JARH00000000.1 | NCBI | Baroncelli et al. 2014 | CFIO |
| *C. fructicola* Naragc5 | ANPB00000000.1 | NCBI | Gan et al. 2013 | CFRU |
| *C. graminicola* M1.001 | Colgr1 | JGI | O'Connell et al. 2012 | CGRA |
| *C. higginsianum* IMI 349063 | Colhig2 | JGI | Zambounis et al. 2016 | CIMI |
| *C. higginsianum* MAFF 305635-RFP | MWPZ00000000.1 | NCBI | This study | CMAF |
| *C. incanum* MAFF 238712 | JTLR00000000.1 | NCBI | Gan et al. 2016 | CINC |
| *C. orbiculare* MAFF 240422 | AMCV00000000.1 | NCBI | Gan et al. 2013 | CORB |
| *Chaetomium globosum* CBS 148.51 | Chagl_1 | JGI | Berka et al. 2011 | CHGL |
| *Eutypa lata* UCR-EL1 | Eutla1 | JGI | Blanco-Ulate et al. 2013 | EUTL |
| *Fusarium graminearum* PH-1 | verison 3 | BROAD | Cuomo et al. 2007 | FUGR |
| *Fusarium oxysporum* f. sp. *lycopersici* 4287 | verison 2 | BROAD | Ma et al. 2010 | FUOX |
| *Leptosphaeria maculans* | Lepmu1 | JGI | Rouxel et al. 2011 | LEPM |
| *Magnaporthe oryzae* 70-15 | version 8 | BROAD | Okagaki et al. 2015 | MGOR |
| *Metarhizium robertsii* ARSEF 23 | Metro1 | JGI | Hu et al. 2014 | METR |
| *Nectria haematococca* MPV1, strain 77-13-4 | Necha2 | JGI | Coleman et al. 2009 | NECH |
| *Neurospora crassa* or74a | verison 10 | BROAD | Galagan et al. 2003 | NCRA |
| *Podospora anserina* S mat+ | Podan2 | JGI | Espagne̅ et al. 2008 | PODA |
| *Saccharomyces cerevisiae* | M3707 | JGI | Brown et al. 2013 | SCER |
| *Sclerotinia sclerotiorum* | verison 2 | BROAD | Amselem et al. 2011 | SCSC |
| *Trichoderma virens* Gv29-8 | ABDF00000000.2 | NCBI | Kubicek et al. 2011 | TRIV |
| *Verticillium dahliae* VdLs.17 | version 1 | BROAD | Klosterman et al. 2011 | VDAH |

Table 3-4. Reciprocal best hits of genes on IMI 349063 minichromosomes to genes of MAFF 305635-RFP.

| MAFF 305635-RFP | | | | | IMI 349063 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| gene ID | *E*-val | q-cov | contig | annotation | gene ID | *E*-val | q-cov | chr | annotation |
| CH35J_012468* | 6.53E-124 | 100 | contig_13 | hypothetical protein | CH63R_14381* | 1.3E-42 | 82 | chr11 | Nudix domain-containing protein |
| CH35J_011880* | 1.8E-62 | 100 | contig_11 | hypothetical protein | CH63R_14383 | 2.68E-17 | 81 | chr11 | ChEC7a |
| CH35J_006565 | 3.94E-125 | 73 | contig_5 | hypothetical protein | CH63R_14408 | 5.5E-125 | 52 | chr11 | hypothetical protein |
| CH35J_006562 | 3.71E-66 | 90 | contig_5 | hypothetical protein | CH63R_14415 | 2.12E-67 | 91 | chr11 | hypothetical protein |
| CH35J_012912 | 4.69E-18 | 58 | contig_23 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_012903 | 4.69E-18 | 58 | contig_21 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_012869 | 4.69E-18 | 58 | contig_16 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_012760 | 4.69E-18 | 58 | contig_14 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_012270 | 4.69E-18 | 58 | contig_12 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_007773 | 4.69E-18 | 58 | contig_6 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_007768 | 4.69E-18 | 58 | contig_5 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_006538 | 4.69E-18 | 58 | contig_4 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_004900 | 4.69E-18 | 58 | contig_3 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_003324 | 4.69E-18 | 58 | contig_3 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_003321 | 4.69E-18 | 58 | contig_2 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_001732 | 4.69E-18 | 58 | contig_2 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_001725 | 4.69E-18 | 58 | contig_1 | hypothetical protein | CH63R_14417 | 1.6E-19 | 90 | chr11 | hypothetical protein |
| CH35J_009746 | 6.27E-76 | 64 | contig_7 | Vegetative incompatibility protein HET-E-1 | CH63R_14428 | 1.07E-75 | 73 | chr11 | Pfs domain-containing protein |
| CH35J_012849 | 7.95E-64 | 100 | contig_16 | hypothetical protein | CH63R_14432 | 6.08E-64 | 100 | chr11 | hypothetical protein |
| CH35J_012848 | 5.34E-31 | 71 | contig_16 | hypothetical protein | CH63R_14448 | 6.14E-31 | 76 | chr11 | hypothetical protein |
| CH35J_012861 | 0 | 98 | contig_16 | hypothetical protein | CH63R_14450 | 0 | 97 | chr11 | hypothetical protein |
| CH35J_012846 | 1.59E-122 | 100 | contig_16 | hypothetical protein | CH63R_14454 | 1.75E-122 | 100 | chr11 | hypothetical protein |
| CH35J_012845 | 0 | 85 | contig_16 | hypothetical protein | CH63R_14455 | 0 | 100 | chr11 | DNA repair helicase |
| CH35J_007405 | 2.89E-171 | 58 | contig_5 | hypothetical protein | CH63R_14460 | 3.72E-177 | 77 | chr11 | Aurora kinase 2 splicing |
| CH35J_012850 | 0 | 93 | contig_16 | hypothetical protein | CH63R_14463 | 0 | 99 | chr11 | Tetratricopeptide repeat domain containing protein |
| CH35J_012847 | 1.56E-81 | 93 | contig_16 | hypothetical protein | CH63R_14465 | 1.68E-81 | 86 | chr11 | hypothetical protein |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CH35J_012866 | 2.17E-100 | 99 | contig_16 | hypothetical protein | CH63R_14467 | 3.81E-100 | 56 | chr11 | hypothetical protein |
| CH35J_000492 | 5.59E-129 | 91 | contig_1 | hypothetical protein | CH63R_14507 | 8.98E-129 | 99 | chr11 | ChEC21 |
| CH35J_012918 | 0 | 100 | contig_25 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_012000 | 0 | 100 | contig_12 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_011916 | 0 | 100 | contig_11 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_011838 | 0 | 100 | contig_11 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_011452 | 0 | 100 | contig_10 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_009008 | 0 | 100 | contig_6 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_007767 | 0 | 100 | contig_5 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_007715 | 0 | 100 | contig_5 | Retrotransposon-derived protein PEG10 | CH63R_14584 | 0 | 100 | chr12 | Gag protein |
| CH35J_010982 | 0 | 67 | contig_9 | hypothetical protein | CH63R_14647 | 0 | 99 | chr12 | Ank-repeat protein mbp1 |
| CH35J_003326 | 0 | 100 | contig_3 | hypothetical protein | CH63R_14651 | 0 | 100 | chr12 | hypothetical protein |

* Effector candidate

Table 3-5. Highly variable effector candidate genes of *C. higginsianum* MAFF 305635-RFP and IMI 349063.

| gene ID | contig/chr | AA length | annotation | category |
|---|---|---|---|---|
| **MAFF 305635-RFP** | | | | |
| CH35J_001046 | contig_1 | 132 | hypothetical protein | frameshift |
| CH35J_002132 | contig_2 | 214 | hypothetical protein | frameshift |
| CH35J_003317 | contig_2 | 87 | hypothetical protein | ≤ 90% query coverage |
| CH35J_003318 | contig_2 | 120 | hypothetical protein | presence/absence |
| CH35J_007515 | contig_5 | 167 | hypothetical protein | presence/absence |
| CH35J_007516 | contig_5 | 173 | hypothetical protein | presence/absence |
| CH35J_010999 | contig_9 | 262 | hypothetical protein | presence/absence |
| CH35J_011924 | contig_11 | 267 | Xyloglucan-specific endo-beta-1,4-glucanase A | presence/absence |
| **IMI 349063** | | | | |
| CH63R_04687 | chr3 | 246 | Cupin domain | frameshift |
| CH63R_05497 | chr4 | 298 | cellulose binding domain-containing | ≤ 90% query coverage |
| CH63R_06433 | chr4 | 230 | Phosphatidylethanolamine-binding protein | frameshift |
| CH63R_08807 | chr6 | 77 | hypothetical protein | ≤ 90% query coverage |
| CH63R_09232 | chr6 | 255 | hypothetical protein | frameshift |
| CH63R_09755 | chr7 | 187 | hypothetical protein | ≤ 90% query coverage |
| CH63R_09757 | chr7 | 168 | gram-positive signal ysirk family | ≤ 90% query coverage |
| CH63R_12049 | chr8 | 230 | Acetylxylan esterase | frameshift |
| CH63R_12296 | chr9 | 288 | Alpha-tubulin suppressor | frameshift |
| CH63R_14384 | chr11 | 73 | hypothetical protein | presence/absence |
| CH63R_14389 | chr11 | 94 | ChEC12a | presence/absence |
| CH63R_14393 | chr11 | 157 | Nudix domain-containing protein | ≤ 90% query coverage |
| CH63R_14470 | chr11 | 291 | hypothetical protein | presence/absence |
| CH63R_14516 | chr11 | 270 | ChEC12 | ≤ 90% query coverage |
| CH63R_14558 | chr12 | 79 | hypothetical protein | ≤ 90% query coverage |
| CH63R_14618 | chr12 | 245 | Fungal specific transcripton factor | presence/absence |
| CH63R_14632 | chr12 | 205 | hypothetical protein | ≤ 90% query coverage |
| CH63R_14648 | chr12 | 173 | ChEC51a | presence/absence |

Table 3-6. Identified sequences related to highly variable effector candidate genes showing ≤ 90% coverage in the other genome.

| Query | | | | Reference | | | | |
|---|---|---|---|---|---|---|---|---|
| gene ID | strain | contig/chr | Annotation | contig/chr | strain | start | end | coding gene |
| CH35J_003317 | MAFF 305635-RFP | contig_2 | hypothetical protein | chr1 | IMI 349063 | 21538 | 21724 | N.A. |
| CH63R_05497 | IMI 349063 | chr4 | cellulose binding domain-containing | contig_3 | MAFF 305635-RFP | 1024678<br>1122651<br>1275187 | 1024729<br>1122711<br>1275892 | N.A.<br>CH35J_003617<br>CH35J_003656 |
| CH63R_08807 | IMI 349063 | chr6 | hypothetical protein | contig_8 | MAFF 305635-RFP | 1295130 | 1294924 | N.A. |
| CH63R_09755 | IMI 349063 | chr7 | hypothetical protein | contig_11 | MAFF 305635-RFP | 1388393 | 1388659 | N.A. |
| CH63R_09757 | IMI 349063 | chr7 | gram-positive signal ysirk family | contig_11 | MAFF 305635-RFP | 1384799 | 1385210 | CH35J_011876 |
| CH63R_14393 | IMI 349063 | chr11 | Nudix domain-containing protein | contig_13 | MAFF 305635-RFP | 833931 | 834329 | CH35J_012468 |
| CH63R_14516 | IMI 349063 | chr11 | ChEC12 | contig_5 | MAFF 305635-RFP | 264402 | 264812 | N.A. |
| CH63R_14558 | IMI 349063 | chr12 | hypothetical protein | contig_3 | MAFF 305635-RFP | 5874120 | 5874326 | CH35J_004872 |
| CH63R_14632 | IMI 349063 | chr12 | hypothetical protein | contig_8 | MAFF 305635-RFP | 397550<br>568029<br>735318 | 397686<br>570087<br>735170 | CH35J_010020<br>CH35J_010070<br>CH35J_010117 |

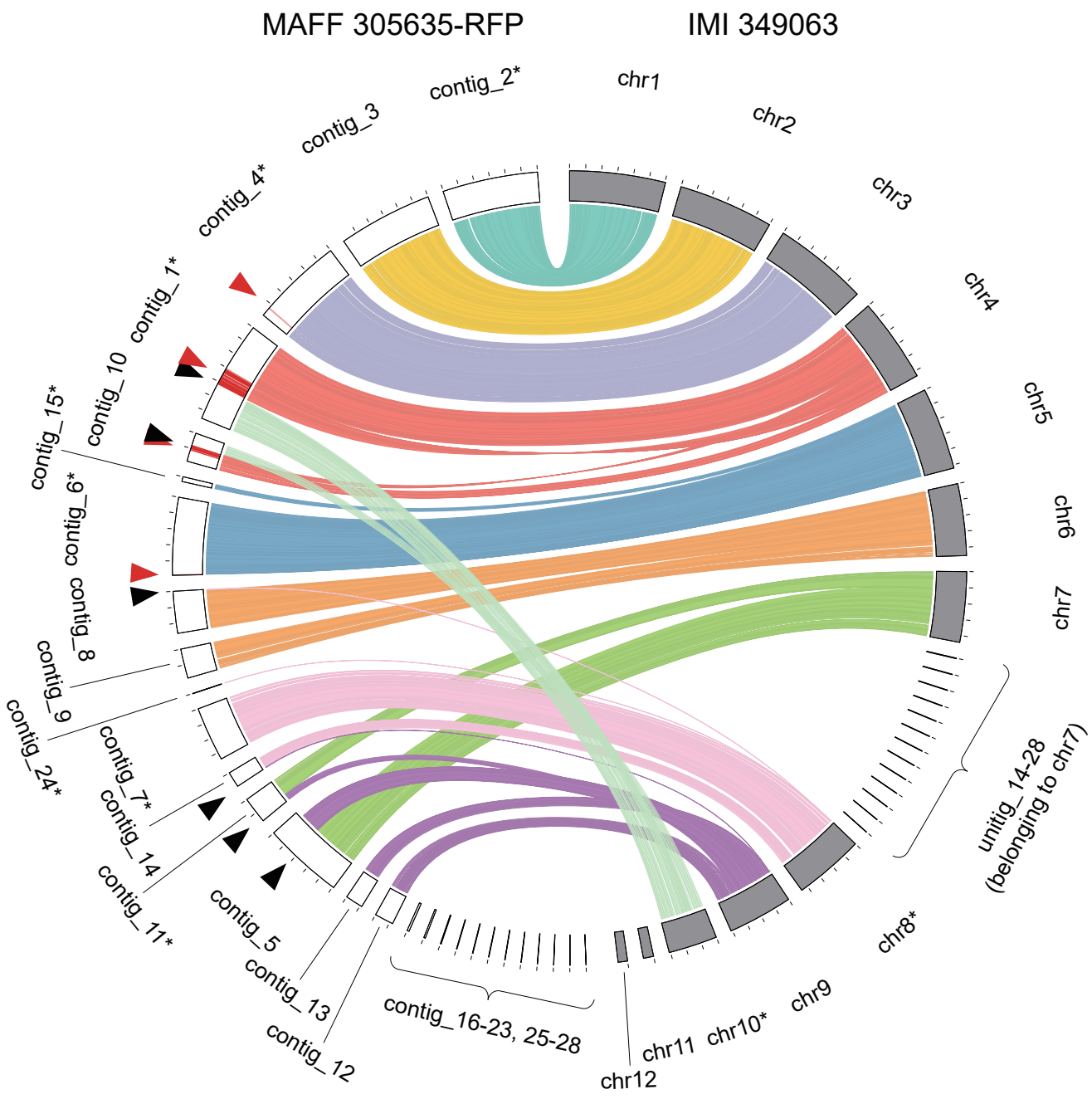



Fig. 3-1. Whole genome alignments between MAFF 305635-RFP and IMI 349063. The outer bands indicate contigs of MAFF 305635-RFP (white) and chromosomes of IMI 349063 (gray), respectively. Syntenic regions (≥99 % identity, ≥15 kb) are linked with different colored ribbons corresponding to a chromosome from the genome assembly of IMI 349063. Black and red arrowheads indicate inter-chromosomal translocations and intra-chromosomal inversions, respectively. Asterisks indicate reverse-complementation of contigs or chromosomes for visual clarity. Ticks on bands represent 1 Mb.

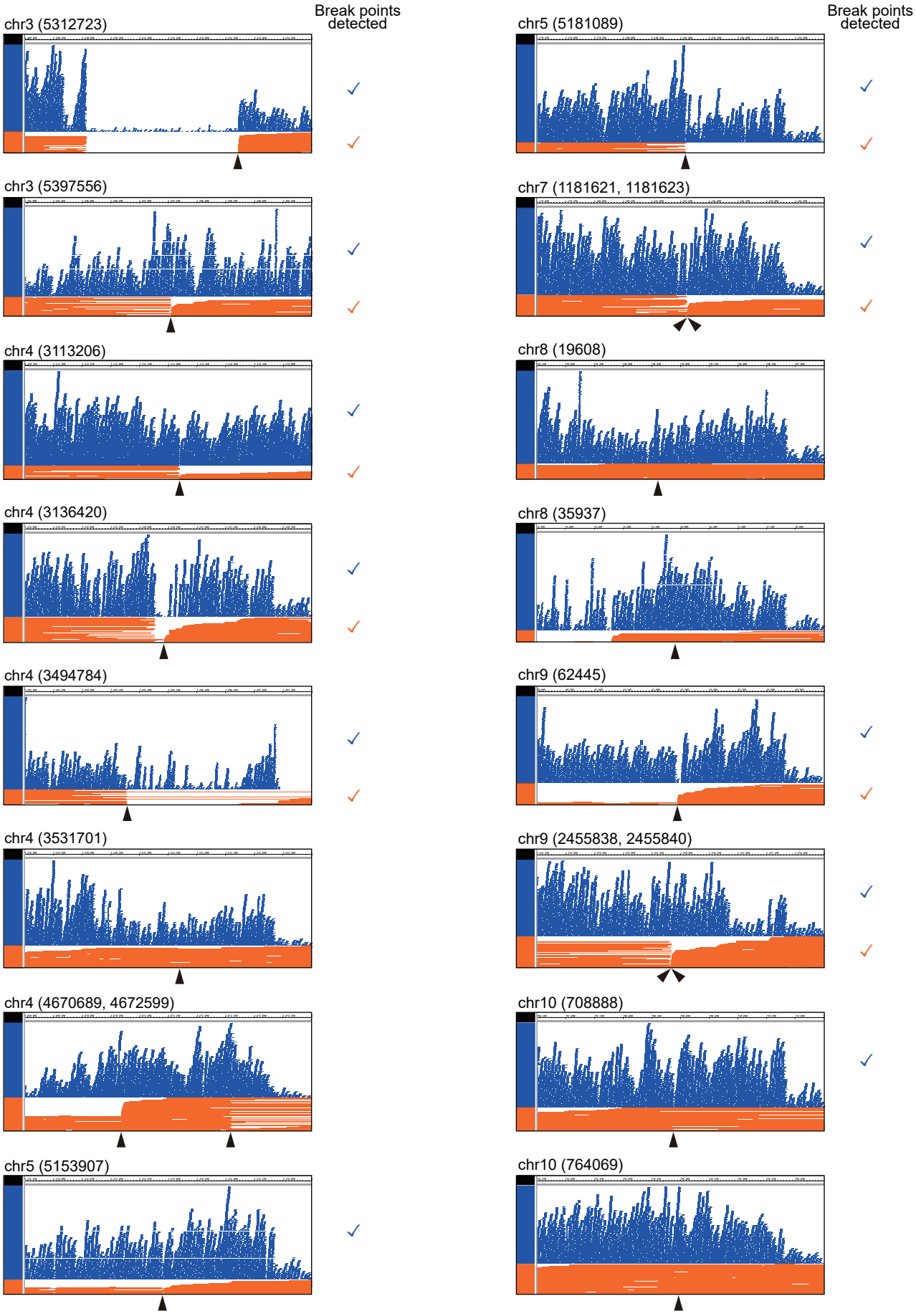



Fig. 3-2. Remapping of MAFF 305635-RFP and MAFF 305635 wild-type raw reads against the genome assembly of IMI 349063. Blue and orange lines represent mapped SOLiD reads derived from MAFF 305635 wild-type and PacBio reads derived from MAFF 305635-RFP, respectively. Arrowheads and numbers in brackets indicate locations of synteny breakpoints. Blue and orange tick signs indicate gaps in mapped SOLiD reads and PacBio reads, respectively. Genomic regions within 5 kb of synteny breakpoints are shown.

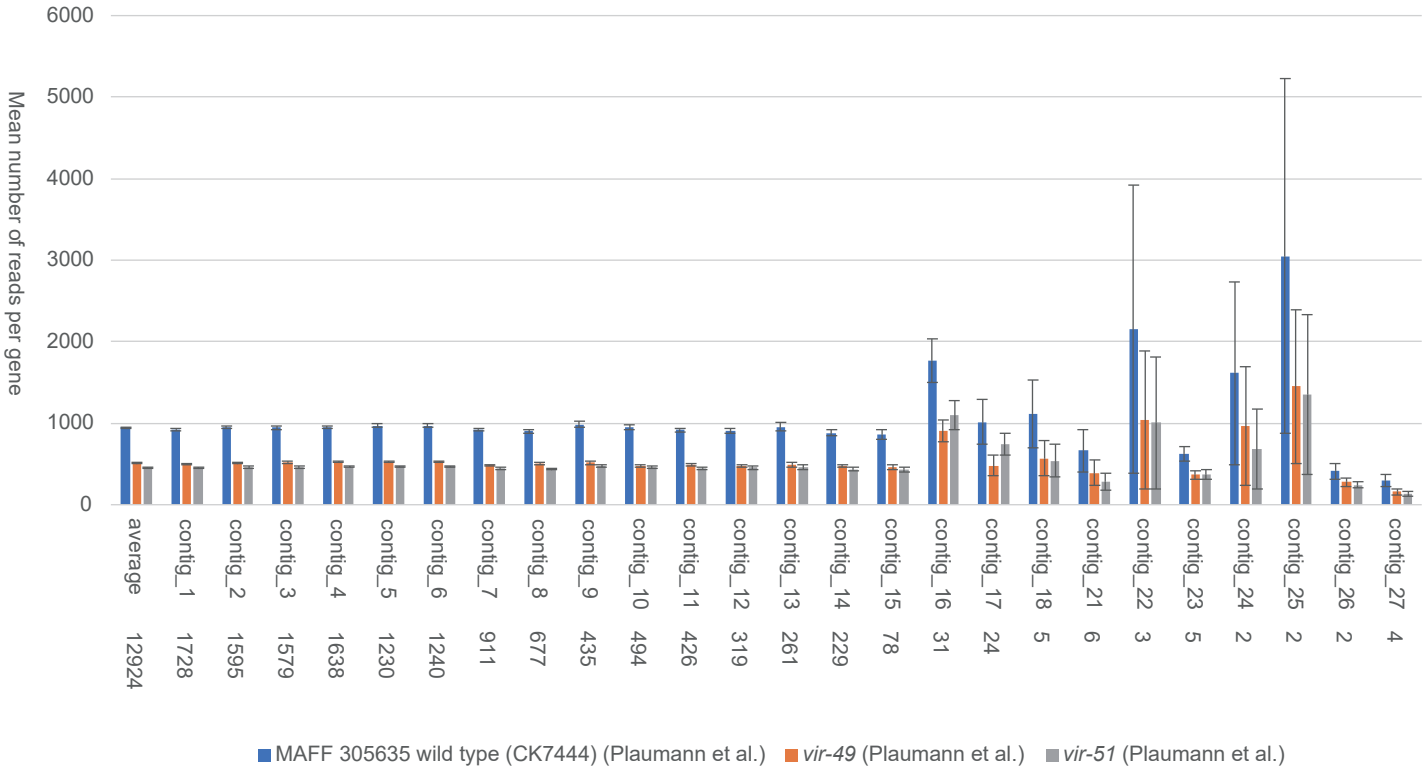



Fig. 3-3. Mean number of reads for MAFF 305635 wild-type (CK7444), *vir-49*, and *vir-51* mapping to each contig from MAFF 305635-RFP. Numbers below contig names indicate the number of genes located on the contig. Numbers below the average indicates the number of all predicted genes in MAFF 305635-RFP. Error bars represent standard errors.

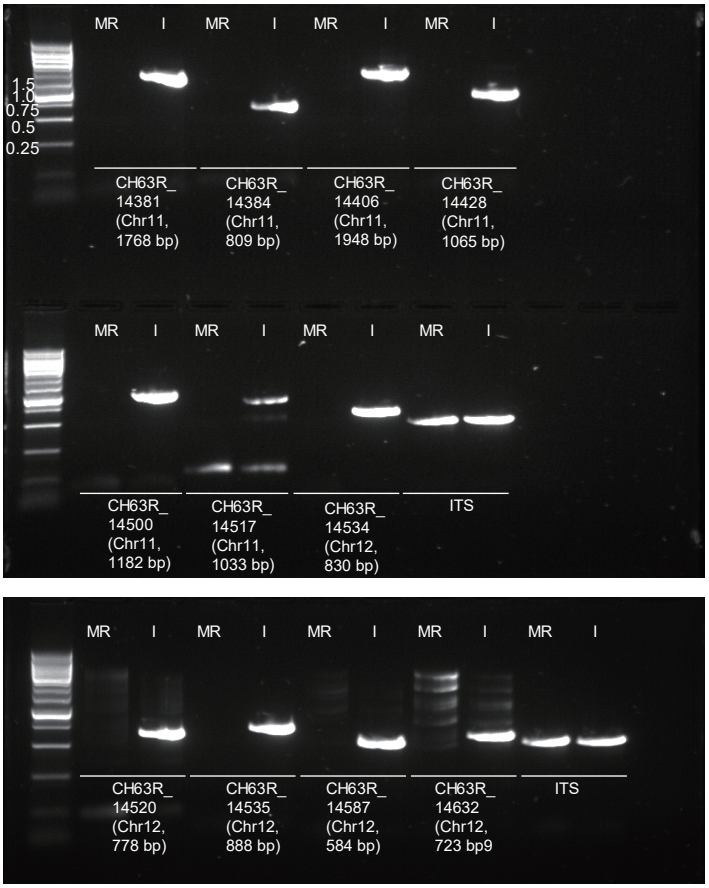



Fig. 3-4. Absence of genes on minichromosomes in MAFF 305635-RFP, assessed by PCR on genomic DNA. MR and I on lanes indicate PCR amplicons from MAFF 305635-RFP and IMI 349063, respectively. Genomic sequences of internal transcribed spacer (ITS) amplified with the primer pair, ITS-1F and ITS-4, are shown as positive controls.

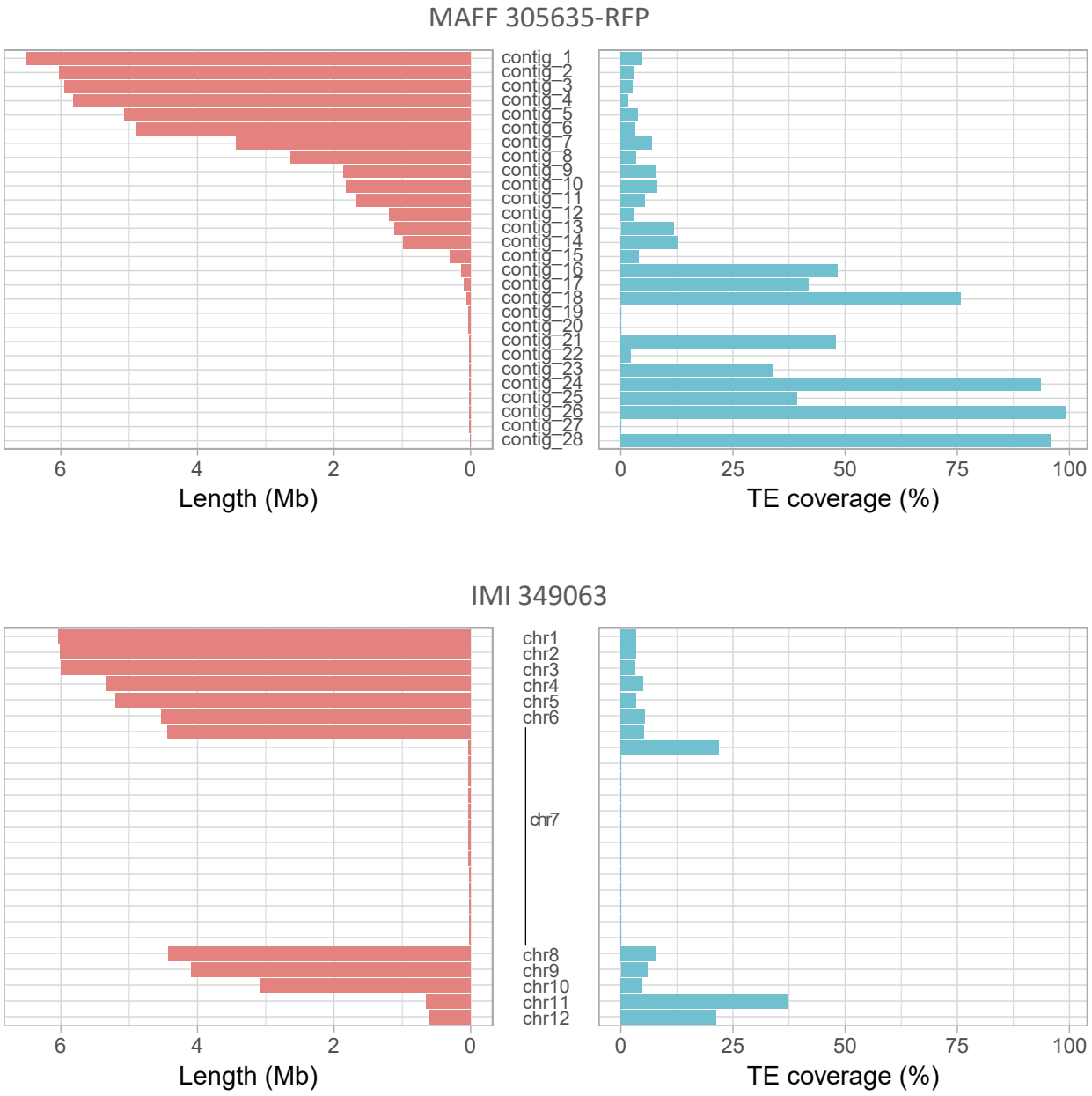



Fig. 3-5. TE coverage of contigs or chromosomes from the genome assemblies of MAFF 305635-RFP and IMI 349063. Coverage was calculated after eliminating the overlap between TEs.

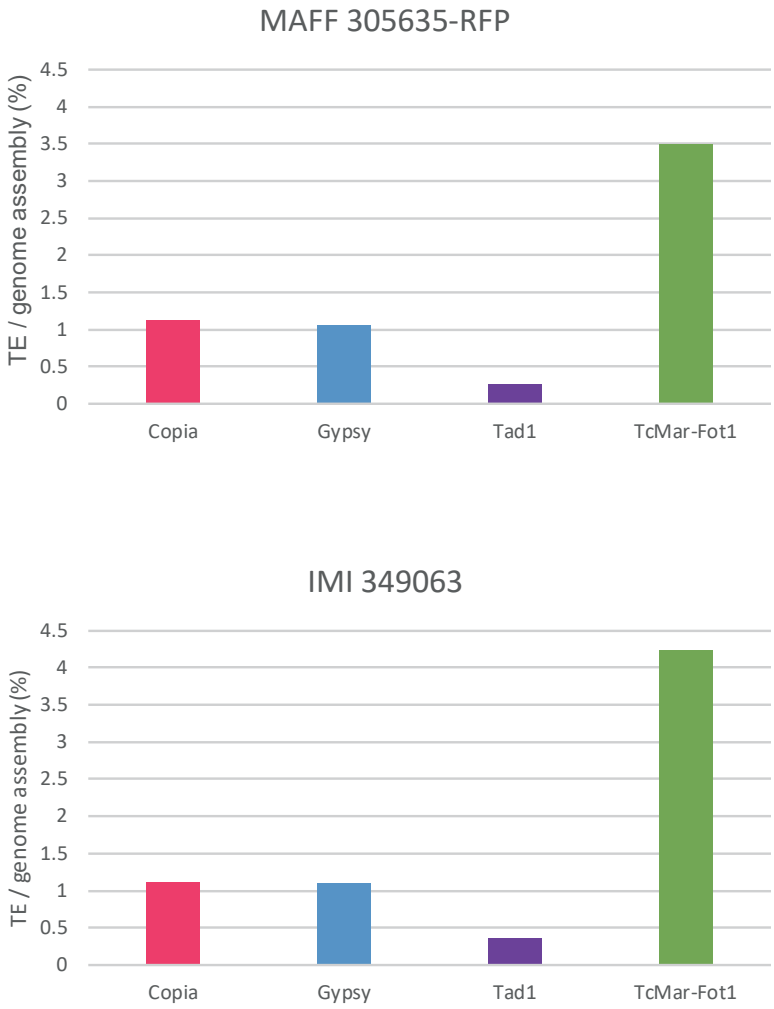



Fig. 3-6. Coverage of four types of TEs per genome assembly. The coverage of each type of TE was calculated by ignoring the overlap between TEs.

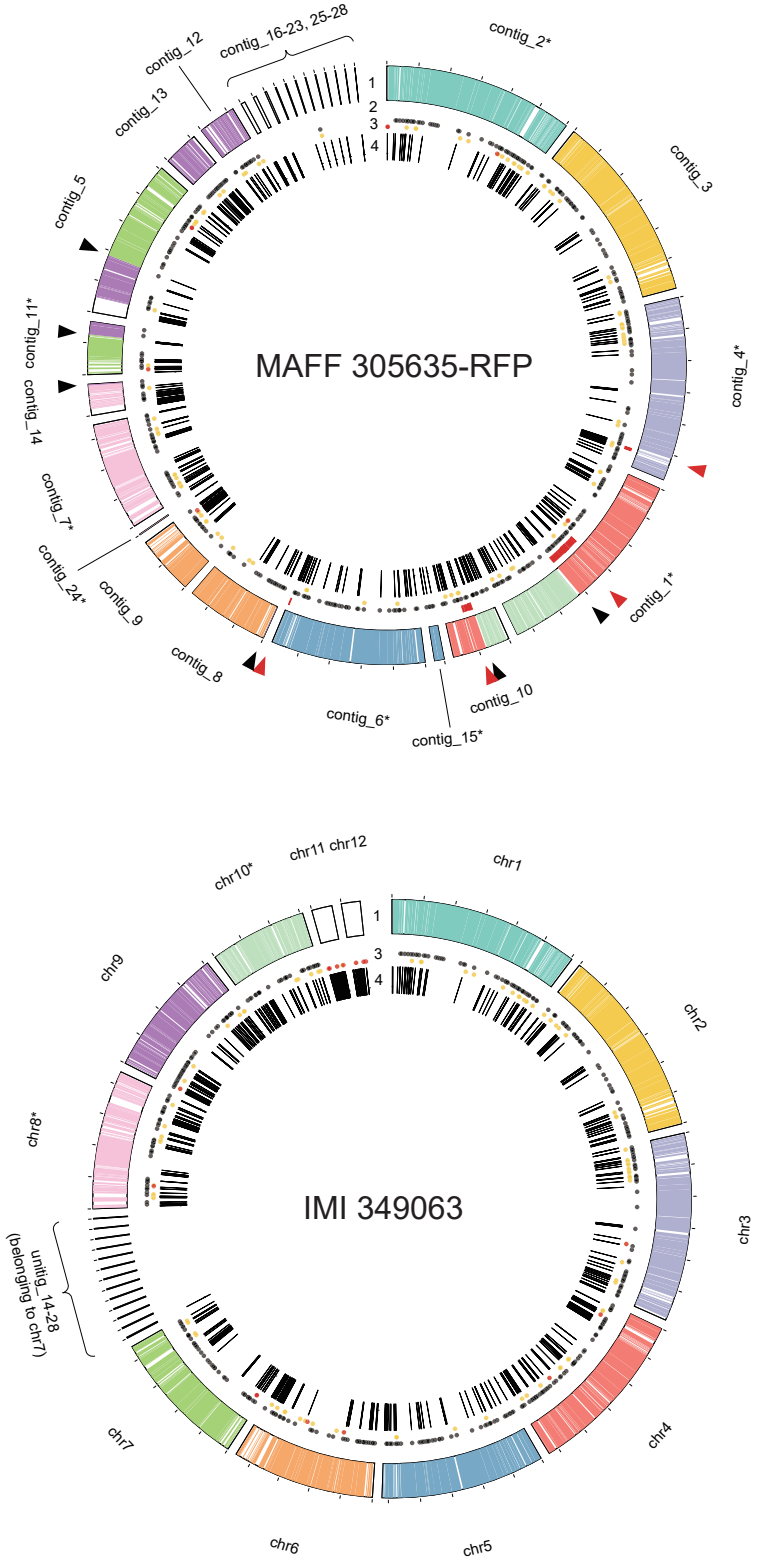



Fig. 3-7. Distribution of genomic features and TEs in *C. higginsianum*. 1: Contigs. Colored regions indicate the presence of syntenic regions in the genome assembly of IMI 349063 (≥ 99% identity, ≥ 15 kb). Colors correspond to chromosomes of IMI 349063 as in Fig. 1. 2: Intra-chromosomal inverted regions. 3: Effector candidate genes. Gray, yellow, and red circles indicate identical, polymorphic, and highly variable effector candidate genes, respectively. 4: TEs. Black and red arrowheads indicate inter-chromosomal translocations and intra-chromosomal inversions, respectively. Asterisks indicate reverse-complementation of contigs for visual clarity in Fig. 1. Ticks on the outer bands represent 1 Mb.


Fig. 3-8. Association between large-scale rearrangements and TEs. Numbers in brackets indicate the location of large-scale rearrangements. Synteny: Rectangles indicate regions with synteny to IMI 349063 (≥ 99% identity, ≥ 15 kb). Colors filling rectangles correspond to chromosomes of IMI 349063 as in Fig. 1. The red line surrounding a rectangle indicates a region of intra-chromosomal inversion. Asterisks represent large-scale rearrangements within 10 kb of the nearest TE.

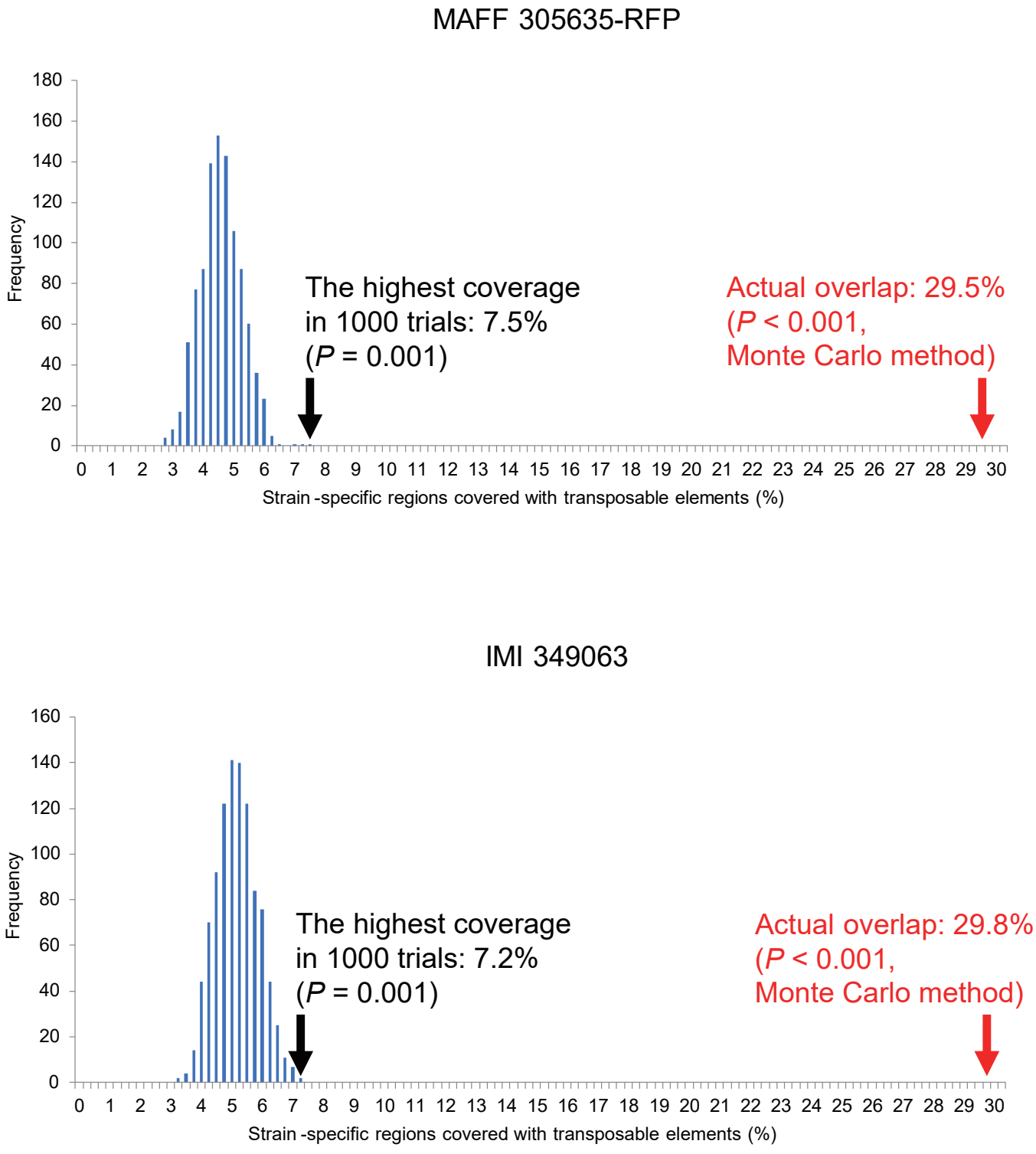
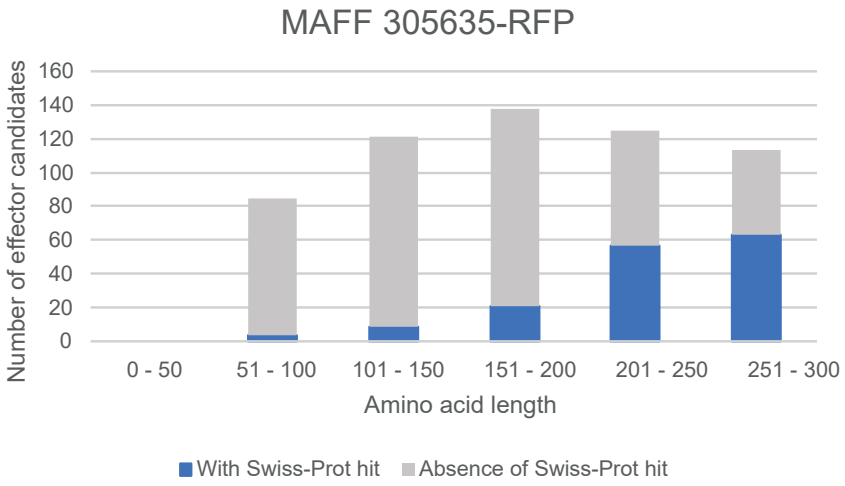



Fig. 3-9. Strain-specific regions showed significant overlap with TEs. Blue columns represent frequency of overlaps between strain-specific regions and TEs in 1,000 trials if TEs were randomly distributed. The observed percentages of overlap between strain-specific regions and TEs are indicated with red arrows.



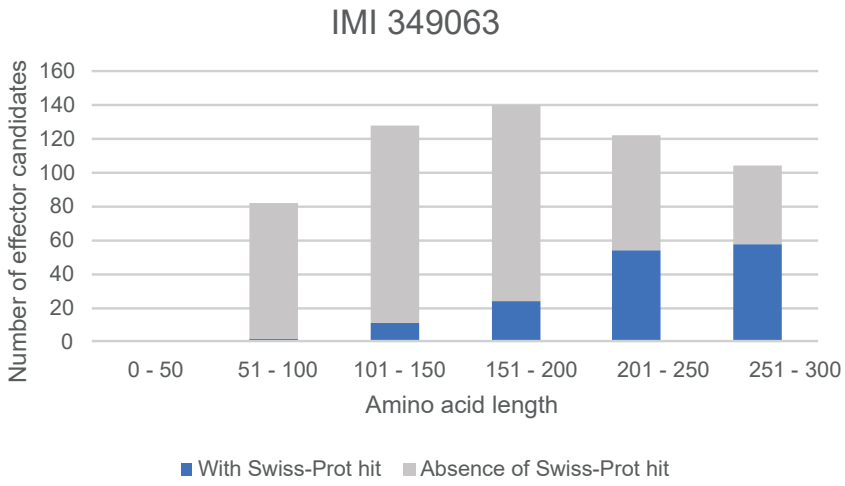




Fig. 3-10. Presence/absence of similarity to sequences in the Swiss-Prot database among effector candidates (BLASTP cutoff *E*-value = $10^{-5}$).

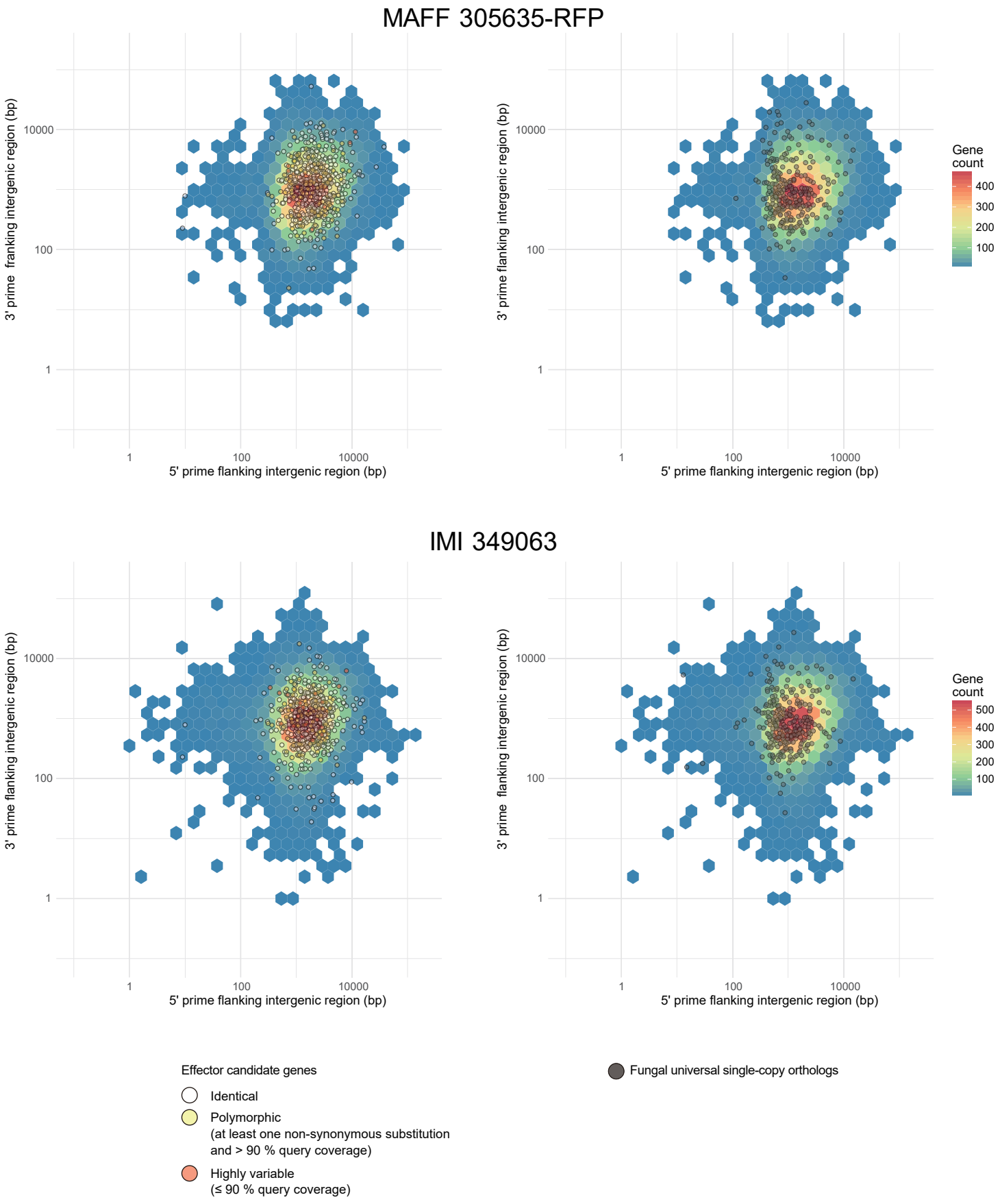



Fig. 3-11. Two-dimensional plots describing 5′ and 3′ flanking intergenic regions of effector candidate genes and fungal universal single-copy orthologs. Left and right plots represent effector candidate genes and fungal universal single-copy orthologs, respectively.

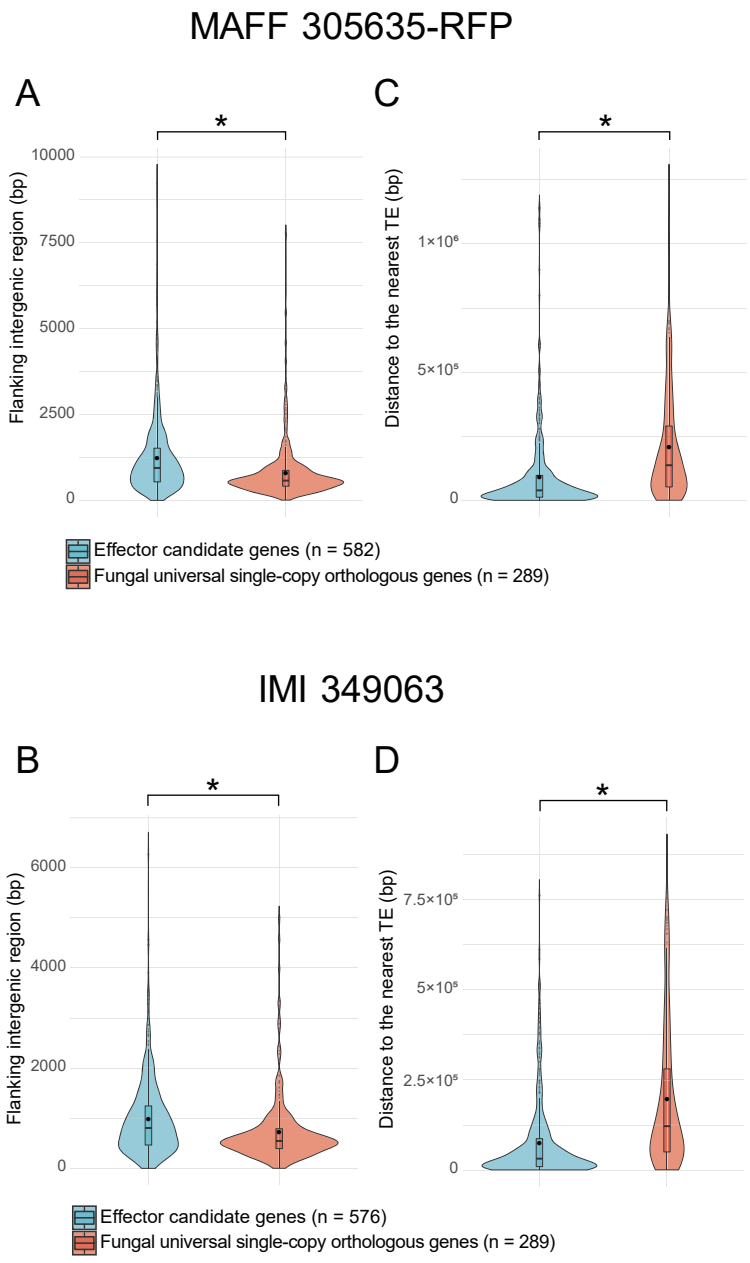



Fig. 3-12. Compartmentalization of effector candidate genes and fungal universal single-copy orthologs. (A, B) Violin plots showing flanking intergenic regions of effector candidate genes and fungal universal single-copy orthologs. (C, D) Violin plots showing distances from effector candidate genes and fungal universal single-copy orthologs to their nearest TEs. Black bars and circles inside violin plots represent the median and mean of each distribution, respectively. Asterisks represent significant differences between effector candidate genes and fungal universal single-copy orthologs ($P < 0.01$, Wilcoxon rank sum test).

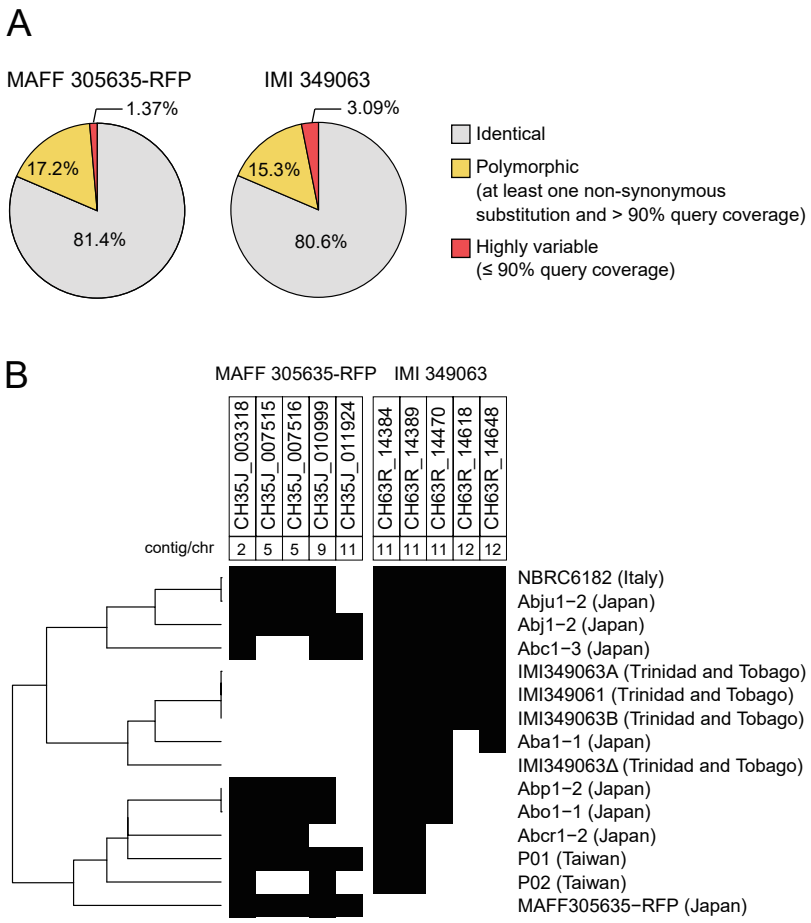



Fig. 3-13. Variations in effector candidate genes between *C. higginsianum* strains. (A) Pie charts showing percentages of effector candidates with different levels of variations. Gray, yellow, and red indicate identical, polymorphic (having at least one non-synonymous substitution and > 90% query coverage), and highly variable (≤ 90% query coverage) effector candidates, respectively. (B) The presence/absence patterns of ten highly variable effector candidate genes among 15 *C. higginsianum* strains. Black and white squares indicate the presence and absence of these genes, respectively. The dendrogram shows hierarchical clusters of 15 strains based on the presence/absence patterns of ten highly variable effector candidate genes.

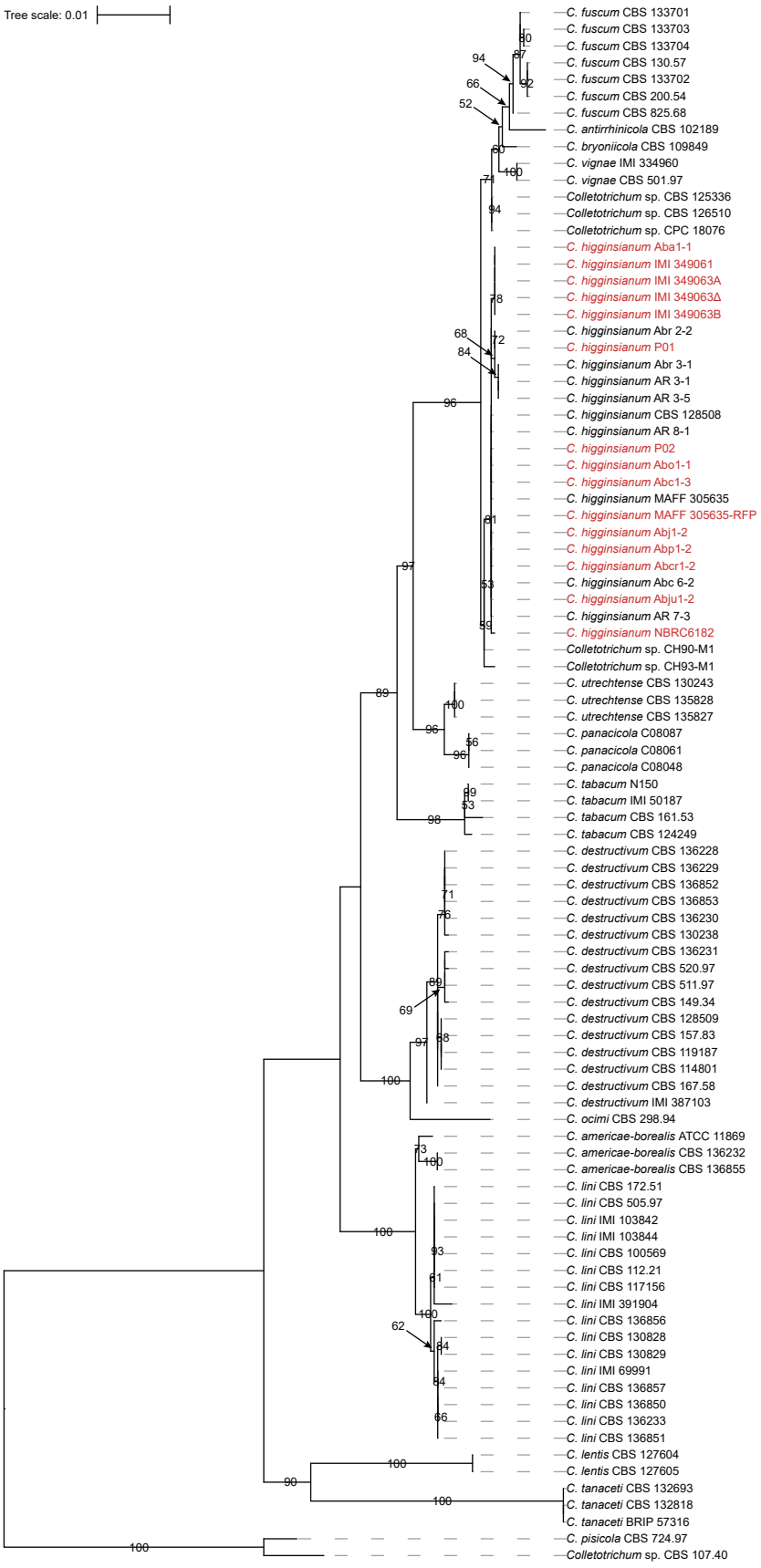



Fig. 3-14. Maximum-likelihood tree showing relationship between 15 different *C. higginsianum* strains with other known species in the *C. destructivum* species complex based on the combined alignment of *actin* (*ACT*), *chitin synthase I* (*CHS-1*), *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*), *histone H3* (*HIS3*), internal transcribed spacer (ITS), and *tubulin-2* (*TUB2*). Red: *C. higginsianum* strains used in this study. Values at the branch points represent percentages of bootstrap support values of 1,000 replicates.

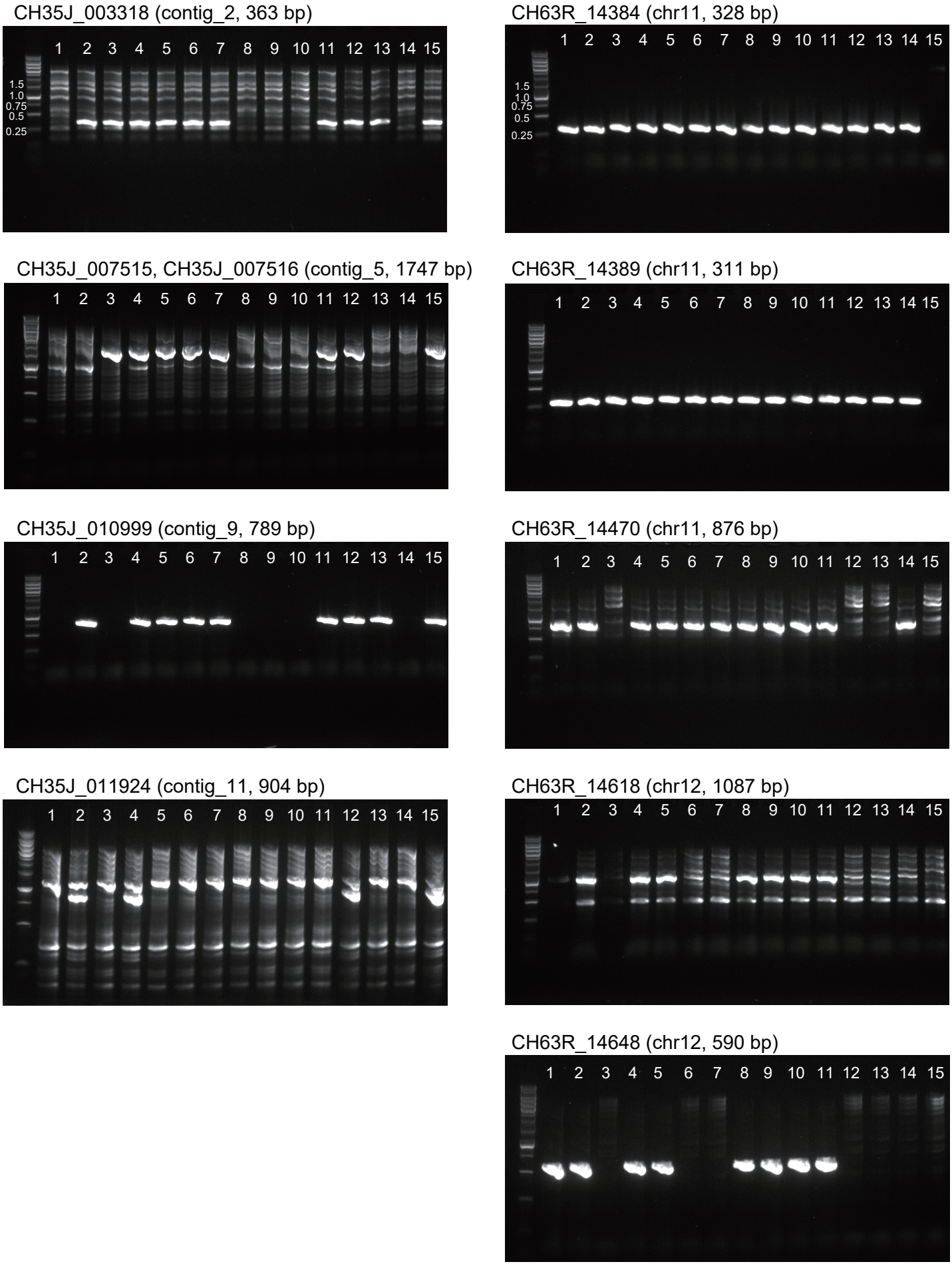



Fig. 3-15. Detection of highly variable effector candidate genes with presence/absence polymorphisms by performing PCR. Numbers on lanes indicate genomic DNAs from each *C. higginsianum* strains. 1: Aba1-1, 2: Abc1-3, 3: Abcr1-2, 4: Abj1-2, 5: Abju1-2, 6: Abo1-1, 7: Abp1-2, 8: IMI 349061, 9: IMI 349063A, 10: IMI 349063B, 11: NBRC 6182, 12: P01, 13: P02, 14: IMI 349063Δ, 15: MAFF 305635-RFP

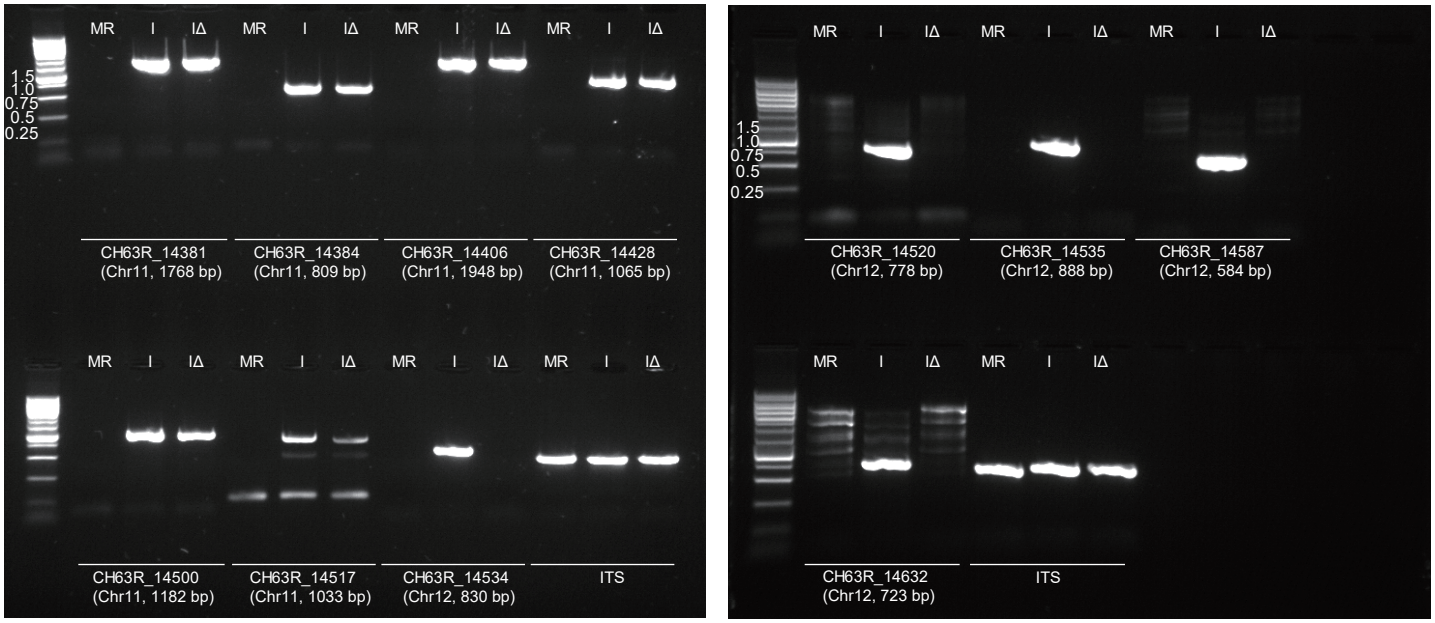



Fig. 3-16. Absence of chromosome 12-encoding genes in IMI 349063Δ, assessed by PCR on genomic DNA. MR, I and IΔ on lanes indicate genomic DNAs from MAFF 305635-RFP, IMI 349063, and IMI 349063Δ, respectively. Genomic sequences of internal transcribed spacer (ITS) amplified with the primer pair ITS-1F and ITS-4 are shown as positive controls.

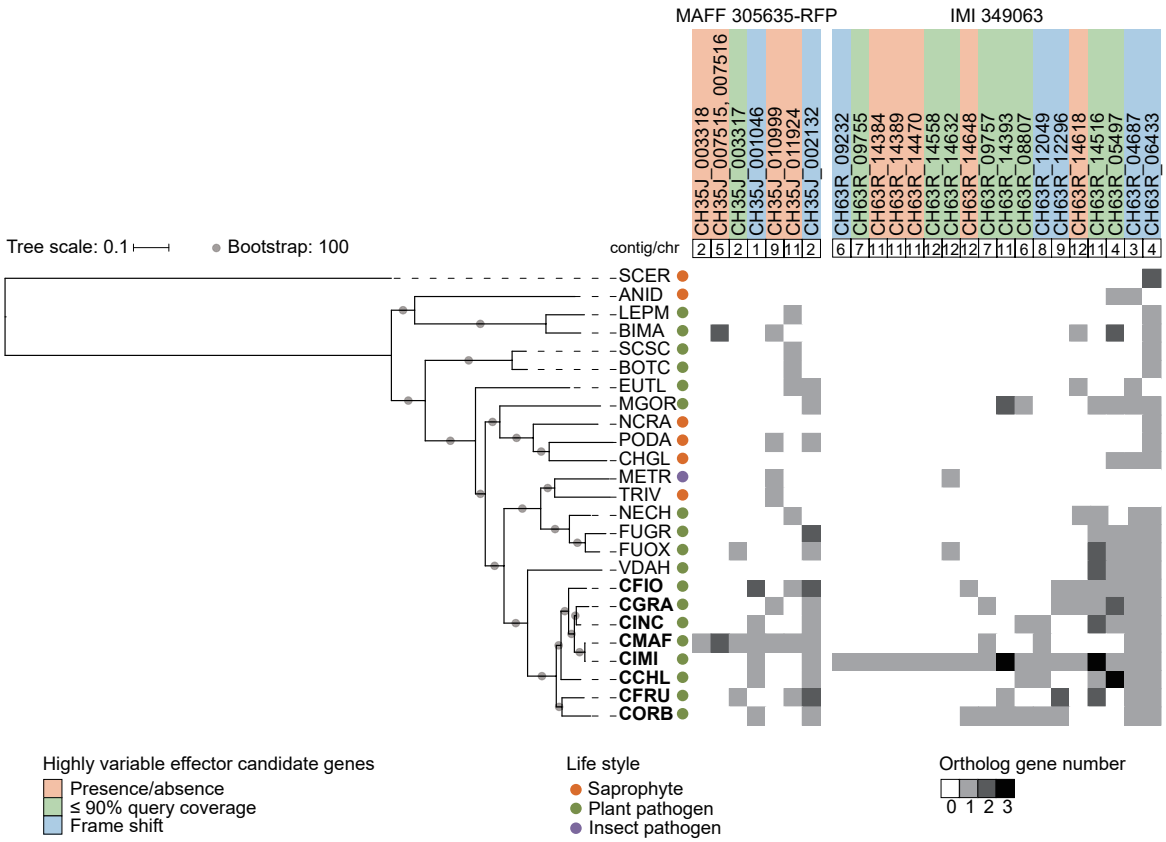



Fig. 3-17. Conservation patterns of highly variable effector candidate genes in Ascomycota. Bold letters indicate *Colletotrichum* fungi. The maximum-likelihood species phylogeny was drawn based on the alignment patterns of single-copy orthologs obtained using OrthoMCL. Bootstrap values are percentages based on 1,000 bootstrap replicates. ANID: *Aspergillus nidulans*, BIMA: *Bipolaris maydis*, BOTC: *Botrytis cinerea*, CCHL: *C. chlorophyti*, CFIO: *C. fioriniae*, CFRU: *C. fructicola*, CGRA: *C. graminicola*, CIMI: *C. higginsianum* IMI 349063, CMAF: *C. higginsianum* MAFF 305635-RFP, CINC: *C. incanum*, CORB: *C. orbiculare*, CHGL: *Chaetomium globosum*, EUTL: *Eutypa lata*, FUGR: *Fusarium graminearum*, FUOX: *Fusarium oxysporum* f. sp. *lycopersici*, LEPM: *Leptosphaeria maculans*, MGOR: *Magnaporthe oryzae*, METR: *Metarhizium robertsii*, NECH: *Nectria haematococca*, NCRA: *Neurospora crassa*, PODA: *Podospora anserine*, SCER: *Saccharomyces cerevisiae*, SCSC: *Sclerotinia sclerotiorum*, TRIV: *Trichoderma virens*, VDAH: *Verticillium dahliae*

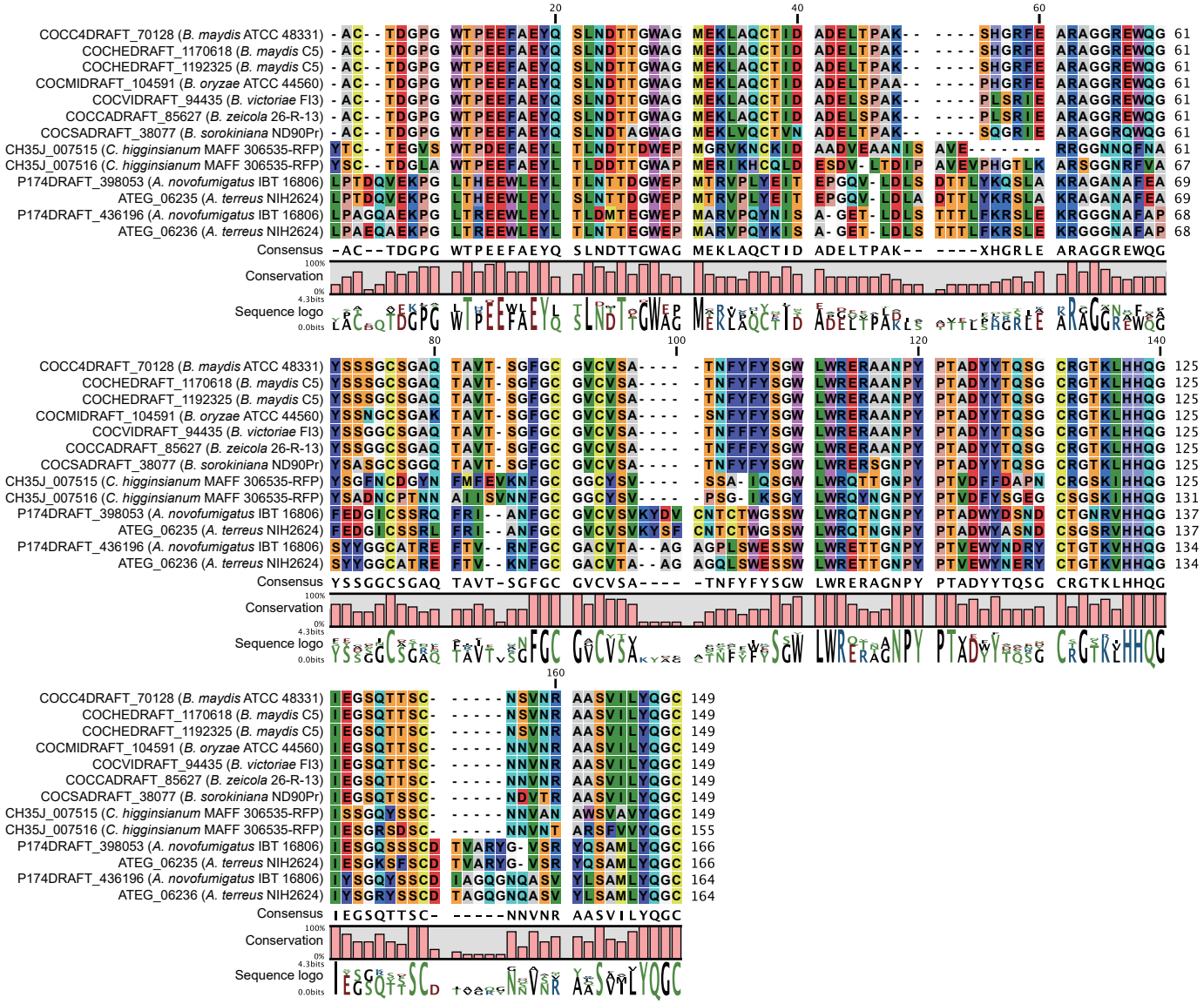



Fig. 3-18. Amino acid sequence alignments of CH35J_007515 and CH35J_007516 and their homologs without their predicted signal peptides.

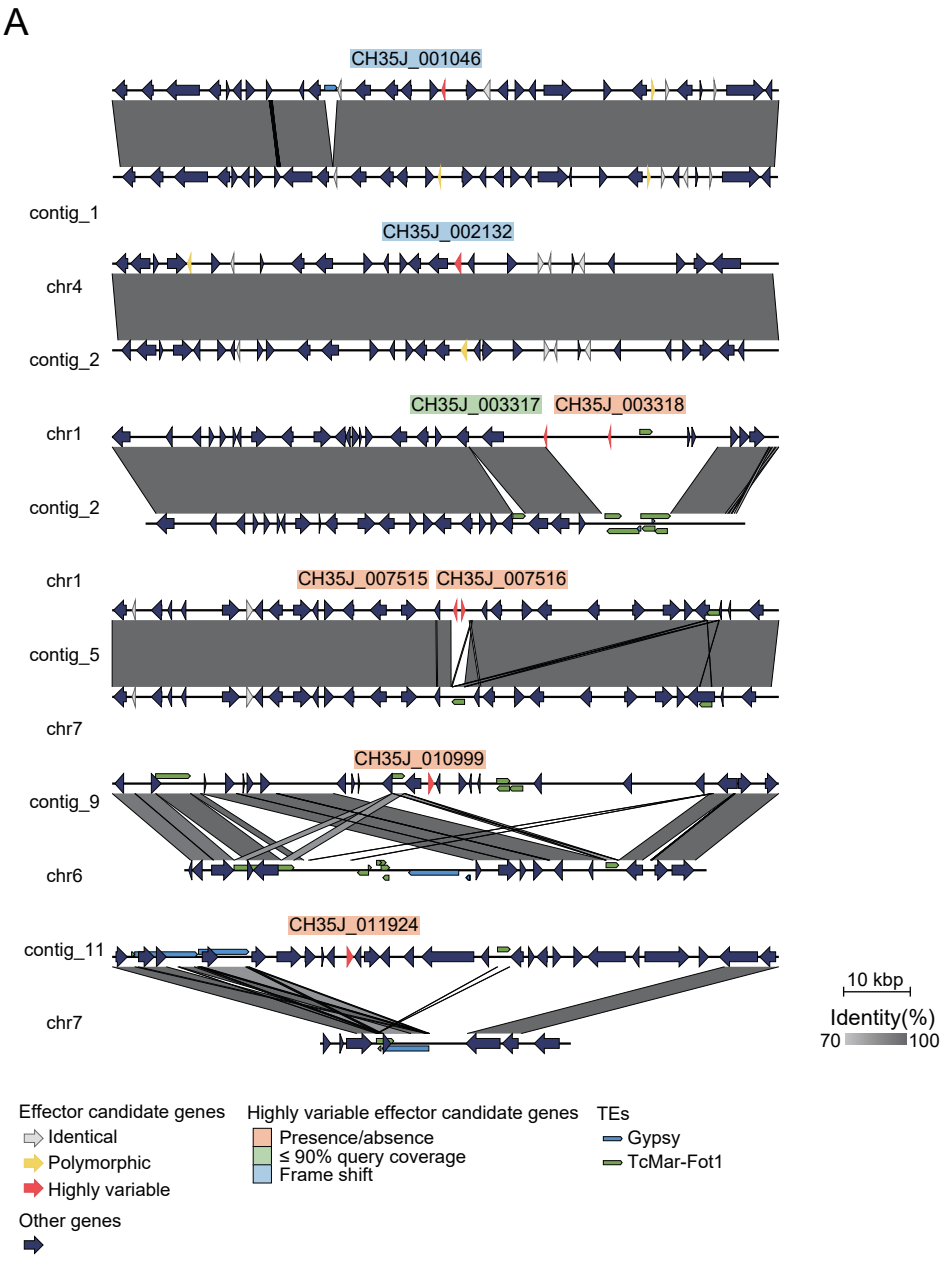
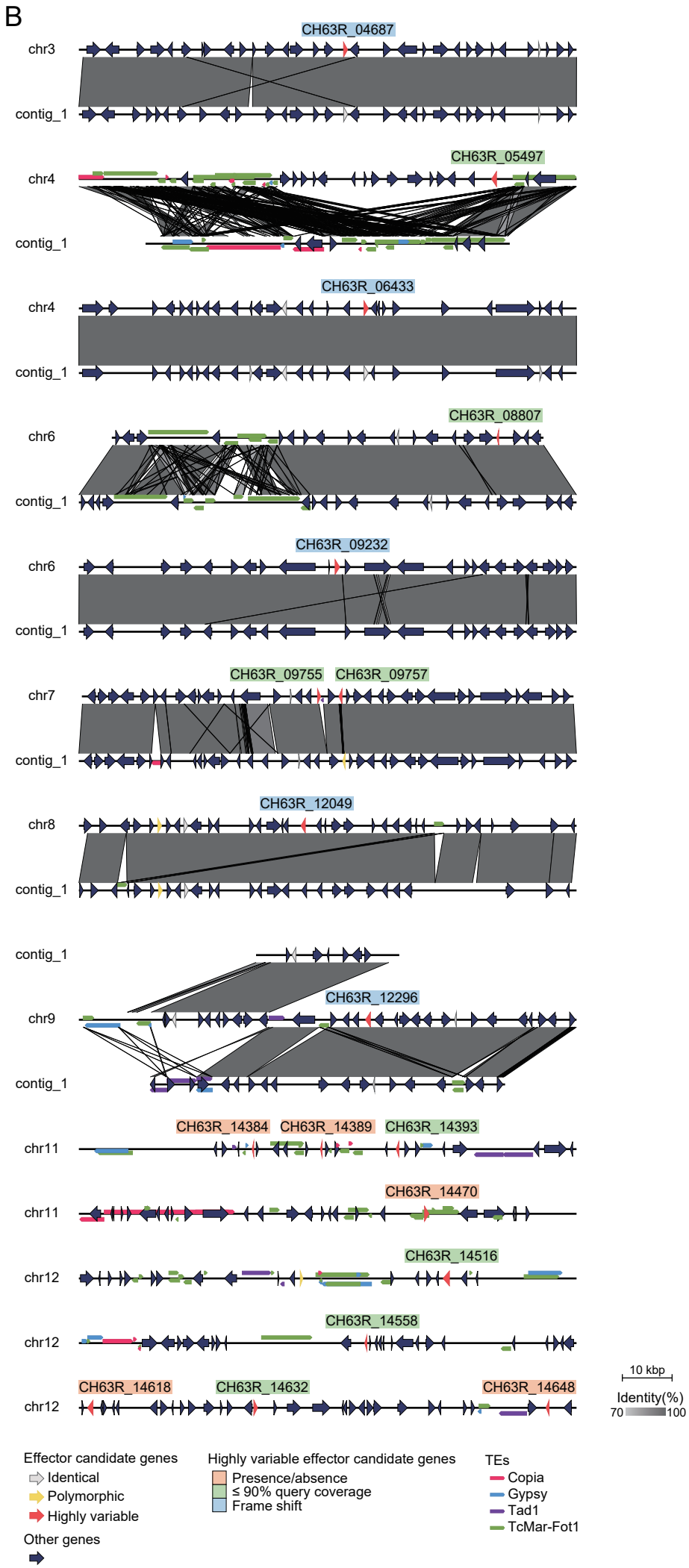



Fig. 3-19. Analysis of genome synteny in regions containing highly variable effector candidate genes. (A) Top: genomic regions containing highly variable effector candidate genes from MAFF 305635-RFP. Bottom: syntenic regions from MAFF 305635-RFP. (B) Top: genomic regions containing highly variable effector candidate genes from IMI 349063. Bottom: Syntenic regions from IMI 349063. Regions containing highly variable effector candidate genes on chromosomes 11 and 12 did not contain syntenic regions in MAFF 305635-RFP.


B
CH63R_04687
chr3
contig_1
CH63R_05497
chr4
contig_1
CH63R_06433
chr4
contig_1
CH63R_08807
chr6
contig_1
CH63R_09232
chr6
contig_1
CH63R_09755
CH63R_09757
chr7
contig_1
CH63R_12049
chr8
contig_1
contig_1
CH63R_12296
chr9
contig_1
CH63R_14384
CH63R_14389
CH63R_14393
chr11
CH63R_14470
chr11
CH63R_14516
chr12
CH63R_14558
chr12
CH63R_14618
CH63R_14632
CH63R_14648
chr12
10 kbp
Identity(%)
70
100
Effector candidate genes
Identical
Polymorphic
Highly variable
Other genes
Highly variable effector candidate genes
Presence/absence
≤ 90% query coverage
Frame shift
TEs
Copia
Gypsy
Tad1
TcMar-Fot1

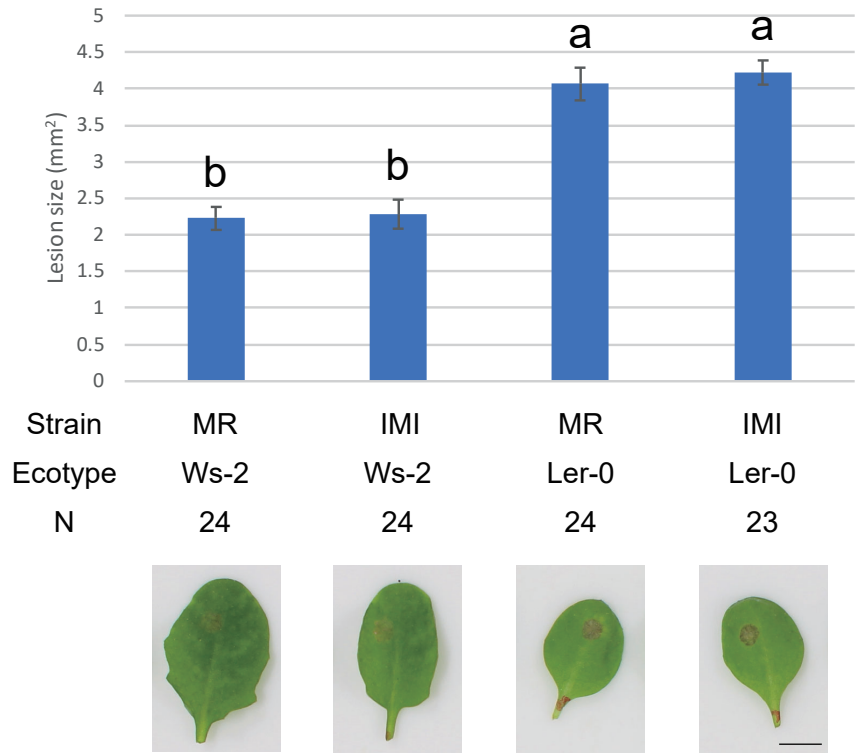



Fig. 3-20. Lesion area measurement assays on *A. thaliana* ecotypes Ws-2 and Ler-0. A representative result from three experiments is shown. MR and IMI indicate MAFF 305635-RFP and IMI 349063, respectively. N represents the number of leaves assessed for each combination. Analysis of variance with Tukey post-hoc honestly significant difference test ($P < 0.05$) was performed. Error bars represent standard errors. Images of representative symptoms at 6 days after infection are shown below. Bar = 5 mm.

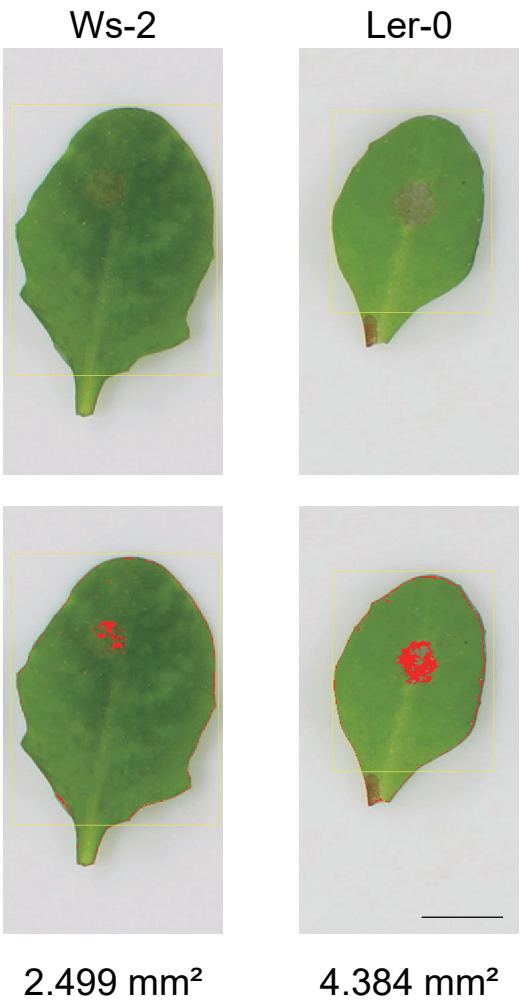



Fig. 3-21. Lesion area measurements using the color threshold function in ImageJ. Images of *A. thaliana* ecotypes Ws-2 and Ler-0 leaves infected with MAFF 305635-RFP at 6 days after infection are shown. Upper panels: original images. Lower panel: detected lesion areas filled with red using the color threshold function in ImageJ. Numbers below the images indicate the actual measured values of lesion area ($mm^2$). Regions of interest are shown as yellow rectangles. Bar = 5 mm.

## Chapter IV: Functional Analysis of CCE1, a Conserved Effector Candidate Among *Colletotrichum* Fungi

*本章については、５年以内に雑誌等で刊行予定のため、非公開。

## Chapter V: Concluding Remarks

Plant pathogenic fungi exhibit great diversity in how they interact with their hosts. The purpose of my Ph.D. research was to understand adaptation of plant pathogenic fungi focusing on the evolution of their effector genes, which play important roles in infection. To achieve this aim, I studied *Colletotrichum* fungi, that have varied their lifestyles to infect distinct host plants, with a range of approaches including bioinformatics, molecular biology, and biochemistry.

In this thesis, I focused on analyzing the genome of *C. higginsianum,* which is widely used in scientific studies as it infects the model plant *Arabidopsis thaliana.* A previously published version of the *C. higginsianum* genome was highly fragmented leading to the possibility of misannotated and/or missing genes. Therefore, in order to obtain reliable genomic information about this pathogen, I sequenced and assembled the genome of MAFF305635-RFP. Using PacBio long reads that can span long tracts of repeat sequences, I could generate a highly contiguous genome assembly. Furthermore, this genome was annotated after assessing two different gene annotation pipelines.

With the newly obtained genomic information, I assessed the conservation of effector candidates from *Colletotrichum* at the genus, species, and strain levels. In Chapter II, I identified *Colletotrichum* genus-specific effector candidates including CCE1 that may aid colonization by targeting conserved components among various host plants of *Colletotrichum*, as well as species-specific effector candidates that may contribute to pathogenicity against individual host plants. In Chapter III, I compared the genome of MAFF305635-RFP against the genome of another *C. higginsianum* strain IMI 349063 and revealed that the two strains of *C. higginsianum* show variations in their effector candidates. These results suggest that individual *Colletotrichum* strains diversify/retain their effector complements to adapt to different niches.

In order to understand how *Colletotrichum* acquire genomic variations in the absence of a sexual cycle, I focused on intra-species genomic variations in *C. higginsianum*. By comparing two closely-related strains of *C. higginsianum*, I discovered large-scale rearrangements and the presence of strain-specific genomic regions in the genome of this species. While previous studies have suggested structural genomic variations in plant pathogenic fungi, I showed direct evidence of the remarkable genomic flexibility for the first time in *Colletotrichum* species. I then tested for an association between genomic variations and transposable elements (TEs). This analysis indicated that TEs contribute to the generation of genomic changes in this clonally proliferating pathogen. This finding implies a potential trade-off between rapid genomic evolution and genomic stability. The genome of *C. higginsianum* could rapidly generate novel genetic alleles through direct and indirect TE contribution; however, a highly plastic genome also reduces synteny which is required for meiosis. Although dynamic genomic changes can induce deleterious effects in organisms, pathogens, which are under strong evolutionary pressure to diversify their effectors to avoid recognition by host immune receptors, could have evolved toward acquiring unstable genomes. Further analysis indicated that *C. higginsianum* has a compartmentalized genome consisting of gene-dense, TE-sparse regions harboring house-keeping genes and gene-sparse, TE-dense regions with more effector candidate genes. As TEs can induce both beneficial and harmful changes in the genome, having such a bipartite genomic structure could allow conservation of house-keeping genes and rapid evolution of effector genes.

To characterize the function of identified effector candidates in more detail, in Chapter IV I conducted functional analysis of CCE1, a conserved effector candidate among *Colletotrichum* that was identified in Chapter II. Transient expression assays revealed that *CCE1* homologs from three *Colletotrichum* species infecting different host plants shared the ability to induce cell death. Furthermore, I identified candidate host plant interactors of CCE1 proteins by performing *in planta*

co-immunoprecipitation. These results provide clues for future studies to assess the function of CCE1.

In this thesis, I analyzed effector candidate gene conservation patterns and the genomic structure of *Colletotrichum*. From this work, I revealed the diversity of effector candidates and a potential mechanism for generating genomic differences in this group of fungi. However, this work raises questions including whether the effector candidates identified contribute to pathogenicity. The characterization of CCE1 as a cell death-inducing effector indicates that at least some of the candidates I identified can affect the host. However, to address the roles of these genes during infection, virulence tests using fungal strains that overexpress or lack effector candidate genes are necessary. The ImageJ-based lesion area quantification method that I established in Chapter III should be useful for precisely determining the virulence effect of each candidate. Another unsolved question is how general extreme variations in genomic structures are among plant pathogenic fungi. Although numerous plant pathogenic fungi genomes have been sequenced to date, there are a limited number of chromosome-level assemblies. Thus, it is of interest to revisit this problem when a greater number of contiguous genome assemblies become available. Overall, this thesis serves as a foundation for further studies on *Colletotrichum* thriving in a variety of niches.

## References


Akcapinar, G.B., Kappel, L., Sezerman, O.U., Seidl-Seiboth, V., 2015. Molecular diversity of LysM carbohydrate-binding motifs in fungi. Curr. Genet. 61, 103–113.

Asai, S., Shirasu, K., 2015. Plant cells under siege: plant immune system versus pathogen effectors. Curr. Opin. Plant Biol. 28, 1–8.

Azevedo, C., Sadanandom, A., Kitagawa, K., Freialdenhoven, A., Shirasu, K., Schulze-Lefert, P., 2002. The RAR1 interactor SGT1, an essential component of R gene-triggered disease resistance. Science 295, 2073–2076.

Bailly-Bechet, M., Haudry, A., Lerat, E., 2014. "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. Mob. DNA 5, 13.

Bao, J., Chen, M., Zhong, Z., Tang, W., Lin, L., Zhang, X., Jiang, H., Zhang, D., Miao, C., Tang, H., Zhang, J., Lu, G., Ming, R., Norvienyeku, J., Wang, B., Wang, Z., 2017. PacBio sequencing reveals transposable elements as a key contributor to genomic plasticity and virulence variation in *Magnaporthe oryzae*. Mol. Plant 10, 1465–1468.

Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6, 11.

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12, 1269–76.

Baroncelli, R., Amby, D.B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., Harrison, R.J., Holub, E., Sukno, S.A., Sreenivasaprasad, S., Thon, M.R., 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. BMC Genomics 17, 555.

Baroncelli, R., Sreenivasaprasad, S., Sukno, S.A., Thon, M.R., Holub, E., 2014. Draft genome sequence of *Colletotrichum acutatum Sensu Lato* (*Colletotrichum fioriniae*). Genome Announc. 2, e00112-14.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–95.

Bhadauria, V., Banniza, S., Vandenberg, A., Selvaraj, G., Wei, Y., 2013. Overexpression of a novel biotrophy-specific *Colletotrichum truncatum* effector, CtNUDIX, in hemibiotrophic fungal phytopathogens causes incompatibility with their host plants. Eukaryot. Cell 12, 2–11.

Birker, D., Heidrich, K., Takahara, H., Narusaka, M., Deslandes, L., Narusaka, Y., Reymond, M., Parker, J.E., O'Connell, R., 2009. A locus conferring resistance to *Colletotrichum higginsianum* is shared by four geographically distinct *Arabidopsis* accessions. Plant J. 60, 602–613.

Cannon, P.F., Damm, U., Johnston, P.R., Weir, B.S., 2012. *Colletotrichum* - current status and future

directions. Stud. Mycol. 73, 181–213.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A.G., Grigoriev, I. V., Stajich, J.E., Ramírez, L., 2016. Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. PLoS Genet. 12, 1–27.

Chaverri, P., Salgado, C., Hirooka, Y., Rossman, A.Y., Samuels, G.J., 2011. Delimitation of *Neonectria* and *Cylindrocarpon* (*Nectriaceae*, *Hypocreales*, *Ascomycota*) and related genera with *Cylindrocarpon*-like anamorphs. Stud. Mycol. 68, 57–78.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10, 563–569.

Chisholm, S.T., Coaker, G., Day, B., Staskawicz, B.J., 2006. Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124, 803–814.

Chuma, I., Isobe, C., Hotta, Y., Ibaragi, K., Futamata, N., Kusaba, M., Yoshida, K., Terauchi, R., Fujita, Y., Nakayashiki, H., Valent, B., Tosa, Y., 2011. Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. PLoS Pathog. 7, e1002147.

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

Croll, D., McDonald, B.A., 2012. The accessory genome as a cradle for adaptive evolution in pathogens. PLoS Pathog. 8, e1002608.

Croll, D., Zala, M., McDonald, B.A., 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLoS Genet. 9, e1003567.

Crouch, J., Beirn, L.A., 2009. Anthracnose of cereals and grasses. Fungal Divers. 2006, 19–44.

Crouch, J., O'Connell, R., Gan, P., Buiate, E., Torres, M.F., Beirn, L., Shirasu, K., Vaillancourt, L., 2014. The genomics of *Colletotrichum*. in: Dean, RA., Lichens-Park, A., Kole, C. (Eds), Genomics of Plant-Associated Fungi: Monocot Pathogens. Springer, Berlin Heidelberg, pp. 69–102.

Damm, U., O'Connell, R.J., Groenewald, J.Z., Crous, P.W., 2014. The *Colletotrichum destructivum* species complex - hemibiotrophic pathogens of forage and field crops. Stud. Mycol. 79, 49–84.

De Jonge, R., Bolton, M.D., Kombrink, A., Van Den Berg, G.C.M., Yadeta, K.A., Thomma, B.P.H.J., 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23, 1271–1282.

de Jonge, R., Peter van Esse, H., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K.,

Zhang, Z., Usami, T., Lievens, B., Subbarao, K. V., Thomma, B.P.H.J., 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc. Natl. Acad. Sci. U. S. A. 109, 5110–5115.

Dean, R., Van Kan, J.A.L., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J., Foster, G.D., 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol. Plant Pathol. 13, 414–430.

Deslandes, L., Rivas, S., 2012. Catch me if you can: bacterial effectors and plant targets. Trends Plant Sci. 17, 644–655.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.-M., Teh, T., Wang, C.-I.A., Ayliffe, M.A., Kobe, B., Ellis, J.G., 2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc. Natl. Acad. Sci. U. S. A. 103, 8888–93.

Dodds, P.N., Rathjen, J.P., 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. Nat. Rev. Genet. 11, 539–548.

Dong, S., Raffaele, S., Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr. Opin. Genet. Dev. 35, 57–65.

Dong, S., Stam, R., Cano, L.M., Song, J., Sklenar, J., Yoshida, K., Bozkurt, T.O., Oliva, R., Liu, Z., Tian, M., Win, J., Banfield, M.J., Jones, A.M.E., van der Hoorn, R.A.L., Kamoun, S., 2014. Effector specialization in a lineage of the Irish potato famine pathogen. Science. 343, 552–555.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.

Eisenhaber, B., Schneider, G., Wildpaner, M., Eisenhaber, F., 2004. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studiesfor *Aspergillus nidulans*, *Candida albicans, Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. J. Mol. Biol. 337, 243–253.

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

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157.

Fabro, G., Steinbrenner, J., Coates, M., Ishaque, N., Baxter, L., Studholme, D.J., Körner, E., Allen, R.L., Piquerez, S.J.M., Rougon-Cardoso, A., Greenshields, D., Lei, R., Badel, J.L., Caillaud, M.C., Sohn, K.H., Van den Ackerveken, G., Parker, J.E., Beynon, J., Jones, J.D.G., 2011. Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. PLoS Pathog. 7, e1002348.

Faino, L., Seidl, M.F., Shi-Kunne, X., Pauper, M., Van Den Berg, G.C.M., Wittenberg, A.H.J., Thomma, B.P.H.J., 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 26, 1091–1100.

Falk, A., Feys, B.J., Frost, L.N., Jones, J.D.G., Daniels, M.J., Parker, J.E., 1999. *EDS1*, an essential component of *R* gene-mediated disease resistance in *Arabidopsis* has homology to eukaryotic lipases. Proc. Natl. Acad. Sci. U. S. A. 96, 3292–3297.

Feldman, D., Kowbel, D.J., Glass, N.L., Yarden, O., Hadar, Y., 2017. A role for small secreted proteins (SSPs) in a saprophytic fungal lifestyle: Ligninolytic enzyme regulation in *Pleurotus ostreatus*. Sci. Rep. 7, 14553.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D.A., Necci, M., Nuka, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C., Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Xenarios, I., Yeh, L.-S., Young, S.-Y., Mitchell, A.L., 2017. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 45, D190–D199.

Fisher, M.C., Henk, D. a., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L., Gurr, S.J., 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484, 186–194.

Flor, H.H., 1971. Current Status of the Gene-For-Gene Concept. Annu. Rev. Phytopathol. 9, 275–296.

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

Freese, N.H., Norris, D.C., Loraine, A.E., 2016. Integrated genome browser: visual analytics platform for genomics. Bioinformatics 32, 2089–2095.

Frey, T.J., Weldekidan, T., Colbert, T., Wolters, P.J.C.C., Hawk, J.A., 2011. Fitness evaluation of *Rcg1*, a locus that confers resistance to *Colletotrichum graminicola* (Ces.) G.W. Wils. using near-isogenic maize hybrids. Crop Sci. 51, 1551–1563.

Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O'Connell, R.J., Narusaka, Y., Takano, Y., Kubo, Y., Shirasu, K., 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol. 197, 1236–1249.

Gan, P., Narusaka, M., Kumakura, N., Tsushima, A., Takano, Y., Narusaka, Y., Shirasu, K., 2016. Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. Genome Biol. Evol. 8, 1467–1481.

Gan, P., Narusaka, M., Tsushima, A., Narusaka, Y., Takano, Y., Shirasu, K., 2017. Draft genome assembly of *Colletotrichum* chlorophyti, a pathogen of herbaceous plants. Genome Announc. 5, e01733-16.

Geib, S.M., Hall, B., Derego, T., Bremer, F.T., Cannoles, K., Sim, S.B., 2018. Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. Gigascience 7, 1–5.

Giraldo, M.C., Valent, B., 2013. Filamentous plant pathogen effectors in action. Nat. Rev. Microbiol. 11, 800–814.

Göhre, V., Spallek, T., Häweker, H., Mersmann, S., Mentzel, T., Boller, T., de Torres, M., Mansfield, J.W., Robatzek, S., 2008. Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB. Curr. Biol. 18, 1824–1832.

Gomes, R.R., Glienke, C., Videira, S.I.R., Lombard, L., Groenewald, J.Z., Crous, P.W., 2013. *Diaporthe*: a genus of endophytic, saprobic and plant pathogenic fungi. Persoonia 31, 1–41.

Grandaubert, J., Lowe, R.G., Soyer, J.L., Schoch, C.L., Van de Wouw, A.P., Fudal, I., Robbertse, B., Lapalu, N., Links, M.G., Ollivier, B., Linglin, J., Barbe, V., Mangenot, S., Cruaud, C., Borhan, H., Howlett, B.J., Balesdent, M.-H., Rouxel, T., 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. BMC Genomics 15, 891.

Griffis, A.H.N., Groves, N.R., Zhou, X., Meier, I., 2014. Nuclei in motion: movement and positioning of plant nuclei in development, signaling, symbiosis, and disease. Front. Plant Sci. 5, 129.

Gross, P., Julius, C., Schmelzer, E., Hahlbrock, K., 1993. Translocation of cytoplasm and nucleus to fungal penetration sites is associated with depolymerization of microtubules and defence gene activation in infected, cultured parsley cells. EMBO J. 12, 1735–1744.

Hacquard, S., Kracher, B., Hiruma, K., Münch, P.C., Garrido-Oter, R., Thon, M.R., Weimann, A., Damm, U., Dallery, J.-F., Hainaut, M., Henrissat, B., Lespinet, O., Sacristán, S., Ver Loren van Themaat, E., Kemen, E., McHardy, A.C., Schulze-Lefert, P., O'Connell, R.J., 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. Nat. Commun. 7, 11362.

Harel, M., Aharoni, A., Gaidukov, L., Brumshtein, B., Khersonsky, O., Meged, R., Dvir, H., Ravelli, R.B.G., McCarthy, A., Toker, L., Silman, I., Sussman, J.L., Tawfik, D.S., 2004. Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. Nat. Struct. Mol. Biol. 11, 412–419.

Hartmann, F.E., Sánchez-Vallet, A., McDonald, B.A., Croll, D., 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11, 1189–1204.

Hatta, R., Ito, K., Hosaki, Y., Tanaka, T., Tanaka, A., Yamamoto, M., Akimitsu, K., Tsuge, T., 2002. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. Genetics 161, 59–70.

He, P., Shan, L., Lin, N.-C., Martin, G.B., Kemmerling, B., Nürnberger, T., Sheen, J., 2006. Specific

bacterial suppressors of MAMP signaling upstream of MAPKKK in *Arabidopsis* innate immunity. Cell 125, 563–575.

Hemetsberger, C., Herrberger, C., Zechmann, B., Hillmer, M., Doehlemann, G., 2012. The *Ustilago maydis* effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity. PLoS Pathog. 8, e1002684.

Hemetsberger, C., Mueller, A.N., Matei, A., Herrberger, C., Hensel, G., Kumlehn, J., Mishra, B., Sharma, R., Thines, M., Hückelhoven, R., Doehlemann, G., 2015. The fungal core effector Pep1 is conserved across smuts of dicots and monocots. New Phytol. 206, 1116–1126.

Henty-Ridilla, J.L., Shimono, M., Li, J., Chang, J.H., Day, B., Staiger, C.J., 2013. The plant actin cytoskeleton responds to signals from microbe-associated molecular patterns. PLoS Pathog. 9, e1003290.

Higaki, T., Kutsuna, N., Sano, T., Kondo, N., Hasezawa, S., 2010. Quantification and cluster analysis of actin cytoskeletal structures in plant cells: role of actin bundling in stomatal movement during diurnal cycles in *Arabidopsis* guard cells. Plant J. 61, 156–165.

Hiruma, K., Onozawa-Komori, M., Takahashi, F., Asakura, M., Bednarek, P., Okuno, T., Schulze-Lefert, P., Takano, Y., 2010. Entry mode-dependent function of an indole glucosinolate pathway in *Arabidopsis* for nonhost resistance against anthracnose pathogens. Plant Cell 22, 2429–2443.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., Stanke, M., 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32, 767–769.

Hogenhout, S. a, Van der Hoorn, R. a L., Terauchi, R., Kamoun, S., 2009. Emerging concepts in effector biology of plant-associated organisms. Mol. Plant. Microbe. Interact. 22, 115–122.

Holt, C., Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 491.

Hubert, D.A., Tornero, P., Belkhadir, Y., Krishna, P., Takahashi, A., Shirasu, K., Dangl, J.L., 2003. Cytosolic HSP90 associates with and modulates the *Arabidopsis* RPM1 disease resistance protein. EMBO J. 22, 5679–5689.

Islam, M.T., Croll, D., Gladieux, P., Soanes, D.M., Persoons, A., Bhattacharjee, P., Hossain, M.S., Gupta, D.R., Rahman, M.M., Mahboob, M.G., Cook, N., Salam, M.U., Surovy, M.Z., Sancho, V.B., Maciel, J.L.N., NhaniJúnior, A., Castroagudín, V.L., Reges, J.T., Ceresini, P.C., Ravel, S., Kellner, R., Fournier, E., Tharreau, D., Lebrun, M.-H., McDonald, B.A., Stitt, T., Swan, D., Talbot, N.J., Saunders, D.G.O., Win, J., Kamoun, S., 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. BMC Biol. 14, 84.

Johnson, B.R., 2018. Taxonomically restricted genes are fundamental to biology and evolution. Front. Genet. 9, 407.

Jones, J.D.G., Dangl, J.L., 2006. The plant immune system. Nature 444, 323–329.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.

Kistler, H.C., Miao, V.P.W., 1992. New modes of genetic change in filamentous fungi. Annu. Rev. Phytopathol. 30, 131–153.

Kleemann, J., Rincon-Rivera, L.J., Takahara, H., Neumann, U., van Themaat, E.V.L., van der Does, H.C., Hacquard, S., Stüber, K., Will, I., Schmalenbach, W., Schmelzer, E., O'Connell, R.J., 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLoS Pathog. 8, e1002643.

Korf, I., 2004. Gene finding in novel genomes. BMC Bioinformatics 5, 59.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Mol. Biol. 305, 567–580.

Kumakura, N., Ueno, A., Shirasu, K., 2018. Establishment of a selection marker recycling system for sequential transformation of the plant-pathogenic fungus *Colletotrichum orbiculare*. Mol. Plant Pathol.

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

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Lee, A.H.-Y., Hurley, B., Felsensteiner, C., Yea, C., Ckurshumova, W., Bartetzko, V., Wang, P.W., Quach, V., Lewis, J.D., Liu, Y.C., Börnke, F., Angers, S., Wilde, A., Guttman, D.S., Desveaux, D., 2012. A bacterial acetyltransferase destroys plant microtubule networks and blocks secretion. PLoS Pathog. 8, e1002523.

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245.

Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–89.

Li, W.F., Costa, L.G., Richter, R.J., Hagen, T., Shih, D.M., Tward, A., Lusis, A.J., Furlong, C.E., 2000. Catalytic efficiency determines the in-vivo efficacy of PON1 for detoxifying organophosphorus compounds. Pharmacogenetics 10, 767–779.

Lingua, G., Fusconi, A., Berta, G., 2001. The nucleus of differentiated root plant cells: modifications induced by arbuscular mycorrhizal fungi. Eur. J. Histochem. 45, 9–20.

Liu, Y., Schiff, M., Marathe, R., Dinesh-Kumar, S.P., 2002. Tobacco Rar1, EDS1 and NPR1/NIM1

like genes are required for N-mediated resistance to tobacco mosaic virus. Plant J. 30, 415–429.
Lomsadze, A., Burns, P.D., Borodovsky, M., 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 42, 1–8.
Ma, L.-J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.-B., Woloshuk, C., Xie, X., Xu, J.-R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A.E., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G.J., Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.-H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.-Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C., Rep, M., 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464, 367–373.
Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., Dubchak, I., 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16, 1046–1047.
McMullan, M., Rafiqi, M., Kaithakottil, G., Clavijo, B.J., Bilham, L., Orton, E., Percival-Alwyn, L., Ward, B.J., Edwards, A., Saunders, D.G.O., Garcia Accinelli, G., Wright, J., Verweij, W., Koutsovoulos, G., Yoshida, K., Hosoya, T., Williamson, L., Jennings, P., Ioos, R., Husson, C., Hietala, A.M., Vivian-Smith, A., Solheim, H., MaClean, D., Fosker, C., Hall, N., Brown, J.K.M., Swarbreck, D., Blaxter, M., Downie, J.A., Clark, M.D., 2018. The ash dieback invasion of Europe was founded by two genetically divergent individuals. Nat. Ecol. Evol. 2, 1000–1008.
Menat, J., Cabral, A.L., Vijayan, P., Wei, Y., Banniza, S., 2012. *Glomerella truncata* : another *Glomerella* species with an atypical mating system. Mycologia 104, 641–649.
Meyer, V., 2008. A small protein that fights fungi: AFP as a new promising antifungal agent of biotechnological value. Appl. Microbiol. Biotechnol. 78, 17-28.
Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., Gurevich, A., 2016. Icarus: visualizer for de novo assembly evaluation. Bioinformatics 32, 3321–3323.
Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.-Y., Bateman, A., Punta, M., Attwood, T.K., Sigrist, C.J.A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D.A., Wu, C.H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P.D., Finn, R.D., 2014. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 43, D213-221.
Möller, M., Stukenbrock, E.H., 2017. Evolution and genome architecture in fungal plant pathogens.

Nat. Rev. Microbiol. 15, 756–771.
Moreno-Hagelsieb, G., Latimer, K., 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics 24, 319–324.
Münch, S., Lingner, U., Floss, D.S., Ludwig, N., Sauer, N., Deising, H.B., 2008. The hemibiotrophic lifestyle of *Colletotrichum* species. J. Plant Physiol. 165, 41–51.
Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., Kimura, T., 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J. Biosci. Bioeng. 104, 34–41.
Narusaka, M., Iuchi, S., Narusaka, Y., 2017. Analyses of natural variation indicates that the absence of RPS4/RRS1 and amino acid change in RPS4 cause loss of their functions and resistance to pathogens. Plant Signal. Behav. 12, e1293218.
Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., Narusaka, Y., 2009. *RRS1* and *RPS4* provide a dual *Resistance*- gene system against fungal and bacterial pathogens. Plant J. 60, 218–226.
Narusaka, Y., Narusaka, M., Park, P., Kubo, Y., Hirayama, T., Seki, M., Shiraishi, T., Ishida, J., Nakashima, M., Enju, A., Sakurai, T., Satou, M., Kobayashi, M., Shinozaki, K., 2004. *RCH1*, a locus in *Arabidopsis* that confers resistance to the hemibiotrophic fungal pathogen *Colletotrichum higginsianum*. Mol. Plant-Microbe Interact. 17, 749–762.
Nelson, B.K., Cai, X., Nebenführ, A., 2007. A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. Plant J. 51, 1126–1136.
O'Connell, R., Herbert, C., Sreenivasaprasad, S., Khatib, M., Esquerré-Tugayé, M.-T., Dumas, B., 2004. A novel *Arabidopsis*-*Colletotrichum* pathosystem for the molecular dissection of plant-fungal interactions. Mol. Plant. Microbe. Interact. 17, 272–282.
O'Connell, R.J., Thon, M.R., Hacquard, S., Amyotte, S.G., Kleemann, J., Torres, M.F., Damm, U., Buiate, E.A., Epstein, L., Alkan, N., Altmüller, J., Alvarado-Balderrama, L., Bauser, C.A., Becker, C., Birren, B.W., Chen, Z., Choi, J., Crouch, J.A., Duvick, J.P., Farman, M.A., Gan, P., Heiman, D., Henrissat, B., Howard, R.J., Kabbage, M., Koch, C., Kracher, B., Kubo, Y., Law, A.D., Lebrun, M.-H., Lee, Y.-H., Miyara, I., Moore, N., Neumann, U., Nordström, K., Panaccione, D.G., Panstruga, R., Place, M., Proctor, R.H., Prusky, D., Rech, G., Reinhardt, R., Rollins, J.A., Rounsley, S., Schardl, C.L., Schwartz, D.C., Shenoy, N., Shirasu, K., Sikhakolli, U.R., Stüber, K., Sukno, S.A., Sweigard, J.A., Takano, Y., Takahara, H., Trail, F., van der Does, H.C., Voll, L.M., Will, I., Young, S., Zeng, Q., Zhang, J., Zhou, S., Dickman, M.B., Schulze-Lefert, P., Ver Loren van Themaat, E., Ma, L.-J., Vaillancourt, L.J., 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat. Genet. 44, 1060–5.
Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in

eukaryotic genomes. Bioinformatics 23, 1061–1067.

Peart, J.R., Cook, G., Feys, B.J., Parker, J.E., Baulcombe, D.C., 2002. An EDS1 orthologue is required for N-mediated resistance against tobacco mosaic virus. Plant J. 29, 569–579.

Petersen, T.N., Brunak, S., Von Heijne, G., Nielsen, H., 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Petre, B., Saunders, D.G.O., Sklenar, J., Lorrain, E., Win, J., Duplessis, E., Kamoun, S., 2015. Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. MPMI 28, 689–700.

Petre, B., Win, J., Menke, F.L.H., Kamoun, S., 2017. Protein–protein interaction assays with effector–GFP fusions in *Nicotiana benthamiana*. In: Periyannan, S. (Ed.), Wheat Rust Diseases. Methods in Molecular Biology. Humana Press, New York, pp. 85–98.

Plaumann, P.-L., Schmidpeter, J., Dahl, M., Taher, L., Koch, C., 2018. A dispensable chromosome is required for virulence in the hemibiotrophic plant pathogen *Colletotrichum higginsianum*. Front. Microbiol. 9, 1005.

Price, A.L., Jones, N.C., Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. Bioinformatics 21, i351–i358.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H.Y., Zody, M.C., Kunjeti, S.G., Donofrio, N.M., Meyers, B.C., Nusbaum, C., Kamoun, S., 2010. Genome evolution following host jumps in the irish potato famine pathogen lineage. Science. 330, 1540–1543.

Raffaele, S., Kamoun, S., 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. 10, 417–430.

Rao, S., Nandineni, M.R., 2017. Genome sequencing and comparative genomics reveal a repertoire of putative pathogenicity genes in chilli anthracnose fungus *Colletotrichum truncatum*. PLoS One 12, e0183567.

Rodríguez-Guerra, R., Ramírez-Rueda, M.-T., Cabral-Enciso, M., nica García-Serrano, M., Lira-Maldonado Ramón Gerardo Guevara-González, Z., González-Chavira, M., Simpson, J., 2005. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. Mycologia 97, 793–803.

Rosenblat, M., Aviram, M., 2009. Paraoxonases role in the prevention of cardiovascular diseases. BioFactors. 35, 98-104.

Sainsbury, F., Thuenemann, E.C., Lomonossoff, G.P., 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol. J. 7, 682–693.

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A., Croll, D., 2018. The genome biology of effector gene evolution in filamentous plant pathogens. Annu. Rev. Phytopathol. 56, 21-40.

Sanz-Martín, J.M., Pacheco-Arjona, J.R., Bello-Rico, V., Vargas, W.A., Monod, M., Díaz-Mínguez, J.M., Thon, M.R., Sukno, S.A., 2016. A highly conserved metalloprotease effector enhances virulence in the maize anthracnose fungus *Colletotrichum graminicola*. Mol. Plant Pathol. 17, 1048–1062.

Sarris, P.F., Duxbury, Z., Huh, S.U., Ma, Y., Segonzac, C., Sklenar, J., Derbyshire, P., Cevik, V., Rallapalli, G., Saucet, S.B., Wirthmueller, L., Menke, F.L.H., Sohn, K.H., Jones, J.D.G., 2015. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. Cell 161, 1089–1100.

Sato, T., Moriwaki, J., 2009. Causal fungi of plant anthracnose (1). Microbiol. Cult. Coll. 25, 27–32.

Saunders, D.G.O., Win, J., Kamoun, S., Raffaele, S., 2014. Two-dimensional data binning for the analysis of genome architecture in filamentous plant pathogens and other eukaryotes. In: Birch, P., Jones, J., Bos, J. (Eds.), Methods in Molecular Biology (Methods and Protocols). Humana Press, Totowa, pp. 29–51.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675.

Schwessinger, B., Roux, M., Kadota, Y., Ntoukakis, V., Sklenar, J., Jones, A., Zipfel, C., 2011. Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. PLoS Genet. 7, e1002046.

Seidl, M.F., Thomma, B.P.H.J., 2014. Sex or no sex: Evolutionary adaptation occurs regardless. BioEssays 36, 335–345.

Shan, X.C., Goodwin, P.H., 2004. Monitoring host nuclear migration and degradation with green fluorescent protein during compatible and incompatible interactions of *Nicotiana tabacum* with *Colletotrichum* species. J. Phytopathol. 152, 454–460.

Shimono, M., Lu, Y.-J., Porter, K., Kvitko, B.H., Henty-Ridilla, J., Creason, A., He, Y., Chang, J.H., Staiger, C.J., Day, B., Conducted, B.H.K., Contrib-, C.J.S., 2016. The *Pseudomonas syringae* type III effector HopG1 induces actin remodeling to promote symptom development and susceptibility during infection. Plant Physiol. 171, 2239–2255.

Shirasu, K., Lahaye, T., Tan, M.W., Zhou, F., Azevedo, C., Schulze-Lefert, P., 1999. A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*. Cell 99, 355–366.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31.

Sohn, K.H., Zhang, Y., Jones, J.D.G., 2009. The *Pseudomonas syringae* effector protein, AvrRPS4, requires in planta processing and the KRVY domain to function. Plant J. 57, 1079–1091.

Spanu, P.D., Panstruga, R., 2017. Editorial: biotrophic plant-microbe interactions. Front. Plant Sci. 8, 192.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stanke, M., Schöffmann, O., Morgenstern, B., Waack, S., 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7, 62.

Sullivan, M.J., Petty, N.K., Beatson, S.A., 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009–1010.

Takahara, H., Hacquard, S., Kombrink, A., Hughes, H.B., Halder, V., Robin, G.P., Hiruma, K., Neumann, U., Shinya, T., Kombrink, E., Shibuya, N., Thomma, B.P.H.J., O'Connell, R.J., 2016. *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. New Phytol. 211, 1323–1337.

Takemoto, D., Jones, D.A., Hardham, A.R., 2006. Re-organization of the cytoskeleton and endoplasmic reticulum in the *Arabidopsis pen1-1* mutant inoculated with the non-adapted powdery mildew pathogen, *Blumeria graminis* f. sp. *hordei*. Mol. Plant Pathol. 7, 553–563.

Tautz, D., Domazet-Lošo, T., 2011. The evolutionary origin of orphan genes. Nat. Rev. Genet. 12, 692–702.

Treangen, T.J., Salzberg, S.L., 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. 13, 36–46.

Vaillancourt, L.J., Hanau, R.M., 1991. A method for genetic analysis of *Glomerella graminicola* (*Colletotrichum graminicola*) from maize. Phytopathology. 81, 530-534.

Wang, J., Li, S., Zhang, Y., Zheng, H., Xu, Z., Ye, J., Yu, J., Wong, G.K.-S., 2003. Vertebrate gene predictions and the problem of large genes. Nat. Rev. Genet. 4, 741–749.

Weir, B.S., Johnston, P.R., Damm, U., 2012. The *Colletotrichum gloeosporioides* species complex. Stud. Mycol. 73, 115–180.

Westerink, N., Brandwagt, B.F., De Wit, P.J.G.M., Joosten, M.H.A.J., 2004. *Cladosporium fulvum* circumvents the second functional resistance gene homologue at the *Cf-4* locus (*Hcr9-4E*) by secretion of a stable avr4E isoform. Mol. Microbiol. 54, 533–545.

White, F.F., Yang, B., Johnson, L.B., 2000. Prospects for understanding avirulence gene function. Curr. Opin. Plant Biol. 3, 291–298.

Wickstead, B., Gull, K., 2011. The evolution of the cytoskeleton. J. Cell Biol. 194, 513–25.

Xiang, T., Zong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W., Li, Y., Tang, X., Zhu, L., Chai, J., Zhou,

J.-M., 2008. *Pseudomonas syringae* effector AvrPto blocks innate immunity by targeting receptor kinases. Curr. Biol. 18, 74–80.

Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J., Kamoun, S., Terauchi, R., 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. Plant Cell 21, 1573–1591.

Yoshino, K., Irieda, H., Sugimoto, F., Yoshioka, H., Okuno, T., Takano, Y., 2012. Cell death of *Nicotiana benthamiana* is induced by secreted protein NIS1 of *Colletotrichum orbiculare* and is suppressed by a homologue of CgDN3. Mol. Plant. Microbe. Interact. 25, 625–36.

Yu, C., Zhang, J., Peterson, T., 2011. Genome rearrangements in maize induced by alternative transposition of reversed *Ac/Ds* termini. Genetics 188, 59–67.

Zampounis, A., Pigné, S., Dallery, J., Wittenberg, A.H.J., Zhou, S., Schwartz, D.C., Thon, M.R., O'Connell, R.J., 2016. Genome sequence and annotation of *Colletotrichum higginsianum*, a causal agent of crucifer anthracnose disease. Genome Announc. 4, e00821-16.

Zhang, J., Yu, C., Pulletikurti, V., Lamb, J., Danilova, T., Weber, D.F., Birehler, J., Peterson, T., 2009. Alternative *Ac/Ds* transposition induces major chromosomal rearrangements in maize. Genes Dev. 23, 755–765.

Zhang, Y., Dorey, S., Swiderski, M., Jones, J.D.G., Centre, J.I., Lane, C., 2004. Expression of *RPS4* in tobacco induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. Plant J. 40, 213–24.

Zhong, Z., Marcel, T.C., Hartmann, F.E., Ma, X., Plissonneau, C., Zala, M., Ducasse, A., Confais, J., Compain, J., Lapalu, N., Amselem, J., McDonald, B.A., Croll, D., Palma-Guerrero, J., 2017. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the *Stb6* resistance gene. New Phytol. 214, 619–631.