

Genomic Plasticity Mediated by Transposable Elements in the Plant Pathogenic Fungus *Colletotrichum higginsianum*

Ayako Tsushima^{1,2}, Pamela Gan², Naoyoshi Kumakura², Mari Narusaka³, Yoshitaka Takano⁴, Yoshihiro Narusaka³, and Ken Shirasu^{1,2,*}

¹Graduate School of Science, The University of Tokyo, Bunkyo, Japan

²Center for Sustainable Resource Science, RIKEN, Yokohama, Japan

³Research Institute for Biological Sciences Okayama, Kaga-gun, Japan

⁴Graduate School of Agriculture, Kyoto University, Kyoto, Japan

*Corresponding author: E-mail: ken.shirasu@riken.jp.

Accepted: April 17, 2019

Data deposition: All associated sequences are deposited at NCBI BioProject accession PRJNA35400. This whole genome shotgun project was deposited at DDBJ/ENA/GenBank under the accession number MWPZ00000000. The version described in this article is version MWPZ01000000.

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, whereas the hosts evolve immune components to recognize the products of these genes. *Colletotrichum higginsianum* (*Ch*), a fungal pathogen with no known sexual morph, infects Brassicaceae plants including *Arabidopsis thaliana*. Previous studies revealed that *Ch* differs in its virulence toward various *Arabidopsis thaliana* ecotypes, indicating the existence of coevolutionary selective pressures. However, between-strain genomic variations in *Ch* have not been studied. Here, we sequenced and assembled the genome of a *Ch* strain, resulting in a highly contiguous genome assembly, which was compared with the chromosome-level genome assembly of another strain to identify genomic variations between strains. We 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). *Ch* 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. Our results reveal mechanisms for generating genomic diversity in this asexual pathogen, which are important for understanding its adaptation to hosts.

Key words: genome rearrangement, effector, transposable element, genome assembly, *Colletotrichum*, plant pathogen.

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 coevolving with their hosts (Raffaele and Kamoun 2012; Möller and 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 and Valent 2013). However, upon recognition by host immune receptors, effectors may also trigger strong immune responses (Dodds and

Rathjen 2010; Asai and Shirasu 2015). Genes encoding effectors often have a higher degree of variation compared with 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 and 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, polymerase chain reaction (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 and Thomma 2014).

Colletotrichum fungi 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* fungi 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, *Colletotrichum 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 coevolved 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 sexual reproduction. Two different strains, MAFF 305635 isolated from *Brassica rapa*

var. *perviridis* (Komatsuna) in Japan and IMI 349063 isolated from *Brassica 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, we identified genomic variations between MAFF 305635-RFP, an MAFF 305635 transformant expressing monomeric red fluorescent protein (RFP) (Hiruma et al. 2010), and IMI 349063. To investigate the genomic differences between the two strains, we sequenced the genome of *C. higginsianum* MAFF 305635-RFP using a PacBio RSII sequencer. We performed whole-genome alignments of this assembly to the chromosome-level assembly of 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, we analyzed their conservation patterns in other ascomycetes and the synteny of genomic regions containing these genes.

Materials and Methods

Fungal Strains

The details of all *C. higginsianum* strains used in this study can be found in [supplementary table S1, Supplementary Material](#) online. To extract genomic DNA, fungal tissue was cultured in potato dextrose broth (BD Biosciences, Franklin Lakes, NJ) 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 1,000 fungal genomes project (<http://1000.fungalgenomes.org>; last accessed on April 26, 2019). Genomic DNA was extracted from other strains using the DNeasy Plant Mini Kit (Qiagen). 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 [supplementary table S2, Supplementary Material](#) online. 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 1 week.

Genome Sequencing and Assembly

Whole genome shotgun sequences were obtained using a PacBio RSII sequencer (Pacific Biosciences, Menlo Park, CA). The filtered subreads (3.47-Gb, 65× average coverage) from

three single-molecule real-time (SMRT) cells were assembled using the RS_HGAP_Assembly.2 protocol in SMRT Analysis v2.3.0 by setting the estimated genome size to 53.4 Mb, which is the reported genome size of *C. higginsianum* IMI 349063 (O'Connell et al. 2012). This whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number MWPZ00000000. The version described in this article is version MWPZ01000000. To evaluate the coverage of gene-coding regions in the assembly, BUSCO v3.0.2 (Simão et al. 2015) was utilized to search Pezizomycotina universal single-copy orthologs using default settings. Additional 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) 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.

Gene Prediction and Functional Annotation

The BRAKER1 (Hoff et al. 2016) pipeline was utilized for gene prediction using the ab initio gene predictors Augustus-3.1.0 (Stanke et al. 2006) and GeneMark-ET v.4.30 (Lomsadze et al. 2014) with additional evidence from transcriptome data derived from IMI 349063 published by O'Connell et al. (2012) (SRR364392.sra, SRR364393.sra, SRR364394.sra, SRR364395.sra, SRR364396.sra, SRR364397.sra, SRR364398.sra, SRR364399.sra, SRR364400.sra, SRR364401.sra, SRR364402.sra, SRR364403.sra, SRR364404.sra, SRR364405.sra, SRR364406.sra, SRR364407.sra, SRR364408.sra, and SRR364409.sra). RNA-seq reads were aligned to the MAFF 305635-RFP assembly using Bowtie 2 version 2.2.6 (Langmead and Salzberg 2012) and TopHat2 v2.1.0 (Kim et al. 2013) with the default settings for both programs. These programs were also utilized for gene prediction. Some gene structures were manually corrected by referring to closely related fungal species. To examine the conservation of IMI 349063 minichromosome-encoded genes in MAFF 305635-RFP, reciprocal best hit analysis was performed with BlastP (cutoff E -value = 10^{-6}) as described by Moreno-Hagelsieb and Latimer (2008) using all the predicted protein sequences from both strains. Only MAFF 305635-RFP proteins that were reciprocal best hits of proteins encoded by genes located on the minichromosomes of IMI 349063 were considered as potential homologs of minichromosome genes.

Whole-Genome Alignments

Whole-genome alignments were performed with nucmer in MUMmer 3.23 (Kurtz et al. 2004) using `-mumreference` (default settings). To remove spurious hits, the alignments were subsequently filtered by length, retaining alignments with $\geq 99\%$ identity and ≥ 15 kb referring to Faino et al. (2016). Shared genomic regions were calculated after combining

overlaps in aligned regions using BEDTools version 2.27.1 (Quinlan and Hall 2010) with the merge option. 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). Whole-genome alignments were also performed in three additional fungal pathogenic fungi, *V. dahliae* (VdLs17 and JR2) (Faino et al. 2015), *Fusarium oxysporum* f. sp. *lycopersici* (4287 and D11) (Ayhan et al. 2018), and *M. oryzae* (BTJP4-1 and BR32) (Win et al. 2019). 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, and SAMN08226881) was performed using Bowtie 2 (Langmead and Salzberg 2012) with default settings.

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 and 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>; last accessed on November 25, 2016). 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, 1,000 trials to model TEs randomly located on the genome were generated using BEDTools version 2.25.0 (Quinlan and 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. As the default settings of nucmer ignores non-unique sequences in the reference, to analyze regions regardless of their uniqueness in the reference strain, we repeated nucmer with the -maxmatch settings and redefined strain-specific regions. Then, Monte Carlo tests were repeated as described above.

Analysis of Genome Compartmentalization

Flanking intergenic regions (FIRs) 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 and 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 <300 amino acids. To assess the overlap between effectors predicted with our pipeline and those predicted by the effector prediction programs EffectorP 1.0 and 2.0 (Sperschneider et al. 2016, 2018), we ran both versions of EffectorP using the default settings. 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 and 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 [supplementary table S3, Supplementary Material](#) online. 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 16 *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 *Colletotrichum destructivum* complex (Damm et al. 2014). The DNA sequences of 16 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 [supplementary table S2, Supplementary Material](#) online. 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 and 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- μ l droplets of conidial suspensions at 5×10^5 conidia ml⁻¹. Plants were maintained at 22 °C with a 10-h photoperiod under 100% humidity conditions after inoculation. The

Table 1Genome-Assembly Statistics of *Colletotrichum higginsianum* MAFF 305635-RFP

	<i>C. higginsianum</i> MAFF 305635-RFP	<i>C. higginsianum</i> IMI 349063
Sequencer	PacBio RS II	PacBio RS II
Total contig length (Mb)	49.8	50.7
Contig number	28	25
N50 contig (Mb)	5.06	5.2
L50 contig	5	5
GC-content (%)	54.61	54.41
Gene space coverage (%) complete/fragmented	99.0/0.3	99.2/0.2
Place of origin	Japan	Trinidad and Tobago
Host	<i>Brassica rapa</i> var. <i>perviridis</i>	<i>Brassica campestris</i>
Reference	This study	Zampounis et al. (2016)

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.

Results

Genome Assembly of *C. higginsianum* Isolated from Japan

The genome of *C. higginsianum* MAFF 305635-RFP was sequenced using a PacBio RSII sequencer. Filtered subreads were assembled using a hierarchical genome-assembly process (Chin et al. 2013). The genome assembly contained 28 contigs with a total length of 49.8 Mb ($N_{50} = 5.06$ and $L_{50} = 5$) (table 1). The completeness of the assembly was evaluated by searching Pezizomycotina universal single-copy orthologs using BUSCO (Simão et al. 2015). This analysis estimated that the assembly included 99.0% complete/0.3% partial assessed loci, indicating that the assembly covered most gene-coding regions. A total of 12,915 protein-coding genes were predicted in the assembly, which was comparable to other closely related *Colletotrichum* fungi, such as *C. graminicola* M1.001, *C. incanum* MAFF 238712, and *C. higginsianum* IMI 349063 possessing 12,006, 11,852, and 14,651 protein-coding genes, respectively (O'Connell et al. 2012; Gan et al. 2016; Zampounis et al. 2016).

Extensive Genomic Variations between *C. higginsianum* Strains

To examine genomic differences between the two sequenced strains of *C. higginsianum*, whole-genome alignments were performed following Faino et al. (2016) (fig. 1). The two strains shared 88.2% in MAFF 305635-RFP and 86.6% in IMI 349063 of their total contig lengths ($\geq 99\%$ identity,

≥ 15 kb). The percentages of shared genomic regions are comparable to other plant pathogenic fungi, such as *V. dahliae* (93.2% in VdLs17; 92.8% in JR2), *Fusarium oxysporum* f. sp. *lycopersici* (93.3% in 4287; 86.0% in D11), and *M. oryzae* (75.9% in BTJP4-1; 81.4% in BR32) when the same cutoffs were applied (supplementary table S4, Supplementary Material online). Whole-genome alignments also revealed that the two *C. higginsianum* strains contained strain-specific regions ($< 99\%$ identity, < 15 kb) with lower sequence similarity compared with the other strain (11.8% in MAFF 305635-RFP and 13.4% in IMI 349063 of the total genomes). It is noted that the amounts of strain-specific regions increase when more stringent cutoffs were applied on whole-genome alignments (supplementary figs. S1 and S2, Supplementary Material online).

Whole-genome alignments also detected 19 synteny breakpoints resulting in large-scale genomic rearrangements between the two strains. The 19 synteny breakpoints included ten large-scale rearrangements between the two strains including six interchromosomal translocations and four intra-chromosomal inversions. 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 (supplementary fig. S3, Supplementary Material online). In addition, mapping of SOLiD reads derived from MAFF 305635 wild-type to the genome assembly of MAFF 305635-RFP also indicated that the RFP expression cassette was inserted into the genome without causing a rearrangement (supplementary fig. S4, Supplementary Material online).

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 (supplementary table S5, Supplementary Material online). 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 with the average coverage of all genes (supplementary fig. S5, Supplementary Material online). 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 whereas PCR products were detected in MAFF 305635 wild-type

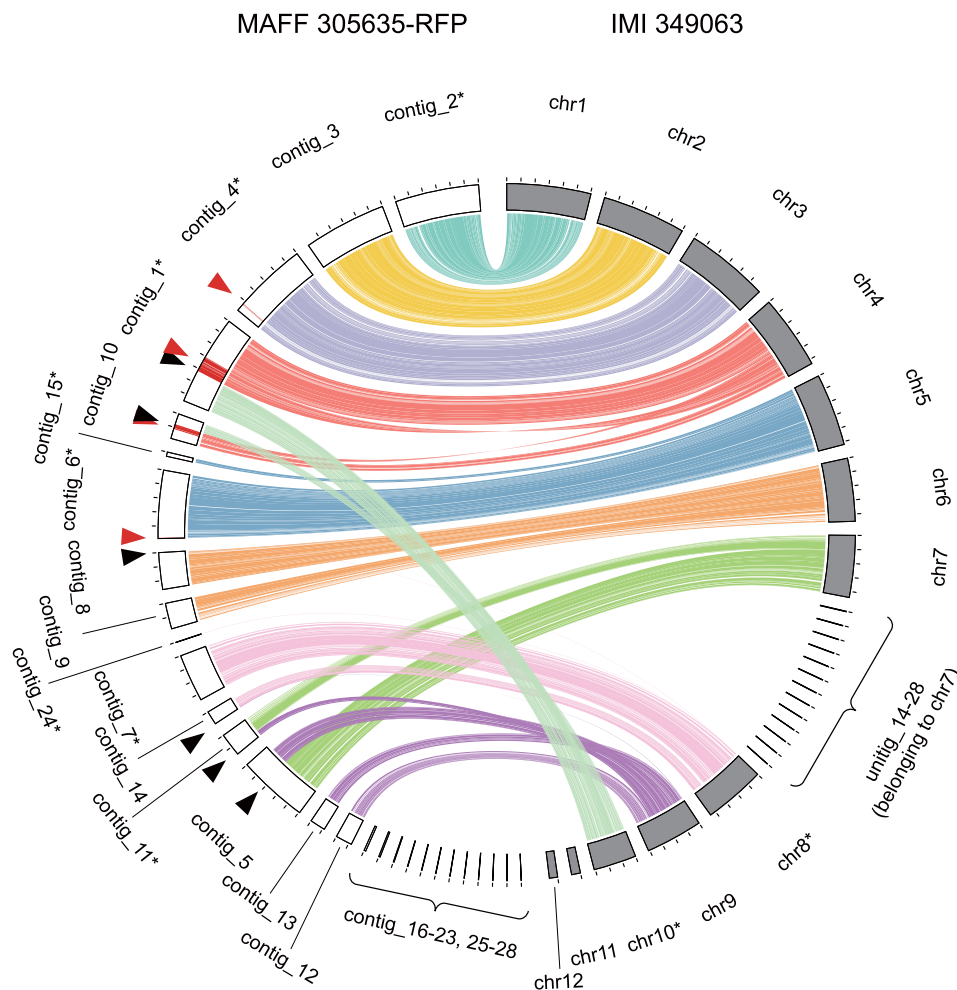


Fig. 1.—Whole-genome alignments between MAFF 305635-RFP and IMI 349063. The outer bands indicate contigs of MAFF 305635-RFP (white) and chromosomes of IMI 349063 (gray), respectively. Syntenic regions ($\geq 99\%$ identity, ≥ 15 kb) are linked with different colored ribbons corresponding to a chromosome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

(supplementary fig. S6, Supplementary Material online). 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 TEs in *C. higginsianum*

TEs are known to contribute to the generation of genomic variations. Thus, we 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 with the rest of the genome (supplementary fig.

S7, Supplementary Material online). 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 TEs and TcMar-Fot1 from Class II TEs. 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) (supplementary fig. S8, Supplementary Material online).

The association between genomic variations and TEs was assessed (fig. 2A and supplementary fig. S9, Supplementary Material online). The results showed that eight of ten large-scale rearrangements were within 10 kb of the nearest TE in MAFF 305635-RFP (fig. 2B and supplementary fig. S10, Supplementary Material online). 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

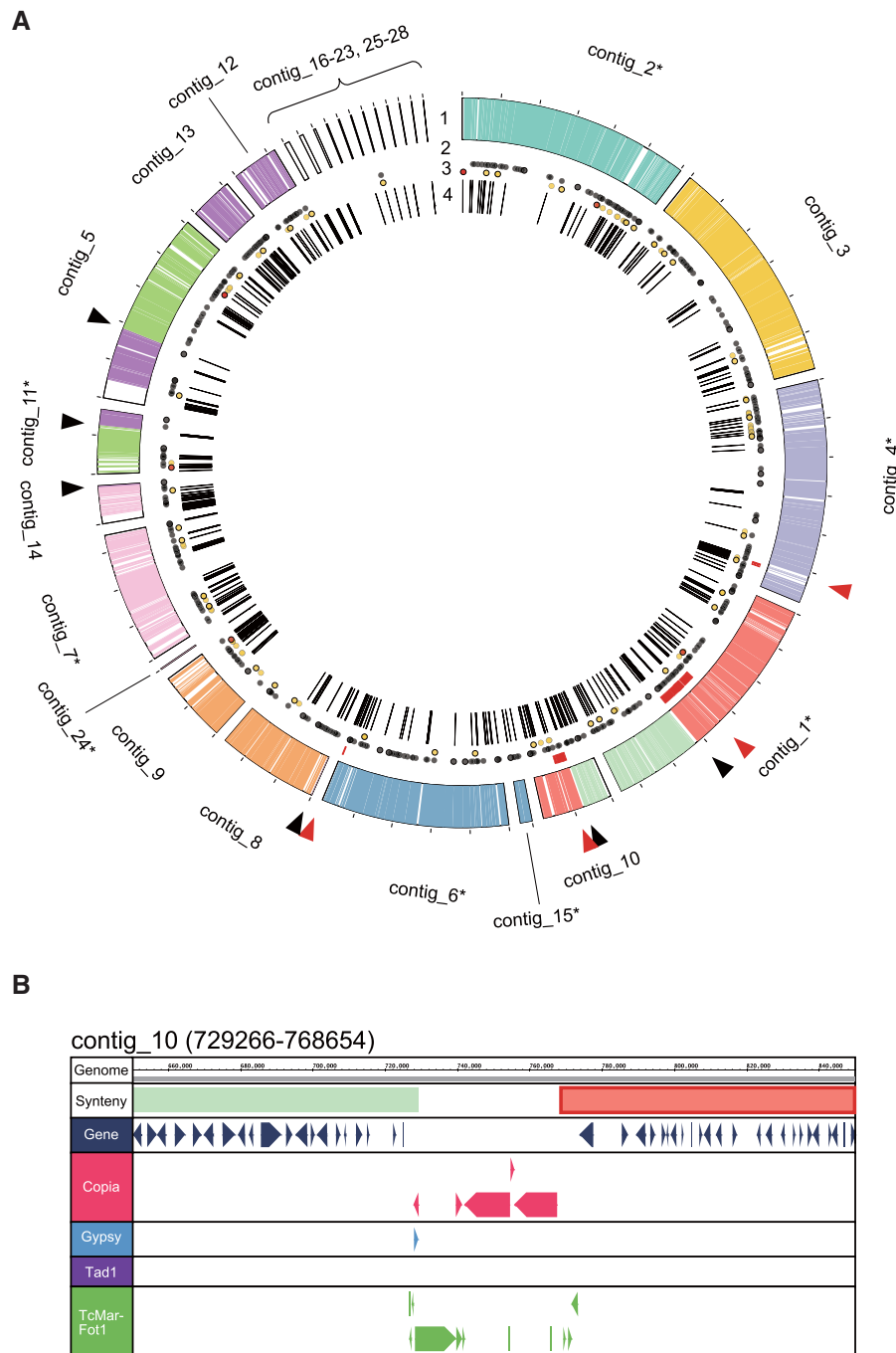


FIG. 2.—Association between genomic variations and TEs in *Colletotrichum higginsianum*. (A) Distribution of genomic features and TEs in the MAFF 305635-RFP genome. (1) Contigs. Colored regions indicate the presence of syntenic regions in the genome assembly of IMI 349063 ($\geq 99\%$ identity, ≥ 15 kb). Colors correspond to chromosomes of IMI 349063 as in figure 1. (2) Intrachromosomal 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 interchromosomal translocations and intrachromosomal inversions, respectively. Asterisks indicate reverse-complementation of contigs for visual clarity in figure 1. Ticks on the outer bands represent 1 Mb. (B) Association between large-scale rearrangements and TEs. A large-scale rearrangement on contig_10 from the genome assembly of MAFF 305635-RFP is shown as an example. Numbers in brackets indicate the location of the large-scale rearrangement. Synteny: pale green and salmon pink rectangles indicate syntenic regions in chromosomes 10 and 4 in IMI 349063 ($\geq 99\%$ identity, ≥ 15 kb), respectively. The red line surrounding a rectangle indicates a region of intrachromosomal inversion.

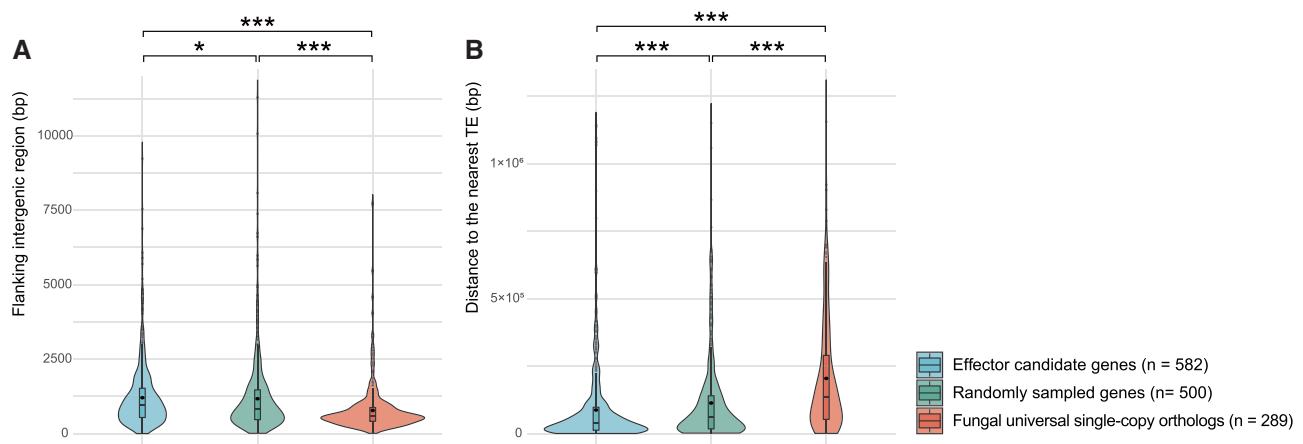


FIG. 3.—Compartmentalization of effector candidate genes and fungal universal single-copy orthologs in the MAFF 305635-RFP genome. (A) Violin plots showing the FIRs of effector candidate genes, randomly sampled genes, and fungal universal single-copy orthologs. (B) Violin plots showing the distances from effector candidate genes, randomly sampled 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 the three categories (* $P < 0.05$, *** $P < 0.001$, Wilcoxon rank sum test with adjustment by Holm's method).

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) (supplementary fig. S11, Supplementary Material online). In this study, we defined strain-specific regions using the default settings of the genome alignment program nucmer, which only considers matches of sequences that are unique in the reference. However, this may result in repeat sequences being overrepresented as strain-specific regions. Therefore, we repeated nucmer with different settings to include sequences that are not unique in the reference as well. In this case, we found that 19.2% in MAFF 305635-RFP and 25.1% in IMI 349063 of strain-specific regions were occupied by TEs, indicating that these rates are still significantly higher than random chance in both genomes ($P < 0.001$, Monte Carlo test) (supplementary fig. S11, Supplementary Material online).

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, we predicted the effector candidates of *C. higginsianum*. Proteins were classified as effector candidates if they were predicted to be secreted proteins with lengths of <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). Using version 1.0 or 2.0 of the effector prediction program EffectorP (Sperschneider et al. 2016, 2018), 378 (64.9%) in MAFF 305635-RFP and 353 (61.3%) in IMI 349063 proteins were also predicted as effector candidates (supplementary tables S6 and S7, Supplementary Material online). 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 (supplementary fig. S12, Supplementary Material online).

Next, we assessed the FIRs of effector candidate genes, fungal universal single-copy orthologs, and randomly sampled genes excluding these two classes to determine if these genes are in gene-dense or gene-poor genomic regions. The distributions of FIRs are significantly different between effector candidate genes, randomly sampled genes, and fungal universal single-copy orthologs in both strains (Wilcoxon rank sum test with adjustment by Holm's method) (fig. 3A and supplementary fig. S13A, Supplementary Material online). Two-dimensional plots describing 5' and 3' FIRs also 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 (supplementary fig. S14, Supplementary Material online). The distances between the three categories of genes and their nearest TEs were also investigated. The distributions of distances from the nearest TEs are significantly different between all pair-wise combinations of the three categories (Wilcoxon rank sum test with adjustment by Holm's method) (fig. 3B and supplementary fig. S13B, Supplementary Material online). 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 or randomly sampled genes and their closest TEs.

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

To provide insight into the adaptive evolution of *C. higginsianum* effectors, variations in effector candidates were

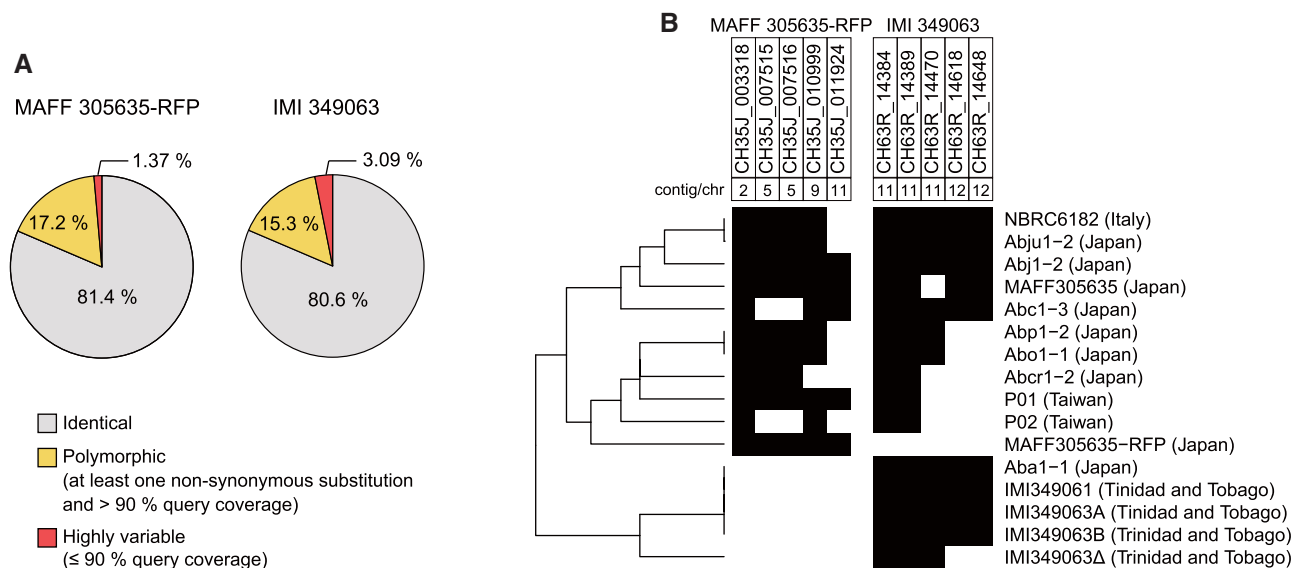


Fig. 4.—Variations in effector candidate genes between *Colletotrichum higginsianum* strains. (A) Pie charts showing percentages of effector candidates with different levels of variations. Gray, yellow, and red indicate identical, polymorphic (having at least one nonsynonymous substitution and >90% query coverage), and highly variable (≤90% query coverage) effector candidates, respectively. (B) The presence/absence patterns of 10 highly variable effector candidate genes among 16 *C. higginsianum* strains. Black and white squares indicate the presence and absence of these genes, respectively. The dendrogram shows hierarchical clusters of 16 strains based on the presence/absence patterns of 10 highly variable effector candidate genes.

investigated (fig. 4A and supplementary tables S6 and S7, Supplementary Material online). 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 (supplementary table S8, Supplementary Material online). 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 16 *C. higginsianum* strains isolated from different geographic locations, including MAFF 305635-RFP and IMI 349063 (supplementary table S1, Supplementary Material online). Molecular phylogenetic analysis, using all available sequences from strains from the *C. destructivum* species complex described by Damm et al. (2014), confirmed that all 16 strains were classified as *C. higginsianum*

(supplementary figure S15, Supplementary Material online). Amplification of these effector candidate genes by PCR revealed that their conservation patterns varied within other strains as well (fig. 4B and supplementary fig. S16, Supplementary Material online). Nine strains isolated in Japan showed different patterns for the presence/absence of these genes. Notably, we found that MAFF 305635 wild-type has four of five highly variable effector candidate genes that were absent in MAFF 305635-RFP. In contrast, three strains from Trinidad and Tobago showed the same presence/absence patterns. In addition, we tested a strain subcultured from IMI 349063 that has lower conidia production rates than MAFF 305635-RFP and that causes less severe infection symptoms than MAFF 305635-RFP and IMI 349063 (supplementary fig. S17, Supplementary Material online). The results showed that this strain appears to lack chromosome 12, as no PCR bands were detected in all tested genes in chromosome 12 (supplementary figs. S16 and S18, Supplementary Material online). Thus, this strain was named IMI 349063Δ.

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

To explore the potential mechanisms of how *C. higginsianum* acquired highly variable effector candidate genes, we investigated the conservation patterns of these genes amongst 25 ascomycetes (fig. 5 and supplementary table S3, Supplementary Material online). The results showed that

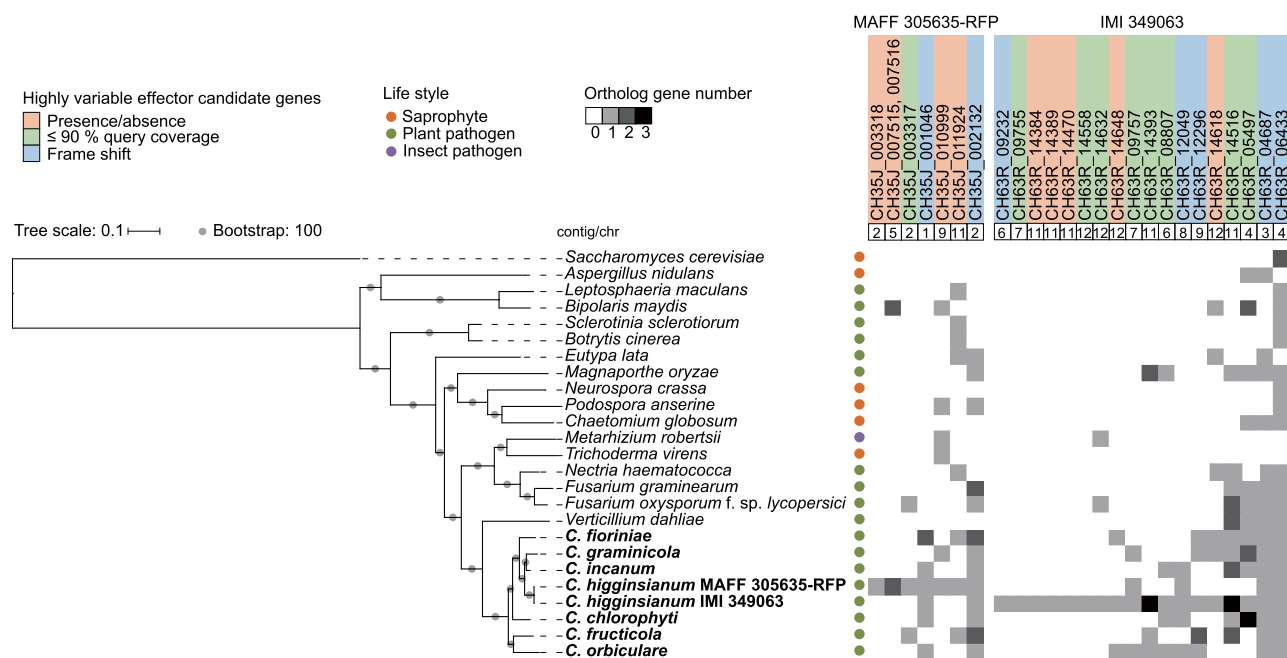


FIG. 5.—Conservation patterns of highly variable effector candidate genes in Ascomycota. Bold letters indicate *Colletotrichum* fungi. The maximum-likelihood species phylogeny was drawn based on the alignment patterns of single-copy orthologs obtained using OrthoMCL. Bootstrap values are percentages based on 1,000 bootstrap replicates.

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 only 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 based on OrthoMCL analysis. Interestingly, the paralogs showed high similarities to proteins found only in *Bipolaris maydis* as shown in figure 5. Querying the protein sequences of the paralogs against the NCBI nucleotide collection revealed that only *Bipolaris maydis*, *Bipolaris oryzae*, *Bipolaris victoriae*, *Bipolaris zeicola*, *Bipolaris sorokiniana*, *Aspergillus novofumigatus*, and *Aspergillus terreus* have similar sequences (TblastN, E -value $\leq 2 \times 10^{-28}$) (supplementary fig. S19, Supplementary Material online).

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

(fig. 6 and supplementary fig. S20, Supplementary Material online). Syntenic regions containing 17 of 26 highly variable effector candidate genes were reconstructed. We found that eight (six in MAFF 305635-RFP and two in IMI 349063) of 17 such genes were in syntenic-disrupted regions associated with TEs. However, nine highly variable effector candidate genes were not associated with syntenic-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, 9 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 (supplementary table S9, Supplementary Material online).

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

To examine the impact of the genetic differences observed between the two sequenced isolates, we assessed the ability

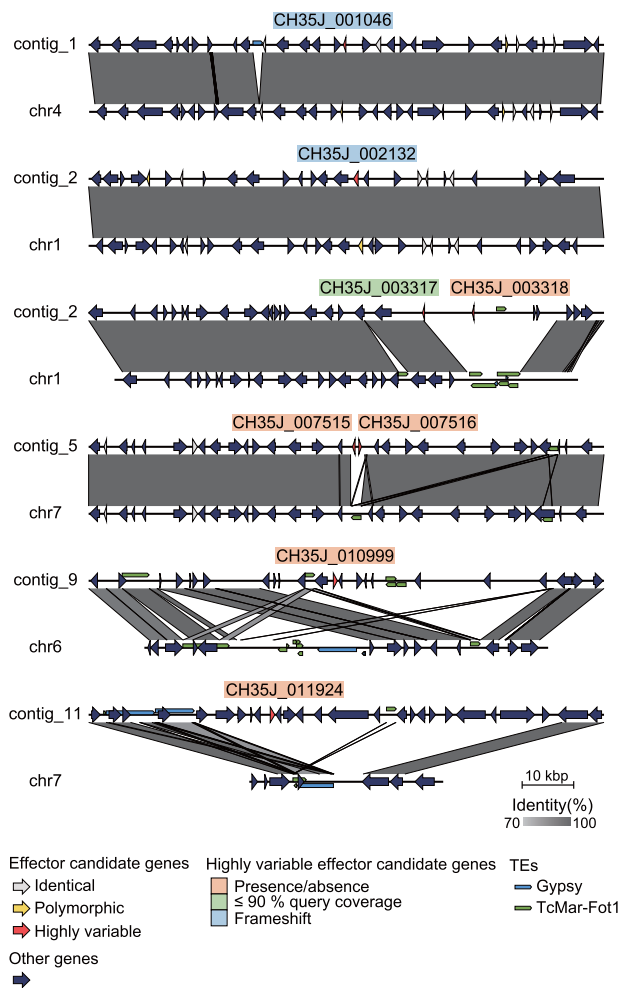


Fig. 6.—Analysis of genome synteny in regions containing highly variable effector candidate genes from MAFF 305635-RFP. Upper: genomic regions containing highly variable effector candidate genes from MAFF 305635-RFP. Lower: syntenic regions from IMI 349063. It is noted that classification of effector candidates (identical, polymorphic, or highly variable) is sometimes inconsistent between orthologs of the two strains due to differences in gene annotation models used. For example, the MAFF 305635-RFP genome has a single nucleotide polymorphism that causes a frameshift of the CH35J_002132 gene model, and thus this gene model is classified as highly variable. However, the orthologous IMI 349063 gene model does not include the frameshifted region and is classified as polymorphic when the IMI 349063 gene model is applied to the MAFF 305635-RFP genome.

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 with Ws-2. However, no significant differences were detected between the strains (supplementary figs. S21 and S22, Supplementary Material online).

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. In this study, we present a highly contiguous genome assembly of a strain of *C. higginsianum*. By performing comparative genomics using this assembly and a chromosome-level assembly of another *C. higginsianum* strain (Zampounis et al. 2016), we show, for the first time in the *Colletotrichum* genus, 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 and Miao 1992). In the *Colletotrichum* genus, few sexual morphs have been described (Vaillancourt and Hanau 1991; Rodríguez-Guerra et al. 2005; Menat et al. 2012) and most species, including *C. higginsianum*, are considered as predominantly asexual. Our results 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.

Our analyses indicate that TEs contribute to the generation of large-scale rearrangements and strain-specific regions in *C. higginsianum*. Seidl and 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 TEs, 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). In contrast to Dallery et al. (2017) who previously reported that Class I TEs are more abundant than Class II in the genome of IMI 349061, the analyses in this study found Class II TEs to be more abundant. However, the overall TE coverage of the IMI 349061 genome was found to be similar (~7%) according to both pipelines, if ignoring the overlaps between TEs. We speculate that nesting of TEs and subsequent divergence of TE sequences after duplication may complicate classification resulting in the discrepancies between the two pipelines.

We 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. This finding is consistent with Dallery et al. (2017) who reported that not only effector candidate genes but also genes in

Downloaded from https://academic.oup.com/gbe/article/11/5/1487/5480666 by guest on 20 April 2024

secondary metabolism clusters that may be involved in pathogenicity tend to be closer to TEs in *C. higginsianum* IMI 349061. These so-called “two-speed genomes” have also been found in other eukaryotic plant pathogens, such as *Phytophthora infestans* and *Leptosphaeria maculans* (Raffaële 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 housekeeping genes from the deleterious effects of TEs.

Our analysis identified 26 highly variable effector candidate genes between two strains of *C. higginsianum*. Further, we observed the presence/absence polymorphisms of ten highly variable effector candidate genes in 16 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 toward 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 noncoding regions (Tautz and Domazet-Lošo 2011).

Our analysis also identified *C. higginsianum* effector candidate genes showing uneven conservation patterns in Ascomycota, such as CH35J_007515 and CH35J_007516. Such conservation patterns 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, we 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.

Although the protoplast transformation used to introduce an RFP expression cassette into MAFF 305635 might cause the loss of its minichromosomes, the loss of minichromosomes from MAFF 305635-RFP may also be due to the unstable nature of minichromosomes in this species, as described by Plaumann et al. (2018). In that study, the authors found that the rate of spontaneous minichromosome loss from in vitro-cultured *C. higginsianum* was more than 1×10^{-4} . Also, the loss of chromosome 12 from IMI 349063 Δ appears to have occurred spontaneously. 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 toward 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). Our 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*, we identified genomic variations in the structure and genes encoding effector candidates and potential mechanisms of altering the genome mediated by TEs in this species. Our results improve the understanding of adaptation driven by genomic evolution in this scientifically and agriculturally important group of plant pathogenic fungi.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported in part by KAKENHI 17H06172 to K.S., the Science and Technology Research Promotion Program for Agriculture, Forestry, Fisheries and Food industry to Y.N., Y.T., JSPS Grant-in-Aid for JSPS Research Fellow to A.T. (17J02983), and JSPS Overseas Challenge Program for Young Researchers to A.T. (201880065). We thank Richard O’Connell for kindly providing *C. higginsianum* strains Aba1-1, Abc1-3, Abcr1-2, Abj1-2, Abju1-2, Abo1-1, Abp1-2, IMI 349061, IMI 349063A, IMI 349063B, NBRC 6182, P01, and

P02. P01 and P02 were used with permission from the original isolator, Miin-Huey Lee. We also thank Yasuyuki Kubo for kindly providing *C. higginsianum* strain IMI 349063Δ. Computations were partially performed on the NIG Supercomputer at the ROIS National Institute of Genetics.

Literature Cited

- Asai S, Shirasu K. 2015. Plant cells under siege: plant immune system versus pathogen effectors. *Curr Opin Plant Biol.* 28:1–8.
- Ayhan DH, López-Díaz C, Di Pietro A, Ma L-J. 2018. Improved Assembly of Reference Genome *Fusarium oxysporum* f. sp. *lycopersici* Strain Fol4287 *Microbiol. Resour. Announc.* 7:e00910-18.
- 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(1):13.
- Bao J, et al. 2017. PacBio sequencing reveals transposable elements as a key contributor to genomic plasticity and virulence variation in *Magnaporthe oryzae*. *Mol Plant* 10(11):1465–1468.
- Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11.
- Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12(8):1269–1276.
- Baroncelli R, Sreenivasaprasad S, Sukno S, Thon MR, Holub E. 2014. Draft genome sequence of *Colletotrichum acutatum sensu lato* (*Colletotrichum fioriniae*). *Genome Announc.* 2:e00112–e00114.
- Baroncelli R, et al. 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics.* 17:555.
- Birker D, et al. 2009. A locus conferring resistance to *Colletotrichum higginsianum* is shared by four geographically distinct *Arabidopsis* accessions. *Plant J.* 60(4):602–613.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
- Castanera R, et al. 2016. Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genet.* 12:1–27.
- Chin C-S, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods.* 10(6):563–569.
- Chuma I, Tosa Y, Taga M, Nakayashiki H, Mayama S. 2003. Meiotic behavior of a supernumerary chromosome in *Magnaporthe oryzae*. *Curr Genet.* 43(3):191–198.
- Chuma I, et al. 2011. Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. *PLoS Pathog.* 7(7):e1002147.
- Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog.* 8(4):e1002608.
- Croll D, Zala M, McDonald BA. 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, et al. 2014. The genomics of *Colletotrichum*. In: Dean RA, Lichens-Park A, Kole C, editors. *Genomics of plant-associated fungi: monocot pathogens*. Berlin/Heidelberg (Germany): Springer. p. 69–102.
- Dallery J-F, et al. 2017. Gapless genome assembly of *Colletotrichum higginsianum* reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. *BMC Genomics.* 18(1):667.
- Damm U, O’Connell RJ, Groenewald JZ, Crous PW. 2014. The *Colletotrichum destructivum* species complex—hemibiotrophic pathogens of forage and field crops. *Stud Mycol.* 79:49–84.
- de Jonge R, et al. 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(13):5110–5115.
- de Jonge R, et al. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res.* 23(8):1271–1282.
- Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet.* 11(8):539–548.
- Dodds PN, et al. 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(23):8888–8893.
- Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev.* 35:57–65.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):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 studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J Mol Biol.* 337(2):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.
- Faino L, et al. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *mBio* 6:e00936–e00915.
- Faino L, et al. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* 26(8):1091–1100.
- Frantzeskakis L, et al. 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 NH, Norris DC, Loraine AE. 2016. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32(14):2089–2095.
- Gan P, et al. 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* 197(4):1236–1249.
- Gan P, et al. 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(5):1467–1481.
- Gan P, et al. 2017. Draft genome assembly of *Colletotrichum chlorophyti*, a pathogen of herbaceous plants. *Genome Announc.* 5:e01733–e01716.
- Giraldo MC, Valent B. 2013. Filamentous plant pathogen effectors in action. *Nat Rev Microbiol.* 11(11):800–814.
- Grandaubert J, et al. 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.
- Hacquard S, et al. 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun.* 7:11362.
- Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME J.* 11(5):1189–1204.
- Hatta R, et al. 2002. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics* 161(1):59–70.
- Hiruma K, et al. 2010. Entry mode-dependent function of an indole glucosinolate pathway in *Arabidopsis* for nonhost resistance against anthracnose pathogens. *Plant Cell* 22(7):2429–2443.

- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32(5):767–769.
- 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(14):3059–3066.
- Kim D, et al. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14(4):R36.
- Kistler HC, Miao V. 1992. New modes of genetic change in filamentous fungi. *Annu Rev Phytopathol.* 30:131–153.
- Kleemann J, et al. 2012. Sequential delivery of host-induced virulence effectors by Appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* 8(4):e1002643.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 305(3):567–580.
- Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5(2):R12.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- 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(W1):W242–W245.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189.
- Lomsadze A, Burns PD, 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, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464(7287):367–373.
- Menat J, Cabral AL, Vijayan P, Wei Y, Banniza S. 2012. *Glomerella truncata*: another *Glomerella* species with an atypical mating system. *Mycologia* 104(3):641–649.
- Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol.* 15(12):756–771.
- Moreno-Hagelsieb G, Latimer K. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24(3):319–324.
- 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(3):e1293218.
- Narusaka Y, et al. 2004. *RCH1*, a locus in *Arabidopsis* that confers resistance to the hemibiotrophic fungal pathogen *Colletotrichum higginsianum*. *Mol Plant Microbe Interact.* 17(7):749–762.
- Narusaka M, et al. 2009. *RRS1* and *RPS4* provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant J.* 60(2):218–226.
- O'Connell R, et al. 2004. A novel *Arabidopsis*–*Colletotrichum* pathosystem for the molecular dissection of plant–fungal interactions. *Mol Plant Microbe Interact.* 17:272–282.
- O'Connell RJ, et al. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat Genet.* 44:1060–1065.
- Petersen TN, Brunak S, Von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8(10):785–786.
- 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 AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
- Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol.* 10(6):417–430.
- Raffaele S, et al. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330(6010):1540–1543.
- Rodríguez-Guerra R, et al. 2005. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. *Mycologia* 97(4):793–803.
- Saunders DGO, 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, editors. *Methods in molecular biology (methods and protocols)*. Totowa (NJ): Humana Press. p. 29–51.
- Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 9(7):671–675.
- Seidl MF, Thomma B. 2014. Sex or no sex: evolutionary adaptation occurs regardless. *BioEssays* 36(4):335–345.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.
- Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol.* 19(9):2094–2110.
- Sperschneider J, et al. 2016. Effector P: predicting fungal effector proteins from secretomes using machine learning. *New Phytol.* 210(2):743–761.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):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 MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27(7):1009–1010.
- Takahara H, et al. 2016. *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. *New Phytol.* 211(4):1323–1337.
- Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet.* 12(10):692–702.
- Vaillancourt LJ, Hanau RM. 1991. A method for genetic analysis of *Glomerella graminicola* (*Colletotrichum graminicola*) from maize. *Phytopathology* 81(5):530–534.
- Win J, et al. 2019. Nanopore sequencing of genomic DNA from *Magnaporthe oryzae* isolates from different hosts. *Zenodo*. doi:10.5281/zenodo.2564950.
- Yoshida K, et al. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* 21(5):1573–1591.
- Yu C, Zhang J, Peterson T. 2011. Genome rearrangements in maize induced by alternative transposition of reversed *Ac/Ds* termini. *Genetics* 188(1):59–67.
- Zampounis A, et al. 2016. Genome sequence and annotation of *Colletotrichum higginsianum*, a causal agent of crucifer anthracnose disease. *Genome Announc.* 4:e00821–e00816.
- Zhang J, et al. 2009. Alternative *Ac/Ds* transposition induces major chromosomal rearrangements in maize. *Genes Dev.* 23(6):755–765.

Associate editor: Yves Van De Peer