The genome architecture of the fungal plant pathogens *Cladosporium fulvum* and *Erysiphe necator* and its relevance to pathogenicity

By

ALEX ZANELLA ZACCARON
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHYLOSOFY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Ioannis Stergiopoulos, Chair

Richard Michelmore

Steven Knapp

Committee in Charge

2024

Abstract

Fungi are diverse eukaryotic microorganisms with pivotal roles in ecosystems, but also notorious pathogens causing significant economic losses. Understanding the mechanisms underlying fungal pathogenicity is crucial for devising effective strategies to mitigate their negative impact. Fungi utilize various mechanisms to infect plants, including secreting effector proteins that promote virulence on susceptible hosts or trigger immune responses (i.e. avirulence) in resistant plants carrying resistance genes. To retain their virulence properties and abolish their avirulence ones, effectors often accumulate mutations in their coding sequence or are entirely deleted from the pathogen genome. Thus, knowing the mutability of effectors is crucial to selecting the right resistance genes when breeding for durable resistance. The tomato pathogen *Cladosporium fulvum* showcases skewed types of mutations in its effector genes to overcome resistance in tomato, which depending on the effector range from point mutations to complete gene deletions. This observation indicates that the type and frequency of mutations accumulating in effectors are to an extent driven by the genes' genomic location and the propensity of these genomic regions to structural variations (SVs). However, the genome architecture of *C. fulvum* and its landscape of SVs were never investigated. Similarly, SVs are thought to be an important source of adaptation in the grape powdery mildew fungus *Erysiphe necator*, economically the most important foliar pathogen on this crop, in which gene duplications have been already associated with the development of resistance to fungicides. In this dissertation, I seek to address these gaps by obtaining high-quality chromosome-level genome assemblies and annotations for *C. fulvum* and *E. necator*, and investigating the extent to which their genomic architecture and SVs contribute to their evolution and pathogenicity. The genome of *C. fulvum* is organized into a variable number of 13 to 15 chromosomes, as two of them are dispensable for fungal growth and pathogenicity. The chromosomes of *C. fulvum* exhibit a peculiar 'checkerboard' pattern of gene-rich/repeat-poor regions, interspersed with gene-poor/repeat-rich regions. Comparisons with an additional five isolates of *C. fulvum* revealed that nearly all SVs corresponded to insertions or deletions in regions rich in transposable elements

(TEs). Notably, three SVs that were likely induced by TEs effected the deletion of the effector genes *Avr9*, *Avr5*, and *Avr4E*, thereby mediating an escape of pathogen recognition by the cognate *Cf-9, Cf-5* and *Cf-4E* resistance genes in tomato. In this dissertation we also investigated the landscape of alternative splicing (AS) events in *C. fulvum* genes during a complete infection cycle. The analysis showed that nearly 40% of the protein-coding genes in *C. fulvum* were AS at some stage during the infection process, suggesting that AS could have a role in finetuning infections of the host. Comparison of the location of the AS genes in the genome of *C. fulvum* revealed that AS genes are more abundant in repeat-rich core chromosomes, and exhibit significant longer 5' intergenic regions richer in repetitive DNA compared to non -AS genes, indicating that the genome organization could have an effect on the occurrence of AS. Our studies on the grape powdery mildew pathogen *E. necator* showed that its genome is organized into 11 chromosomes which do not exhibit large-scale compartmentalization into gene-rich and repeat-rich regions. A total of 13.1% of the genes in *E. necator* were predicted to be duplicated and were particularly enriched for genes encoding candidate effectors. Comparative analysis among six isolates of *E. necator* revealed a total of 122 genes that varied in their copy numbers. One of these varied from 1 to 31 copies and encoded a putative secreted carboxylesterase (CE), which is a member of a novel family of CEs that is unique to powdery mildew fungi. Next to the nuclear genome, the organization of the mitochondrial genome of *E. necator* and that of other powdery mildew fungi were also analyzed. Comparative genomics among *E. necator* and three other species of powdery mildew fungi revealed a wide variation of mitochondrial genome sizes, ranging from 109.8 kb in *B. graminis* f. sp. *tritici* to 332.2 kb in *G. cichoracearum*, which has the largest mitochondrial genome of a fungal pathogen reported to date. Finally, the introns of the cytochrome *b* gene, which encodes for the target site of QoI fungicides, of powdery mildew fungi contained rare open reading frames encoding reverse transcriptases that were likely acquired horizontally. Collectively, the results presented in this dissertation reveal new evolutionary aspects of the genomes of *C. fulvum* and *E. necator*, and highlight the importance of genome organization and genomic structural variations in overcoming host resistance.

Acknowledgments

I am grateful to many colleagues who supported me throughout my academic journey and made the completion of this dissertation possible. First, I express my gratitude to my major dissertation advisor, Prof. Ioannis Stergiopoulos, for his guidance and dedication in elevating the quality of my work. Under his mentorship, I gained much knowledge and evolved into a more proficient academic. I am also grateful for the support of current and past members of the Stergiopoulos lab, including Dr. Anastasios Samaras, Prof. Jorge T. De Souza, and Dr. Li-Hung Chen, with whom I had the pleasure to collaborate throughout my doctoral studies. I also thank Dr. Sunil Yadav and Sotirios Pilafidis for their wonderful companionship.

I extend my gratitude to Dr. Shree Thapa for his friendship and advice while sharing the 'bioinformatics office' with me during the first years of my PhD. I also thank the Stergiopoulos lab collaborators from the FRAME networks group for their support, particularly Tara Neill, who did an extraordinary job extracting enough DNA from an obligate biotrophic plant pathogen for sequencing.

I am also grateful to the personnel from IGG, notably Najwa Marrush, Victoria Torres, Prof. Sean Burgess, and Prof. David Segal, who worked wonders in keeping the group functional and contributed to a good experience not only for me but also for other IGG students. I also thank Jordan Dade for assisting me in finding a source of income as a teaching assistant in the Food Science and Technology department when funding became scarce.

I am deeply grateful to Prof. Richard Michelmore and Prof. Steven Knapp for kindly agreeing to serve on my dissertation committee.

Finally, my deepest appreciation goes to my family, my parents Oscar and Nilsa, and my siblings Marcio, Grasi, and Michelle. Their love and support made this long journey thousands of miles away from home much more pleasant.

Table of Contents

List of Figures

List of Tables

Chapter 1

The dynamics of fungal genome organization and its impact on host adaptation and antifungal resistance

Alex Z. Zaccaron
Ioannis Stergiopoulos

Leading author statement

I contributed to the conceptualization of this chapter, performed all analyses, generated all figures, contributed to the investigation and interpretation, wrote the original draft, and contributed to the review and editing of the final draft.

―――――――――――――――

A modified version of this chapter was submitted to the Journal of Genetics and Genomics (submission ID: JGG-D-24-00832).

1.1 The positive and negative impacts of fungi on our ecosystem and society

Fungi comprise a diverse kingdom of eukaryotic microorganisms with pivotal roles in nature, including in decomposition and nutrient cycling in terrestrial ecosystems and the establishment of mutualistic or parasitic associations with other organisms. Fungi also exhibit remarkable diversity, ranging from obligate intracellular species of Microsporidia (Wadi et al., 2023) to the largest organism on Earth, *Armillaria ostoyae*, which spans nearly 9.5 km^2 and is estimated to be 8,650 years old (Scheppke, 2023). To date, at least 155,000 species of fungi have been described but the total number of fungal species is estimated to 2.5 million or higher, thereby making fungi the second largest kingdom of eukaryotes after animals, with an estimated 8.5 million species of invertebrates (Niskanen et al., 2023). Despite the staggering number of fungal species, only an estimated 5% to 8% of them have been described, which makes fungi proportionally the largest group of eukaryotes with uncharacterized species (Niskanen et al., 2023).

Fungi directly influence both the environment and our society. In the pharmaceutical industry, fungi are a valuable source of natural bioactive compounds that can find a wide-range of applications, including as immunosuppressants (Strader et al., 2011), antimicrobials (Kaaniche et al., 2019), anti-inflammatory agents (Teimoori-Boghsani et al., 2020), antihyperlipidemic (Manzoni and Rollini, 2002), and even as anticancer drugs (Kanokmedhakul et al., 2012). Fungi are also gaining attention as a sustainable source of eco-friendly and cost-effective materials that find uses in a wide range of industries, including as substitutes of bovine leather, textiles, and high-performance paper (Gandia et al., 2021). In forests and fields, approximately 90% of all plants derive benefits from fungi through mycorrhizal symbiosis (Bonfante and Genre, 2010), in which the fungus acquires nutrition made available by the plant, while the fungal mycelia act as an extension of the plants' root system, thereby increasing nutrient and water absorption from the soil (Bonfante and Genre, 2010). Mycorrhizae fungi can also be used as biofertilizer that stimulates microbial activity in the rhizosphere, which further promotes transfer of nutrients from the soil into the plant and increases plant health and yield (Benami et al., 2020).

Despite the multitude of benefits that fungi provide across various ecosystems, they can also pose serious threats to plant, human, and animal health. In agriculture, it has been estimated that up to 200 USD billion are lost every year due to fungal diseases that reduce crop yields (Elliott, 2022). The global annual burden of human fungal diseases is also concerningly high (Denning, 2024). Estimates are that every year fungi infect a billion people, causing 1.5 million deaths worldwide (Bongomin et al., 2017). One of the reasons for this staggering death toll is the drastic increase of immunocompromised individuals since the 1950s, due to the use of immunosuppressive agents, cancer chemotherapy, and the advent of diseases that negatively impact the immune system, such as AIDS (Fisher et al., 2020; Rokas, 2022). Among fungi infecting vertebrates, chytridiomycosis caused by species of *Batrachochytrium* has thus far driven to near extinction 90 amphibian species and caused over 90% population reductions in another 124 species, thereby making it the most devastating fungal disease ever recorded on vertebrates (Scheele et al., 2019) and the one that eradicated the most biodiversity on the planet (Stokstad, 2019). These records make it evident that fungi are one of the most important groups of pathogens threatening food supply as well as human and animal health.

A remarkable feature of fungi is their ability to rapidly adapt to adverse conditions in response to selection pressure. Key to this rapid adaptation is their efficient reproductive strategies, that can be sexual and/or asexual, coupled with mutations in their genomes that extend from point mutations affecting a single nucleotide (Bentham et al., 2021), to large-scale changes such as whole-chromosome duplications (Gorkovskiy and Verstrepen, 2021). Recent advances in whole-genome sequencing technologies allow the in-depth study of genomes at the chromosome level. These advances have identified mutations and characterized peculiar patterns of fungal genome organization that elucidate many aspects of the evolution of fungi and their adaptation to different hosts and environmental conditions (Auxier et al., 2023; van Westerhoven et al., 2023; Winter et al., 2018; Yildirir et al., 2022).

This review delves into fundamental aspects of fungal genomes, exploring their chromosomal organization and mechanisms of evolution. Emphasis is on ascomycete fungi but examples or also drawn as needed from other fungal phyla as well. Finally, this chapter concludes with detailed review of the genomes of the tomato pathogen *Cladosporium fulvum* and the grape pathogen *Erysiphe necator*, as these two fungal pathogens are the focus in this dissertation.

1.2 Current state of fungal genome assemblies

To date, least 21,848 fungal genome assemblies of species from nearly all major fungal lineages have been obtained and made publicly available in genome repositories. The most comprehensive repository for fungal genomes is GenBank from the National Center for Biotechnology Information (NCBI), containing a total of 17,789 fungal genomes, followed by *MycoCosm* (Grigoriev et al., 2014), containing 2,554 fungal genomes, and *EnsemblFungi* (Kersey et al., 2010), containing 1,505 fungal genomes. Of the 17,789 fungal genome assemblies in GenBank, 407 (2.3%) represent gapless chromosome level assemblies, 1,092 (6.1%) are chromosome level too but contain gaps or unplaced contigs, and the remaining 16,290 (91.6%) are assembled at scaffold or contig level. The fungal genome assemblies available in GenBank have a heavily skewed taxonomic distribution. Specifically, 95.2% of all fungal genome assemblies are from species of the subkingdom Dikarya (phyla Basidiomycota and Ascomycota), followed by 3.5% from species of the Zygomycota (phyla Mucoromycota and Zoopagomycota), and only 1.3% from zoosporic fungi (phyla Chytridiomycota, Blastocladiomycota, Microsporidia, and Cryptomycota) (Table 1.1). The fungal genome assemblies considered complete also show skewed taxonomic distribution. From the 407 fungal genomes completed, 306 (75.2%) are from species of the Ascomycota, 95 (23.3%) are from species of the Basidiomycota, while the remaining 6 are from species of the other six fungal phyla. An impressive 80% of all fungal genome assemblies available in GenBank are from species of the Ascomycota, which included the species with the largest number of genomes available. These are *Saccharomyces cerevisiae* (*n*=1,458 genomes), followed by the plant pathogens *Fusarium oxysporum* (*n*=415 genomes) and *Magnaporthe* *oryzae* (*n*=377 genomes). As expected, Ascomycota also has the largest number of 4,635 species with sequenced genomes, followed by 1,407 species from Basidiomycota and 225 species from Mucoromycota with sequenced genomes. In contrast, a mere 151 species of zoosporic fungi have their genomes sequenced. Combined, a total of 6608 fungal species have sequenced genomes available in GenBank (Table 1.1).

Table 1.1: Summary of the number of genome assemblies for fungal species available in GenBank. The table shows the total number of genome assemblies, reported statuses of assemblies, and the number of species that have sequenced genomes. Numbers were obtained as of April 26th, 2024.

Phylum	Genome	Genome assembly level			Species with
	assemblies	Complete	Chromosome	Contig/scaffold	genome assemblies
Basidiomycota	2579	95	89	2395	1407
Ascomycota	14356	306	989	13061	4635
Mucoromycota	404	1	8	395	225
Zoopagomycota	212	0	0	212	190
Chytridiomycota	114	0	0	114	72
Blastocladiomycota	10	0	0	10	8
Microsporidia	103	5	6	92	69
Cryptomycota	3	0	0	3	2

1.3 Differences in fungal genome composition and its impact on the genome architecture

1.3.1 Size of fungal genomes

The remarkable taxonomic diversity in the fungal kingdom is also reflected in the large diversity in fungal genome sizes. An analysis of 1,994 fungal genome assemblies with gene annotation from diverse fungal species available in GenBank, showed that fungal genome sizes range from 2.2 Mb to 1.1 Gb, with a mean size of 40.8 Mb and a median size of 35.4 Mb (Fig 1.1A). In comparison, the mean size of the human genome is estimated at 3.05 Gb (Nurk et al., 2022), and the mean size of the green plants genomes is estimated at 1.21 Gb (Kress et al., 2022), meaning that the mean fungal genome size is 75 times and 30 times smaller, respectively. The large differences in fungal genome sizes are only partially reflected in the species' divergence time and evolutionary relationships. In this respect, zoosporic fungi from Microsporidia and

Cryptomycota, which stand at the basal branch of the fungal kingdom (Spatafora et al., 2017), have on average the smallest genome sizes with a mean size of 7.9 Mb and 9.5 Mb, respectively (Wadi et al., 2023) (Fig 1.1A). Included in Microsporidia is the mammalian pathogen *Encephalitozoon cuniculi*, which has a genome size of only 2.9 Mb, one of the smallest genomes recorded thus far in eukaryotes (Katinka et al., 2001; Wadi et al., 2023). Some Microsporidia species, however, have considerably larger genome sizes, as is the case for example the mosquito parasite *Edhazardia aedis* with a genome size of 51.3 Mb (Haag et al., 2020), and the gammarids parasite *Dictyocoela muelleri* with a genome size of 41.9 Mb (Cormier et al., 2021). Such genome sizes are comparable to those from species of their sister phyla of Blastocladiomycota and Chytridiomycetes (Spatafora et al., 2017), which have a mean genome size of 32.0 Mb and 39.0 Mb respectively. Similar observations regarding the disparity in genome sizes can be made when considering species in lineages of so-called terrestrial fungi, which include the Zygomycetes and the Dikarya (Spatafora et al., 2017). For instance, members of Mucoromycota have genome sizes ranging from 19.5 Mb in *Bifiguratus adelaidae* (Torres-Cruz et al., 2017) to 773.1 Mb in *Gigaspora margarita* (Venice et al., 2020), and a mean genome size of 76.0 Mb. In an analogous way, genome sizes within the Ascomycota vary greatly ranging from 7.3 Mb in *Pneumocystis wakefieldiae* (Cissé et al., 2021) to 192.8 Mb in *Tuber magnatum* (Murat et al., 2018), and a mean genome size of 37.8 Mb. In general, among the Ascomycota, the unicellular yeasts in the subphylum of Saccharomycotina have the smallest genomes as compared to the filamentous fungi in Pezizomycotina (Spatafora et al., 2017). For example, the baker's yeast *S. cerevisiae*, one of the most widely used model organisms for genetic research, has a genome size of only 12.07 Mb (Fig 1.1A), while members of the Pichiaceae family also have considerable small genome sizes of approximately 9 Mb (Hanson et al., 2021; Ramezani-Rad et al., 2003). In contrast, some fungi within the phylum of Basidiomycota can have genomes that surpass the one billion base pair mark (Fig 1.1A). This is the case, for instance, for some fungi within the subphylum of Pucciniomycotina that cause rust diseases in plants. Notably, recent efforts to assemble the genome of the soybean rust pathogen *Phakopsora pachyrhizi*, which typically has two haploid nuclei in its cells (i.e., dikaryotic), revealed a genome size of 1.1 Gb (Gupta et al., 2023). Similarly, the genome assembly of the myrtle rust pathogen *Austropuccinia psidii* revealed a haploid genome size of 1.0 Gb (Tobias et al., 2021). In general, Basidiomycete fungi, which include most mushrooms (Agaricomycotina) and fungi that cause rust (Pucciniomycotina) and smut (Ustilagomycotina) diseases on plants, have the largest genomes with a median of 41.6 Mb and mean of 51.4 Mb. Taken together, the above examples illustrate the vast disparity that exists in genome sizes among different fungal species.

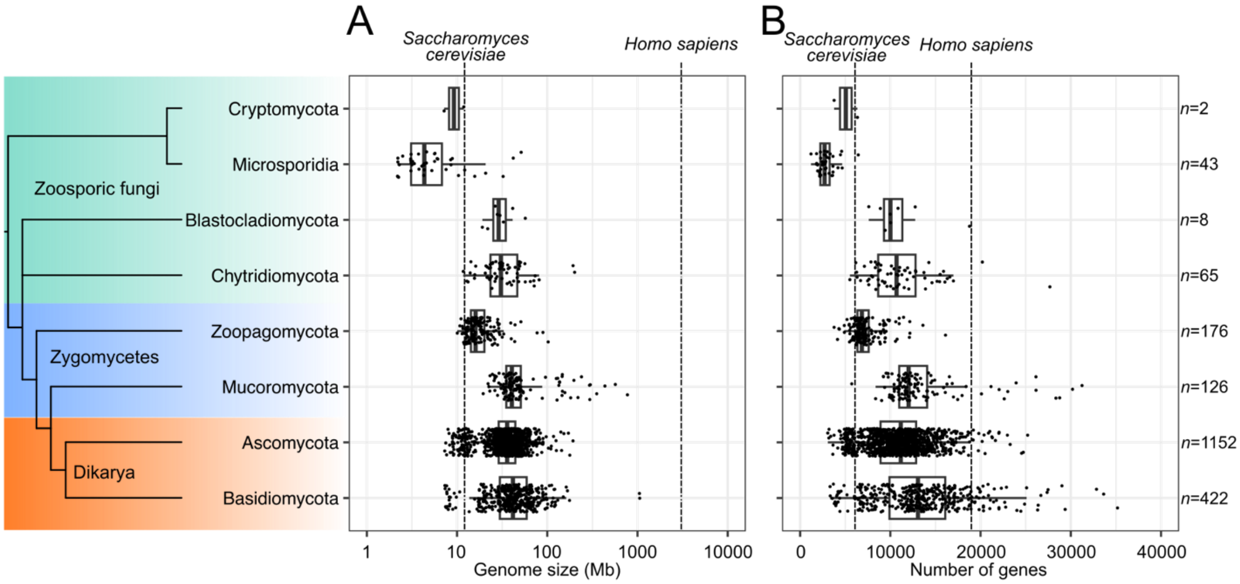


Figure 1.1: Distribution of (A) the genome sizes, and (B) number of genes in fungal genomes. The boxplots show the estimated genome sizes (A) and the number of predicted protein-coding genes (B) in 1,994 fungal species from 8 phyla. A cladogram organizing the fungal phyla (Spatafora et al., 2017) is shown on the left. For comparison, values for Saccharomyces cerevisiae (genome size=12.07 Mb; 6,014 genes) and Homo sapiens (genome size=3.05 Gb; 19,116 genes) are indicated with vertical lines. Fungal genomes were obtained from GenBank in November 2023.

1.3.1 Diversity of fungal gene content

Fungal genomes typically contain a large and diverse arsenal of genes from various functional categories. When comparing the number of protein-coding genes across different fungal phyla then, perhaps not surprisingly, there is a positive correlation (r^2=0.67) between the number of predicted genes and genome sizes, indicating that larger fungal genomes harbor more genes (Fig 1.2A). Overall, the total gene count in fungi ranges from as few as 1,201 genes in the genome of the Microsporidium species *Dictyocoela roeselum* to over 30,000 predicted protein-coding genes, higher than the predicted number of 19,116 protein-coding genes in the human genome (Piovesan et al., 2019). Notable examples of gene-rich fungi are the mycorrhizae fungi *Gigaspora rosea* and *Sphaerobolus stellatus,* which contain 31,243 and 35,177 genes, respectively, the highest number of protein-coding genes reported thus far in fungal genomes (Kohler et al., 2015). The average number of protein-coding genes among fungi is 11,128 (median=11,142 genes) (Fig 1.1B), but a large distribution in gene numbers exists between and within the different fungal phyla. Species within Basidiomycetes have the largest sets of genes (mean=13,241, median=13,057), followed by Mucoromycota (mean=13,681, median=12,017) and Ascomycota (mean=11,018, median=11,156). In contrast, species of the Microsporidia have the smallest sets of genes (mean=2,871, median=2,709) (Fig 1.1B).

Although at broader phylogenetic scales, it is difficult to specify the sources of such large differences in gene numbers, in general these are mainly due to increases in rates of gene duplication or gene loss in gene families that affect key traits such as nutritional lifestyles and adaptation to diverse hosts or environments (Taylor et al., 2017). For instance, species of powdery mildew, which are obligate biotrophic plant pathogenic fungi within the Ascomycetes phylum, are known to harbor a reduced number of circa (ca.) 7,000 genes (Wu et al., 2018). This is mainly the result of losses in gene families such as hydrolyzing enzymes and phytotoxins that would be injurious to the obligate biotrophic lifestyle of these pathogens (Spanu et al., 2010; Spanu, 2012). Recent studies have revealed that throughout their evolutionary history, fungi have periodically experienced drastic changes in their gene content. Early fungi, for instance, underwent constant losses of protist genes, combined with episodic expansions in genes encoding extracellular proteins, transcription factors, and proteins associated with nutrient uptake and growth (Merényi et al., 2023; Ocaña-Pallarès et al., 2022). These observations highlight how the evolutionary path of fungi has been marked by gradual gene losses and gene turnover.

Compared to other eukaryotic organisms, fungi appear to have more compact genomes with higher gene densities. For example, the *S. cerevisiae* genome has an estimated gene density of 498 protein-coding genes per Mb, while humans have on average 6.3 protein-coding genes per Mb (Fig 1.2C). At the species level, the distribution of gene densities in fungal genomes can vary widely, thereby in part shaping their taxon-specific differences in genome sizes and architecture. In this respect, species of Microsporidia which occupy the basis of the fungal tree of life and have some of the smallest genomes, exhibit the highest distribution in gene densities ranging from as low as 82 genes/Mb in *Edhazardia aedis* to 874 genes/Mb in *Encephalitozoon intestinalis*, with mean of 575 genes/Mb (median=644 genes/Mb). The distribution of gene densities in other fungal phyla are narrower (Fig 1.2C). For example, gene density within Ascomycota ranges from 49 to 824 genes/Mb (mean=318 genes/Mb; median=317 genes/Mb), and within Basidiomycota it ranges from 103 to 693 genes/Mb (mean=319 genes/Mb; median=323 genes/Mb).

Fungal genomes are further typified by an overall low abundance and length of introns per gene. Richness and length of spliceosomal introns are key traits that contribute to the phenotypic complexity of an organism and their ability to respond to selection pressure, by affecting traits such as gene expression, splicing efficiency, proteome diversity, and mutational rates in exons (Girardini et al., 2023). In fungal genomes, the mean intron size is 106 bp with median size of 86 bp, and the mean number of introns per gene is 2.3 with median of 2 (Fig 1.3B). In comparison, the average size of introns in protein-coding genes from the human genome is 6,938 bp (median=1,747 bp) and average of 10.3 introns per gene (median=8.0) (Piovesan et al., 2019). However, considerable variation among different fungal phyla is observed. Overall, species of the Microsporidia have the lower intron size (mean=30 bp, median=0 bp) and introns per gene (mean=0.1 intron, median=0 intron). In contrast, species of the Zoopagomycota have the longest introns (mean=199 bp, median=188 bp), and species of the Basidiomycota have the higher number of introns per gene (mean=4.3 introns, median=4.7 introns). Because the small introns and low intron density reflects small size of gen es, in the same physical space that humans have one gene, fungi are expected to have several genes (Fig 1.2B).

Notably, the typical fungal genome carrying approximately 11,000 genes has similar size as compared to the smallest human chromosome (chromosome 21; 47 Mb) carrying only 238 protein-coding genes (Fig 1.4). Taken together, high gene densities and low spliceosomal intron densities are two archetypal features of fungal genomes at a broader phylogenetic scale which further underlie their architecture and evolutionary potential.

The diversity in the number and density of genes in fungal genomes is also accompanied by contrasting arsenals of genes involved in overcoming physical barriers, as well as manipulating and retrieving nutrients from the environment. A distinctive feature of fungi is that they digest their food extracellularly and subsequently absorb it through their hyphae (Harley, 1971). As such, their genomes typically encode powerful arsenals of genes encoding secreted digestive enzymes that hydrolyze complex molecules to simpler ones that the fungus can absorb (Harley, 1971; Kües, 2015; Meyer et al., 2020). These enzymes include carbohydrate-active enzymes (CAZymes), lipases, and proteases that breakdown carbohydrates, lipids, and proteins, respectively. Fungi also produce a wide array of secondary metabolites (SMs) with diverse biological activities that support important ecological functions, including as antimicrobials against competitors and predators in their niche, as virulence factors during host infections, as mediators of mutualist symbiotic relationships, and several others (Bräse et al., 2009). Plant pathogenic fungi are also known to secrete during host infections a diverse arsenal of so-called effector proteins, which promote infections and host colonization by typically interfering with the host's immune system and by reprogramming its metabolism (Stergiopoulos and de Wit, 2009). However, effectors may also be recognized by host immune receptor proteins, thereby leading to effector-triggered immunity (ETI) and avirulence (Pradhan et al., 2021). Thus, effectors are both virulence and avirulence factors in disease, depending on the absence or presence of cognate immune receptors in the host, respectively (Stergiopoulos and de Wit, 2009).

A study revealed contrasting numbers of genes predicted to encode secreted and effector proteins among 22 genomes of fungal pathogens (Wang et al., 2022). The number of genes encoding secreted proteins (mean=1,146, median=1,094) varied from 344 in the human pathogen *Histoplasma capsulatum* to 2,015 in the plant pathogen *Colletotrichum gloeosporioides*, and the number of candidate effector genes (mean=227, median=185) varied from 48 in *H. capsulatum* to 618 in the rust pathogen *Puccinia graminis*. Using functional gene annotations available in *MycoCosm* (Grigoriev et al., 2014), a recent study also reported contrasting numbers of genes from different functional categories among fungi with different lifestyles (Dort et al., 2023). Genes encoding CAZymes were more abundant in pathogenic fungi (mean=529, median=528) compared to non-pathogenic fungi (mean=454, median=417). Pathogenic fungi also had larger arsenals of genes encoding key enzymes for SM production (mean=31, median=30) and transporters (mean=1,002, median=930) compared to the arsenals of genes encoding key enzymes for SM production (mean=24, median=21) and transporters (mean=909, median=871) in non-pathogenic fungi. Wide variation in gene numbers were also observed within pathogenic fungi. The arsenals of genes in obligate biotrophs encoding CAZymes (mean=400, median=417), key enzymes for SM production (mean=4, median=4), and transporters (mean=687, median=647) were much smaller compared to the arsenals of genes in hemi-biotrophs encoding CAZymes (mean=707, median=730), key enzymes for SM production (mean=42, median=44), and transporters (mean=1,363, median=1,166). Differences in the numbers of genes from different functional categories has also been used as the basis for promising machine learning-based efforts to identify and predict lifestyle of pathogenic fungi based on their gene content (Dort et al., 2023; Haridas et al., 2020; Thomas et al., 2023).

Overall, the contrasting difference in numbers of genes from different functional categories reflect large-scale gain or loss of genes throughout the evolution of many fungal lineages to adapt to different niches or lifestyles.

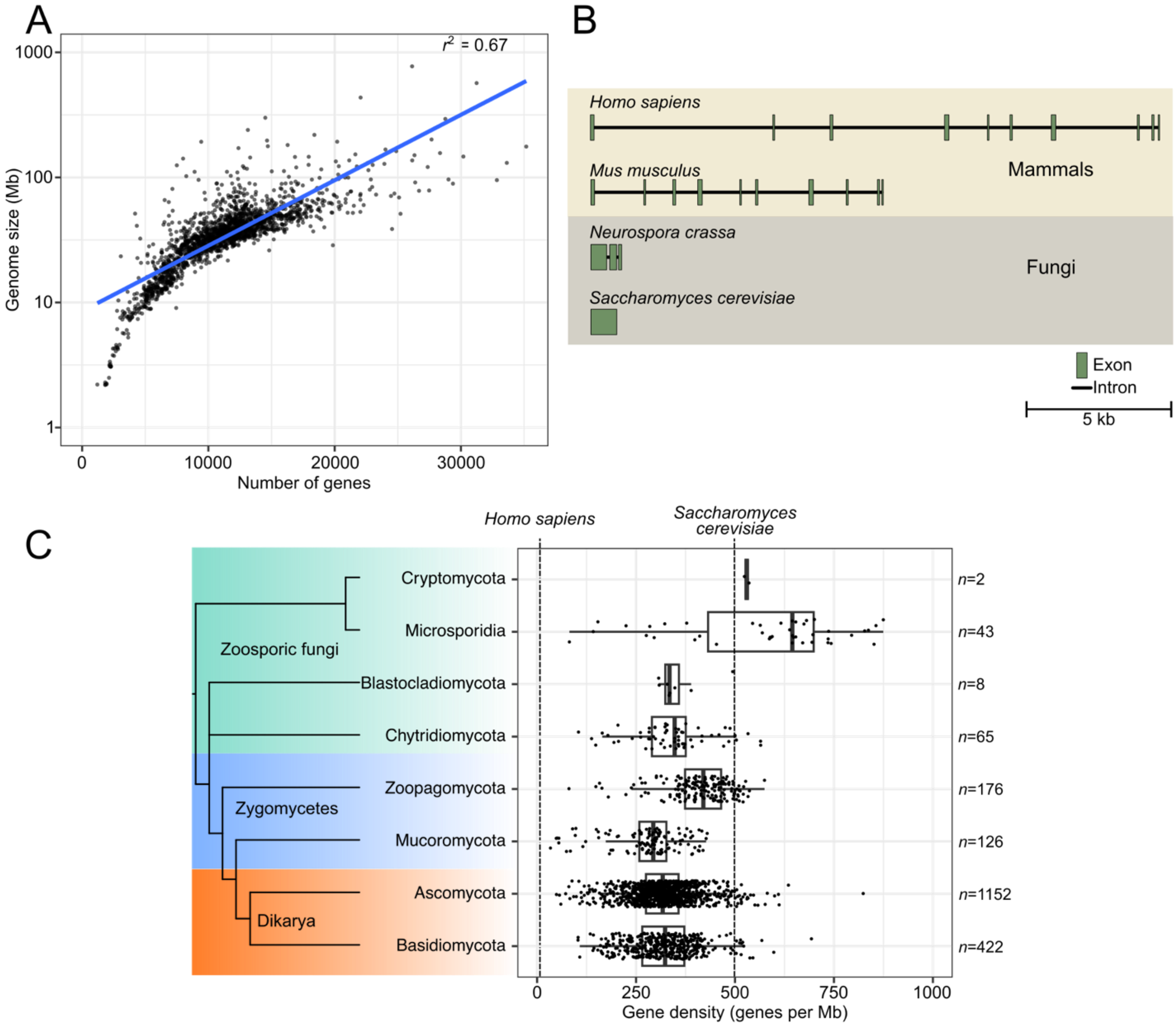


Figure 1.2: Fungal genomes have high gene density and show positive correlation between size and number of genes. (A) Scatter plot showing correlation between size of genomes and number of protein-coding genes in 1994 fungal genomes. (B) Comparison of the structures of orthologous genes encoding the coatomer subunit epsilon in *Homo sapiens* (XP_047294075.1), *Mus musculus* (NP_067513.1), *Neurospora crassa* (XP_961614.1), and *Saccharomyces cerevisiae* (NP_012189.2). (C) Box plots showing the distribution of gene density, i.e., number of protein-coding genes per Mb of the genome for 1994 fungal genomes from 8 phyla. A cladogram organizing the fungal phyla (Spatafora et al., 2017) is shown on the left. For comparison, gene densities for *Homo sapiens* (6.3 genes per Mb) and *S. cerevisiae* (498 genes per Mb) are shown with vertical lines.

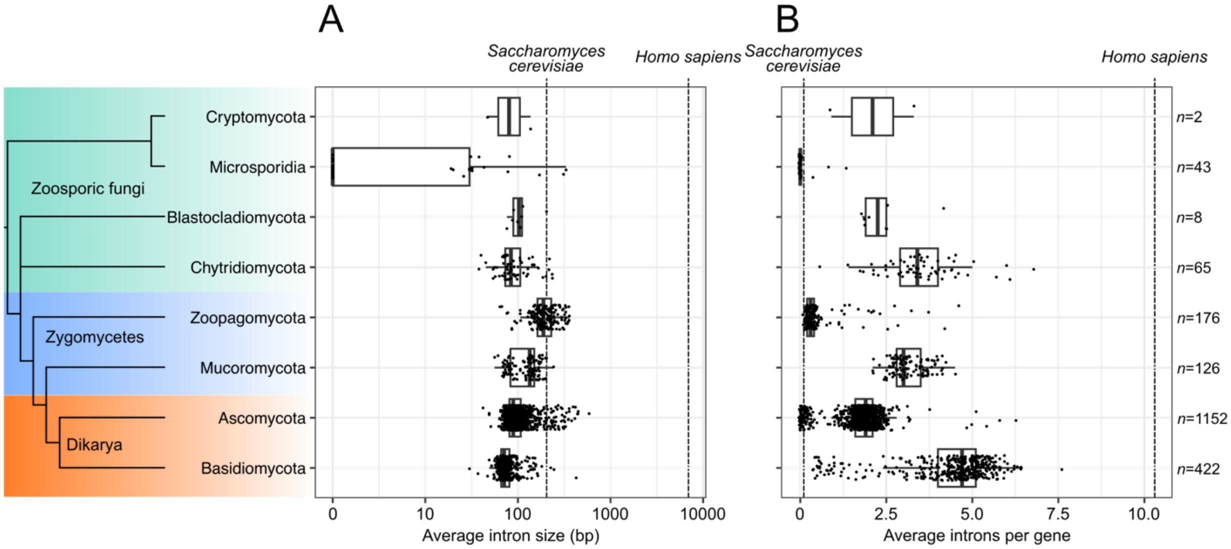


Figure 1.3: Fungal genomes have low intron content. The boxplots show (A) the average sizes of introns and (B) the average numbers of introns per predicted protein-coding gene in 1994 fungal species from 8 phyla. A cladogram organizing the fungal phyla (Spatafora et al., 2017) is shown on the left. For comparison, values for *Saccharomyces cerevisiae* (average intron size=204 bp; 0.1 intron per gene) and *Homo sapiens* (average intron size=6938 bp; 10.3 introns per gene) are indicated with vertical lines. Fungal genomes were obtained from GenBank in November 2023.

1.3.2 Presence of transposable elements in fungal genomes

A major component that drastically impacts the genome organization of fungi is the presence of transposable elements (TEs). TEs are DNA sequences that can excise or transcribe from their location of origin and reintegrate in a different location in the genome (Levin and Moran, 2011). They are regarded as "selfish" genetic elements, and their proliferation makes them ubiquitous in the genomes of eukaryotes. TEs fall into two main classes based on their mechanism of transposition, namely retrotransposons (class 1 transposons) that mobilize through a 'copy-and-paste' mechanism by transcription in RNA that is reverse transcribed, and DNA transposons (class 2 transposons) that mobilize through a "cut-and-paste" mechanism by excising and inserting themselves from the original to a new location in the genome (Levin and Moran, 2011).

Next to the gene and intron densities, the abundance of TEs in fungal genome also largely determines their sizes. Indeed, a study reported a strong positive correlation (r^2=0.71) between TE content and size of 16 genomes from 13 fungal species (Castanera et al., 2016). A positive correlation (r^2=0.58) was also observed between number of TEs per isolate and size of the genomes of 284 isolates of the wheat pathogen *Zymoseptoria tritici* (Oggenfuss et al., 2021). As the case of size and gene content, the amount of TEs in fungal genomes vary widely. TEs can comprise as little as 0.1% of the genome, as for example in the wheat pathogen *F. graminearum* (Cuomo et al., 2007), to more than 90% of the genome, as reported for the rust pathogens *P. pachyrhizi* (Gupta et al., 2023) and *A. psidii* (Tobias et al., 2021). These numbers contrast with the amount of TEs present in the human genome, which is estimated to be at approximately 45% (Lander et al., 2001). Moreover, in fungi, the amount of TEs can differ significantly even among closely related species. A notable example are the fungal pathogens *C. fulvum* and *D. septosporum* that are estimated to have diverged ca. 20 million years ago but have large difference in TE content (Ohm et al., 2012). For instance, while the genome of *C. fulvum* has an estimated 39% of TE content, the genome of *D. septosporum* is composed of only 3% of TEs (De Wit et al., 2012). Another notable example has been reported within Erysiphaceae, a family of fungal pathogens that that cause powdery mildew diseases in dicots and monocots. The genome of the grass powdery mildew pathogen *B. graminis* f. sp. *tritici* is composed of 85% TEs (Müller et al., 2019), while the genome of the Asian oak tree powdery mildew pathogen *Parauncinula polyspora* is composed of no more than 8.5% TEs (Frantzeskakis et al., 2019b). These observations indicate that different fungal lineages can undergo major proliferation of TEs throughout their evolution.

Recently, a novel group of DNA transposons known as *Starships* have been discovered in fungal genomes (Gluck-Thaler et al., 2022). Fungal *Starship* TEs are giant mobile elements that can surpass 100 kb in size and serve as vehicles for horizontal gene transfer between and within fungal species (Bucknell and McDonald, 2023; Gluck-Thaler et al., 2022). During their mobilization, *Starship* elements can carry along "cargo" genes that have the potential to increase fitness of the host. A *Starship* element called *Horizon*, for instance, carried and promoted the horizontal transfer of the *ToxA* gene between the wheat pathogens

Pyrenophora tritici-repentis, *Parastagonospora nodorum*, and *Bipolaris sorokiniana* (McDonald et al., 2019). ToxA is a virulence protein that induces necrosis on wheat lines that carry the susceptibility gene called *Tsn1* (Faris et al., 2010). Another example of a *Starship* element is *HEPHAESTUS* (*Hϕ*) that was discovered in the fungus *Paecilomyces variotii* and is hypothesized to have been acquired horizontally from *Penicillium fuscoglaucum* (Urquhart et al., 2022). *Hϕ* comprise a large gene cluster containing a diverse set of 39 genes, 14 of which have putative roles in metal tolerance. Indeed, *Hϕ* has been shown to confer tolerance to five metal/metalloid ions (arsenate, cadmium, copper, lead, and zinc) in *P. variotii* (Urquhart et al., 2022). Lastly, two *Starship* elements known as *Wallaby* and *ARISTAEUS* present in genomes of *Penicillium* spp. are hypothesized to provide fitness advantages against other fungi (Urquhart et al., 2023). *Wallaby* contains a gene encoding an antifungal protein, known for inhibiting the growth of competitors (Ropars et al., 2015), and both *Wallaby* and *ARISTAEUS* contain genes predicted to belong to a superfamily of fungal effectors known as Hce2 (for Homologs of *C. fulvum* Ecp2) that might confer an advantage in microbial competition and plant virulence (Stergiopoulos et al., 2012).

Taken together, these observations indicate that TEs, next to size and gene content, vary widely among fungal species, and are a key factor affecting the genome organization of fungal genomes.

1.4 Fungal karyotypes, ploidy levels, and chromosome organization

1.4.1 Fungal karyotypes and key structural elements of fungal chromosomes

Similar to other eukaryotes, fungal genomes are organized into multiple chromosomes located in the nuclear envelope (Wieloch, 2006). As with other fungal traits, the number of chromosomes in fungal genomes varies considerably, ranging from as low as 3 in the fission yeast *Schizosaccharomyces pombe* (Wood et al., 2002) up to a remarkable 33 chromosomes in the arbuscular mycorrhizal fungus *Rhizophagus irregularis* (Yildirir et al., 2022) (Fig 1.4). Large variation in the number of chromosomes is also observed even between species of the same genus. A notable example is found in species of *Fusarium* spp. for which the number of chromosomes vary from 4 in the wheat pathogen *F. graminearum* to 15 in the tomato pathogen *F. oxysporum* f. sp. *lycopersici* (Cuomo et al., 2007; King et al., 2015; King et al., 2015; Ma et al., 2010; Waalwijk et al., 2018) (Fig 1.4). The 4 chromosomes of *F. graminearum* exhibit syntenic regions to multiple chromosomes of *F. verticillioides* and *F. oxysporum* f. sp. *lycopersici* (Cuomo et al., 2007; Waalwijk et al., 2018). This observation led to the conclusion that the drastic difference in number of chromosomes in the *Fusarium* genus is attributed to chromosome fusion events in ancestor species. Another observation that supports the hypothesis of chromosome fusion is the biased distribution of SNPs in the *F. graminearum* chromosomes, which contain much higher frequencies of SNPs at telomeric and subtelomeric regions, but also clusters of SNPs at central chromosomal regions that are the putative junctions of ancestral fusions (Cuomo et al., 2007). The advantages of fused chromosomes in *F. graminearum* are still elusive. However, the putative chromosome fusion junctions and other regions of high SNP frequency were enriched in genes encoding secreted proteins, major facilitator transporters, and cytochrome P450 enzymes specifically expressed during plant infection (Cuomo et al., 2007). Thus, one possibility is that chromosome fusion events provided advantages for the pathogen during interaction with the host.

The idiosyncrasy of fungi as eukaryotic organisms also extends to key structural elements of their chromosomes, including in their telomere and centromere dynamics. Naturally, the ends of linear fungal chromosomes have telomeres, which are formed by non-coding tandem copies of simple sequence motifs. Telomeres serve two primary functions, i.e. to prevent the loss of terminal sequences from the lagging DNA strand during replication, and to prevent the chromosome ends from being treated as double-strand breaks. Filamentous fungi typically have the vertebrate-type of telomeric repeat (5'-TTAGGG-3'). However, there are exceptions, particularly among yeasts, for which the size of the telomeric repeats vary from 8 to 25 bp, and their sequence can vary even among species of the same genus (Lue, 2010). The telomeric repeat of the well-studied baker's yeast *S. cerevisiae* is unusual, formed by 300 +/-30 bp of $C_{1-3}A$ (Shampay et al., 1984). In comparison, different telomeric sequences are present among *Candida* species. In particular, the telomeric sequences of *C. albicans* is 5'-TGTACGGATGTCTAACTTCTTGG-3' (McEachern and Haber, 2006), *C. glabrata* is 5'-CTGTGGGGTCTGGGTG-3' (McEachern and Blackburn, 1994), and *C. orthopsilosis* is 5'-GGTTAGGATGTAGACAATACTGC-3' (Gunisova et al., 2009). The maintenance of telomeric repeats by the telomerase complex is regarded essential to prevent cell senescence (Osterhage and Friedman, 2009). However, failure in telomere maintenance has been proposed to play adaptive roles in eukaryotes (Mason and McEachern, 2018). Notably, a study that characterized chromosomal responses to failed telomere maintenance in the wheat blast fungus pathogen *Magnaporthe oryzae* reported for the first time 'spontaneous telomere failure' (Rahnama et al., 2021). In different strains of *M. oryzae,* telomere-associated rearrangements induced by failures in telomere maintenance were not as detrimental as one would expected, but instead likely provide adaptive benefits by driving rapid accumulation of sequence polymorphisms (Rahnama et al., 2021).

In addition to telomeres, other regions of chromosomes important for cell division are the centromeres, where spindle microtubules attach to facilitate proper segregation of sister chromatids. The yeast *S. cerevisiae* has point centromeres of approximately 125 nucleotides in size that are wrapped around single nucleosomes (Guin et al., 2020; Sullivan et al., 2001). However, filamentous fungi typically have regional centromeres that extend to thousands of base pairs (Guin et al., 2020). For example, in *F. graminearum*, centromeres vary in size between 56 kb and 65 kb (King et al., 2015), whereas in *Neurospora crassa*, they vary in size from 174 kb to 287 kb (Smith et al., 2011). In the grass powdery mildew pathogens *Blumeria graminis* f. sp. *tritici*, *B. graminis* f. sp. *hordei* and *B. graminis* f. sp. *triticale*, centromeric regions cover large segments of the chromosomes, extending up to 2.4 Mb in size (Frantzeskakis et al., 2018; Müller et al., 2019; Müller et al., 2021). The centromeric regions of grass powdery mildew pathogens are almost completely composed of repetitive DNA, are poorly conserved between species, lacks recombination, and are almost absent of predicted protein-coding genes. These observations highlight a remarkable variability of the organization of centromeric regions in fungal genomes, and that they can largely affect the overall organization of the genome.

Fungi exhibit an interesting arrangement of chromatin that conforms with the so-called Rabl configuration (Torres et al., 2023) (Fig 1.5A). During interphase, centromeres of all chromosomes cluster at the nuclear periphery, while the chromosome arms extend towards the opposite end of the nucleus (Fig 1.5B). Chromosome conformation capture techniques, such as Hi-C, have been successfully used to demonstrate the Rabl chromatin configuration across many fungal species (Galazka et al., 2016; Kim et al., 2017; Li Cheng-Xi et al., 2022; Seidl et al., 2020; Winter et al., 2018), highlighting its ubiquity in fungi. It has been hypothesized that the Rabl configuration is required in fungal species for proper gene regulation through the formation of chromatin loops (Torres et al., 2023). However, more research is needed to reveal the true extent to which the Rabl configuration contributes to the functioning and maintenance of fungal genomes.

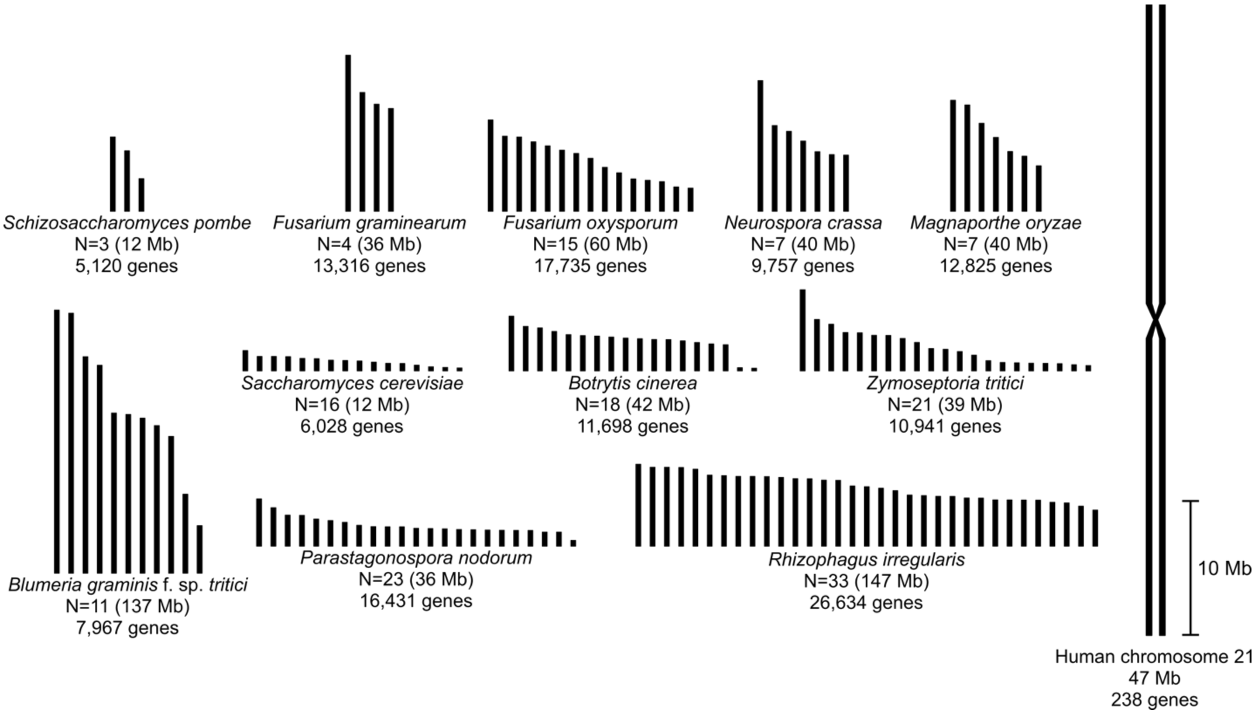


Figure 1.4: Karyotypes of 11 fungal species compared to the human chromosome 21. Chromosomes are represented as rectangles with size in scale. The figure depicts high karyotype diversity among fungal species, and how small their chromosomes are compared to the smallest human chromosome.

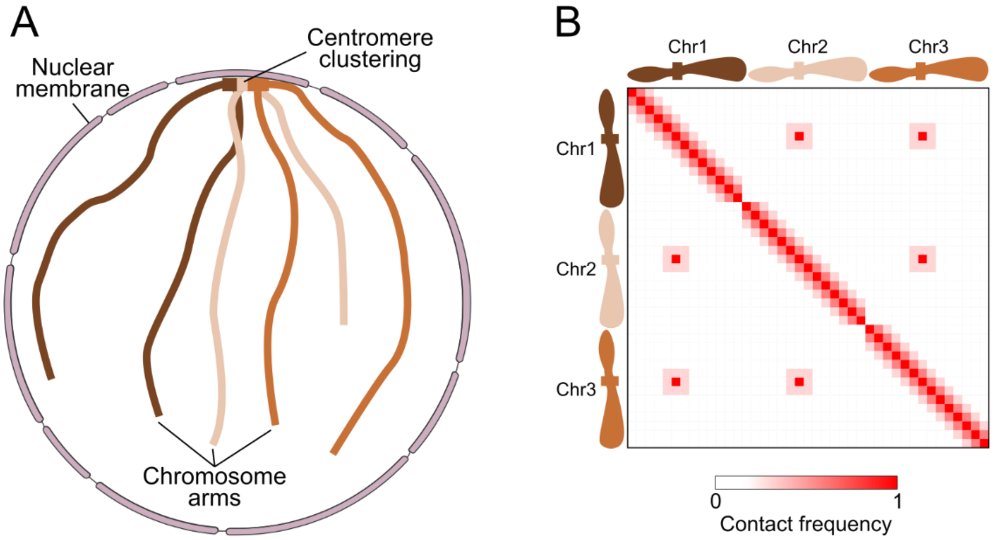


Figure 1.5: Representation of the Rabl chromatin configuration frequently observed in fungal genomes. (A) Centromeric regions of the chromosomes are clustered in the periphery of the nucleus. The chromosome arms extend toward the opposite end of the nucleus. (B) Representation of an expected contact frequency matrix of the chromosomes is shown in (A). Intrachromosomal high contact frequency is observed for regions that are physically close in the genome. The figure was generated using images from https://bioicons.com.

1.4.2 Fungal ploidy

Filamentous fungi are predominantly haploid organisms as their nuclei contain a single copy of each chromosome. However, many fungal species may also switch to and exist as stable diploid or polyploid organisms, and the change in ploidy levels is often associated with adaptation to adverse environmental conditions (Gerstein et al., 2006; Hose et al., 2015; Selmecki et al., 2006; Sunshine et al., 2015). Diploid, polyploid, and aneuploid species, for instance, are particularly frequent among yeast-like fungi of the Saccharomycetes. In the baker's yeast *S. cerevisiae*, diploidization is a relatively common phenomenon that can be advantageous under stress conditions (Harari et al., 2018). In a comprehensive study that analyzed the ploidy level of 794 natural strains of *S. cerevisiae*, researchers observed that only 1% of the strains were haploids, whereas 87% were diploids, and 12% were polyploids, including tri-, tetra- or pentaploid (Peter et al., 2018). In the same study, researchers also observed that diploid strains had significantly higher fitness compared to haploid and polyploid strains.

Polyploidy and ploidy changes are also frequently observed in human pathogenic fungi (Vande Zande et al., 2023). A typical example is *C. albicans*, one of the most common opportunistic human fungal pathogens, which is primarily a diploid species (Noble and Johnson, 2007). Even though its diploid nature is believed to be due to the occurrence of recessive lethal mutations in its genome (Whelan and Soll, 1982), *C. albicans* can also exist in a viable haploid state (Hickman et al., 2013) and may further change ploidy and become tetraploid through a parasexual cycle (Miller and Johnson, 2002). A study investigating the virulence and genome stability of tetraploid populations of *C. albicans* in the nematode host *Caenorhabditis elegans* observed a rapid transition of *C. albicans* to a diploid state (Smith et al., 2022). This transition was accompanied by an increase in virulence followed by a subsequent loss of virulence. A different study reported spontaneous switches between diploid and haploid states of isolates of the human pathogen *C. glabrata* both *in vivo* and *in vitro* conditions (Zheng Qiushi et al., 2022). This same study also showed that *C. glabrata* isolates with different ploidy exhibited differences in tolerance against the azole fungicide itraconazole, with haploid individuals frequently more tolerant. Overall, these observations indicate that ploidy shifts could act as a strategy to promote genetic diversity and benefit fungal pathogens to rapidly adapt to the changes in the environment.

1.4.3 Dispensable chromosomes

The observation that in some fungal species certain chromosomes show presence/absence variation among isolates of the species has led to the classification of chromosomes into core and dispensable ones (Mehrabi et al., 2017). Specifically, while core chromosomes are present in all isolates of a species, dispensable or accessory chromosomes are differentially present among its isolates, thus making these genetic elements presumably non-essential for the organism survival (Mehrabi et al., 2017). To date, dispensable chromosomes have been reported in the genomes of at least 25 fungal species (Bertazzoni et al., 2018; Goodwin et al., 2011; Wang et al., 2019; Yang et al., 2020). Such dispensable chromosomes typically have a non-Mendelian mode of transmission and features that distinguish them from core chromosomes, including being small in size (< 2 Mb), containing high amounts of repetitive DNA and a small number of genes that mostly encode hypothetical proteins (Komluski et al., 2022).

One of the earliest reports of the presence of a fungal dispensable chromosome was in the pea pathogen *Fusarium solani* (Miao et al., 1991). This fungus contains a small 1.6 Mb dispensable chromosome harboring the gene *Pda6* that mediates detoxification of the phytoalexin pisatin produced by pea plants. Many subsequent instances of dispensable chromosomes have been observed in *Fusarium* species, such as *F. oxysporum*, *F. asiaticum* and *F. fujikuroi* (Galazka and Freitag, 2014; Ma et al., 2010), establishing it as the most extensively researched fungal genus in this regard (Yang et al., 2020). Another plant pathogen reported to contain dispensable chromosomes is the wheat pathogen *Zymoseptoria tritici*, which has a total of 13 core chromosomes and a remarkable set of 8 dispensable chromosomes, thereby making it one of the largest complements of accessory chromosomes reported to date among fungi (Goodwin et al., 2011). A recent study analyzed over one thousand *Z. tritici* genomes and reported that chromosomes 14, 16, and 18 were missing in more than half of the genomes analyzed, indicating a dynamic genome organization among populations with possibly rapid gain and loss of chromosomes (Feurtey et al., 2023). Moreover, the instability of the dispensable chromosomes of *Z. tritici* is not only the result of presence/absence variation, but also whole-chromosome fusion, as reported for chromosomes 15 and 16 (Badet et al., 2020). However, a clear effect on fitness of the fungus has not yet been demonstrated for all its dispensable chromosomes.

Even though the presence of dispensable chromosomes has been reported in many fungal species, their origin remains largely elusive. However, it is generally accepted that these chromosomes originate by intra-species horizontal transfer or are formed by fragments of the core chromosome following extensive rearrangements and fusions (de Jonge et al., 2013). Moreover, the function of dispensable chromosomes in most fungal species remains unknown (Bertazzoni et al., 2018), which makes their persistence within fungal populations intriguing.

1.5 The dynamics of fungal genome architecture and organization

1.5.1 Repeat-Induced Point mutations and GC content distribution

The presence and proliferation of transposable elements (TEs) are key factors affecting the organization of fungal genomes. TEs are shown to partake in several biological processes, including in the regulation, pseudogenization, deletion, horizontal transfer, and duplication of genes in fungal genomes (Mat Razali et al., 2019). However, TEs are also directly associated with the so-called repeat-induced point (RIP) mutations, one of the main sources of mutations in fungal genomes. First described in the fungus *N. crassa* (Selker and Garrett, 1988), RIP is a premeiotic mechanism that so far has been reported only in fungi, and which induces transition nucleotide substitutions (C-to-T or the complement G-to-A) in duplicated genomic sequences, with a strong bias toward CpA-to-TpA (or the complement TpG-to-TpA) dinucleotides (Hane and Oliver, 2008; Selker, 1990). TEs can rapidly accumulate RIP mutations and thereby become inactive, which led to the assumption that RIP evolved in fungi as a defense mechanism against the deleterious effects of TE proliferation (Clutterbuck, 2011; Selker, 1990; Selker, 2002). The transition nucleotide substitutions induced by RIP considerably decreases the GC content of repetitive regions of fungal genomes, leading to the formation of AT-rich regions commonly referred to as 'AT-isochores' (Testa et al., 2016). The AT-isochores intersperse with non-repetitive regions that typically have higher GC content. These regions with contrasting GC content results in bimodal GC distributions that are frequently observed in fungal genomes. For example, the genome of the hazelnut pathogen *Anisogramma anomala* has a bimodal GC distribution with peaks at 39.9% GC and 57.1% GC (Cohen et al., 2024). Bimodal GC distributions have also observed in the genome of the mycoparasitic fungus *Ampelomyces quisqualis*, with peaks at 33.9% GC and 66.1% GC (Huth et al., 2021), and in the genome of *L. maculans*, with peaks at 33.9% and 51.0% (Rouxel et al., 2011).

The mechanism that induces RIP mutations is currently not well understood. However, in fungi, RIP occurs almost exclusively during sexual reproduction, after cell fusion (plasmogamy) and before nuclear fusion (karyogamy) that leads to meiosis (Galagan and Selker, 2004) (Fig 1.6). Thus, RIP is an important source of mutations for fungi that undergo sexual reproduction. For example, a study reported that in the wheat pathogen *Z. tritici* the meiotic mutation rate was three orders of magnitude higher compared to its mitotic mutation rate, and that 78% of the point mutations during meiosis were caused by RIP (Komluski et al., 2023). Because RIP can occur at each cycle of sexual reproduction, RIP mutations can accumulate within just a few generations, thus resulting in high rates of mutations that can be from the highest among non-viral organisms (Wang et al., 2020).

Even though RIP targets repeat regions of the genome, it can also induce point mutations in single-copy regions that are adjacent to repetitive DNA, thus intensifying its impact on fungal genomes. This phenomenon is referred to as RIP leakage because mutations seem to ‘spill’ from neighboring repetitive regions into single-copy regions. Proof of RIP leakage has been provided for the genomes of *N. crassa* (Irelan et al., 1994) and *L. maculans* (Rouxel et al., 2011), but despite the lack of supporting evidence, RIP leakage is hypothesized to occur in many other fungal species as well. Moreover, it is widely accepted that due to RIP leakage, genes in TE-rich regions are expected to accumulate more point mutations compared to genes in TE-poor regions. This assumption supports a beneficial organization of fungal genomes into compartments of repeat-rich and repeat-poor regions, as described in the next section.

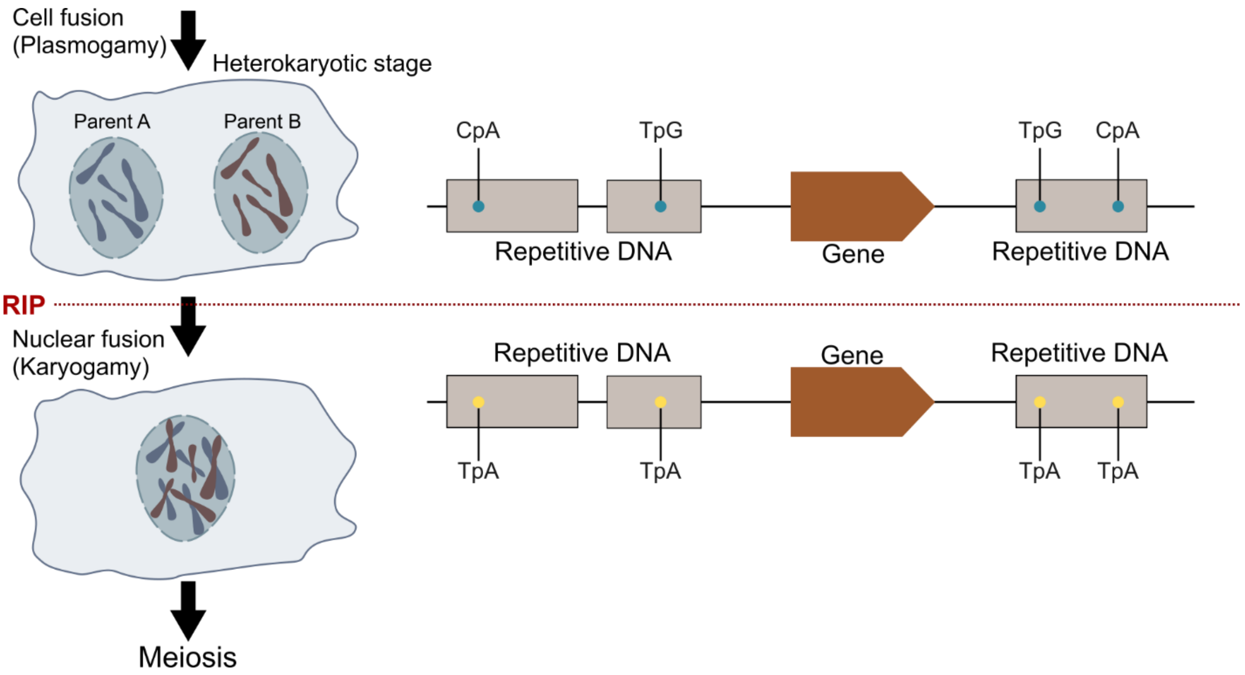


Figure 1.6: Repeat-induced point (RIP) mutations occur during sexual reproduction in fungi. The figure depicts the stage during sexual reproduction where RIP mutations can occur, i.e., after plasmogamy and before karyogamy. RIP typically induces transversion nucleotide substitution CpA-to-TpA (or the reverse complement TpG-to-TpA) in repetitive regions of the genome. Figure generated using images from https://bioicons.com.

1.5.2 Genome compartmentalization

The proliferation of TEs in the genomes of fungal pathogens can drastically change the overall architecture of the genomes. Even though TEs can mobilize randomly, they often accumulate preferentially in regions that already have high TE content (Cook et al., 2020; Faino et al., 2016). One explanation for this is that TEs inserted in gene-rich regions are more likely to cause deleterious effects by disrupting genes that contribute to the organismal fitness (Torres et al., 2020) and thus, selection against insertion of TEs in gene-rich regions is expected to occur in higher frequencies. The uneven distribution of TEs results in a distinctive compartmentalization of fungal genomes into gene-rich and TE-poor regions interspersed with gene-poor and TE-rich regions. This compartmentalization can be observed by comparing the size of intergenic gene regions, as genes located in gene-rich regions have overall small intergenic sizes, whereas genes in gene-poor regions have long intergenic sizes inflated by insertion of TEs.

The genome compartmentalization by the uneven distribution of TEs has been observed in many fungal and oomycete pathogens (Dong et al., 2015; Faino et al., 2016; Feurtey et al., 2020; Fletcher et al., 2022; Treindl et al., 2023; Wacker et al., 2023), and is the basis for the so-called ‘two-speed genome’ model of evolution in fungi (Dong et al., 2015). In this model, TE-rich regions of the genome are assumed to evolve faster as compared to TE-poor regions. This difference in the speed of evolution is attributed to structural variations caused by TE mobility and point mutations generated by RIP leakage, and which are expected to be enriched in TE-rich regions of fungal genomes (Faino et al., 2016; Plissonneau et al., 2018; Raffaele et al., 2010). The ‘two-speed genome’ model of evolution is thought to speed up the evolution of fungal pathogens and their ability to overcome host defenses, as genes encoding effector proteins important for pathogenicity are frequently enriched in TE-rich regions of the genome (Frantzeskakis et al., 2019a; Rouxel et al., 2011).

Genome compartmentalization in fungi is more often reported for plant pathogens. However, a recent study showed that the genome of the amphibian fungal pathogen *Batrachochytrium salamandrivorans* also exhibits compartmentalization, demonstrating that animal fungal pathogens can too have compartmentalized genomes (Wacker et al., 2023). The gene-sparse compartment of the genome of *B. salamandrivorans* is enriched in genes encoding candidate effectors, secreted M36 metalloproteases, and genes with evidence of positive selection. This same study proposed a set of statistical methods based on enrichment tests and discrete-time pattern Markov chain calculations that can be applied in other genomes as an attempt to standardize the classification of the two-speed model of evolution, thus setting a basis for using pre-defined statistics to support different speeds of evolution in fungal genomes.

While the genomes of many fungal and oomycete pathogens have been shown to exhibit a bipartite architecture, other genomes exhibit a different configuration. The observation of varying degrees of compartmentalization in fungal genomes led to the broad classification of one-compartment, two-compartment, and multi-compartment genomes (Frantzeskakis et al., 2019a). One-compartment genomes have been reported for powdery mildew fungi (Frantzeskakis et al., 2018; Frantzeskakis et al., 2019b; Müller et al., 2019), the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* (Schwessinger et al., 2018), and the wheat tan spot pathogen *P. tritici-repentis* (Gourlie et al., 2022). The genomes of these fungi do not exhibit evident compartmentalization and it is instead hypothesized that the evolution of these fungi, including to an extent their adaptation on their host, is governed by copy-number variations of genes. In these genomes, gene copies can accumulate mutations independently of each other, thus accelerating sequence diversification. Moreover, there could be beneficial dosage effects resulting from the combined expression of gene copies.

A multi-compartment genome architecture has been reported for species of *Fusarium*. In this model, the genome is composed of multiple compartments that likely evolve at different speeds (Frantzeskakis et al., 2019a). In addition to dispensable chromosomes, core chromosomes may have long segments that exhibit presence/absence variation among isolates (Vlaardingerbroek et al., 2016). This is the case, for example, in *F. oxysporum*, in which the three smallest core chromosomes (11, 12, and 13) are referred to as the 'fast-core' genome because, due to chromosomal rearrangements, they exhibit higher divergence and lower synteny (Fokkens et al., 2018). However, the overall higher divergence of these chromosomes is not associated with presence of TEs. Instead, the 'fast-core' genome in *F. oxysporum* is enriched with histone H3 lysine 27 trimethylation (H3K27me3) markers (Fokkens et al., 2018), which originated the hypothesis of an association between chromatin state and higher frequency of *de novo* mutations. Another example of an association between chromatin state and compartmentalization has been reported for the arbuscular mycorrhizal fungus *Rhizophagus irregularis* (Yildirir et al., 2022). Using chromatin conformation capture, Yildirir and coauthors revealed that the genome of *R. irregularis* exhibits a 'checkerboard' pattern composed of two compartments, consisting of regions of open chromatin (euchromatin) interspersed with regions of closed chromatin (heterochromatin) (Yildirir et al., 2022). When comparing the two, then the euchromatic regions are gene-dense with higher transcription activity, while the heterochromatic regions are more repeat-dense with lower transcription activity and higher rates of chromosomal rearrangements.

Interestingly, the heretochromatic regions were also enriched in genes upregulated during host colonization, possibly triggered by signals from the host (Yildirir et al., 2022).

In summary, these studies demonstrate that the presence and distribution of TEs play a crucial role in shaping the genomic architecture of fungal pathogens, influencing their adaptive strategies, and pathogenic traits.

1.6 Structural variations in fungal genomes and their impact on host adaptation and antifungal resistance

1.6.2 Small-scale structural variations promoting host adaptation and antifungal resistance

Structural variations (SVs) refer to mutations in the genome that impact the organization of the chromosomes or segments of the chromosomes. Specifically, SVs are deletion, insertion, duplication, inversion, or translocation events that affect more than 50 bp (Abel et al., 2020). In humans, SVs are the underlying cause of various genetic disorders and diseases. Among the most notable examples is the Down syndrome, which is caused by the trisomy of chromosome 21. Additionally, conditions such as Fragile X syndrome, Huntington's disease, and Machado-Joseph disease are attributed to the expansion of trinucleotide repeats in the genes *FMR1*, *HTT*, and *ATXN3*, respectively (Hollox et al., 2022). In contrast to their potential deleterious effects, SVs in the human genome have also been associated, among others, with evolutionary processes such as speciation (Hollox et al., 2022), brain morphology (Dennis et al., 2012), and dietary changes (Carpenter et al., 2015).

In fungi, SVs may play key roles in adaptation to adverse conditions and are often associated with resistance to antifungal drugs and pathogenicity (Gorkovskiy and Verstrepen, 2021). One of the earliest reported cases of SVs in fungi was observed in *S. cerevisiae*, in which it mediated its adaptation to nutrient limitation (Brown et al., 1998). In this study, Brown and co-authors performed an experimental evolution of *S. cerevisiae* through 450 generations cultured under glucose-limiting conditions. Compared to the founding strain, the evolved strain increased its fitness in the glucose-limiting condition by producing 2-fold more biomass and exhibiting a 1.8-fold increase in the expression of the high affinity *HXT6* and *HXT7* hexone transporters in its genome. Further analysis showed the duplication of the *HXT6* and *HXT7* genes, which are located next to each in the genome *S. cerevisiae* likely contributed to the dosage effect and the increase in fitness in the glucose-limited condition. The high similarity of the genes *HXT6* and *HXT7* (99% identity) combined with their tandem arrangement in the genome led to the hypothesis that *HXT6* and *HXT7* were duplicated through non-allelic recombination, specifically unequal crossing over, which occurs when homologous sequences are not properly paired during meiosis or mitosis.

A notable example of SVs leading to an increase of tolerance to fungicides was reported in the grape powdery mildew pathogen *Erysiphe necator*, in which the duplication of a 10 kb fragment carrying the gene *Cyp51* was associated with resistance to azoles (Jones et al., 2014). Upon the screening of 89 isolates collected in vineyards, Jones and co-authors observed remarkable variation in the copy numbers of *Cyp51*, which varied from 2 to 14, and a strong positive correlation between *Cyp51* copy numbers and fungal growth under fungicide treatment (r^2=0.81). There was also a strong positive correlation between copy numbers and expression of *Cyp51* (r^2=0.95), suggesting that resistance to fungicide derives, in part, from a dosage effect of the combined expression of all *Cyp51* copies in *E. necator* isolates. SVs affecting regulatory regions of fungal genes also play important roles in gaining tolerance to fungicides. In this regard, the insertion of TEs into *cis*-regulatory regions of the genome can lead to changes in gene expression levels, thereby influencing a number of trains such as the development of fungicide resistance, when the insertion is in genes that encode for the fungicides' targets or genes whose products mediate detoxification of xenobiotics (Chen et al., 2017; Steinhauer et al., 2019; Sun et al., 2013). Notably, in the stone fruit pathogen *Monilinia fructicola*, insertion of a TE-like element named 'Mona' in the promoter region of the gene *Cyp51* increased its expression and provide resistance against the azole fungicide propiconazole (Chen et al., 2017). Gain of resistance against propiconazole has also been observed in an isolate of *Z. tritici* that had a TE 229 bp upstream of a gene encoding a SNARE domain protein and 286 bp upstream of a gene encoding a flavin amine oxidoreductase (Oggenfuss et al., 2021).

Mobilization of TEs can also favor fungal pathogens by their insertion into the regulatory or coding regions of effector genes, thereby leading to changes in their expression or to their pseudogenization (Kang et al., 2001; Wu et al., 2015). For example, insertion of a TE in the coding region of the *M. oryzae* avirulence gene *AvrPi9* disrupts this gene, thus preventing AvrPi9 recognition by its cognate resistance protein Pi9 in rice (Wu et al., 2015). Moreover, complete deletion of effector genes *Avr9*, *Avr5*, and *Avr4E* in the tomato pathogen *C. fulvum* (Stergiopoulos and de Wit, 2009), *Avr-Pita* in the rice blast fungus *M. oryzae* (Chuma et al., 2011), *Ave1* in the vascular wilt-causing fungus *Verticillium dahliae* (Faino et al., 2016), and the candidate effector *Zt_8_609* in *Z. tritici* (Hartmann et al., 2017) have also been associated with the presence or mobilization of TEs in their vicinities. The complete deletion of dispensable effector genes follows a similar purpose as their pseudogenization, which is to evade effector-triggered immunity mediated by cognate resistance genes in the host.

In contrast to deletion of effector genes, SVs are also hypothesized to act in favor of fungal pathogens by inducing the duplication of effector genes. In a recent study, van Westerhoven and co-authors analyzed the pan-genome of 69 strains of *F. oxysporum* that causes wilt of banana (van Westerhoven et al., 2023). They observed that segmental duplications occur specifically in accessory regions and sub-telomeric regions of the genome of different *Fusarium* spp. The identified segmental duplications expanded the repertoire of genes encoding candidate effectors, including of the effector *SIX1* that is required for full virulence of *Fusarium* on banana. The impact of duplication of effector genes on virulence of *Fusarium* spp. is not yet clear, but these observations support that segmental duplications are a major driver of evolution in *Fusarium* spp. Notable cases of duplication of candidate effector genes also take place in fungal pathogens that cause powdery mildew (Müller et al., 2019; Pedersen et al., 2012; Sharma et al., 2019). The grass powdery mildew pathogens *B. graminis* f. sp. *tritici* and *B. graminis* f. sp. *hordei* have remarkable arsenals of over 400 paralogous genes encoding candidate effectors similar to ribonucleases (Frantzeskakis et al., 2018; Müller et al., 2019; Pedersen et al., 2012; Seong and Krasileva, 2023). These genes, referred to as RNase-like proteins associated with haustoria (RALPHs), are the result of many duplication events followed by sequence diversification, which resulted in at least 15 distinct families of RALPHs that are hypothesized to target different host proteins (Cao et al., 2023; Maruta et al., 2023). The RALPH CSEP0064 inhibits the degradation of host ribosomal RNA by binding to host RNA (Pennington et al., 2019), CSEP0105 and CSEP0162 promote plant disease by interacting with host chaperone proteins (Ahmed et al., 2016), and CSEP0027 binds to a barley catalase and alters its subcellular localization (Yuan et al., 2021).

In summary, small-scale SVs in fungal genomes play pivotal roles in adaptation, resistance to antifungal agents, and pathogenicity. Understanding the intricate interplay between SVs and fungal biology can elucidate evolutionary processes and provide insights for combating fungal diseases.

1.6.1 Aneuploidy and their impact on virulence and antifungal resistance

Aneuploidy refers to changes in chromosome numbers from the parental karyotypes that occur after mis-segregation during meiosis or mitosis, typically due to the failure of homologous chromosomes to properly separate during cell division (i.e., nondisjunction) (Gilchrist and Stelkens, 2019). In human fungal pathogens, cases of aneuploidy that exhibit differences in virulence and resistance to antifungal drugs have been reported (Bing et al., 2020; Hirakawa et al., 2017; Li et al., 2015; Yang Feng et al., 2021). For example, strains of *C. albicans* carrying three copies of chromosome 1 were avirulent against mouse, differently from closely related strains carrying two copies this chromosome (Chen et al., 2004).

Duplication of chromosomes 1 and 4 of *Cryptococcus neoformans* (Sionov et al., 2010), duplication of chromosomes 5 and R of *C. albicans* (Li et al., 2015; Selmecki et al., 2006), duplication of chromosome 5 of *C. auris* (Bing et al., 2020), and duplication of chromosome E of *C. glabrata* (Ksiezopolska et al., 2021) have been associated with increase in resistance to azole fungicides that target ergosterol biosynthesis. In all these aneuploidy cases, the increase in tolerance to the fungicide is attributed to dosage effects resulting from the duplication of genes located in the duplicated chromosomes. These genes include *Cyp51* (*Erg11* in yeast) encoding the target of azole fungicides, as well as genes likely involved in xenobiotic detoxification such as transcription factors and transporters.

In contrast to yeast-like fungi, cases of aneuploidy-mediated adaptation to adverse conditions are rarely reported in filamentous fungi, possibly due to the increased deleterious effects of aneuploidy in these organisms and thus the lower frequency of such SVs. However, notable examples are reported in the literature. For instance, a recent study reported aneuploidy-mediated tolerance to the antifungal drug voriconazole in the opportunistic fungus *A. flavus* that causes aspergillosis (Omer et al., 2023). After constant exposure to non-lethal doses of voriconazole, resulting colonies of *A. flavus* gained tolerance to voriconazole and exhibited either a complete duplication of chromosome 8 or a partial duplication of chromosome 3. Gain of voriconazole resistance in aneuploidy individuals of *A. flavus* was likely the result of the duplication of the *AtrR* gene that is located in the region of chromosome 3 that was duplicated. In *A. fumigatus*, *AtrR* is an essential determinant of tolerance to azoles fungicides by regulating the *Cyp51* gene (Sanjoy et al., 2019). Thus, it is possible that the duplication of *AtrR* in *A. flavus* upregulated *Cyp51* as well, thereby leading to tolerance to voriconazole. Finally, no obvious association of duplication of genes in chromosome 8 of *A. flavus* and tolerance to voriconazole could be established.

In addition to resistance to fungicides, aneuploidy in filamentous fungi has also been hypothesized to promote pathogenicity. In the pine needle tree pathogen *D. septorum*, full or partial duplication of chromosomes 5, 11, 13, and 14 in strain ALP3 has been associated with increase in the production of the dothistromin, a toxin required for full virulence of the fungus on pine needle trees (Bradshaw et al., 2019). The aneuploid strain ALP3 was reported to produce four to seven times more dothistromin compared to the 18 other strains analyzed. However, the genes involved in dothistromin biosynthesis were not located in the duplicated chromosome segment and based on this observation, the authors hypothesized that the duplicated segments harbored genes involved in the secretion of dothistromin, such as efflux transporters.

In summary, an increasing body of research is revealing evidence of aneuploidy across phylogenetically distant fungal species. Furthermore, these studies support that aneuploidy in fungi serves as a rapid and transient avenue for adaptive evolution.

1.6.2 Impact of dispensable chromosomes on host adaptation

In some fungal species, the presence of dispensable chromosomes is required for pathogenicity (Akagi Yasunori et al., 2009; Ma et al., 2010; Wang et al., 2003). One of the earliest demonstrations of dispensable chromosomes mediating pathogenicity was reported for the tomato pathogen *F. oxysporum* f. sp. *lycopersici*. Indeed, when an isolate (Fol007) pathogenic towards tomato that contained chromosome 14 was co-incubated with an isolate (Fo-47) non-pathogenic towards tomato and that did not have chromosome 14, then some of the resulting isolates that had the Fo-47 karyotype, had gained the ability to infect tomato after horizontally acquiring chromosome 14 (Ma et al., 2010). Gain of pathogenicity was most likely due to genes located in chromosome 14, including the secreted in xylem (*SIX*) genes that are shown to be virulence factors in this species (Houterman et al., 2009). The function of the *SIX* genes is still elusive but it has been postulated that they modulate hormonal pathways or defense response cascades in tomato.

In the wheat blast fungus pathogen *Magnaporthe oryzae*, occurrence of dispensable chromosomes is known since the early 1990s (Talbot et al., 1993). In this pathogen, dispensable chromosomes are hypothesized to be important for virulence or niche adaptation, and are an important source of genetic variability for asexually reproducing lineages (Langner et al., 2021; Liu et al., 2024). Notably, in a recent study, researchers observed under natural field conditions the horizontal transfer of the dispensable chromosome mChr. The transfer happened from a lineage of isolates causing disease in Indian goosegrass (*Eleusine indica*) to the lineage of isolates causing disease in rice (*Oryza sativa*), and as many as nine horizontal transfer events of mChr are thought to have taken place over the past three centuries (Barragan et al., 2024). These observations led to the speculation that blast fungus populations infecting wild grasses serve as genetic reservoirs for clonal populations infecting crops, increasing their genetic diversity through recurrent horizontal chromosome transfers (Barragan et al., 2024).

Another interesting aspect of dispensable chromosomes is that their presence may affect the fungal lifestyle. This has been elegantly demonstrated for the rhizospheric soil fungus *Stagonosporopsis rhizophilae* (Wei et al., 2023), where researchers uncovered a 0.6 Mb dispensable chromosome that was likely acquired horizontally from close related species. It is shown that mutants of *S. rhizophilae* that do not have the dispensable chromosome exhibit increased melanization and shift the lifestyle of the fungus toward beneficial interactions with poplar plants. Thus, the dispensable chromosome in *S. rhizophilae* is thought to act as a suppressor of symbiosis and although the mechanism underlying such change in the lifestyle of the fungus remains elusive, it is speculated that it relates to changes in gene expression and/or the genes within the disposable chromosome. The hypothesis is supported by the fact that loss of the dispensable chromosome in *S. rhizophilae* triggered a genome-wide transcriptional reprogramming, most likely because of changes in the chromatin accessibility of core chromosomes. Moreover, the dispensable chromosome contained 44 genes upregulated during interaction with the host. Among these genes there was a type 1 polyketide synthase gene likely involved in the biosynthesis of secondary metabolites. Other genes in the dispensable chromosome included genes encoding candidate effectors, although these genes showed no evidence of upregulation during interaction with the host.

These reports highlight the dynamic roles of dispensable chromosomes, extending from fungal pathogenicity to fungal lifestyle, and shed light on their multifaceted contributions to fungal ecology and adaptation.

1.7 Fungal mitochondrial genomes

Many genomic studies in fungi focus on their nuclear genome to unravel the molecular mechanisms underlaying evolution and adaptation. In contrast, fungal mitochondrial genomes are often understudied, although they are of fundamental importance to organismal metabolism and proper functioning. Mitochondria are essential organelles that partake in the generation of adenosine triphosphate (ATP) in cells and other various metabolic and cellular processes, including iron metabolism and programmed cell death (Chan, 2006). In fungi, mitochondria are also involved in pathogenicity and acquirement of resistance to fungicides that block the production of ATP by interfering with the electron transfer chain (Black et al., 2021). Mitochondria have their own genome (mtDNA) and, at least in mammals, follow a non-Mendelian form of inheritance as they are almost always uniparentally inherited from the mother. However, fungi exhibit more diverse patterns of mitochondrial genome inheritance. For instance, in *S. cerevisiae* and *S. pombe*, mtDNA is inherited from both parents, following subsequent segregation by budding and fission (Berger and Yaffe, 2000) but in basidiomycete fungi, both uniparental and biparental mtDNA inheritance have been reported (Xu and Wang, 2015). Among filamentous ascomycete fungi, mtDNA is predominantly inherited uniparentally, although from either parent (Xu and Wang, 2015). Each mating type of ascomycete fungi can produce “male” gametes (microconidia) or “female” gametes (ascogonia) and during sexual crosses, mtDNA is inherited predominantly from the ascogonia (Xu and Wang, 2015). A recent study showed that the progeny of hybridization between the ascomycete fungi *Ceratocystis fimbriata* and *C. eucalypticola,* and between *C. fimbriata* and *C. manginecans,* inherited mtDNA from both parental species (van der Walt et al., 2023), indicating that biparental inheritance of mtDNA occurs in filamentous fungi. Despite these studies, there is currently a lack of research on mtDNA genetics in fungi, potentially leading to an underestimation of the full scope of its inheritance patterns.

Another important aspect of fungal mtDNA is that it can harbor mutations that are associated with virulence and fungicide resistance. Notably, in a study that evolved a strain of the pathogenic yeast *C. neoformans* by serial passage through the wax moth *Galleria mellonella*, the resulting strain acquired an insertion in the promoter region of the mitochondrial gene *nad1* that led to its overexpression and higher ATP production, and which further correlated with an increase in virulence against mice (Merryman et al., 2020). Resistance to quinone outside inhibitor fungicides that target complex III of the electron transfer chain is frequently associated with the mutations in the cytochrome *b* gene (*cytb*) that lead to the amino acid substitutions p.G143A, p.G137R, or p.F129L (Yin et al., 2023). However, the presence of an intron neighboring codon 143 of *cytb* has been reported to prevent the emergence of the p.G143A mutation that confers almost complete tolerance to QoIs (Banno et al., 2009; Luo et al., 2010). Collectively, these observations support the idea that the organization of fungal mitochondrial genomes influences virulence and the ability to gain resistance to fungicides.

Genomic analyses of fungal mitochondria have been gaining attention over the last few years. For example, as of February 2024, there are 692 mt genomes in the NCBI's Organelle Genome Resources database, 301 (43%) of which were added since 2020. Fungal mt genomes harbor a standard set of genes encoding proteins involved in the electron transport chain and oxidative phosphorylation, as well as sets of genes encoding transfer RNAs and ribosomal subunits. However, fungal mt genomes also exhibit remarkable diversity when it comes to size, and presence/absence of introns and mobile genetic elements. Notably, within the yeast genus *Metschnikowia*, the size of the mt genomes varies from 21 kb to 187 kb, with intronic content ranging from zero to 66% of the entire mt genome (Lee et al., 2020). The diversity of mt genomes within *Metschnikowia* also extends to chromatin organization, with some species having circular mtDNA while others having linear mtDNA containing telomeric repeats (Lee et al., 2020). Further genomic studies will help shed into light the true diversity of fungal mtDNA and the advantage of such dynamic organization of this organelle genome.

1.8 The tomato leaf mold pathogen *Cladosporium fulvum*

Cladosporium fulvum (syn. *Passalora fulva*, *Fulvia fulva*) is a haploid (hemi-)biotrophic fungal pathogen member of the Ascomycetes (Dothideomycetes; Capnodiales) that causes tomato leaf mold (Thomma et al., 2005). The disease is more prevalent in greenhouses and high tunnels with high relative humidity that is typically required for conidia germination and disease onset (Thomma et al., 2005). Over the last 40 years, this fungus has been a valuable model for the study of plant-microbe interactions (De Wit et al., 2009; de Wit, 2016). Notably, the first TE in filamentous fungi was identified in *C. fulvum* (McHale et al., 1989), and the first fungal effector gene (*Avr9*) was cloned from this pathogen (van Kan et al., 1991). During host colonization, *C. fulvum* secretes a plethora of effector proteins in the tomato leaf apoplast, many of which are already known to be recognized by corresponding tomato resistance proteins encoded by the so-called *Cf* genes (Thomma et al., 2005; de Wit, 2016). A total of 24 *Cf* genes (*Cf-1* to *Cf-24*) have thus far been identified and mapped to 10 out of the 12 chromosomes of the tomato genome (Zhao et al., 2022). All of the described *Cf* genes encode receptor-like proteins composed of an extracellular leucine-rich repeat region, a transmembrane region, and a short cytoplasmic domain with no signaling function (Mesarich et al., 2023; Thomma et al., 2005).

The interaction between *C. fulvum* and tomato is driven by the gene-for-gene model (Flor, 1971). In this model, the product of a resistance gene in the plant (*Cf* in tomato) recognizes the product of a matching avirulence gene in the fungus (*Avr/Ecp* in *C. fulvum*), thereby triggering a hypersensitive response (HR) that halts the infection. In compatible interactions, however, this recognition does not take place and thus infections can proceed. Based on these two types of interactions, *C. fulvum* isolates are classified into physiological races that are assigned according to the resistance *Cf* genes that they can overcome, resulting in compatible interactions. For example, isolates of *C. fulvum* Race 2.5.9 can overcome and cause disease in tomato plants carrying the resistance genes *Cf-2*, *Cf-5*, and *Cf-9* (Chua et al., 1998; de Wit, 2016).

To date, at least 10 *C. fulvum* effectors (Avr2, Avr4, Avr4E, Avr5, Avr9, Ecp1, Ecp2, Ecp4, Ecp5, and Ecp6) have been shown to be avirulence determinants in different tomato accessions with matching *Cf* genes in tomato (de Wit, 2016). Despite the ongoing study of effector genes from *C. fulvum* since the 1980s, molecular functions have been assigned to only three of them, namely Avr2, Avr4, and Ecp6. The Avr2 effector inhibits cysteine proteases that are required for basal defense in tomato (van Esse et al., 2008). The Avr4 effector binds to the chitin layer of the fungal cell wall, protecting it against hydrolysis by host chitinases (van den Burg et al., 2006). Finally, the Ecp6 effector binds to chitin oligomers released from the fungal cell wall, preventing recognition of *C. fulvum* by chitin receptors in tomato (Bolton et al., 2008; De Jonge et al., 2010). Many other effector-like proteins are secreted by *C. fulvum* (Mesarich et al., 2018), however their function remains elusive.

Incompatible interactions between tomato and *C. fulvum* due to effector recognition by the host increase the selection pressure on the effector-coding genes towards avoidance of recognition. As expected, many *C. fulvum* effector genes were shown to harbor mutations, some of which are associated with overcoming Cf proteins. Interestingly, some effector genes under selection pressure from cognate Cf resistance proteins tend to accumulate certain types of mutations. For example, *C. fulvum* Race 4 isolates overcome Cf-4-mediate resistance almost exclusively through point mutations in the coding sequence of *Avr4* gene that lead to amino caid substitutions of preferential cysteine residues in the encoded protein (Stergiopoulos et al., 2007; de Wit, 2016). In contrast, Race 2 isolates overcome Cf-2 mediated resistance though point mutations that lead to frameshifts in *Avr2,* and Race 9 isolates overcome Cf-9-mediated resistance through the deletion of the *Avr9* gene (Stergiopoulos et al., 2007; de Wit, 2016). Similarly, Race 5 and Race 4E isolates overcome Cf-5- and Cf-4E-mediated resistance frequently through the deletion of *Avr5* and *Avr4E* (Stergiopoulos et al., 2007; de Wit, 2016). Possible explanations for such skewed frequency in types of mutations include the dispensability of the effector genes for tomato colonization and the location of the effector genes in the genome. The later hypothesis was proposed when the first reference genome of *C.*

fulvum was obtained (De Wit et al., 2012). In this study, the authors observed that *Avr9* is in a region of the genome rich in TEs, which could facilitate gene loss through SVs mediated by TEs. In comparison, *Avr4* is located in a region of the genome with low abundance of TEs, wherein gene deletion could be less likely to occur by chance. These observations suggest that genes encoding effectors exhibit different modes of evolution that is largely influenced by their location in the genome. Based on this, one hypothesis is that the location of effector genes in the genome of *C. fulvum* can be used to predict their typical mode of evolution to overcome their matching resistance genes in tomato. In this context, the investigation of the landscape of SVs in the genome of *C. fulvum*, which has not been done thus far, can help elucidate the extent to which SVs are important for pathogenicity and adaptive evolution of *C. fulvum*.

1.9 The grape powdery mildew pathogen *Erysiphe necator*

Erysiphe necator (syn. *Uncinula necator*) is a haploid, obligate biotroph fungal pathogen member of the Ascomycetes (Leotiomycetes, Erysiphales) that causes grape powdery mildew, one of the most severe and widespread diseases in western vineyards (Gadoury et al., 2012). The pathogen can infect all green tissues of the plant, thus drastically reducing quality of the berries and leading to yield losses. The disease is of particular importance in California, where 99% of the commercially grown table grapes in the USA are produced (California table grape commission, 2024). In 2022, the grape acreage in California was estimated at 742,000, which includes 127,000 from table grapes and 615,000 from wine grape (California Department of Food and Agriculture, 2023b). In the same year, grapes ranked 2nd among the most valued commodities in California, with an estimated year-round value of USD$ 5.54 billion, after dairy products with a predicted year-round value of USD$ 10.4 billion (California Department of Food and Agriculture, 2023a).

Because most commercial grapevine cultivars are highly susceptible to *E. necator*, significant financial resources are allocated annually to manage grape powdery mildew (Feechan et al., 2011; Gadoury et al., 2012). In 2011, it was estimated that approximately $189 million was spent in California alone for grape powdery mildew control, constituting a substantial 37% of the total gross production value and making it the highest costs for control among all grape diseases (Fuller et al., 2014). For this reason, ongoing efforts to develop resistant cultivars through breeding that aim to introduce resistant genes from closely related species in the Vitaceae into *V. vinifera* are underway.

Thus far, at least 16 loci have been associated with resistance against grape powdery mildew, located in eight out of the 19 chromosomes of the *V. vinifera* genome (Sosa-Zuniga et al., 2022). These loci are named *Ren* or *Run*, which are acronyms of Resistance to *Erysiphe/Uncinula necator*, and comprehend *Ren1, Ren1.2, Ren2, Ren3, Ren4, Ren5, Ren6, Ren7, Ren8, Ren9, Ren10, Ren11, Run1, Run1.2, Run2.1, and Run2.2*. The exact genes that confer resistance to grape powdery mildew are unknown in most of the *Run* and *Ren* loci. However, they typically trigger HR leading subsequently to programmed cell death, a typical outcome of effector-triggered immunity (ETI). In addition, *Run1* has been shown to contain a family of seven consecutive *R* genes encoding proteins containing a Toll/Interleukin-1 receptor domain (Feechan et al., 2013). Thus, it is believed that *Run* and *Ren* loci contain genes encoding R proteins that activate ETI upon recognition of effectors from *E. necator*. Currently there is a lack of functional studies to characterize effectors from *E. necator*, and interaction with R proteins was never shown. Thus far, only one effector (CSEP080) from *E. necator* has been functionally studied and shown to suppress HR in *Nicotiana benthamiana* (Mu et al., 2023).

Population genetics studies performed using isolates collected in Europe and Australia identified two genetically distinct populations of *E. necator*, named A and B (Amrani and Corio-Costet, 2006; Núñez et al., 2006; Péros et al., 2005). Although population A shows less genetic diversity than population B, the biological relevance of these populations is not clear as a trait that effectively differentiate them remains elusive. By sequencing three nuclear genes (i.e., internal transcribed spacer, beta-tubulin, and translation elongation factor 1-alpha) from 146 isolates of *E. necator* representing USA, Europe, and Australia, a study showed that population A dominates the East Coast in the USA, while individuals from population B are found in Europe, Australia, and in the West Coast in the USA (Brewer and Milgroom, 2010). This same study showed that East USA was the richest region in distinct haplotypes of *E. necator*, thus suggesting that this pathogen originated in the East USA and was subsequently introduced into Europe, possibly in the mid-1800s. From Europe, *E. necator* was likely introduced into other regions, including the Mediterranean, Australia, and West USA, which explains the similar haplotypes found in these regions. USA as the place of origin of *E. necator* is also supported by North American grape varieties, which are typically more resistant to the pathogen, possibly resulted from a longer co-evolutionary history (Cadle-Davidson et al., 2011). However, estimation of the divergence of populations A and B have not been reported, and genome-wide comparisons among individuals from these two populations have not been performed yet.

The first whole-genome sequencing and assembly for *E. necator* isolate C-strain resulted in a highly fragmented reference genome due to a high content of repetitive DNA that is difficult to assemble (Jones et al., 2014). Nonetheless, this same study revealed a positive correlation between copy number of the gene *Cyp51* and resistance to sterol demethylase inhibitor fungicides, indicating that structural variations, specifically copy number variations, can be a mechanism of adaptation in *E. necator*. However, the genome architecture and landscape of SVs mediating adaptation in *E. necator* have yet to be explored.

1.10 Scope and outline of the dissertation

In this dissertation, I describe advances in our understanding of fungal nuclear and mitochondrial genome organization and how these contribute to pathogen adaptation to their hosts and antifungal agents, by generating chromosome-level nuclear genome assemblies as well as complete mitochondrial genome assemblies, for the tomato (hemi-)biotrophic pathogen *Cladosporium fulvum* and the obligate pathogen of grape *Erysiphe necator*. Specifically, I sought to answer the following questions:

- How are the nucleal and mitochondrial genomes of *C. fulvum* and *E. necator* organized and are there any prominent differences between the two pathogens?

- To what extent repeats and transposable elements shape the nuclear and mitochondrial genome architectures of *C. fulvum* and *E. necator*?
- What is the landscape of genomic structural variations in *C. fulvum* and *E. necator*, and how these contribute to host adaptation and the development of fungicide resistance?
- To what extent does alternative splicing during infection of the host affects virulence-associated genes and increases proteome diversity in the tomato pathogen *C. fulvum*?

To achieve my aims and answer these questions, I applied advanced bioinformatics techniques to analyze next generation sequencing data of the fungal pathogens *C. fulvum* and *E. necator*. Summaries of the chapters in this dissertation are shown next.

In **Chapter 2,** I present a near-complete reference genome for *C. fulvum* obtained by combining third-generation sequencing with Hi-C chromatin capture technologies. The resulting assembly contains 67.2 Mb organized into 14 chromosomes, one of which was found to be dispensable. The *C. fulvum* genome is composed of 49.7% repetitive DNA and shows a peculiar 'checkerboard' pattern composed of gene-dense, repeat-poor regions interspersed by gene-sparse, repeat-rich regions. A total of 39.2% of the genome of *C. fulvum* was predicted to be affected by RIP mutations and signatures of RIP leakage was observed more abundantly in the dispensable chromosome. A total of 345 candidate effector genes were identified, predominantly located in gene-sparse regions, supporting the 'two-speed genome' model of evolution. Effector genes, including *Avr9*, *Avr5*, and *Avr4E*, that are typically deleted to overcome matching resistance genes in tomato were located in repeat-rich regions that likely promote their deletion. The results indicate that *C. fulvum* evolved a dynamic genome architecture that may provide advantages for host adaptation.

In **Chapter 3,** I present an extensive comparative analysis at the chromosome-level of five isolates of *C. fulvum*. The five isolates shared 13 core chromosomes, while 2 chromosomes exhibited presence/absence variation. The accessory chromosomes were significantly smaller in size, and one carried pseudogenized copies of two effector genes. Analysis of the structural variation (SV) landscape of *C. fulvum* revealed that nearly all SVs in the genome of this pathogen were insertions or deletions of regions rich in transposable elements (TEs), and had no evident effect on predicted genes. However, three SVs likely resulting from non-homologous recombination events mediated by the presence of TEs led to the deletion of the effector genes *Avr9*, *Avr5*, and *Avr4E*. These findings unveil novel evolutionary facets in the genome of *C. fulvum* and offer new perspectives on the significance of SVs in fungal plant pathogens to overcome host resistance.

In **Chapter 4,** I analyzed the patterns of alternative splicing (AS) occurring in the tomato pathogen *C. fulvum* during a complete infection cycle of its tomato host. The results indicated that 40% of all the genes were predicted to undergo AS, which is at the high end compared to other fungi. Moreover, 59% of the AS genes encoded multiple proteins, which were predicted to contribute to 31% of the proteome diversity in *C. fulvum*. The AS genes in *C. fulvum* were more abundant in repeat-rich core chromosomes, and exhibited significantly longer 5' intergenic regions richer in repetitive DNA compared to non-AS genes, suggesting that the genome organization may influence the occurrence of AS. Finally, estimation of transcript expression revealed genes involved in basic cellular functions, as well as genes encoding candidate effectors differentially spliced during interaction with tomato. This study revealed for the first time the AS landscape of *C. fulvum*, and provide clues for the importance of AS in fungal pathogens.

In **Chapter 5**, I present a chromosome-scale genome assembly for *E. necator* obtained by combining third-generation sequencing with Hi-C chromatin capture technologies. The resulting 81.1 Mb genome assembly is organized into 34 scaffolds, 11 of which represent complete chromosomes. The 62.7% predicted TE content was evenly distributed among chromosomes with no evident clustering, except within pericentromeric regions, which were almost entirely composed of TEs with signatures of recent proliferation. Younger gene duplicates exhibited more relaxed selection pressure and were more frequently located in tandem compared to older gene duplicates. The gene *HI914_00624* encoding a putative secreted carboxylesterase exhibited the highest variation in copy number, between 1 and 31, among six isolates analyzed. I further demonstrated that this gene belongs to a putative new family of inactive carboxylesterases unique to powdery mildew fungi. This study sheds light on the genomic architectural characteristics of *E. necator*, and demonstrates to what extent gene duplication and distribution of TEs shape its genome organization.

In **Chapter 6**, I present a reference mitochondrial genome for *E. necator*, the first for a powdery mildew species. In this chapter, I show that the mitochondrial genome of *E. necator* contains a total of 70 introns, making it one of the richest in introns mitochondrial genomes among fungi. I also demonstrate that *E. necator* has an atypical bicistronic-like expression of the mitochondrial genes *atp6* and *nad3*, and exhibits notably high mitochondria copy number per cell, as evidenced by an analysis of five isolates. Together, these findings provide new perspectives on the structure of mitochondrial genomes of powdery mildew pathogens.

In **Chapter 7**, I present an extensive comparative genomic analysis of the mitochondrial genomes of *E. necator* and three other powdery mildew pathogens. The analysis revealed that the mitochondrial genomes of powdery mildew pathogens have similar gene content, but vary remarkably in size ranging from 109.8 kb bp in *B. graminis* f. sp. *tritici* to 332.2 kb in *G. cichoracearum*, which is the largest mitochondrial genome of a fungal pathogen reported to date. Moreover, the mitochondrial genomes of *B. graminis* f. sp. *tritici* and *G. cichoracearum* have bimodal GC content distributions not previously observed in fungal mitochondrial genomes. Finally, the cytochrome *b* genes of the powdery mildew pathogens analyzed were exceptionally rich in introns, and harbored rare open reading frames encoding reverse transcriptases that were likely acquired horizontally.

Lastly, in **Chapter 8**, I provide a comprehensive conclusion that synthesizes the main discoveries outlined in this dissertation. This conclusion underscores their relevance to th e fungal genetics and plant pathology fields while also suggesting potential avenues for future research.

1.11 Data availability

Values used to generate statistics and illustrations are available as Supplementary Tables at https://zenodo.org/records/11211529. Specifically, the complete list of all 17,789 genome assemblies summarized in Table 1.1 is shown in Table 1.S1. Values used to generate the figures Fig 1.1, Fig 1.2, and Fig 1.3 are shown in Table 1.S2. The GenBank or RefSeq accession numbers of the genomes shown in Fig 1.4 are listed in Table 1.S3.

1.12 References

Abel, H.J., Larson, D.E., Regier, A.A., Chiang, C., Das, I., Kanchi, K.L., et al. (2020) Mapping and characterization of structural variation in 17,795 human genomes. *Nature*, 583, 83–89. https://doi.org/10.1038/s41586-020-2371-0.

Ahmed, A.A., Pedersen, C. & Thordal-Christensen, H. (2016) The Barley Powdery Mildew Effector Candidates CSEP0081 and CSEP0254 Promote Fungal Infection Success. *PLOS ONE*, 11, e0157586. https://doi.org/10.1371/journal.pone.0157586.

Akagi Yasunori, Akamatsu Hajime, Otani Hiroshi, & Kodama Motoichiro (2009) Horizontal Chromosome Transfer, a Mechanism for the Evolution and Differentiation of a Plant-Pathogenic Fungus. *Eukaryotic Cell*, 8, 1732–1738. https://doi.org/10.1128/ec.00135-09.

Amrani, L. & Corio-Costet, M.-F. (2006) A single nucleotide polymorphism in the β-tubulin gene distinguishing two genotypes of *Erysiphe necator* expressing different symptoms on grapevine. *Plant pathology*, 55, 505–512.

Auxier, B., Debets, A.J.M., Stanford, F.A., Rhodes, J., Becker, F.M., Reyes Marquez, F., et al. (2023) The human fungal pathogen *Aspergillus fumigatus* can produce the highest known number of meiotic crossovers. *PLOS Biology*, 21, e3002278. https://doi.org/10.1371/journal.pbio.3002278.

Badet, T., Oggenfuss, U., Abraham, L., McDonald, B.A. & Croll, D. (2020) A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biology*, 18, 12. https://doi.org/10.1186/s12915-020-0744-3.

Banno, S., Yamashita, K., Fukumori, F., Okada, K., Uekusa, H., Takagaki, M., et al. (2009) Characterization of QoI resistance in *Botrytis cinerea* and identification of two types of mitochondrial cytochrome b gene. *Plant Pathol.*, 58, 120–129. https://doi.org/10.1111/j.1365-3059.2008.01909.x.

Barragan, A.C., Latorre, S.M., Malmgren, A., Harant, A., Win, J., Sugihara, Y., et al. (2024) Multiple horizontal mini-chromosome transfers drive genome evolution of clonal blast fungus lineages. *bioRxiv*, 2024.02.13.580079. https://doi.org/10.1101/2024.02.13.580079.

Benami, M., Isack, Y., Grotsky, D., Levy, D. & Kofman, Y. (2020) The Economic Potential of Arbuscular Mycorrhizal Fungi in Agriculture. *Grand challenges in fungal biotechnology*, 239–279.

Bentham, A.R., Petit-Houdenot, Y., Win, J., Chuma, I., Terauchi, R., Banfield, M.J., et al. (2021) A single amino acid polymorphism in a conserved effector of the multihost blast fungus pathogen expands host-target binding spectrum. *PLOS Pathogens*, 17, e1009957. https://doi.org/10.1371/journal.ppat.1009957.

Berger, K.H. & Yaffe, M.P. (2000) Mitochondrial DNA inheritance in *Saccharomyces cerevisiae*. *Trends in Microbiology*, 8, 508–513. https://doi.org/10.1016/S0966-842X(00)01862-X.

Bertazzoni, S., Williams, A.H., Jones, D.A., Syme, R.A., Tan, K.-C. & Hane, J.K. (2018) Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic Fungi. *Molecular Plant-Microbe Interactions*, 31, 779–788.

Bing, J., Hu, T., Zheng, Q., Muñoz, J.F., Cuomo, C.A. & Huang, G. (2020) Experimental evolution identifies adaptive aneuploidy as a mechanism of fluconazole resistance in *Candida auris*. *Antimicrobial Agents and Chemotherapy*, 65, 10–1128.

Black, B., Lee, C., Horianopoulos, L.C., Jung, W.H. & Kronstad, J.W. (2021) Respiring to infect: Emerging links between mitochondria, the electron transport chain, and fungal pathogenesis. *PLOS Pathogens*, 17, e1009661. https://doi.org/10.1371/journal.ppat.1009661.

Bolton, M.D., Van Esse, H.P., Vossen, J.H., De Jonge, R., Stergiopoulos, I., Stulemeijer, I.J., et al. (2008) The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. *Molecular microbiology*, 69, 119–136.

Bonfante, P. & Genre, A. (2010) Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications*, 1, 48. https://doi.org/10.1038/ncomms1046.

Bongomin, F., Gago, S., Oladele, R.O. & Denning, D.W. (2017) Global and Multi-National Prevalence of Fungal Diseases—Estimate Precision. *Journal of Fungi*, 3. https://doi.org/10.3390/jof3040057.

Bradshaw, R.E., Sim, A.D., Chettri, P., Dupont, P.-Y., Guo, Y., Hunziker, L., et al. (2019) Global population genomics of the forest pathogen *Dothistroma septosporum* reveal chromosome duplications in high dothistromin-producing strains. *Molecular plant pathology*, 20, 784–799.

Bräse, S., Encinas, A., Keck, J. & Nising, C.F. (2009) Chemistry and Biology of Mycotoxins and Related Fungal Metabolites. *Chemical Reviews*, 109, 3903–3990. https://doi.org/10.1021/cr050001f.

Brewer, M.T. & Milgroom, M.G. (2010) Phylogeography and population structure of the grape powdery mildew fungus, *Erysiphe necator*, from diverse Vitis species. *BMC Evolutionary Biology*, 10, 1–13. https://doi.org/10.1186/1471-2148-10-268.

Brown, C.J., Todd, K.M. & Rosenzweig, R.F. (1998) Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Molecular Biology and Evolution*, 15, 931–942. https://doi.org/10.1093/oxfordjournals.molbev.a026009.

Bucknell, A.H. & McDonald, M.C. (2023) That's no moon, it's a Starship: Giant transposons driving fungal horizontal gene transfer. *Molecular Microbiology*.

Burg, H.A. van den, Harrison, S.J., Joosten, M.H., Vervoort, J. & Wit, P.J. de (2006) *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. *Molecular Plant-Microbe Interactions*, 19, 1420–1430.

Cadle-Davidson, L., Chicoine, D.R. & Consolie, N.H. (2011) Variation within and among Vitis spp. for foliar resistance to the powdery mildew pathogen *Erysiphe necator*. *Plant Disease*, 95, 202–211.

California Department of Food and Agriculture (2023a) California Agricultural Production Statistics. Available at https://www.cdfa.ca.gov/statistics/. Last accessed Feb 20, 2024.

California Department of Food and Agriculture (2023b) Grape Acreage Report. Available at www.nass.usda.gov. Last accessed Feb 20, 2024.

California table grape commission (2024) Grapes from California. Available https://www.grapesfromcalifornia.com/. Last accessed Feb 20, 2024.

Cao, Y., Kümmel, F., Logemann, E., Gebauer, J.M., Lawson, A.W., Yu, D., et al. (2023) Structural polymorphisms within a common powdery mildew effector scaffold as a driver of coevolution with cereal immune receptors. *Proceedings of the National Academy of Sciences*, 120, e2307604120. https://doi.org/10.1073/pnas.2307604120.

Carpenter, D., Dhar, S., Mitchell, L.M., Fu, B., Tyson, J., Shwan, N.A.A., et al. (2015) Obesity, starch digestion and amylase: association between copy number variants at human salivary (*AMY1*) and pancreatic (*AMY2*) amylase genes. *Human Molecular Genetics*, 24, 3472–3480. https://doi.org/10.1093/hmg/ddv098.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., et al. (2016) Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles Feschotte, C. (Ed.). *PLOS Genetics*, 12, e1006108. https://doi.org/10.1371/journal.pgen.1006108.

Chan, D.C. (2006) Mitochondria: dynamic organelles in disease, aging, and development. *Cell*, 125, 1241–1252. https://doi.org/10.1016/j.cell.2006.06.010.

Chen, S., Yuan, N., Schnabel, G. & Luo, C. (2017) Function of the genetic element 'Mona' associated with fungicide resistance in *Monilinia fructicola*. *Molecular Plant Pathology*, 18, 90–97. https://doi.org/10.1111/mpp.12387.

Chen, X., Magee, B.B., Dawson, D., Magee, P.T. & Kumamoto, C.A. (2004) Chromosome 1 trisomy compromises the virulence of *Candida albicans*. *Molecular Microbiology*, 51, 551–565. https://doi.org/10.1046/j.1365-2958.2003.03852.x.

Chua, N. –H., Hetherington, A.M., Hooley, R., Irvine, R.F., Thomas, C.M., Dixon, M.S., et al. (1998) Genetic and molecular analysis of tomato *Cf* genes for resistance to *Cladosporium fulvum*. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353, 1413–1424. https://doi.org/10.1098/rstb.1998.0296.

Chuma, I., Isobe, C., Hotta, Y., Ibaragi, K., Futamata, N., Kusaba, M., et al. (2011) Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus Magnaporthe oryzae and related species. *PLoS Pathogens*, 7, e1002147.

Cissé, O.H., Ma, L., Dekker, J.P., Khil, P.P., Youn, J.-H., Brenchley, J.M., et al. (2021) Genomic insights into the host specific adaptation of the *Pneumocystis* genus. *Communications biology*, 4, 305.

Clutterbuck, A.J. (2011) Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genetics and Biology*, 48, 306–326.

Cohen, A.B., Cai, G., Price, D.C., Molnar, T.J., Zhang, N. & Hillman, B.I. (2024) The massive 340 megabase genome of *Anisogramma anomala*, a biotrophic ascomycete that causes eastern filbert blight of hazelnut. *BMC Genomics*, 25, 347. https://doi.org/10.1186/s12864-024-10198-1.

Cook, D.E., Kramer, H.M., Torres, D.E., Seidl, M.F. & Thomma, B.P.H.J. (2020) A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen Weigel, D. and Tyler, B.M. (Eds.). *eLife*, 9, e62208. https://doi.org/10.7554/eLife.62208.

Cormier, A., Chebbi, M.A., Giraud, I., Wattier, R., Teixeira, M., Gilbert, C., et al. (2021) Comparative Genomics of Strictly Vertically Transmitted, Feminizing Microsporidia Endosymbionts of Amphipod Crustaceans. *Genome Biology and Evolution*, 13, evaa245. https://doi.org/10.1093/gbe/evaa245.

Cuomo, C.A., Güldener, U., Xu, J.-R., Trail, F., Turgeon, B.G., Di Pietro, A., et al. (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*, 317, 1400–1402.

De Jonge, R., Peter van Esse, H., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., et al. (2010) Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *science*, 329, 953–955.

De Wit, P.J., Joosten, M.H., Thomma, B.H. & Stergiopoulos, I. (2009) Gene for gene models and beyond: the *Cladosporium fulvum*-Tomato pathosystem. Plant relationships. Springer, pp. 135–156.

De Wit, P.J., Van Der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K.A., Aerts, A.L., et al. (2012) The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLoS genetics*, 8, e1003088.

Denning, D.W. (2024) Global incidence and mortality of severe fungal disease. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(23)00692-8.

Dennis, M.Y., Nuttle, X., Sudmant, P.H., Antonacci, F., Graves, T.A., Nefedov, M., et al. (2012) Evolution of Human-Specific Neural *SRGAP2* Genes by Incomplete Segmental Duplication. *Cell*, 149, 912–922. https://doi.org/10.1016/j.cell.2012.03.033.

Dong, S., Raffaele, S. & Kamoun, S. (2015) The two-speed genomes of filamentous pathogens: waltz with plants. *Current opinion in genetics & development*, 35, 57–65.

Dort, E.N., Layne, E., Feau, N., Butyaev, A., Henrissat, B., Martin, F.M., et al. (2023) Large-scale genomic analyses with machine learning uncover predictive patterns associated with fungal phytopathogenic lifestyles and traits. *Scientific Reports*, 13, 17203. https://doi.org/10.1038/s41598-023-44005-w.

Elliott, S. (2022) Food Security: How Do Crop Plants Combat Pathogens?

Esse, H.P. van, Van't Klooster, J.W., Bolton, M.D., Yadeta, K.A., Baarlen, P. van, Boeren, S., et al. (2008) The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *The Plant Cell*, 20, 1948–1963.

Faino, L., Seidl, M.F., Shi-Kunne, X., Pauper, M., Berg, G.C. van den, Wittenberg, A.H., et al. (2016) Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research*, 26, 1091–1100.

Faris, J.D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., et al. (2010) A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proceedings of the National Academy of Sciences*, 107, 13544–13549.

Feechan, A., Anderson, C., Torregrosa, L., Jermakow, A., Mestre, P., Wiedemann-Merdinoglu, S., et al. (2013) Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia rotundifolia* identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. *The Plant Journal*, 76, 661–674. https://doi.org/10.1111/tpj.12327.

Feechan, A., Kabbara, S. & Dry, I.B. (2011) Mechanisms of powdery mildew resistance in the *Vitaceae* family. *Molecular Plant Pathology*, 12, 263–274. https://doi.org/10.1111/j.1364-3703.2010.00668.x.

Feurtey, A., Lorrain, C., Croll, D., Eschenbrenner, C., Freitag, M., Habig, M., et al. (2020) Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*. *BMC Genomics*, 21, 588. https://doi.org/10.1186/s12864-020-06871-w.

Feurtey, A., Lorrain, C., McDonald, M.C., Milgate, A., Solomon, P.S., Warren, R., et al. (2023) A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nature Communications*, 14, 1059. https://doi.org/10.1038/s41467-023-36674-y.

Fisher, M.C., Gurr, S.J., Cuomo, C.A., Blehert, D.S., Jin, H., Stukenbrock, E.H., et al. (2020) Threats posed by the fungal kingdom to humans, wildlife, and agriculture. *MBio*, 11, 10–1128.

Fletcher, K., Shin, O.-H., Clark, K.J., Feng, C., Putman, A.I., Correll, J.C., et al. (2022) Ancestral Chromosomes for Family Peronosporaceae Inferred from a Telomere-to-Telomere Genome Assembly of *Peronospora effusa*. *Molecular Plant-Microbe Interactions®*, 35, 450–463. https://doi.org/10.1094/MPMI-09-21-0227-R.

Flor, H.H. (1971) Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology*, 9, 275–296. https://doi.org/10.1146/annurev.py.09.090171.001423.

Fokkens, L., Shahi, S., Connolly, L.R., Stam, R., Schmidt, S.M., Smith, K.M., et al. (2018) The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. *BioRxiv*, 465070.

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., 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. https://doi.org/10.1186/s12864-018-4750-6.

Frantzeskakis, L., Kusch, S. & Panstruga, R. (2019a) The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology*, 20, 3–7.

Frantzeskakis, L., Németh, M.Z., Barsoum, M., Kusch, S., Kiss, L., Takamatsu, S., et al. (2019b) The *Parauncinula polyspora* draft genome provides insights into patterns of gene erosion and genome expansion in powdery mildew fungi. *mBio*, 10, e01692-19.

Fuller, K.B., Alston, J.M. & Sambucci, O.S. (2014) The value of powdery mildew resistance in grapes: evidence from California. *Wine Economics and Policy*, 3, 90–107.

Gadoury, D.M., Cadle-Davidson, L., Wilcox, W.F., Dry, I.B., Seem, R.C. & Milgroom, M.G. (2012) Grapevine powdery mildew (*Erysiphe necator*): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. *Molecular Plant Pathology*, 13, 1–16.

Galagan, J.E. & Selker, E.U. (2004) RIP: the evolutionary cost of genome defense. *Trends in Genetics*, 20, 417–423. https://doi.org/10.1016/j.tig.2004.07.007.

Galazka, J.M. & Freitag, M. (2014) Variability of chromosome structure in pathogenic fungi—of 'ends and odds.' *Host–microbe interactions: fungi/parasites/viruses*, 20, 19–26. https://doi.org/10.1016/j.mib.2014.04.002.

Galazka, J.M., Klocko, A.D., Uesaka, M., Honda, S., Selker, E.U. & Freitag, M. (2016) *Neurospora* chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Research*, 26, 1069–1080.

Gandia, A., Brandhof, J.G. van den, Appels, F.V.W. & Jones, M.P. (2021) Flexible Fungal Materials: Shaping the Future. *Trends in Biotechnology*, 39, 1321–1331. https://doi.org/10.1016/j.tibtech.2021.03.002.

Gerstein, A.C., Chun, H.-J.E., Grant, A. & Otto, S.P. (2006) Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS genetics*, 2, e145.

Gilchrist, C. & Stelkens, R. (2019) Aneuploidy in yeast: Segregation error or adaptation mechanism? *Yeast*, 36, 525–539. https://doi.org/10.1002/yea.3427.

Girardini, K.N., Olthof, A.M. & Kanadia, R.N. (2023) Introns: the "dark matter" of the eukaryotic genome. *Frontiers in Genetics*, 14, 1150212.

Gluck-Thaler, E., Ralston, T., Konkel, Z., Ocampos, C.G., Ganeshan, V.D., Dorrance, A.E., et al. (2022) Giant Starship elements mobilize accessory genes in fungal genomes. *Molecular biology and evolution*, 39, msac109.

Goodwin, S.B., Ben M'Barek, S., Dhillon, B., Wittenberg, A.H., Crane, C.F., Hane, J.K., et al. (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS genetics*, 7, e1002070.

Gorkovskiy, A. & Verstrepen, K.J. (2021) The Role of Structural Variation in Adaptation and Evolution of Yeast and Other Fungi. *Genes*, 12. https://doi.org/10.3390/genes12050699.

Gourlie, R., McDonald, M., Hafez, M., Ortega-Polo, R., Low, K.E., Abbott, D.W., et al. (2022) The pangenome of the wheat pathogen *Pyrenophora tritici-repentis* reveals novel transposons associated with necrotrophic effectors ToxA and ToxB. *BMC Biology*, 20, 239. https://doi.org/10.1186/s12915-022-01433-w.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic acids research*, 42, D699–D704.

Guin, K., Sreekumar, L. & Sanyal, K. (2020) Implications of the Evolutionary Trajectory of Centromeres in the Fungal Kingdom. *Annual Review of Microbiology*, 74, 835–853. https://doi.org/10.1146/annurev-micro-011720-122512.

Gunisova, S., Elboher, E., Nosek, J., Gorkovoy, V., Brown, Y., Lucier, J.-F., et al. (2009) Identification and comparative analysis of telomerase RNAs from *Candida* species reveal conservation of functional elements. *Rna*, 15, 546–559.

Gupta, Y.K., Marcelino-Guimarães, F.C., Lorrain, C., Farmer, A., Haridas, S., Ferreira, E.G.C., et al. (2023) Major proliferation of transposable elements shaped the genome of the soybean rust pathogen *Phakopsora pachyrhizi*. *Nature Communications*, 14, 1–16.

Haag, K.L., Pombert, J.-F., Sun, Y., Albuquerque, N.R.M. de, Batliner, B., Fields, P., et al. (2020) Microsporidia with Vertical Transmission Were Likely Shaped by Nonadaptive Processes. *Genome Biology and Evolution*, 12, 3599–3614. https://doi.org/10.1093/gbe/evz270.

Hane, J.K. & Oliver, R.P. (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC bioinformatics*, 9, 1–12.

Hanson, S.J., Cinnéide, E.Ó., Salzberg, L.I., Wolfe, K.H., McGowan, J., Fitzpatrick, D.A., et al. (2021) Genomic diversity, chromosomal rearrangements, and interspecies hybridization in the *Ogataea polymorpha* species complex. *G3 Genes|Genomes|Genetics*, 11, jkab211. https://doi.org/10.1093/g3journal/jkab211.

Harari, Y., Ram, Y., Rappoport, N., Hadany, L. & Kupiec, M. (2018) Spontaneous changes in ploidy are common in yeast. *Current Biology*, 28, 825–835.

Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A., et al. (2020) 101 Dothideomycetes genomes: a test case for predicting lifestyles and emergence of pathogens. *Studies in mycology*, 96, 141–153.

Harley, J. (1971) Fungi in ecosystems. *Journal of Ecology*, 59, 653–668.

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

Hickman, M.A., Zeng, G., Forche, A., Hirakawa, M.P., Abbey, D., Harrison, B.D., et al. (2013) The 'obligate diploid' *Candida albicans* forms mating-competent haploids. *Nature*, 494, 55–59.

Hirakawa, M.P., Chyou, D.E., Huang, D., Slan, A.R. & Bennett, R.J. (2017) Parasex Generates Phenotypic Diversity de Novo and Impacts Drug Resistance and Virulence in *Candida albicans*. *Genetics*, 207, 1195–1211. https://doi.org/10.1534/genetics.117.300295.

Hollox, E.J., Zuccherato, L.W. & Tucci, S. (2022) Genome structural variation in human evolution. *Focus issue: Studying genetic variation through an evolutionary lens*, 38, 45–58. https://doi.org/10.1016/j.tig.2021.06.015.

Hose, J., Yong, C.M., Sardi, M., Wang, Z., Newton, M.A. & Gasch, A.P. (2015) Dosage compensation can buffer copy-number variation in wild yeast. *Elife*, 4, e05462.

Houterman, P.M., Ma, L., Van Ooijen, G., De Vroomen, M.J., Cornelissen, B.J.C., Takken, F.L.W., et al. (2009) The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *The Plant Journal*, 58, 970–978. https://doi.org/10.1111/j.1365-313X.2009.03838.x.

Huth, L., Ash, G.J., Idnurm, A., Kiss, L. & Vaghefi, N. (2021) The "Bipartite" Structure of the First Genome of *Ampelomyces quisqualis*, a Common Hyperparasite and Biocontrol Agent of Powdery Mildews, May Point to Its Evolutionary Origin from Plant Pathogenic Fungi. *Genome Biology and Evolution*, 13, evab182. https://doi.org/10.1093/gbe/evab182.

Irelan, J.T., Hagemann, A.T. & Selker, E.U. (1994) High frequency repeat-induced point mutation (RIP) is not associated with efficient recombination in Neurospora. *Genetics*, 138, 1093–1103. https://doi.org/10.1093/genetics/138.4.1093.

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K.C.H., McGuire, B., Gubler, W.D., et al. (2014) Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. *BMC Genomics*, 15, 1081. https://doi.org/10.1186/1471-2164-15-1081.

Jonge, R. de, Bolton, M.D., Kombrink, A., Berg, G.C. van den, Yadeta, K.A. & Thomma, B.P. (2013) Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome research*, 23, 1271–1282.

Kaaniche, F., Hamed, A., Abdel-Razek, A.S., Wibberg, D., Abdissa, N., El Euch, I.Z., et al. (2019) Bioactive secondary metabolites from new endophytic fungus *Curvularia*. sp isolated from *Rauwolfia macrophylla*. *PloS one*, 14, e0217627.

Kan, J.A. van, Van den Ackerveken, G. & De Wit, P. (1991) Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. *Mol. Plant-Microbe Interact*, 4, 52–59.

Kang, S., Lebrun, M.H., Farrall, L. & Valent, B. (2001) Gain of Virulence Caused by Insertion of a Pot3 Transposon in a *Magnaporthe grisea* Avirulence Gene. *Molecular Plant-Microbe Interactions®*, 14, 671–674. https://doi.org/10.1094/MPMI.2001.14.5.671.

Kanokmedhakul, S., Lekphrom, R., Kanokmedhakul, K., Hahnvajanawong, C., Bua-Art, S., Saksirirat, W., et al. (2012) Cytotoxic sesquiterpenes from luminescent mushroom *Neonothopanus nambi*. *Tetrahedron*, 68, 8261–8266.

Katinka, M.D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., et al. (2001) Genome sequence and gene compaction of the eukaryote parasite </i>Encephalitozoon cuniculi</i>. *Nature*, 414, 450–453. https://doi.org/10.1038/35106579.

Kersey, P.J., Lawson, D., Birney, E., Derwent, P.S., Haimel, M., Herrero, J., et al. (2010) Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic acids research*, 38, D563–D569.

Kim, S., Liachko, I., Brickner, D.G., Cook, K., Noble, W.S., Brickner, J.H., et al. (2017) The dynamic three-dimensional organization of the diploid yeast genome Ren, B. (Ed.). *eLife*, 6, e23623. https://doi.org/10.7554/eLife.23623.

King, R., Urban, M., Hammond-Kosack, M.C.U., Hassani-Pak, K. & Hammond-Kosack, K.E. (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics*, 16, 544. https://doi.org/10.1186/s12864-015-1756-1.

Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., et al. (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics*, 47, 410–415. https://doi.org/10.1038/ng.3223.

Komluski, J., Habig, M. & Stukenbrock, E.H. (2023) Repeat-Induced Point Mutation and Gene Conversion Coinciding with Heterochromatin Shape the Genome of a Plant-Pathogenic Fungus. *Mbio*, e03290-22.

Komluski, J., Stukenbrock, E.H. & Habig, M. (2022) Non-Mendelian transmission of accessory chromosomes in fungi. *Chromosome Research*, 1–13. https://doi.org/10.1007/s10577-022-09691-8.

Kress, W.J., Soltis, D.E., Kersey, P.J., Wegrzyn, J.L., Leebens-Mack, J.H., Gostel, M.R., et al. (2022) Green plant genomes: What we know in an era of rapidly expanding opportunities. *Proceedings of the National Academy of Sciences*, 119, e2115640118.

Ksiezopolska, E., Schikora-Tamarit, M.À., Beyer, R., Nunez-Rodriguez, J.C., Schüller, C. & Gabaldón, T. (2021) Narrow mutational signatures drive acquisition of multidrug resistance in the fungal pathogen *Candida glabrata*. *Current Biology*, 31, 5314-5326.e10. https://doi.org/10.1016/j.cub.2021.09.084.

Kües, U. (2015) Fungal enzymes for environmental management. *Environmental biotechnology • Energy biotechnology*, 33, 268–278. https://doi.org/10.1016/j.copbio.2015.03.006.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860–921. https://doi.org/10.1038/35057062.

Langner, T., Harant, A., Gomez-Luciano, L.B., Shrestha, R.K., Malmgren, A., Latorre, S.M., et al. (2021) Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLOS Genetics*, 17, e1009386. https://doi.org/10.1371/journal.pgen.1009386.

Lee, D.K., Hsiang, T., Lachance, M.-A. & Smith, D.R. (2020) The strange mitochondrial genomes of Metschnikowia yeasts. *Current Biology*, 30, R800–R801. https://doi.org/10.1016/j.cub.2020.05.075.

Levin, H.L. & Moran, J.V. (2011) Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, 12, 615–627.

Li Cheng-Xi, Liu Lin, Zhang Ting, Luo Xue-Mei, Feng Jia-Xun, & Zhao Shuai (2022) Three-Dimensional Genome Map of the Filamentous Fungus *Penicillium oxalicum*. *Microbiology Spectrum*, 10, e02121-21. https://doi.org/10.1128/spectrum.02121-21.

Li, X., Yang, F., Li, D., Zhou, M., Wang, X., Xu, Q., et al. (2015) Trisomy of chromosome R confers resistance to triazoles in *Candida albicans*. *Medical Mycology*, 53, 302–309. https://doi.org/10.1093/mmy/myv002.

Liu, S., Lin, G., Ramachandran, S.R., Daza, L.C., Cruppe, G., Tembo, B., et al. (2024) Rapid mini-chromosome divergence among fungal isolates causing wheat blast outbreaks in Bangladesh and Zambia. *New Phytologist*, 241, 1266–1276. https://doi.org/10.1111/nph.19402.

Lue, N.F. (2010) Plasticity of telomere maintenance mechanisms in yeast. *Trends in biochemical sciences*, 35, 8–17.

Luo, C.-X., Hu, M.-J., Jin, X., Yin, L.-F., Bryson, P.K. & Schnabel, G. (2010) An intron in the cytochrome b gene of *Monilinia fructicola*. mitigates the risk of resistance development to QoI fungicides. *Pest Manag. Sci.*, 66, 1308–1315. https://doi.org/10.1002/ps.2016.

Ma, L.-J., Van Der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.-J., Di Pietro, A., et al. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, 464, 367–373.

Manzoni, M. & Rollini, M. (2002) Biosynthesis and biotechnological production of statins by filamentous fungi and application of these cholesterol-lowering drugs. *Applied microbiology and biotechnology*, 58, 555–564.

Maruta, N., Outram, M.A. & Kobe, B. (2023) Mildew RALPHs up in arms with cereals. *Proceedings of the National Academy of Sciences*, 120, e2311817120. https://doi.org/10.1073/pnas.2311817120.

Mason, J.M.O. & McEachern, M.J. (2018) Chromosome ends as adaptive beginnings: the potential role of dysfunctional telomeres in subtelomeric evolvability. *Current Genetics*, 64, 997–1000. https://doi.org/10.1007/s00294-018-0822-z.

Mat Razali, N., Cheah, B.H. & Nadarajah, K. (2019) Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. *International journal of molecular sciences*, 20, 3597.

McDonald, M.C., Taranto, A.P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., et al. (2019) Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *MBio*, 10, e01515-19.

McEachern, M.J. & Blackburn, E.H. (1994) A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proceedings of the National Academy of Sciences*, 91, 3453–3457.

McEachern, M.J. & Haber, J.E. (2006) Break-induced replication and recombinational telomere elongation in yeast. *Annu. Rev. Biochem.*, 75, 111–135.

McHale, M.T., Roberts, I.N., Talbot, N.J. & Oliver, R.P. (1989) Expression of reverse transcriptase genes in *Fulvia fulva*. *Mol Plant Microbe Interact*, 2, 165–168.

Mehrabi, R., Mirzadi Gohari, A. & Kema, G.H.J. (2017) Karyotype Variability in Plant-Pathogenic Fungi. *Annual Review of Phytopathology*, 55, 483–503. https://doi.org/10.1146/annurev-phyto-080615-095928.

Merényi, Z., Krizsán, K., Sahu, N., Liu, X.-B., Bálint, B., Stajich, J.E., et al. (2023) Genomes of fungi and relatives reveal delayed loss of ancestral gene families and evolution of key fungal traits. *Nature Ecology & Evolution*, 7, 1221–1231. https://doi.org/10.1038/s41559-023-02095-9.

Merryman, M., Crigler, J., Seipelt-Thiemann, R. & McClelland, E. (2020) A mutation in *C. neoformans* mitochondrial NADH dehydrogenase results in increased virulence in mice. *Virulence*, 11, 1366–1378. https://doi.org/10.1080/21505594.2020.1831332.

Mesarich, C.H., Barnes, I., Bradley, E.L., Rosa, S. de la, Wit, P.J.G.M. de, Guo, Y., et al. (2023) Beyond the genomes of *Fulvia fulva* (syn. *Cladosporium fulvum*) and *Dothistroma septosporum*: New insights into how these fungal pathogens interact with their host plants. *Molecular Plant Pathology*, 24, 474–494. https://doi.org/10.1111/mpp.13309.

Mesarich, C.H., Ökmen, B., Rovenich, H., Griffiths, S.A., Wang, C., Karimi Jashni, M., et al. (2018) Specific hypersensitive response–associated recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato. *Molecular plant-microbe interactions*, 31, 145–162.

Meyer, V., Basenko, E.Y., Benz, J.P., Braus, G.H., Caddick, M.X., Csukai, M., et al. (2020) Growing a circular economy with fungal biotechnology: a white paper. *Fungal biology and biotechnology*, 7, 1–23.

Miao, V.P., Covert, S.F. & VanEtten, H.D. (1991) A fungal gene for antibiotic resistance on a dispensable ("B") chromosome. *Science*, 254, 1773–1776.

Miller, M.G. & Johnson, A.D. (2002) White-opaque switching in <i>Candida albicans<i> is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell*, 110, 293–302.

Morin, E., Miyauchi, S., San Clemente, H., Chen, E.C.H., Pelin, A., Providencia, I. de la, et al. (2019) Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytologist*, 222, 1584–1598. https://doi.org/10.1111/nph.15687.

Mu, B., Teng, Z., Tang, R., Lu, M., Chen, J., Xu, X., et al. (2023) An effector of *Erysiphe necator* translocates to chloroplasts and plasma membrane to suppress host immunity in grapevine. *Horticulture Research*, 10, uhad163. https://doi.org/10.1093/hr/uhad163.

Müller, M.C., Kunz, L., Graf, J., Schudel, S. & Keller, B. (2021) Host adaptation through hybridization: genome analysis of triticale powdery mildew reveals unique combination of lineage-specific effectors. *Molecular Plant-Microbe Interactions*, 34, 1350–1357. https://doi.org/10.1094/MPMI-05-21-0111-SC.

Müller, M.C., Praz, C.R., Sotiropoulos, A.G., Menardo, F., Kunz, L., Schudel, S., et al. (2019) A chromosome-scale genome assembly reveals a highly dynamic effector repertoire of wheat powdery mildew. *New Phytologist*, 221, 2176–2189. https://doi.org/10.1111/nph.15529.

Murat, C., Payen, T., Noel, B., Kuo, A., Morin, E., Chen, J., et al. (2018) Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle. *Nature ecology & evolution*, 2, 1956–1965.

Niskanen, T., Lücking, R., Dahlberg, A., Gaya, E., Suz, L.M., Mikryukov, V., et al. (2023) Pushing the Frontiers of Biodiversity Research: Unveiling the Global Diversity, Distribution, and Conservation of Fungi. *Annual Review of Environment and Resources*. https://doi.org/10.1146/annurev-environ-112621-090937.

Noble, S.M. & Johnson, A.D. (2007) Genetics of *Candida albicans*, a diploid human fungal pathogen. *Annu. Rev. Genet.*, 41, 193–211.

Núñez, Y., Gallego, J., Ponz, F. & Raposo, R. (2006) Analysis of population structure of *Erysiphe necator* using AFLP markers. *Plant Pathology*, 55, 650–656.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., et al. (2022) The complete sequence of a human genome. *Science*, 376, 44–53. https://doi.org/10.1126/science.abj6987.

Ocaña-Pallarès, E., Williams, T.A., López-Escardó, D., Arroyo, A.S., Pathmanathan, J.S., Bapteste, E., et al. (2022) Divergent genomic trajectories predate the origin of animals and fungi. *Nature*, 609, 747–753. https://doi.org/10.1038/s41586-022-05110-4.

Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F.E., Singh, N.K., Abraham, L., et al. (2021) A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen Weigel, D., Mirouze, M., Joly-Lopez, Z., and Quadrana, L. (Eds.). *eLife*, 10, e69249. https://doi.org/10.7554/eLife.69249.

Ohm, R.A., Feau, N., Henrissat, B., Schoch, C.L., Horwitz, B.A., Barry, K.W., et al. (2012) Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS pathogens*, 8, e1003037.

Omer, B., Sudharsan, S., Di, G., Adi, D.-F., Varda, Z., T., D.M., et al. (2023) Aneuploidy Formation in the Filamentous Fungus *Aspergillus flavus* in Response to Azole Stress. *Microbiology Spectrum*, 11, e04339-22. https://doi.org/10.1128/spectrum.04339-22.

Omrane, S., Sghyer, H., Audéon, C., Lanen, C., Duplaix, C., Walker, A.-S., et al. (2015) Fungicide efflux and the *MgMFS1* transporter contribute to the multidrug resistance phenotype in *Zymoseptoria tritici* field isolates. *Environmental Microbiology*, 17, 2805–2823. https://doi.org/10.1111/1462-2920.12781.

Osterhage, J.L. & Friedman, K.L. (2009) Chromosome End Maintenance by Telomerase. *Journal of Biological Chemistry*, 284, 16061–16065. https://doi.org/10.1074/jbc.R900011200.

Pedersen, C., Themaat, E.V.L. van, McGuffin, L.J., Abbott, J.C., Burgis, T.A., Barton, G., et al. (2012) Structure and evolution of barley powdery mildew effector candidates. *BMC Genomics*, 13, 694.

Pennington, H.G., Jones, R., Kwon, S., Bonciani, G., Thieron, H., Chandler, T., et al. (2019) The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA Dodds, P.N. (Ed.). *PLOS Pathogens*, 15, e1007620. https://doi.org/10.1371/journal.ppat.1007620.

Péros, J.-P., Troulet, C., Guerriero, M., Michel-Romiti, C. & Notteghem, J.-L. (2005) Genetic variation and population structure of the grape powdery mildew fungus, *Erysiphe necator*, in southern France. *European Journal of Plant Pathology*, 113, 407–416.

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018) Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556, 339–344.

Piovesan, A., Antonaros, F., Vitale, L., Strippoli, P., Pelleri, M.C. & Caracausi, M. (2019) Human protein-coding genes and gene feature statistics in 2019. *BMC research notes*, 12, 1–5.

Plissonneau, C., Hartmann, F.E. & Croll, D. (2018) Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology*, 16, 5. https://doi.org/10.1186/s12915-017-0457-4.

Pradhan, A., Ghosh, S., Sahoo, D. & Jha, G. (2021) Fungal effectors, the double edge sword of phytopathogens. *Current Genetics*, 67, 27–40. https://doi.org/10.1007/s00294-020-01118-3.

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., et al. (2010) Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*, 330, 1540–1543.

Rahnama, M., Wang, B., Dostart, J., Novikova, O., Yackzan, D., Yackzan, A., et al. (2021) Telomere roles in fungal genome evolution and adaptation. *Frontiers in genetics*, 12, 676751.

Ramezani-Rad, M., Hollenberg, C.P., Lauber, J., Wedler, H., Griess, E., Wagner, C., et al. (2003) The *Hansenula polymorpha* (strain CBS4732) genome sequencing and analysis. *FEMS Yeast Research*, 4, 207–215. https://doi.org/10.1016/S1567-1356(03)00125-9.

Rokas, A. (2022) Evolution of the human pathogenic lifestyle in fungi. *Nature Microbiology*, 7, 607–619. https://doi.org/10.1038/s41564-022-01112-0.

Ropars, J., Vega, R.C.R. de la, López-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, É., et al. (2015) Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Current Biology*, 25, 2562–2569.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., Van de Wouw, A.P., Couloux, A., et al. (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature communications*, 2, 1–10.

Sanjoy, P., Mark, S., Heredge, T.G., Hong, L., Daisuke, H., Katsuya, G., et al. (2019) AtrR Is an Essential Determinant of Azole Resistance in *Aspergillus fumigatus*. *mBio*, 10, 10.1128/mbio.02563-18. https://doi.org/10.1128/mbio.02563-18.

Scheele, B.C., Pasmans, F., Skerratt, L.F., Berger, L., Martel, A., Beukema, W., et al. (2019) Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science*, 363, 1459–1463. https://doi.org/10.1126/science.aav0379.

Scheppke, J. (2023) Humongous Fungus. Available at https://www.oregonencyclopedia.org/articles/humongous-fungus-armillaria-ostoyae/. Last accessed Feb 20, 2024.

Schwessinger, B., Sperschneider, J., Cuddy, W.S., Garnica, D.P., Miller, M.E., Taylor, J.M., et al. (2018) A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus Puccinia striiformis f. sp. tritici reveals high interhaplotype diversity. *MBio*, 9, 10–1128.

Seidl, M.F., Kramer, H.M., Cook, D.E., Fiorin, G.L., Berg, G.C.M. van den, Faino, L., et al. (2020) Repetitive elements contribute to the diversity and evolution of centromeres in the fungal genus *Verticillium*. *mBio*, 11, e01714-20. https://doi.org/10.1128/mBio.01714-20.

Selker, E.U. (1990) Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annual review of genetics*, 24, 579–613.

Selker, E.U. (2002) Repeat-induced gene silencing in fungi. *Advances in genetics*, 46, 439–450.

Selker, E.U. & Garrett, P.W. (1988) DNA sequence duplications trigger gene inactivation in *Neurospora crassa*. *Proceedings of the National Academy of Sciences*, 85, 6870–6874.

Selmecki, A., Forche, A. & Berman, J. (2006) Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science*, 313, 367–370.

Seong, K. & Krasileva, K.V. (2023) Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nature Microbiology*, 8, 174–187. https://doi.org/10.1038/s41564-022-01287-6.

Shampay, J., Szostak, J.W. & Blackburn, E.H. (1984) DNA sequences of telomeres maintained in yeast. *Nature*, 310, 154–157.

Sharma, G., Aminedi, R., Saxena, D., Gupta, A., Banerjee, P., Jain, D., et al. (2019) Effector mining from the *Erysiphe pisi* haustorial transcriptome identifies novel candidates involved in pea powdery mildew pathogenesis. *Mol. Plant Pathol.*, 20, 1506–1522. https://doi.org/10.1111/mpp.12862.

Sionov, E., Lee, H., Chang, Y.C. & Kwon-Chung, K.J. (2010) *Cryptococcus neoformans* Overcomes Stress of Azole Drugs by Formation of Disomy in Specific Multiple Chromosomes. *PLOS Pathogens*, 6, e1000848. https://doi.org/10.1371/journal.ppat.1000848.

Smith, A.C., Rizvi, H., Hickman, M.A. & Morran, L.T. (2022) ≈ze of Tetraploid *Candida albicans* Evolved in Nematode Hosts. *Frontiers in Fungal Biology*, 3.

Smith, K.M., Phatale, P.A., Sullivan, C.M., Pomraning, K.R. & Freitag, M. (2011) Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. *Molecular and Cellular Biology*, 31, 2528–2542. https://doi.org/10.1128/MCB.01285-10.

Sosa-Zuniga, V., Vidal Valenzuela, Á., Barba, P., Espinoza Cancino, C., Romero-Romero, J.L. & Arce-Johnson, P. (2022) Powdery Mildew Resistance Genes in Vines: An Opportunity to Achieve a More Sustainable Viticulture. *Pathogens*, 11. https://doi.org/10.3390/pathogens11060703.

Spanu, P.D. (2012) The genomics of obligate (and nonobligate) biotrophs. *Annual review of Phytopathology*, 50, 91–109.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., et al. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, 330, 1543–1546. https://doi.org/http//doi.org/10.1126/science.1194573.

Spatafora, J.W., Aime, M.C., Grigoriev, I.V., Martin, F., Stajich, J.E. & Blackwell, M. (2017) The fungal tree of life: from molecular systematics to genome-scale phylogenies. *The fungal kingdom*, 1–34.

Steinhauer, D., Salat, M., Frey, R., Mosbach, A., Luksch, T., Balmer, D., et al. (2019) A dispensable paralog of succinate dehydrogenase subunit C mediates standing resistance towards a subclass of SDHI fungicides in *Zymoseptoria tritici*. *PLOS Pathogens*, 15, e1007780. https://doi.org/10.1371/journal.ppat.1007780.

Stergiopoulos, I., De Kock, M.J., Lindhout, P. & De Wit, P.J. (2007) Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. *Molecular Plant-Microbe Interactions*, 20, 1271–1283.

Stergiopoulos, I., Kourmpetis, Y.A., Slot, J.C., Bakker, F.T., De Wit, P.J. & Rokas, A. (2012) In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Molecular Biology and Evolution*, 29, 3371–3384.

Stergiopoulos, I. & Wit, P.J. de (2009) Fungal effector proteins. *Annual review of phytopathology*, 47, 233–263.

Stokstad, E. (2019) This fungus has wiped out more species than any other disease.

Strader, C.R., Pearce, C.J. & Oberlies, N.H. (2011) Fingolimod (FTY720): a recently approved multiple sclerosis drug based on a fungal secondary metabolite. *Journal of natural products*, 74, 900–907.

Sullivan, B.A., Blower, M.D. & Karpen, G.H. (2001) Determining centromere identity: cyclical stories and forking paths. *Nature Reviews Genetics*, 2, 584–596. https://doi.org/10.1038/35084512.

Sun, X., Xu, Q., Ruan, R., Zhang, T., Zhu, C. & Li, H. (2013) PdMLE1, a specific and active transposon acts as a promoter and confers *Penicillium digitatum* with DMI resistance. *Environmental Microbiology Reports*, 5, 135–142. https://doi.org/10.1111/1758-2229.12012.

Sunshine, A.B., Payen, C., Ong, G.T., Liachko, I., Tan, K.M. & Dunham, M.J. (2015) The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS biology*, 13, e1002155.

Talbot, N.J., Salch, Y.P., Ma, M. & Hamer, J.E. (1993) Karyotypic variation within clonal lineages of the rice blast fungus, *Magnaporthe grisea*. *Applied and Environmental Microbiology*, 59, 585–593.

Taylor, J.W., Branco, S., Gao, C., Hann-Soden, C., Montoya, L., Sylvain, I., et al. (2017) Sources of fungal genetic variation and associating it with phenotypic diversity. *The Fungal Kingdom*, 635–655.

Teimoori-Boghsani, Y., Ganjeali, A., Cernava, T., Müller, H., Asili, J. & Berg, G. (2020) Endophytic fungi of native Salvia abrotanoides plants reveal high taxonomic diversity and unique profiles of secondary metabolites. *Frontiers in microbiology*, 10, 3013.

Testa, A.C., Oliver, R.P. & Hane, J.K. (2016) OcculterCut: A Comprehensive Survey of AT-Rich Regions in Fungal Genomes. *Genome Biology and Evolution*, 8, 2044–2064. https://doi.org/10.1093/gbe/evw121.

Thomas, G., Stoner, O., Costa, F. & Ames, R.M. (2023) Fungal Pathogen Gene Selection for Predicting the Onset of Infection Using a Multi-Stage Machine Learning Approach. *bioRxiv*, 2023–09.

Thomma, B.P., Van Esse, H.P., Crous, P.W. & Wit, P.J. de (2005) *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. *Molecular plant pathology*, 6, 379–393.

Tobias, P.A., Schwessinger, B., Deng, C.H., Wu, C., Dong, C., Sperschneider, J., et al. (2021) *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. *G3*, 11, jkaa015.

Torres, D.E., Oggenfuss, U., Croll, D. & Seidl, M.F. (2020) Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews*.

Torres, D.E., Reckard, A.T., Klocko, A.D. & Seidl, M.F. (2023) Nuclear genome organization in fungi: from gene folding to Rabl chromosomes. *FEMS Microbiology Reviews*, 47, fuad021. https://doi.org/10.1093/femsre/fuad021.

Torres-Cruz, T.J., Billingsley Tobias, T.L., Almatruk, M., Hesse, C.N., Kuske, C.R., Desirò, A., et al. (2017) *Bifiguratus adelaidae*, gen. et sp. nov., a new member of Mucoromycotina in endophytic and soil-dwelling habitats. *Mycologia*, 109, 363–378.

Treindl, A.D., Stapley, J., Croll, D. & Leuchtmann, A. (2023) Two-speed genomes of *Epichloe* fungal pathogens show contrasting signatures of selection between species and across populations. *Molecular Ecology*, n/a, e17242. https://doi.org/10.1111/mec.17242.

Urquhart, A.S., Chong, N.F., Yang, Y. & Idnurm, A. (2022) A large transposable element mediates metal resistance in the fungus *Paecilomyces variotii*. *Current Biology*, 32, 937–950.

Urquhart, A.S., Vogan, A.A., Gardiner, D.M. & Idnurm, A. (2023) *Starships* are active eukaryotic transposable elements mobilized by a new family of tyrosine recombinases. *Proceedings of the National Academy of Sciences*, 120, e2214521120.

Vande Zande, P., Zhou, X. & Selmecki, A. (2023) The Dynamic Fungal Genome: Polyploidy, Aneuploidy and Copy Number Variation in Response to Stress. *Annual Review of Microbiology*, 77, 341–361.

Venice, F., Ghignone, S., Salvioli di Fossalunga, A., Amselem, J., Novero, M., Xianan, X., et al. (2020) At the nexus of three kingdoms: the genome of the mycorrhizal fungus *Gigaspora margarita* provides insights into plant, endobacterial and fungal interactions. *Environmental microbiology*, 22, 122–141.

Vlaardingerbroek, I., Beerens, B., Schmidt, S.M., Cornelissen, B.J. & Rep, M. (2016) Dispensable chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. *Molecular Plant Pathology*, 17, 1455–1466.

Waalwijk, C., Taga, M., Zheng, S.-L., Proctor, R.H., Vaughan, M.M. & O'Donnell, K. (2018) Karyotype evolution in *Fusarium*. *IMA Fungus*, 9, 13–26. https://doi.org/10.5598/imafungus.2018.09.01.02.

Wacker, T., Helmstetter, N., Wilson, D., Fisher, M.C., Studholme, D.J. & Farrer, R.A. (2023) Two-speed genome evolution drives pathogenicity in fungal pathogens of animals. *Proceedings of the National Academy of Sciences*, 120, e2212633120.

Wadi, L., El Jarkass, H.T., Tran, T.D., Islah, N., Luallen, R.J. & Reinke, A.W. (2023) Genomic and phenotypic evolution of nematode-infecting microsporidia. *PLOS Pathogens*, 19, e1011510. https://doi.org/10.1371/journal.ppat.1011510.

Walt, D. van der, Steenkamp, E.T., Wingfield, B.D. & Wilken, P.M. (2023) Evidence of Biparental Mitochondrial Inheritance from Self-Fertile Crosses between Closely Related Species of Ceratocystis. *Journal of Fungi*, 9. https://doi.org/10.3390/jof9060686.

Wang, C., Skrobek, A. & Butt, T.M. (2003) Concurrence of losing a chromosome and the ability to produce destruxins in a mutant of *Metarhizium anisopliae*. *FEMS Microbiology Letters*, 226, 373–378.

Wang, L., Sun, Y., Sun, X., Yu, L., Xue, L., He, Z., et al. (2020) Repeat-induced point mutation in *Neurospora crassa* causes the highest known mutation rate and mutational burden of any cellular life. *Genome biology*, 21, 1–23.

Wang, M., Fu, H., Shen, X.-X., Ruan, R., Rokas, A. & Li, H. (2019) Genomic features and evolution of the conditionally dispensable chromosome in the tangerine pathotype of *Alternaria alternata*. *Molecular plant pathology*, 20, 1425–1438.

Wang, Y., Wu, J., Yan, J., Guo, M., Xu, L., Hou, L., et al. (2022) Comparative genome analysis of plant ascomycete fungal pathogens with different lifestyles reveals distinctive virulence strategies. *BMC Genomics*, 23, 34. https://doi.org/10.1186/s12864-021-08165-1.

Wei, H., Zhong, Z., Li, Z., Zhang, Y., Stukenbrock, E.H., Tang, B., et al. (2023) Loss of the accessory chromosome converts a pathogenic tree root fungus into a mutualistic endophyte. *Plant Communications*, 100672.

Westerhoven, A. van, Aguilera-Galvez, C., Nakasato-Tagami, G., Shi-Kunne, X., Dijkstra, J., Martinez de la Parte, E., et al. (2023) Segmental Duplications Drive the Evolution of Accessory Regions in a Major Crop Pathogen. *bioRxiv*, 2023–06.

Whelan, W.L. & Soll, D.R. (1982) Mitotic recombination in *Candida albicans*: recessive lethal alleles linked to a gene required for methionine biosynthesis. *Molecular and General Genetics MGG*, 187, 477–485.

Wieloch, W. (2006) Chromosome visualisation in filamentous fungi. *Journal of Microbiological Methods*, 67, 1–8. https://doi.org/10.1016/j.mimet.2006.05.022.

Winter, D.J., Ganley, A.R., Young, C.A., Liachko, I., Schardl, C.L., Dupont, P.-Y., et al. (2018) Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. *PLoS genetics*, 14, e1007467.

Wit, P.J. de (2016) *Cladosporium fulvum* effectors: weapons in the arms race with tomato. *Annual review of phytopathology*, 54, 1–23.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., et al. (2002) The genome sequence of Schizosaccharomyces pombe. *Nature*, 415, 871–880.

Wu, J., Kou, Y., Bao, J., Li, Y., Tang, M., Zhu, X., et al. (2015) Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector *AvrPi9* that triggers *Pi9*-mediated blast resistance in rice. *New Phytologist*, 206, 1463–1475.

Wu, Y., Ma, X., Pan, Z., Kale, S.D., Song, Y., King, H., et al. (2018) Comparative genome analyses reveal sequence features reflecting distinct modes of host-adaptation between dicot and monocot powdery mildew. *BMC Genomics*, 19, 705. https://doi.org/10.1186/s12864-018-5069-z.

Xu, J. & Wang, P. (2015) Mitochondrial inheritance in basidiomycete fungi. *Special Issue: Fungal sex and mushrooms – A credit to Lorna Casselton*, 29, 209–219. https://doi.org/10.1016/j.fbr.2015.02.001.

Yang Feng, Gritsenko Vladimir, Lu Hui, Zhen Cheng, Gao Lu, Berman Judith, et al. (2021) Adaptation to Fluconazole via Aneuploidy Enables Cross-Adaptation to Amphotericin B and Flucytosine in *Cryptococcus neoformans*. *Microbiology Spectrum*, 9, e00723-21. https://doi.org/10.1128/Spectrum.00723-21.

Yang, H., Yu, H. & Ma, L.-J. (2020) Accessory chromosomes in *Fusarium oxysporum*. *Phytopathology*, 110, 1488–1496.

Yildirir, G., Sperschneider, J., Malar C, M., Chen, E.C., Iwasaki, W., Cornell, C., et al. (2022) Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytologist*, 233, 1097–1107.

Yin, Y., Miao, J., Shao, W., Liu, X., Zhao, Y. & Ma, Z. (2023) Fungicide Resistance: Progress in Understanding Mechanism, Monitoring, and Management. *Phytopathology®*, 113, 707–718. https://doi.org/10.1094/PHYTO-10-22-0370-KD.

Yuan, H., Jin, C., Pei, H., Zhao, L., Li, X., Li, J., et al. (2021) The Powdery Mildew Effector CSEP0027 Interacts With Barley Catalase to Regulate Host Immunity. *Frontiers in Plant Science*, 12.

Zhao, T., Pei, T., Jiang, J., Yang, H., Zhang, H., Li, J., et al. (2022) Understanding the mechanisms of resistance to tomato leaf mold: A review. *Horticultural Plant Journal*, 8, 667–675.

Zheng Qiushi, Liu Jing, Qin Juanxiu, Wang Bingjie, Bing Jian, Du Han, et al. (2022) Ploidy Variation and Spontaneous Haploid-Diploid Switching of *Candida glabrata* Clinical Isolates. *mSphere*, 7, e00260-22. https://doi.org/10.1128/msphere.00260-22.

1.13 Supplementary materials

1.13.1 Supplementary tables

Table 1.S1: Fungal genome assemblies available in GenBank. The table shows 17,789 genome assemblies from fungal species available in GenBank as of April 26th, 2024. The accession IDs, assembly level statuses, taxonomy ID, and organism lineages are presented. This table is available at https://zenodo.org/records/11211529.

Table 1.S2: Information about 1994 annotated fungal genomes available in GenBank. The table shows accession IDs, taxonomy IDs, organism names, phyla, genome size, number of predicted protein-coding genes, gene density (genes per Mb), average size of introns, and average number of introns per gene, of 1994 fungal genomes available in GenBank as of April 26th, 2024. This table is available at https://zenodo.org/records/11211529.

Table 1.S3: Accession numbers of the genomes or chromosomes shown in the main figure Fig 1.4.

NCBI accession	Organism name
GCA_000002495.2	*Magnaporthe oryzae* 70-15
GCA_000002945.2	*Schizosaccharomyces pombe* 972h-
GCA_000143535.4	*Botrytis cinerea* B05.10
GCA_000146045.2	*Saccharomyces cerevisiae* S288C
GCA_000219625.1	*Zymoseptoria tritici* IPO323
GCA_000240135.3	*Fusarium graminearum* PH-1
GCA_008452785.1	*Parastagonospora nodorum* SN15
GCA_020716725.1	*Rhizophagus irregularis* DAOM-197198
GCA_900519115.1	*Blumeria graminis* f. sp. *tritici* v3.16
GCF_000149955.1	*Fusarium oxysporum* f. sp. *lycopersici* 4287
GCF_000182925.2	*Neurospora crassa* OR74A
NC_000021.9	Human chromosome 21

Chapter 2

A chromosome-scale genome assembly of the tomato pathogen *Cladosporium fulvum* reveals a compartmentalized genome architecture and the presence of a dispensable chromosome

Alex Z. Zaccaron
Li-Hung Chen
Anastasios Samaras
Ioannis Stergiopoulos

Leading author statement

I contributed to the conceptualization of this chapter, performed all analyses, generated most of the figures, contributed to the investigation and interpretation, wrote the original draft, and contributed to the review and editing of the final draft.

This chapter was published in: *Microbial Genomics*, April 2022, volume 8, number 4 (https://doi.org/10.1099/mgen.0.000819).

Abstract

Cladosporium fulvum is a fungal pathogen that causes leaf mold of tomato. The reference genome of this pathogen was released in 2012 but its high repetitive DNA content prevented a contiguous assembly and further prohibited the analysis of its genome architecture. In this study, we combined third generation sequencing technology with the Hi-C chromatin conformation capture technique, to produce a high-quality and near complete genome assembly and gene annotation of a Race 5 isolate of *C. fulvum*. The resulting genome assembly contained 67.17 Mb organized into 14 chromosomes (Chr1-to-Chr14), all of which were assembled telomere-to-telomere. The smallest of the chromosomes, Chr14, is only 460 kb in size and contains 25 genes that all encode hypothetical proteins. Notably, PCR assays revealed that Chr14 was absent in 19 out of 24 isolates of a world-wide collection of *C. fulvum*, indicating that Chr14 is dispensable. Thus, *C. fulvum* is currently the second species of Capnodiales shown to harbor dispensable chromosomes. The genome of *C. fulvum* Race 5 is 49.7% repetitive and contains 14,690 predicted genes with an estimated completeness of 98.9%, currently one of the highest among the Capnodiales. Genome structure analysis revealed a compartmentalized architecture composed of gene-dense and repeat-poor regions interspersed with gene-sparse and repeat-rich regions. Nearly 39.2% of the *C. fulvum* Race 5 genome is affected by Repeat-Induced Point (RIP) mutations and evidence of RIP leakage toward non-repetitive regions was observed in all chromosomes, indicating the RIP plays an important role in the evolution of this pathogen. Finally, 345 genes encoding candidate effectors were identified in *C. fulvum* Race 5, with a significant enrichment of their location in gene-sparse regions, in accordance with the “two-speed genome” model of evolution. Overall, the new reference genome of *C. fulvum* presents several notable features and is a valuable resource for studies in plant pathogens.

2.1 Introduction

Rapid advances in whole genome sequencing technologies over the last two decades have enabled the sequencing and comparative genome analysis of a wide range of fungal plant pathogens (Haridas et al. 2020). However, the majority of fungal genomes sequenced so far have mostly utilized short-read sequencing technologies which, while they excel at characterizing small DNA polymorphisms in populations, they have limited power when it comes to analyzing repetitive genomic regions, phasing alleles, and inferring transposable element (TE) mobilization patterns and insertion sites (Salzberg and Yorke 2005; Treangen and Salzberg 2012). Such limitations, however, can significantly impede the study of higher-order genomic architectural features, such as segmental sequence duplications, translocations, copy number variations, and other chromosomal variations that can potentially affect several aspects of microbial life, including fitness, virulence, and adaptation to adverse environmental conditions (Castanera et al. 2016; Guo et al. 2020; Möller et al. 2018; Muszewska et al. 2019; Plissonneau et al. 2017). These shortcomings can be overcome by third generation long-read sequencing and chromosome conformation capture (3C)-based techniques, which enable the *de novo* assembly of genomes to often chromosome level, thus allowing in-depth studies of genome architectures.

The increasing number of fungal genomes assembled to near-chromosome level has already revealed large differences in their architecture and organization, often even among genomes of phylogenetically closely-related fungal species or strains within a species (Covo 2020; Haridas et al. 2020; Miyauchi et al. 2020; Möller and Stukenbrock 2017). One of the most prominent features of the dynamic nature of fungal genomes is the presence of dispensable chromosomes in some species. Also known as accessory or B chromosomes, these chromosomes show a non-Mendelian mode of inheritance and they are present in some but not all individuals of a population (Houben et al. 2014). In fungi, accessory chromosomes are typically small (< 2 Mb), lack homology to chromosomes of phylogenetically closely-related species, are rich in repetitive DNA, and harbor a small number of genes that may or may not increase fitness of the organism (Bertazzoni et al. 2018). For instance, eight dispensable chromosomes have been described in the wheat pathogen *Zymoseptoria tritici* but a clear effect on fitness of the fungus has yet to be demonstrated for most of these chromosomes (Goodwin et al. 2011; Möller et al. 2018). In contrast, accessory chromosomes in *Fusarium* spp. and *Alternaria* spp. are enriched in effector genes and genes for the biosynthesis of host-selective toxins that can be transferred horizontally (Li et al. 2020; Ma et al. 2010; Tsuge et al. 2013; Wang et al. 2019; Witte et al. 2021). For example, *F. oxysporum* f. sp. *radicis-cucumerinum* has an accessory chromosome (chr^{RC}) rich in repetitive DNA and candidate effector genes, one of which (*SIX6*) contributes to virulence toward cucumber (van Dam et al. 2017). Moreover, a non-pathogenic strain of *F. oxysporum* f. sp. *radicis-cucumerinum* turned pathogenic toward cucumber by acquiring chr^{RC} through horizontal chromosome transfer (van Dam et al. 2017). Thus, accessory chromosomes can potentially allow rapid adaptation of fungal pathogens to hosts.

Next to accessory chromosomes, various other structural variations have been observed in fungal genomes. For instance, species of Dothideomycetes, the largest class of the fungal kingdom that includes many economically important plant pathogens (Wijayawardene et al. 2017), show a particularly intriguing phenomenon of mesosynteny, which refers to the conservation of the gene content between species but in a randomized order and orientation on homologous chromosomes, presumably caused by extensive intrachromosomal and rare interchromosomal rearrangements (Hane et al. 2011; Ohm et al. 2012). Although genes in fungal genomes can undergo extensive reshuffling via intrachromosomal rearrangements, genome compartmentalization has been observed as well, particularly with reference to virulence and pathogenicity-related genes, such as those encoding effector proteins (Dong et al. 2015; Möller and Stukenbrock 2017; Raffaele and Kamoun 2012; Stergiopoulos and de Wit 2009). This is the case for example with smut fungi (Dutheil et al. 2016; Navarrete et al. 2021), *Colletotrichum* spp. (Gan et al. 2021; Tsushima et al. 2019), *Verticillium* spp. (Depotter et al. 2019), and *Leptosphaeria* spp. (Grandaubert et al. 2014; Rouxel et al. 2011), among others (Frantzeskakis et al. 2019; Plissonneau et al. 2017), in which effectors genes are found clustered within their genomes. Such clusters are often embedded in subtelomeric parts of the chromosomes, and other dynamic and fast-evolving regions of the genome that are generally characterized by a low gene density, low GC-content, and high abundance of repetitive DNA and TEs (Frantzeskakis et al. 2019; Plissonneau et al. 2017). This compartmentalized genome architecture characterized by repeat-rich regions enriched with candidate effector genes, and repeat-poor regions harboring mostly housekeeping genes, has given rise to the so-called “two-speed genome” model, which is thought to facilitate the rapid evolution and adaptive diversification of genes co-localizing in repeat-rich regions (Croll and McDonald 2012; Dong et al. 2015; Raffaele and Kamoun 2012). TEs have a profound impact on the evolution and the genomic architecture of such regions, as their activity promotes genetic variation and phenotypic diversity (Castanera et al. 2016; Muszewska et al. 2019; Schrader and Schmitz 2019; Iida et al. 2015; Stergiopoulos et al. 2007a). For example, gain-of-virulence was observed in a strain of the rice pathogen *Magnaporthe oryzae* that carried a nonfunctional copy of the effector gene *AvrPi9*, which was disrupted by the insertion of a TE in its coding sequence (Wu et al. 2015). Horizontal transfer of the virulence gene *ToxA* mediated by TEs surrounding this gene has been reported among the wheat fungal pathogens *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, and *Bipolaris sorokiniana* (McDonald et al. 2019). Repetitive DNA has also been shown to accelerate genome evolution, particularly of effector genes, by the spillage of Repeat-Induced Point (RIP) mutations (Rouxel et al. 2011). RIP is a premeiotic defense mechanism specific to fungal genomes that hypermutates repetitive DNA by inducing C-to-T changes during sexual reproduction (Clutterbuck 2011; Selker 1990; Selker and Garrett 1988). Effector genes in *Leptosphaeria maculans* were shown to carry RIP mutations that “leaked” from repeats in close physical proximity (Rouxel et al. 2011). These examples indicate that a genome architecture characterized by the co-localization of genes important for pathogenicity or host adaptation and repetitive DNA, potentially enhances pathogen virulence and adaptation to hosts (Croll and McDonald 2012; Dong et al. 2015; Möller and Stukenbrock 2017).

Cladosporium fulvum (syn. *Passalora fulva*, syn. *Fulvia fulva*) is a non-obligate biotrophic fungal pathogen (Ascomycetes; Dothideomycetes; Capnodiales) and the causal agent of the tomato leaf mold (Thomma et al. 2005). Although the disease is nowadays mostly of local interest to some parts of the world, the pathogen has been extensively used as a model species to study plant-microbe interactions (De Wit et al. 2009; Joosten and De Wit 1999) and is the first fungus from which an avirulence (*Avr*) gene was ever cloned (van Kan et al. 1991). To date at least 12 effectors, i.e. Avr2, Avr4, Avr4E, Avr5, Avr9, Ecp1, Ecp2, Ecp2-2, Ecp2-3, Ecp4, Ecp5, and Ecp6 have been cloned from this pathogen and are shown to be avirulence determinants in tomato accessions with matching *Cf* resistance genes (de Wit 2016), although many more effector encoding genes have been identified *in silico* in its genome (De Wit et al. 2012). The first and so far only genome of *C. fulvum* was assembled and annotated nearly a decade ago (De Wit et al. 2012) and although the genome of *C. fulvum* isolate 0WU has provided ample insights into the biology of the fungus, its high repetitive DNA content prohibited a contiguous assembly based on short-read sequencing. Consequently, the resulting highly fragmented assembly (De Wit et al. 2012) prevented the study of the genome architecture of *C. fulvum* and the mapping of its genes to chromosomes.

In this study, we combined the PacBio single-molecule real-time (SMRT) sequencing technology (Eid et al. 2009) with Hi-C chromatin conformation capture (Lieberman-Aiden et al. 2009) to obtain a high-quality and nearly complete genome assembly for *C. fulvum* isolate Race 5 Kim (hereafter *C. fulvum* Race 5) (Stergiopoulos et al. 2007b). The resulting assembly contains 14 chromosomes (Chr1-to-Chr14), ten of which have been assembled telomere-to-telomere. Genomic analyses revealed a compartmentalized genome architecture composed of gene-dense regions interspersed with repeat-rich regions. PCR assays revealed that Chr14, the smallest of the chromosomes in *C. fulvum*, is present in a few but missing in several *C. fulvum* isolates, indicating that this chromosome is dispensable. The new reference genome of *C. fulvum* presented herein is a considerable improvement over the previous reference genome of isolate 0WU (De Wit et al. 2012) and is a valuable resource for future functional and comparative genomic studies.

2.2 Results

2.2.1 A chromosome-level and nearly complete genome assembly of *C. fulvum* Race 5

To produce a high-quality genome assembly for *C. fulvum* Race 5, a workflow was employed that combined PacBio reads, Illumina reads, and Hi-C chromatin conformation capture (Fig 2.S1). Initially, the genome of *C. fulvum* Race 5 was sequenced on a PacBio Sequel I platform, which produced a total of 583,199 reads (8.5 Gbp of data) with an average read length of 14,638 bp (~125x coverage). The assembler Canu (Koren et al. 2017) assembled the PacBio reads into 43 contigs, with a total length of 80.5 Mb. Of these, 29 contigs were removed from the assembly because they matched to bacterial genomes (24 contigs with a combined size of 12.9 Mb), were contained within other contigs, or were formed from a single PacBio read (four contigs with a combined size of 186.4 kb) (Table 2.S1), or corresponded to the mitochondrial genome of *C. fulvum* (one contig of 179.1 kb in size). To further improve the quality of the assembly, two rounds of po lishing were carried out on Pilon with 97.4M Illumina reads that were obtained, which performed 692 changes in the first round and four changes in the second round. The resulting assembly of *C. fulvum* Race 5 contained 14 contigs totaling 67.17 Mb in size, with an L_{50} of 5 and an N_{50} of 5.7 Mb. Notably, these assembly contiguity metrics are a considerable improvement over the previous reference genome of *C. fulvum* isolate 0WU (Table 2.1). Further genome size estimation based on k-mer counting of contamination-free Illumina reads indicated a genome size of 66.53 Mb, which is in agreement with the size of the obtained assembly. To verify the integrity of the assembly and identify potential misassemblies, the Illumina reads were trimmed and mapped to the polished contigs. Mapping was realized at an alignment rate of 99.67%, with 99.3% of the reads properly paired and 90.21% uniquely mapped. Of these, 91.76% mapped to the nuclear chromosomes and 7.9% mapped to the mitochondrial contig. Next, the PacBio reads were mapped to the assembled contigs and structural variants (SVs) were called with Sniffles (Sedlazeck et al. 2018), which identified only two possible SVs in the nuclear contigs. However, these SVs were not supported by the Illumina reads (Fig 2.S2). Moreover, analysis of the Illumina and PacBio coverage revealed only three possible collapsed regions located in three different contigs, with one such collapsed region co-localizing with the 18S-5.8S-28S rDNA locus (Fig 2.S3). Collectively, these results indicate that the obtained assembly of *C. fulvum* Race 5 is essentially free of major misassembly errors. As a final verification step, the obtained Hi-C data were used to evaluate the correctness of the assembly. A total of 80M Hi-C reads were produced and aligned to the genome of *C. fulvum* Race 5 with an alignment rate of 98.4%. Interaction intensity by proximity ligation supported that the 14 contigs were distinct chromosomes of *C. fulvum* Race 5 with no visible misassemblies (Fig 2.S4A). Thus, the 14 assembled contigs represent individual chromosomes (Chr) and henceforth will be referred to as Chr1-to-Chr14, according to their size, from largest to smallest.

Table 2.2: Genome assembly statistics of *Cladosporium fulvum* Race 5 compared to the previous reference genome assembly of *C. fulvum* isolate 0WU.

Assembly statistics	*C. fulvum* Race 5	*C. fulvum* 0WU
Assembly size (bp)	67,169,167	61,113,266
Scaffolds	14	4865
Contigs	14	5715
Scaffold N_{50}	5,777,465	56,512
Scaffold L_{50}	5	250
Scaffold N_{90}	3,311,397	5928
Scaffold L_{90}	11	1756
Longest scaffold	11,362,290	530,628
GC (%)	48.94	48.78
Number of gaps	0	850
Gapped bases	0	320,683

2.2.2 Overall properties of the *C. fulvum* Race 5 chromosomes

The 14 assembled chromosomes of *C. fulvum* Race 5 vary in size from 0.46 Mb to 11.36 Mb (Fig 2.1 and Table 2.2). Notably, at 11.36 Mb, Chr1 is considerably larger in size than the other chromosomes, which are at most 7.0 Mb long. All assembled chromosomes have the canonical telomeric repeat 5′-TTAGGG-3′ at both ends, indicating that they were assembled end-to-end (Fig 2.S4B and Table 2.2). As previously reported in other fungi, the Hi-C interaction frequency indicated the approximate location of centromeres

(Varoquaux et al. 2015). In *C. fulvum*, centromeres are putatively located proximal to chromosome ends (Fig 2.S4B). Specifically, Chr1, Chr2, Chr5, Chr8, Chr9, Chr10, Chr11, and Chr12 are acrocentric, whereas Chr3, Chr4, Chr6, Chr7, Chr13, and Chr14 are submetacentric (Levan et al. 1964). The distribution of protein-coding genes and of repetitive DNA throughout the genome (described in detail in later subsections) revealed an idiosyncratic compartmentalized pattern of gene-rich, repeat-poor, and high GC regions that are interspersed by gene-poor, repeat-rich, and low GC regions (Fig 2.1). However, Chr14 is an exception; at 460 kb, this chromosome is composed of nearly 80% repeats and harbors only 25 predicted genes of an unknown function. Such characteristics are typical of dispensable chr omosomes in fungi (Bertazzoni et al. 2018), suggesting that Chr14 could be dispensable.

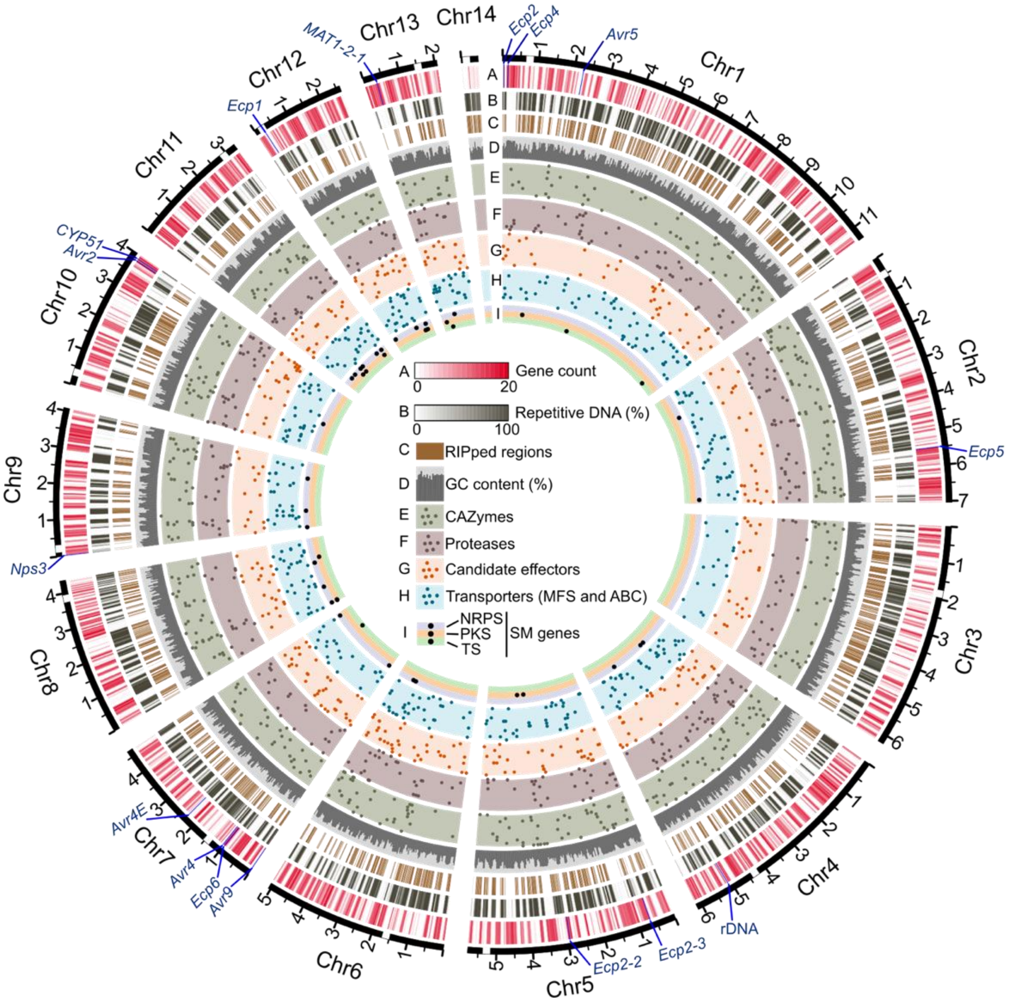


Figure 2.7: Chromosomes of *Cladosporium fulvum* Race 5. The circos plot shows the assembled chromosomes (solid black lines) with tracks representing (A) protein-coding genes, (B) repetitive DNA content, (C) regions affected by Repeat-Induced Point (RIP) mutations, (D) GC content from 30% to 60%, (E) the location of genes encoding carbohydrate-active enzymes (CAZymes), (F) the location of genes encoding proteases, (G) the location of genes encoding candidate effectors, (H) the location of genes encoding transporters from the major facilitator superfamily (MFS) and ATP-binding cassette (ABC) family, and (I) the location of genes encoding key enzymes for secondary metabolism, i.e. non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), and terpene synthases (TS). Gene locations are represented by points, and points in tracks (E, F, G, H) were randomly distributed on the perpendicular axis. The location of a few genes of general interest is indicated in the outermost track. These genes include previously described avirulence (*Avr*) and extracellular protein encoding genes (*Ecp*), the 18S-5.8S-28S rDNA, the mating type 2 (MAT1-2) locus, and *CYP51* that encodes the target enzyme of demethylation inhibitor (DMI) fungicides. The approximate location of each centromere is indicated with a white rectangle on the outermost axis and major tick marks represent Mb. Gene count (A), repetitive DNA (B), and GC content (D) were determined using a sliding window of 30 kb. The figure shows that the chromosomes of *C. fulvum* Race 5 are composed of gene-rich and repeat-poor regions that are interspersed with gene-poor and repeat-rich regions, in accordance with the "two-speed genome" model. The figure also shows that a large

portion of the genome of *C. fulvum* Race 5 is affected by RIP mutations and that genes involved in secondary metabolism are preferentially located in smaller chromosomes.

Table 2.3: Statistics of the chromosomes of Cladosporium fulvum Race 5. Copy number of telomeric repeats at the chromosomes' immediate ends are indicated when present.

Chromo- some	Size (bp)	GC (%)	Genes	Gene density (genes/Mb)	Repeats (%)	Left telomere	Right telomere
Chr1	11,362,290	48.85	2322	204.4	52.5	(CCCTAA)x9	(TTAGGG)x7
Chr2	7,036,032	49.18	1577	224.1	48.6	(CCCTAA)x19	(TTAGGG)x7
Chr3	6,232,865	47.22	1040	166.9	61.6	(CCCTAA)x8	(TTAGGG)x21
Chr4	6,141,308	49.91	1491	242.8	44.6	(CCCTAA)x15	(TTAGGG)x7
Chr5	5,777,465	49.22	1237	214.1	49.8	(CCCTAA)x12	(TTAGGG)x9
Chr6	5,061,772	48.88	1137	224.6	50.2	(CCCTAA)x7	(TTAGGG)x7
Chr7	4,686,795	48.33	951	202.9	54.2	(CCCTAA)x6	(TTAGGG)x12
Chr8	4,340,606	48.48	913	210.3	53.9	(CCCTAA)x9	(TTAGGG)x7
Chr9	4,070,492	49.83	1027	252.3	42.4	(CCCTAA)x8	(TTAGGG)x13
Chr10	4,017,737	48.77	851	211.8	49.3	(CCCTAA)x7	(TTAGGG)x15
Chr11	3,311,397	49.6	850	256.7	40.9	(CCCTAA)x32	(TTAGGG)x10
Chr12	2,606,583	50.21	698	267.8	38.3	(CCCTAA)x8	(TTAGGG)x5
Chr13	2,063,147	50.12	571	276.8	36.0	(CCCTAA)x12	(TTAGGG)x11
Chr14	460,486	45.63	25	54.3	77.9	(CCCTAA)x26	(TTAGGG)x5

2.2.3 Repetitive regions in the genome of *C. fulvum* Race 5 are heavily affected by RIP

De novo annotations of the repeats present in the genome of *C. fulvum* Race 5 with RepeatModeler v1.0.11 revealed that nearly half of its genome (33.4 Mb; 49.7%) is composed of interspersed repetitive DNA, in agreement with the estimated repeat content of 47.2% for isolate 0WU (De Wit et al. 2012). Most repeats can be classified as retrotransposons (27.5 Mb; 40.9% of the genome), and are more abundant than DNA transposons (1.2 Mb; 1.8% of the genome) and unclassified repeats (4.7 Mb; 7.0% of the genome) (Table 2.S2). Moreover, 82% of the regions putatively covered by transposable elements are composed of three abundant families, i.e. the LINE Tad1 family (16.3% of the genome), the LTR Gypsy family (15.4% of the genome), and the LTR Copia family (9.2% of the genome). Comparative analysis of the repeats with their corresponding consensus sequences revealed that almost all repetitive sequ ences (29.8 Mb; 44.4% of the genome) have an overall low nucleotide sequence divergence of less than 10% (Fig 2.2A). Indeed, most repeats exhibit a sequence divergence between 3% and 5%, whereas a considerable amount of the genome

(3 Mb; 4.5%) is covered by repeats with very low divergence of less than 1%, suggesting that these regions likely correspond to recently proliferated repeat families. To confirm these results, a similar analysis was conducted using the newer version of RepeatModeler v2.0.2 (Flynn et al. 2020). Indeed, both RepeatModeler v1 and RepeatModeler v2 predicted similar amounts of repetitive DNA, 49.7 and 49.8 %, respectively (Table 2.S2). Moreover, in accordance with the findings obtained with RepeatModeler v1, analysis of the repeats identified with RepeatModeler v2 indicated that almost all repeats have divergence of less than 10%, and most repeats have divergence between 3 and 5 % (Fig 2.S5).

A sliding window analysis further revealed that 39.2% of the genome of *C. fulvum* Race 5 is affected by RIP mutations (i.e. RIPed) (Clutterbuck 2011; Selker and Garrett 1988), in agreement with the previous estimate of 42.4% for isolate 0WU (De Wit et al. 2012). In addition, 1,532 Large RIP Affected Regions (LRARs) longer than 4 kb in size and with an average size of 16.7 kb could be detected. In comparison, a recent survey which analyzed the occurrence of RIP in 58 fungal species indicated that all had less than 30% of their genomes RIPed and contained at most 482 LRARs (Van Wyk et al. 2021). This indicates that *C. fulvum* is among the fungal species affected the most by RIP. However, depending on their repeat content, the chromosomes of *C. fulvum* Race 5 are RIPed at different extents, with Chr14 and Chr3 affected the most (70.4% and 53.0%, respectively), and Chr12 and Chr13 affected the least (28.0% and 28.6%, respectively) (Table 2.S3). These differences in RIP levels among the chromosomes can be attributed to their differences in repetitive DNA content, as a positive correlation existed among the two (correlation coefficient $R = 0.99$; p-value = 6.8E-11) (Fig 2.S6). When considering only repetitive regions, then between 73.1% and 85.3% of the repeats are affected by RIP (Fig 2.2B). Evidence of RIP leakage toward non-repetitive regions was also observed in all chromosomes. Specifically, from Chr1 to Chr13, between 2.5 kb and 32.7 kb of single-copy regions are RIPed, with RIP levels ranging from 0.1% to 1.6% of their unmasked bases. However, in the mini-chromosome Chr14, 24.5 kb of single-copy regions show evidence of RIP, with RIP levels corresponding to a larger percentage (24.1%) of unmasked bases in this chromosome (Fig 2.2B and Table 2.S3). This indicates that Chr14 is heavily affected by RIP and that RIP leakage occurs more frequently in Chr14 than in the other 13 chromosomes.

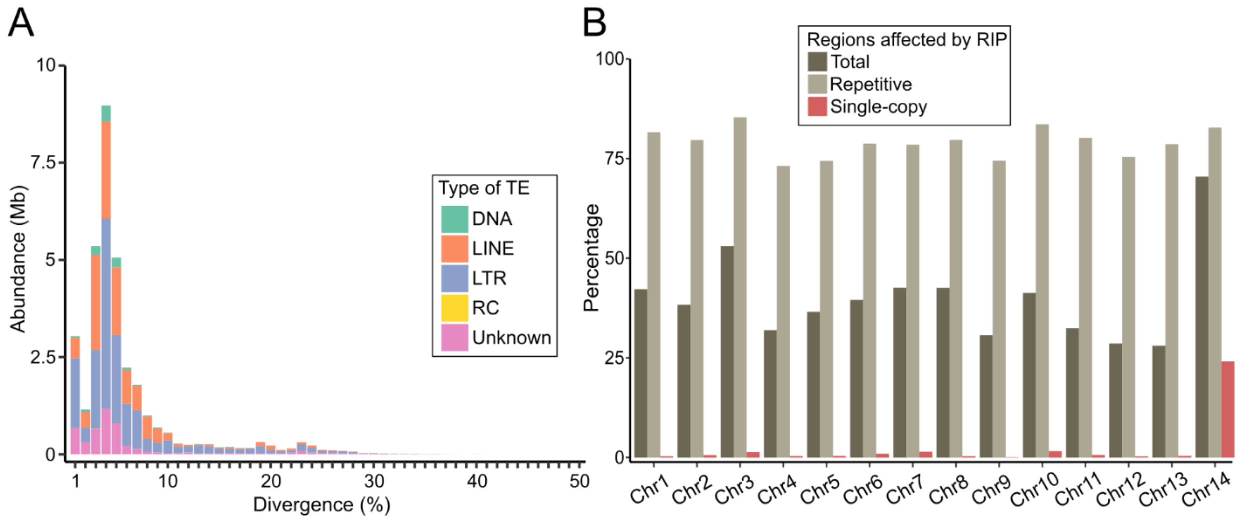


Figure 2.8: The repetitive DNA landscape of *Cladosporium fulvum* Race 5. (A) Bar plot showing the number of bases covered by predicted transposable elements (TEs) of different (sub)classes, i.e. DNA transposons (DNA), long interspersed nuclear elements (LINE), long terminal repeats (LTR), rolling-circles (RC), and unclassified TEs. The x-axis shows the divergence of repeats from the consensus sequences. The figure shows that the genome of *C. fulvum* Race 5 is abundant in repeats with an overall low divergence. (B) Bar plot showing the percentage of regions in the chromosomes of *C. fulvum* Race 5 that are predicted to be affected by Repeat-Induced Point (RIP) mutations. The figure shows that RIP affects approximately 40% of all chromosomes and that RIP predominates in repeat-rich regions. However, single-copy regions are also predicted to be affected by RIP and particularly in the mini-chromosome 14 (Chr14).

2.2.4 An accurate and fairly complete gene annotation of *C. fulvum* Race 5 assisted by pooled RNA-seq data

To assist gene predictions and to further obtain evidence of gene expression, RNA was isolated from *C. fulvum* Race 5 grown in different stress conditions and sequenced on an Illumina NovaSeq 6000 platform (PE150 format) at a high depth. The 326.2 M reads obtained by transcriptome sequencing (98.5 Gbp of data) mapped to the genome at an alignment rate of 97.1%. A total of 12,822 transcripts were assembled, which were then used together with other gene-supporting evidence (see Methods) to predict 14,690 genes. Analysis of the completeness of gene space using BUSCO genes revealed that the genome assembly and annotation of *C. fulvum* Race 5 was 98.9% complete, which is higher than the completeness (95.9%) of the *C. fulvum* isolate 0WU (Table 2.3). Notably, compared to other 39 annotated genomes of Capnodiales available at NCBI, only *Z. tritici* isolate ST99_3D1 has currently a higher completeness (99.1%) than *C. fulvum* Race 5 at the protein level (Fig 2.S7). The average size of the 14,690 predicted genes in the genome of *C. fulvum* Race 5 is 1,375 bp, with 5,474 (37.3%) genes being single-exon and 9,216 (62.7%) genes being multi-exon. Of the later ones, 7,843 (85.1%) genes have at least one splice-site supported by five or more RNA-seq reads, and 6,935 (75.2%) genes have all splice-sites supported by five or more RNA-seq reads, indicating accurate exon-intron structure prediction.

Functional annotations with InterProScan indicated that of the 14,690 genes predicted in the genome of *C. fulvum* Race 5, 6,316 (43.0%) genes encode proteins with a conserved Pfam domain in their primary sequence. Homology searches performed with eggNOG (evolutionary genealogy of genes: non-supervised orthologous groups) further assigned 8,884 (60.5%) genes to a KOG (eukaryotic orthologous groups) category and 1,686 (11.4%) genes to a KEGG (Kyoto encyclopedia of genes and genomes) Orthology (KO) ID group. Gene categories with relevance to fungal pathogens were also annotated. Specifically, 42 genes were predicted to encode key enzymes for the biosynthesis of secondary metabolites, including 11 polyketide synthases (PKSs), 15 non-ribosomal peptide synthetases (NRPSs), 11 NRPS-like fragments, one PKS-NRPS hybrid, and four terpene synthases (TSs) (Table 2.S4). A total of 488 genes encoding CAZymes were identified as well, including 260 glycoside hydrolases (GHs), 106 glycosyltransferases (GTs), 30 carbohydrate esterases (CEs), 74 auxiliary activity enzymes (AAs), nine polysaccharide lyases (PLs), and nine CAZymes with single carbohydrate binding domains (CBMs) (Table 2.S5). The genome of *C. fulvum* Race 5 further contains 359 proteases, divided into 177 serine, 83 metallo, 62 cysteine, 19 threonine, 14 aspartic, one asparagine, and three inhibitory peptidases (Table 2.S6). A total of 2,287 genes encoding putative transporters were also identified (Table 2.S7), with the most abundant family of transporters being the major facilitator superfamily (MFS) (*n* = 382), followed by the nuclear pore complex family (NPC) (*n* =

121), the pore-forming NADPH-dependent 1-acyldihydroxyacetone phosphate reductase family (*n* = 93), the equilibrative nucleoside (ENT) family (*n* = 72), and the ATP-binding cassette (ABC) transporter family (*n* = 57). Finally, the secretome of *C. fulvum* Race 5 is predicted to consist of 1,320 proteins, including 229 CAZymes, 77 proteases, and 345 candidate effectors (Table 2.S8). When mining the genome for candidate effector genes, it was observed that several genes (*n* = 35) previously described as candidate effectors in *C. fulvum* (Chang et al. 2016) were absent in the gene annotation of *C. fulvum* Race 5. Most of these genes (*n* = 31) were also absent in the reference annotation of *C. fulvum* 0WU, indicating that they are difficult to annotate based solely on *ab initio* predictions. Therefore, these genes were obtained from NCBI and manually annotated in the genome of *C. fulvum* Race 5. Lastly, another 69 previously described effector (e.g. *Avr2, Avr4, Avr9*) or candidate effector genes (e.g. *Ecp7, Ecp8, Ecp9-1*) in *C. fulvum* had their annotation verified and manually adjusted when needed based on mapped RNA-seq reads.

Table 2.4: Gene annotation statistics of *Cladosporium fulvum* Race 5 compared to the previous reference genome annotation of *C. fulvum* isolate 0WU.

Gene annotation statistics	*C. fulvum* Race 5	*C. fulvum* 0WU
Number of genes	14,690	14,127
Number of single-exon gene	5474	5411
Number of multi-exon gene	9216	8716
Mean exons per gene	2.1	2.2
Total gene length (bp)	20,211,715	19,998,515
Total exon length (bp)	18,870,323	18,300,722
Total intron length (bp)	1,358,240	1,714,406
Mean gene length (bp)	1375	1415
Mean cds length (bp)	1284	1295
Mean exon length (bp)	598	595
Mean intron length (bp)	80	103
Complete BUSCOs (%)	98.9	95.9
Complete single-copy BUSCOs (%)	98.8	95.6
Complete duplicated BUSCOs (%)	0.1	0.3
Fragmented BUSCOs (%)	0.3	0.7
Missing BUSCOs (%)	0.8	3.4

2.2.5 Distribution of genes within the chromosomes of *C. fulvum* Race 5

The 14 chromosomes assembled in *C. fulvum* Race 5 vary in gene content (Table 2.2), with the mini-chromosome Chr14 being markedly different from the other 13 chromosomes in that it harbors only hypothetical genes. Notably, smaller chromosomes also have higher density of key genes involved in secondary metabolism (i.e. PKSs, NRPSs, and TSs) and genes encoding MFS transporters, whereas larger chromosomes have higher densities of genes encoding ABC transporters and BUSCO genes (Fig 2.3A, Fig 2.3B and Fig 2.S8). Genes encoding CAZymes, proteases, secreted proteins, and candidate effectors are not preferentially located in smaller or larger chromosomes (Fig 2.3B).

When considering the distribution of genes on the chromosomes then, contrary to expectations, there was no substantial reduction in gene content in subtelomeric regions. Specifically, 143 genes were identified within 25 kb of the telomeric repeats (Table 2.S9), representing a density of 204.3 genes per Mb, which is slightly below the gene density for the entire genome (218.7 genes per Mb). Interestingly, most of these genes (*n* = 90) encode hypothetical proteins. Conversely, BUSCO genes are underrepresented in subtelomeric regions (p-value = 1.7E-5), whereas the density in these regions of genes encoding proteases and key enzymes for secondary metabolism is similar to that of the whole genome (Table 2.S10). Notably, one NRPS gene, previously described as *Nps3* (Collemare et al. 2014), is the leftmost gene in Chr9, located at a distance of 15,648 bp from the telomere (Fig 2.1). In contrast to the gene categories described above, candidate effector genes are significantly enriched in subtelomeric regions (p -value = 0.002) and exhibit a higher density (16.7 genes per Mb) in these regions, compared to the whole genome (5.1 genes per Mb) (Table 2.S10). Among the effector genes present in subtelomeric regions is the avirulence gene *Avr9,* which is located 6,545 bp upstream of the left -hand side telomere in Chr7, and nine other genes that encode the candidate effectors Ecp13, Ecp25, Ecp37, Ecp47, CE10, and CE29. Finally, another gene that is considerably close to a telomere, i.e. 123.7 kb from right -hand side of the telomere in Chr10, is *CYP51* that is involved in ergosterol biosynthesis and is the target of the demethylase inhibitor (DMI) fungicides (Fig 2.1).

Two mating-type idiomorphs, designated MAT1-1 and MAT1-2, have been described in *C. fulvum* (Stergiopoulos et al. 2007b). In accordance to previous reports (Stergiopoulos et al. 2007b), a search for mating-type genes in the genome of *C. fulvum* Race 5 identified a *MAT1-2-1* gene encoding a DNA-binding domain of the high-mobility group (HMG), but not the alpha-domain-encoding gene *MAT1-1-1*. The *MAT1-2-1* gene is located in Chr13, the second smallest chromosome, and is flanked by the hypothetical genes *ORF1-1-2* and *ORF1-2-2* (Stergiopoulos et al. 2007b). In most Ascomycetes, *MAT* genes are typically flanked by the *Apn2* and *SLA2* genes, but in *C. fulvum* only *Apn2* is located proximal to *MAT1-2-1*, whereas *SLA2* is near the opposite end of Chr13 at a distance of 1.43 Mb from *MAT1-2-1* (Fig 2.S9).

By querying the rDNA from *Neurospora crassa* (FJ360521) with BLASTn, the 18S-5.8S-28S rDNA of *C. fulvum* was identified in Chr4. Five 5,579 bp identical copies of this gene, tandemly arranged in a 44.5 kb locus could be assembled but because this region is collapsed in the assembly, the rDNA copy number in the assembly is underestimated. Indeed, by mapping the Illumina reads to one of the rDNA copies and normalizing the coverage by the median coverage of all genes, then the estimated rDNA copy number in *C. fulvum* Race 5 is 42, which is comparable to the predicted number of copies in other Ascomycetes (Lofgren et al. 2019).

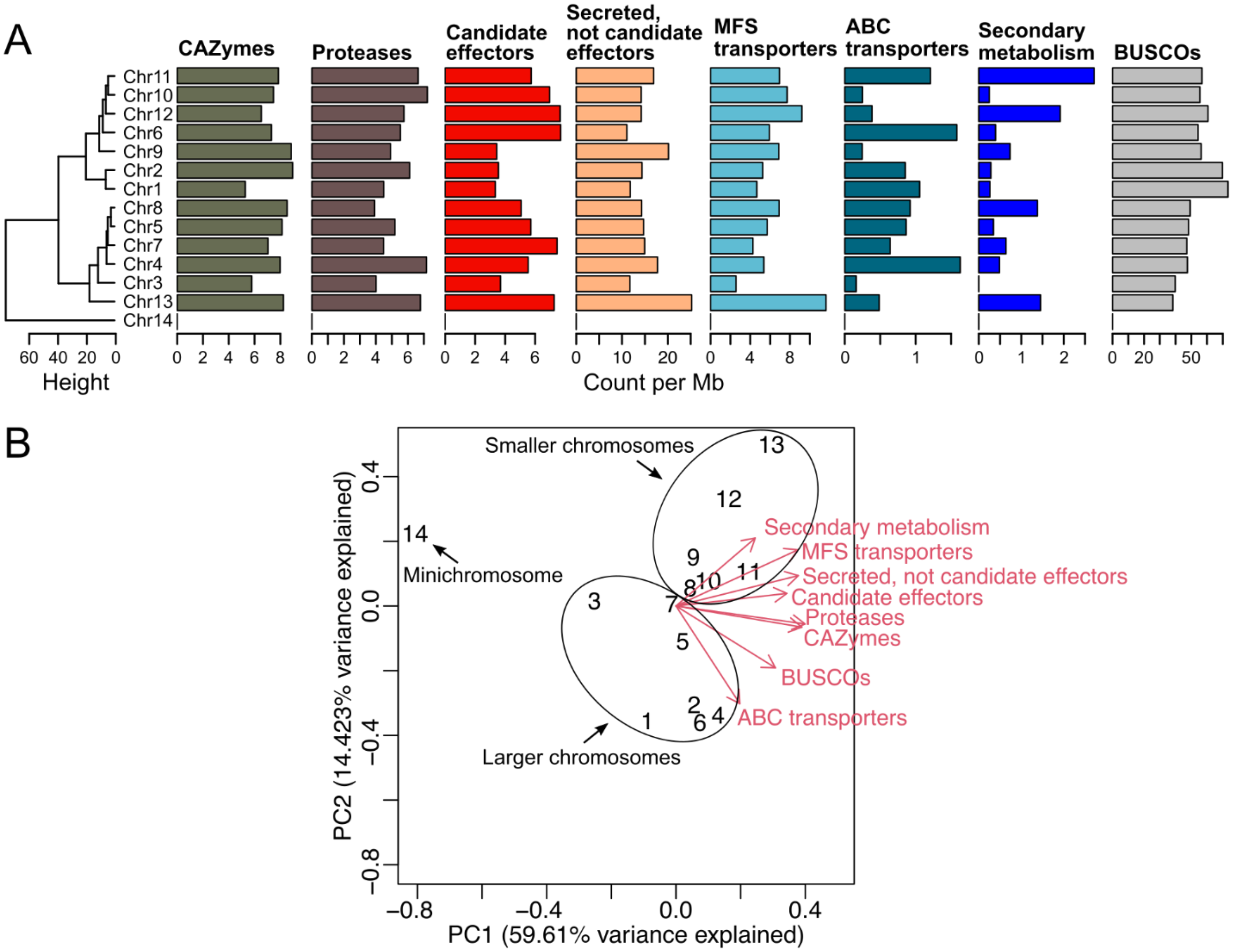


Figure 2.9: Large and small chromosomes of *Cladosporium fulvum* Race 5 differ in gene content. (A) Overall gene density, i.e. counts per million base pairs, of specific categories of genes in each chromosome. Chromosome were grouped using a hierarchical clustering performed using the complete method based on the Euclidean distances. (B) Principal component analysis biplot grouping the chromosomes based on gene density. The mini-chromosome 14 (14) appears far from the others, as this chromosome contains only hypothetical genes. Chromosomes 1 to 13 can be organized into two groups based on their size. One group contains chromosome 1 to chromosome 7 (the seven largest chromosomes) which, overall, have higher densities of genes encoding ATP-binding cassette (ABC) transporters and BUSCO genes. Chromosome 8 to chromosome 13 have, overall, higher densities of genes encoding key enzymes for secondary metabolism and major facilitator superfamily (MFS) transporters.

2.2.6 The genome of *C. fulvum* Race 5 is compartmentalized into gene clusters flanked by repeat-rich intergenic regions

Further examination of the organization of the chromosomes of *C. fulvum* Race 5 (Fig 2.1) indicated that genes and repetitive regions are unevenly distributed on them, thus resulting in a parallel skewed distribution of the size of their intergenic regions. The median size of the intergenic regions is 646 bp, which is considerably small compared to the average size for the entire genome (3,196 bp) and the size of the 5% (*n* = 735) longest intergenic regions, which ranges from 4,127 bp to 278,721 bp. Interestingly, in contrast to the large percentage of repetitive DNA in the genome of *C. fulvum* Race 5, the majority of the intergenic regions (*n* = 12,901; 87.8%) are free of repeats. However, from the 509 intergenic regions longer than 10 kb, 475 are composed of more than 80% repeats, whereas all 243 intergenic regions longer than 60 kb are composed of at least 88.9% repeats (Fig 2.4A). These observations indicate that repeats are generally clustered instead of being dispersed throughout the genome of *C. fulvum* and that repeats are the major component of long intergenic regions.

An analysis of the sizes of the intergenic regions among all genes revealed a clear pattern of gene-rich and gene-sparse regions, in accordance with the two-speed genome model (Dong et al. 2015) (Fig 2.4B). Several candidate effector genes were located in gene-sparse regions, including the previously described avirulence genes *Avr9*, *Avr4E,* and *Avr5,* and the extracellular protein encoding genes *Ecp1, Ecp5*, and *Ecp7,* all of which are flanked by some of the longest intergenic regions present in the genome (Fig 2.4C). In contrast, *Avr2, Avr4*, *Ecp2, Ecp2-2, Ecp2-3, Ecp4,* and *Ecp6* are located in gene-rich regions. It should be noted that the coding sequence of *Avr5* is disrupted in *C. fulvum* Race 5 by a 2-bp deletion, which causes a frameshift that leads to a premature stop codon (Fig 2.S10). Genes present in gene-sparse regions typically have one long and one short intergenic region (Fig 2.4B). This observation is consistent with a pattern of gene clustering in which, on the one hand, genes in gene-sparse regions are flanking clusters of genes and, on the other hand, the gene clusters are separated from each other by long intergenic regions. Based on this, the genes of *C. fulvum* Race 5 were organized into clusters, where a cluster is the largest set of consecutive genes in the genome such that the distance between any pair of adjacent genes is less than a defined threshold. Using different distance threshold values (i.e. between 1-to-20 kb), then the number of gene clusters identified in *C. fulvum* somewhat stabilizes to approximately 500 at threshold distances of over 5 kb (Fig 2.4D and Table 2.S11). For example, at the distance thresholds of 8 kb and 20 kb, there are 531 and 449 gene clusters, containing on average 27.7 and 32.7 genes, respectively. Notably, intergenic regions within gene clusters are typically short and have a low repetitive DNA content of less than 3% on average. In contrast, intergenic regions between gene clusters are long and highly enriched with repetitive DNA of more than 90% on average (Fig 2.4E and Fig 2.4F). This indicates that gene clusters are almost free of repetitive DNA, whereas genes that are flanking gene clusters are next to highly repetitive regions, and therefore likely more prone to mutations induced by transposon activity.

Although the majority (93.3%) of the genes have both intergenic regions shorter than 8kb, there are 990 genes (6.7%) with intergenic regions longer than 8 kb up- or downstream of their coding sequence. Of these, 61 genes have intergenic regions longer than 8 kb at both sides, and thus correspond to single-gene clusters (i.e. isolated genes), whereas 929 genes are flanked by an intergenic region longer than 8 kb at only one side, and thus correspond to cluster-flanking genes. Among the 990 genes with long intergenic regions, only 391 (39.5%) encode proteins with a Pfam domain in their primary structure, whereas the rest 599 (60.5%) genes encode hypothetical proteins. An enrichment analysis based on a hypergeometric test showed that at an adjusted p-value < 0.05, five Pfam domains are overrepresented in the subset of 391 genes. The three most significant of these Pfam domains (adjusted p-value < 4E-4) are commonly associated with transposons and represent an endonuclease-reverse transcriptase domain (PF14529), a CHRromatin Organisation MOdifier domain (PF00385), and a reverse transcriptase domain (PF00078). Indeed, these domains are commonly present in predicted gene models that overlap masked regions of the genome (Table 2.S12). The other two significant Pfam domains are a velvet factor domain (PF11754; adjusted p-value < 8E-4) and a mycotoxin biosynthesis protein UstYa domain (PF11807; adjusted p-value = 0.01). Finally, of the 990 genes present in gene-sparse regions, 142 encode secreted proteins, 66 of which are candidate effectors. A hypergeometric test showed that candidate effectors are significantly enriched (p-value = 6.6E-15) within the set of 990 genes. In contrast, genes encoding for secreted proteins that are not candidate effectors and other gene categories were not significantly enriched (Table 2.S13). It should be noted that, although gene-sparse regions are enriched for genes encoding candidate effectors, the majority (*n* = 279, 80.9%) of candidate effectors are still present in gene clusters rather than gene-sparse regions (*n* = 66, 19.1%).

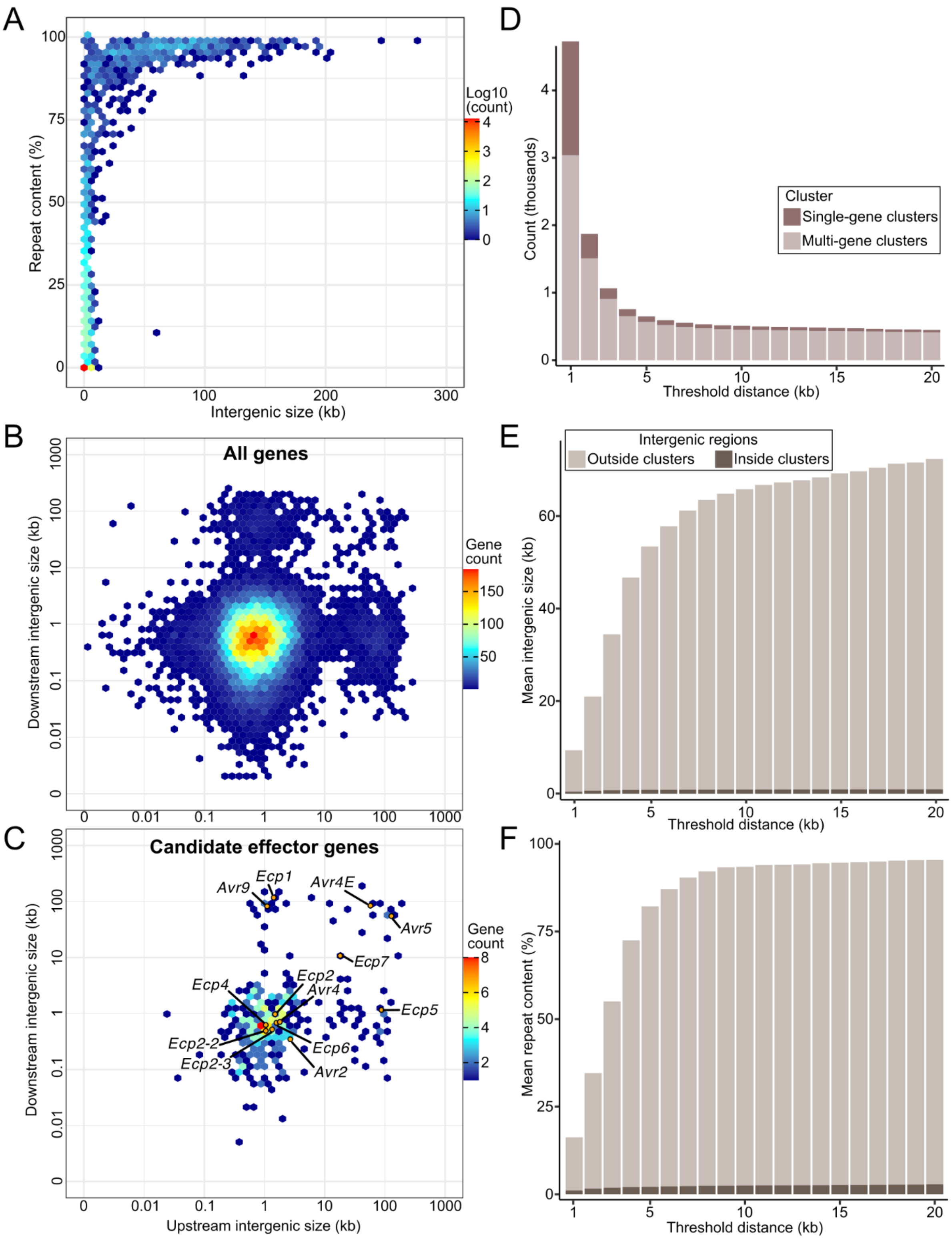
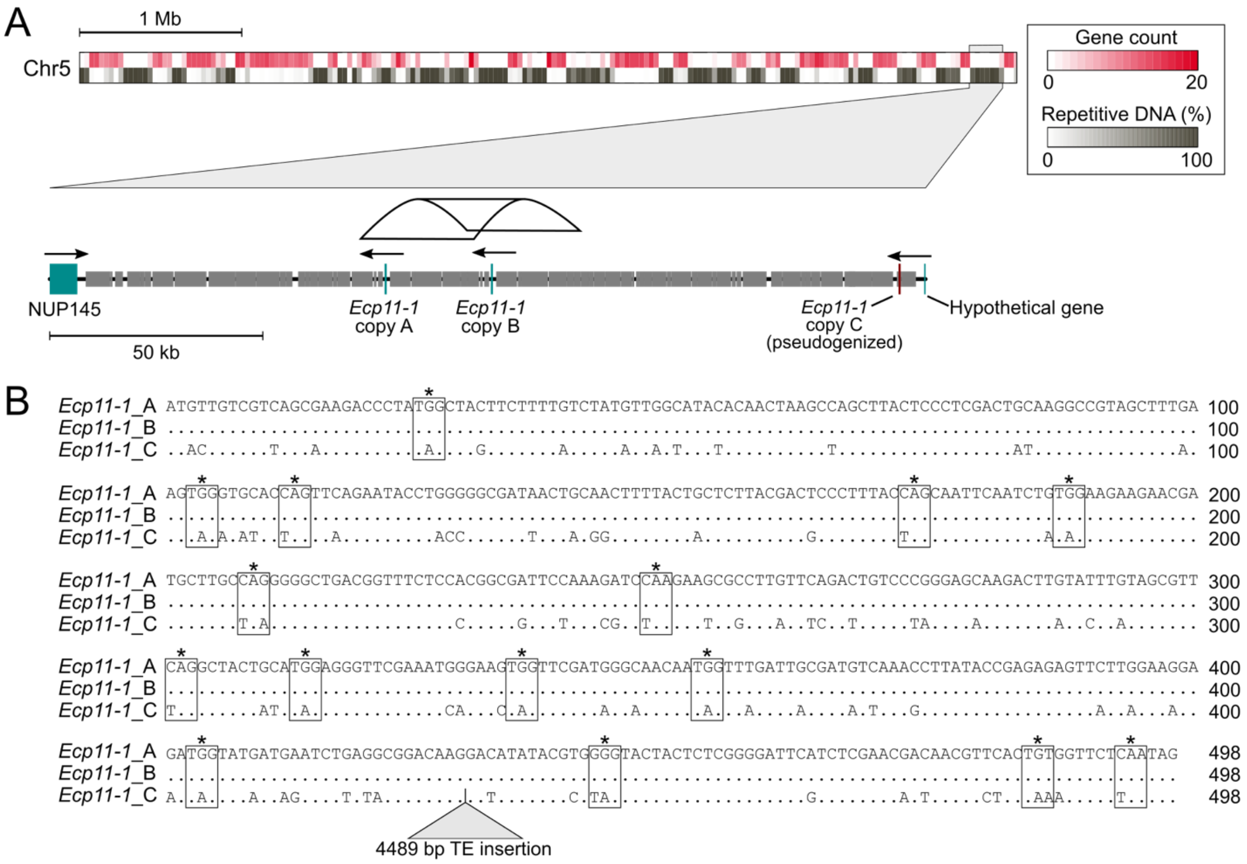


Figure 2.10: Compartmentalization of the genome of *Cladosporium fulvum* Race 5. (A) Heat map showing the number in log10 scale of intergenic sequences relative to their size and repeat content. This heat map shows that the genome of *C. fulvum* Race 5 is abundant in small, non-repetitive intergenic regions, whereas nearly all intergenic regions larger than 50 kb in size are almost entirely composed of repeats. This

indicates that repeats in the genome of *C. fulvum* Race 5 are clustered and form long intergenic regions. (B, C) Heat maps showing the number of genes with the corresponding up- and downstream intergenic sizes on the x- and y-axis, respectively. The heat map in panel (A) includes all genes (*n* = 14,690), and the heat map in panel (B) includes only genes encoding candidate effectors (*n* = 345). Previously described avirulence genes (*Avr*) and the extracellular protein encoding genes (*Ecp*) are indicated with points. The heat maps show that there are several candidate effector genes located in gene-sparse regions of the *C. fulvum* Race 5 genome and they typically have at least one neighboring gene in close proximity, either up- or downstream of their coding sequence. However, the majority of candidate effector genes either follow the trend of all genes, i.e. located in gene-rich regions (e.g. *Avr2, Avr4*, *Ecp2, Ecp2-2, Ecp2-3, Ecp4,* and *Ecp6),* or are located next to some of the longest intergenic regions of the genome, i.e. in gene-sparse regions (e.g. *Avr4E, Avr5, Avr9, Ecp1, Ecp5*, and *Ecp7*). (D, E, F) Bar plots showing the statistics of gene clustering based on different threshold intergenic distance values (i.e. 1 to 20 kb) among genes within the same cluster. The bar plot in panel (D) shows the number of clusters identified, divided into single-gene clusters and clusters with more than one gene. The bar plots in panels (E, F) show the mean intergenic region size and percentage of repetitive DNA within intergenic regions outside clusters, i.e. flanking clusters, and inside clusters, respectively. The figures show that the genes of *C. fulvum* Race 5 can be organized into approximately 500 clusters separated by long intergenic regions that are rich in repeats, whereas intergenic regions inside clusters are poor in repeats.

2.2.7 Gene duplication analysis reveals two identical copies of an *Avr3Lm*-like candidate effector gene in *C. fulvum* Race 5

One advantage long read-based assemblies have is that they allow identification of nearly identical copies of genes, likely caused by recent duplication events that would otherwise be collapsed in short read-based assemblies. To search for recently duplicated genes in the genome of *C. fulvum* Race 5, the coding sequences of the 14,690 genes predicted in it were clustered with *cd-hit-est,* using a minimum identity of 90%. Twenty multi-gene clusters were identified, containing a total 59 genes (Table 2.S14). Of the 20 multi-gene clusters, 12 clusters included genes similar to transposable elements, six clusters contained genes that encode hypothetical proteins, and one cluster included two genes similar to the proton-dependent oligopeptide transporter family, although one of the two copies is truncated and likely non-functional. The remaining cluster included two identical copies (copy A and copy B) of the candidate effector gene *Ecp11-1*, which is homologous (36.9% identity at the amino acid level, e-value = 4E-24) to the avirulence gene *AvrLm3* from the Dothideomycete *L. maculans* (Balesdent et al. 2002). Notably, *Ecp11-1* is the only duplicated candidate effector gene retained in the genome of *C. fulvum* Race 5. Both copy A and copy B of *Ecp11-1* are intronless and are tandemly arranged in a repeat-rich region in Chr5 (Fig 2.5A). To rule out the possibility that the two copies are the result of an assembly artifact, the PacBio reads were mapped to the genome and the region was confirmed to be free of misassemblies (Fig 2.S11). Based on a self-alignment performed with NUCmer, the identified duplication is 26.6 kb long, with 99.2% alignment identity, and contains copy A and copy B of *Ecp11-1*, plus 5.4 kb and 20.6 kb down- and upstream of the two copies, respectively. A third copy (copy C) of *Ecp11-1* was identified 94.7 kb upstream of copy A (Fig 2.5A). This third copy is predicted to be affected by RIP because its entire coding sequence resides within RIPped regions, and 70 (75%) of the 93 point mutations compared to copy A and B are C:G to T:A transitions, which resulted in several premature stop codons. Therefore, *Ecp11-1* copy C is pseudogenized likely due to the accumulation of RIP-like mutations. In addition, a LINE-like transposable element of 4,489 bp is inserted toward the 3'-end of the pseudogenized *Ecp11-1* (Fig 2.5B). Overall, the small number of duplicated genes identified in the genome of *C. fulvum* Race 5 and the pseudogenization of one of the three copies of *Ecp11-1*, suggest that duplicated genes in *C. fulvum* are often lost, possibly due to the accumulation of mutations.



```
                                   *
Ecp11-1_A ATGTTGTCGTCAGCGAAGACCCTATGGCTACTTCTTTTGTCTATGTTGGCATACACAACTAAGCCAGCTTACTCCCTCGACTGCAAGGCCGTAGCTTTGA 100
Ecp11-1_B .................................................................................................... 100
Ecp11-1_C ..AC......T...A..........A....G.......A.....A..A.T...T..........T.................AT..............A. 100

             *        *                                                           *              *
Ecp11-1_A AGTGGGTGCACCAGTTCAGAATACCTGGGGGCGATAACTGCAACTTTTACTGCTCTTACGACTCCCTTTACCAGCAATTCAATCTGTGGAAGAAGAACGA 200
Ecp11-1_B .................................................................................................... 200
Ecp11-1_C ...A.A.AT..T....A.........ACC......T...A.GG........A..........G........T.............A.A............ 200

                  *                                      *
Ecp11-1_A TGCTTGCCAGGGGGCTGACGGTTTCTCCACGGCGATTCCAAAGATCCAAGAAGCGCCTTGTTCAGACTGTCCCGGGAGCAAGACTTGTATTTGTAGCGTT 300
Ecp11-1_B .................................................................................................... 300
Ecp11-1_C .......T.A..................C.....G...T...CG..T.....T..G...A..TC..T.....TA....A.......A..C..A....... 300

           *           *                    *                 *
Ecp11-1_A CAGGCTACTGCATGGAGGGTTCGAAATGGGAAGTGGTTCGATGGGCAACAATGGTTTGATTGCGATGTCAAACCTTATACCGAGAGAGTTCTTGGAAGGA 400
Ecp11-1_B .................................................................................................... 400
Ecp11-1_C T........AT..A.............CA...C.A.......A..A......A...A....A....A.T...G.................A..A...A.. 400

             *                                      *                                         *        *
Ecp11-1_A GATGGTATGATGAATCTGAGGCGGACAAGGACATATACGTGGGGTACTACTCTCGGGGATTCATCTCGAACGACAACGTTCACTGTGGTTCTCAATAG   498
Ecp11-1_B ..................................................................................................   498
Ecp11-1_C A..A....A..AG....T.TA........|..T.......C.TA...................G........A.T......CT..AAA.....T.....   498
```

Figure 2.11: Segmental duplication of the gene encoding the candidate effector Ecp11-1 in *Cladosporium fulvum* Race 5. (A) A zoomed-in region near the end of chromosome 5 (Chr5). The zoomed-in region is 205.5 kb in size and shows the location of two identical copies of *Ecp11-1* (copy A and copy B) and an additional pseudogenized copy (copy C) of the gene, surrounded by repetitive DNA. Genes are represented as boxes with arrows indicating their transcriptional orientation. Repetitive regions are represented as smaller boxes. The duplicated segment that contains the two functional copies of *Ecp11-1* (copy A and copy B) is shown with a ribbon. The three copies of *Ecp11-1* are located between *NUP145*, encoding a component of the nuclear pore complex, and a hypothetical gene. (B) Global alignment of the coding sequences of the three *Ecp11-1* copies (copies A, B, and C). Codons mutated to a stop codon are indicated with boxes. The location of an insertion of a 4.5 kb transposable element in copy C is shown with a triangle.

2.2.8 Genome comparison between *C. fulvum* isolates Race 5 and 0WU reveals hundreds of genes missing in the previous reference genome assembly of *C. fulvum* 0WU

The first and so far only genome assembly of *C. fulvum* (isolate 0WU) was published nearly a decade ago (De Wit et al. 2012) and since then, it has been used as a reference for comparative genome analyses. However, the assembly of isolate 0WU is highly fragmented into 4,865 scaffolds. In order to compare the genomes of *C. fulvum* isolates Race 5 and 0WU, the scaffolds of isolate 0WU were split at gapped regions and the resulting 5,715 contigs were then mapped to the genome of isolate Race 5. Out of the 5,715 contigs, 5,713 mapped to *C. fulvum* Race 5 chromosomes, covering 57,086,585 bp, or 84.9% of the assembly. Coverage of most chromosomes was between 80% and 90%, with almost all genes present in both isolates (Table 2.S15). However, Chr12 was an exception, as the contigs of isolate 0WU covered only 53% of this chromosome (Fig 2.S12A). In addition, when the gene sequences of *C. fulvum* Race 5 were queried with BLASTn (e-value < 1e-5) against the genome assembly of isolate 0WU, then 352 genes had no BLASTn hit, and thus were missing from the assembly of isolate 0WU (Table 2.S16). Notably, of the 352 genes, 348 are located in Chr12, corresponding to half (49.8%) of all predicted genes in this chromosome (Fig 2.S12A). The missing genes are unlikely dispensable because 74 of them are universal single copy orthologs (i. e. BUSCOs) that are conserved among Capnodiales. Among the 352 missing genes, there were 51 putative transporters, 30 secreted proteins, nine candidate effectors, eight CAZymes, and eight proteases (Table 2.S16). Whole-genome sequencing data of *C. fulvum* 0WU is currently not publicly available to confirm the absence of these genes from its genome. However, upon analysis of *in vitro* RNA-seq data of *C. fulvum* 0WU (SRR1171044, SRR1171045, and SRR1171046), we observed that from the 352 genes missing, 241 show clear evidence of expression (transcripts per million [TPM] >2) in all three different RNA -seq datasets (Fig 2.S12B and Table 2.S16). This indicates that most of the 352 genes are present in *C. fulvum* 0WU but they are missing in its reference genome assembly.

2.2.9 Comparison of the *C. fulvum* Race 5 genome assembly with other Dothideomycetes indicates a core set of 13 mesosyntenic chromosomes

The genome of the pine tree pathogen *D. septosporum*, a phylogenetically close relative of *C. fulvum*, has been previously assembled at chromosome-scale (De Wit et al. 2012). Both species have the same number of predicted chromosomes (*n* = 14), but the size of their genomes varies markedly, ranging from 30.2 Mb for *D. septosporum* isolate NZE10 to 67.1 Mb for *C. fulvum* Race 5 (Fig 2.6A). In order to compare the genome organization of the two species, their genomes were aligned with PROmer, which produced 17,337 aligned segments with an average length of 1,000 bp and an average identity at the amino acid level of 73.8% (Fig 2.6A and Fig 2.6B). The whole-genome alignment revealed a clear pattern of mesosynteny, consisting of a small number of inter-chromosomal translocations and a large number of intra-chromosomal rearrangements. This confirmed previous observations that suggested that the gen omes of *C. fulvum* race 5 and *D. septosporum* NZE10 are mesosyntenic (De Wit et al. 2012). This pattern is common within Dothideomycetes (Hane et al. 2011; Ohm et al. 2012) and indicates that gene content of both species is conserved within chromosomes, although the order of genes is not well conserved. Out of the 14 chromosomes present in each species, nine (i.e. Chr1, Chr2, Chr5, Chr6, Chr7, Chr8, Chr9, Chr11, and Chr13 with reference to the *C. fulvum* Race 5 chromosomes) have a one-to-one match, whereas of the five remaining chromosomes, four chromosomes of *C. fulvum* Race 5 (i.e. Chr3, Chr4, Chr10, and Chr12) match two or more chromosomes of *D. septosporum* NZE10. Finally, the mini-chromosome Chr14 of *C. fulvum* Race 5 has no matches to *D. septosporum* NZE10 chromosomes, whereas Chr14 of *D. septosporum* NZE10 matches Chr4 of *C. fulvum* Race 5 (Fig 2.6A and Fig 2.6B). A similar pattern of mesosynteny was also observed when the genome of *C. fulvum* was aligned with the genomes of the phylogenetically close relatives *Septoria musiva*, *Z. tritici*, *Pseudocercospora fijiensis,* and *Cercospora beticola* (Fig 2.S13). Notably, none of these has matches to Chr14 of *C. fulvum*, further supporting that this chromosome could be dispensable.

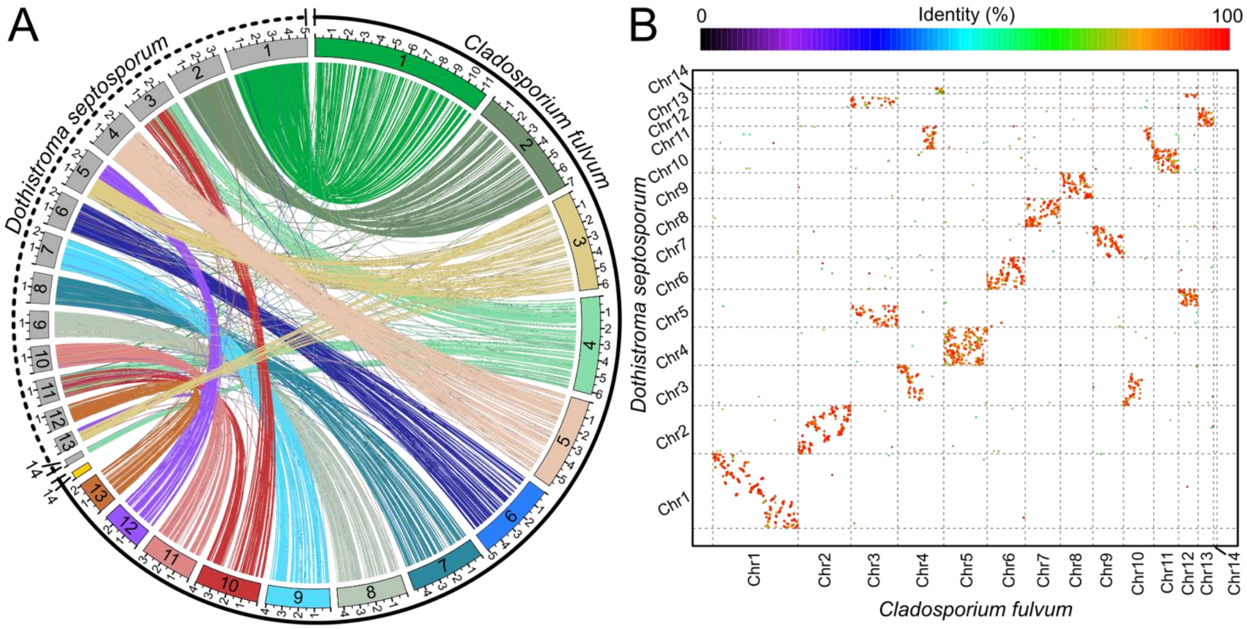


Figure 2.12: Mesosynteny between *Cladosporium fulvum* Race 5 and *Dothistroma septosporum* NZE10. The figure shows the whole-genome alignment produced with PROmer based on the six-frame translation of the genomes of *C. fulvum* Race 5 and of *D. septosporum* isolate NZE10. (A) Circos plot showing the collinearity between the chromosomes of the two species. Ribbons are based on nucleotide identity. (B) Dot-plot showing syntenic relations between the chromosomes of the two species. Chromosomes are numbered and dots are color-coded for percent nucleotide identity. The plots show a pattern of mesosynteny between *C. fulvum* Race 5 and *D. septosporum* NZE10, in which gene content is largely conserved within chromosomes, with few interchromosomal rearrangements, whereas gene order is not conserved. The alignment with the *D. septosporum* NZE10 genome produced high identity values, as shown in panel (B), which contrasts with the large difference in genome size compared to *C. fulvum* Race 5, as shown in the circos plot in panel (A). The mini-chromosome 14 (Chr14) of *C. fulvum* Race 5 had no matches with the genome of *D. septosporum* NZE10, supporting the hypothesis that this chromosome is dispensable. Chromosomes were named based on their size, from the longest to the smallest one, and are shown in scale in the circos plot, with tick labels indicating Mb. Whole-genome alignments of *C. fulvum* Race 5 with other Capnodiales are shown in Fig 2.S13.

2.2.10 The mini-chromosome Chr14 of *C. fulvum* is dispensable

The small size and low gene density of the mini-chromosome Chr14 led us to hypothesize that it might be a dispensable chromosome. To test this hypothesis, a collection of 24 isolates that were obtained from different locations around the world and in a span of over 40 years (Stergiopoulos et al. 2007b) was analyzed for the presence or absence of Chr14. This was done using primers designed to capture eight genes located towards the 5'-end (one gene), middle (five genes), and 3'-end (two genes) of this mini-chromosome (Table 2.S17 and Fig 2.S14A). PCR amplifications revealed that Chr14 was present in only five isolates, i.e. isolates Race 5, 0WU, 2, IMI Argent 358077, and Turk 1a, out of the 24 isolates examined (Table 2.4 and Fig 2.S14B). Interestingly, Chr14 was present in five out of the eight MAT1-2 isolates present in the collection, whereas none of the twelve MAT1-1 isolates had this chromosome. These results indicate that Chr14 is dispensable, and that it is more likely to be present in MAT1-2 than in MAT1-1 isolates. To further confirm dispensability of Chr14, DNA samples of isolates 2, IMI Argent 358077, and Turk 1a, for which Chr14 was predicted to be present, were pooled (pool 1) and sequenced with Illumina technology. In addition, DNA samples for isolates IPO 2.4.8.9.11 Polen, IPO 249 France, and 2.5, for which Chr14 was predicted to be absent, were pooled (pool 2) and sequenced with Illumina technology as well. Sequenced reads were mapped to the genome of *C. fulvum* Race 5 at a mapping rate of 94%, and whole-genome coverage depths were estimated to be 59x for pool 1 and 80x for pool 2. Further coverage analysis revealed that 1.1 M reads from pool 1 and 0.26 M reads from pool 2 mapped to Chr14 (Fig 2.S15a). As expected, all genes predicted in Chr14 exhibited high levels of coverage of reads from pool 1 (Fig 2.S15b). In contrast, almost all predicted genes in Chr14 exhibited practically no coverage of reads from pool 2 (Fig 2.S15b). The only exception was the gene CLAFUR5_14645, which is duplicated in *C. fulvum* Race 5 with two identical copies, one in Chr14 and the other in Chr1. These results further support that Chr14 is indeed dispensable. The function of the genes in Chr14 remains elusive. Also, more isolates will have to be analyzed for the presence/absence of Chr14 in their genome before a causal connection could be made between this mini-chromosome and one of the mating-type idiomorphs of *C. fulvum*.

Table 2.5: *Cladosporium fulvum* isolates examined for the presence or absence of the mini-chromosome 14 (Chr14). The isolates selected were previously described in (Stergiopoulos et al. 2007b).

No.	Isolate	Race	Mating type	Origin	Year of collection	Presence of Chr14
1	2	2	MAT1-2	Netherlands	Unknown	Yes
2	2.4	2.4	MAT1-1	Netherlands	1971	No
3	2.4.5	2.4.5	MAT1-1	Netherlands	1977	No
4	2.4.5.9.11 IPO	2.4.5.9.11	MAT1-2	Netherlands	Unknown	No
5	2.4.5.9	2.4.5.9	MAT1-1	Netherlands	1980	No
6	2.4.8.11	2.4.8.11	MAT1-1	Netherlands	Unknown	No
7	2.4.9.11	2.4.9.11	MAT1-1	Poland	Unknown	No
8	IPO 2.4.5.9 (60787)	2.4.5.9	MAT1-2	Netherlands	Unknown	No
9	IPO 2.4.8.9.11 Polen	2.4.8.9.11	MAT1-1	Poland	Unknown	No
10	2.5.9	2.5.9	MAT1-1	France	1987	No
11	4	4	MAT1-1	Netherlands	1971	No
12	IPO 249 France	2.4E	MAT1-1	France	Unknown	No
13	Can 38	4.4E	MAT1-1	USA	1962	No
14	IMI Argent 358077	0	MAT1-2	Argentina	1991	Yes
15	Turk 1a	2	MAT1-2	Turkey	2005	Yes
16	L25	Unknown	Unknown	France	Unknown	No
17	2021-002	Unknown	Unknown	France	Unknown	No
18	18-A6	2.9	MAT1-2	Japan	Unknown	No
19	Can54b	2.3-1	Unknown	Unknown	Unknown	No
20	2.5	2.5	MAT1-1	Bulgaria	Unknown	No
21	Croatia 7	Unknown	Unknown	Croatia	2006	No
22	0WU	0	MAT1-2	Netherlands	Unknown	Yes
23	Race 4	4	MAT1-1	Netherlands	1971	No
24	Race 5	5	MAT1-2	France	1979	Yes

2.3 Discussion

In this study, by combining long-read sequencing and Hi-C chromatin conformation capture data, we successfully produced a nearly complete genome assembly and a high-quality gene annotation for the TE-rich genome of *C. fulvum* Race 5. The assembly showed that the genome of this isolate consists of 13 core chromosomes and a dispensable mini-chromosome. The large percentage of Illumina reads that could be properly mapped to the final assembly and the high BUSCO completeness indicated that only a small portion of the genome, which may or may not contain genes, remains unassembled. The genome of *C. fulvum* Race 5 presented herein is a considerable improvement over the previous reference genome of *C. fulvum* 0WU (De Wit et al. 2012). Among the improvements worth noting, is the assembly of half of the genes in Chr12 that were somehow missed in the genome assembly of *C. fulvum* 0WU and of the long repetitive intergenic regions that previously could not be assembled. Moreover, the genome size of *C. fulvum* 0WU was underestimated during its sequencing, which translated into a reduced coverage of just 21-fold instead of the 32-fold coverage calculated previously, thus making it more likely that some genomic regions in isolate 0WU were not assembled due to low coverage. Such misassemblies caused by shallow sequencing coverage are expected to be randomly scattered throughout the genome and not to be concentrated in a particular chromosome, as was the case with Chr12. To this end, the reason why the gene space in Chr12 was unassembled in *C. fulvum* 0WU remains unknown.

The genomes of Dothideomycete fungi often display large and more than tenfold differences in size, with the increase in genome sizes typically instigated by their invasion by repetitive DNA and TEs (Haridas et al. 2020). High variability in repeat content is observed even among phylogenetically close-related species, as for example is the case in the *L. maculans*-*L. biglobosa* species complex, in which repeat content ranges from 3.9% in the 30.2 Mb genome of *L. biglobosa* 'canadensis' strain J154, to 35.5% in the 45.1 Mb genome of *L. maculans* 'brassicae' strain v23.1.3 (Grandaubert et al. 2014). In a similar way, species of *Pseudocercospora* spp. in the Sigatoka disease complex also exhibit large differences in genome size as a result of a high variation in their repetitive DNA content, which ranges from 35.7% in the 53.79 Mb genome of *P. eumusae,* to 62.2% in the 82.77 Mb genome of *P. musae* (Chang et al. 2016). Contrasting genome sizes driven by differences in repeat and TE content were also observed between *C. fulvum* and *D. septosporum* (De Wit et al. 2012). Predicted TEs within the same family have an overall low divergence in *C. fulvum* Race 5, which is consistent with the hypothesis of a recent proliferation of TEs (Frantzeskakis et al. 2018), possibly after divergence from *D. septosporum*. This could potentially explain why the genome of *C. fulvum* is markedly larger with expanded repeat content as compared to the genome of *D. septosporum* (De Wit et al. 2012). The overall low repeat divergence also supports recent proliferation of TEs. However, this observation contrasts with high levels of RIP mutations identified in the genome, which are expected to increase TE diversity. One explanation is that *C. fulvum* rarely undergoes sexual reproduction (De Wit et al.

2012; Stergiopoulos et al. 2007b), thus decreasing the speed at which RIP mutations accumulate over time. Further comparative genomic studies can provide insights into these contradictory observations.

Although TEs can potentially spread to any region of the genome, in fungal genomes they are typically unevenly distributed and tend to accumulate in gene-sparse regions where they often cluster with other TEs (Muszewska et al. 2019). One explanation for this non-random distribution of TEs is that their disruptive effects upon insertion proximally to genes triggers purifying selection, which in turn will favor purging these elements from the population (Torres et al. 2020). In plant pathogens, mutations or epigenetic modifications caused by the activity of TEs are major drivers of genetic variability associated with their evolution, genome plasticity, virulence, and host adaptation (Lorrain et al. 2021; Muszewska et al. 2019; Oggenfuss et al. 2021). In this context, the uneven distribution of TEs across their genomes leads to their compartmentalization into TE-rich and TE-poor regions, in which TE-rich regions can accumulate mutations faster than TE-poor regions due to TE activity. This peculiar genome architecture is referred to as the "two-speed genome" model (Croll and McDonald 2012; Dong et al. 2015; Raffaele and Kamoun 2012) or the "dynamic compartmentalization" (Frantzeskakis et al. 2019). In this model, TE-rich regions can provide a favorable environment to induce a fast evolutionary rate of virulence-related genes, which then can provide advantages for host adaptation. In agreement with the two-speed genome model, our analysis revealed that repetitive DNA sequences in the genome of *C. fulvum* Race 5 are clustered and form long, repeat-rich intergenic regions that intersperse gene-dense regions. Notably, gene-sparse regions were enriched with candidate effector genes, but not with other gene categories, suggesting that genome compartmentalization in *C. fulvum* could be a driver of virulence and adaptation to different host genotypes, although most of candidate effectors are not in gene-sparse, TE-rich regions. Indeed, the loss through complete gene deletion of certain effector genes from populations of *C. fulvum*, including for example of the *Avr4E*, *Avr5,* and *Avr9* genes, has been linked to overcoming their cognate resistance genes in tomato (Mesarich et al. 2014; Van den Ackerveken et al. 1992; Westerink et al. 2004). Previous genomic analysis indicated co-localization of effector genes (e.g., *Avr4E*, *Avr5*, and *Avr9*) and repetitive regions in *C. fulvum* 0WU (De Wit et al. 2012). This observation led to the hypothesis that the loss of these effectors is mediated by structural variations induced by neighboring repeats (De Wit et al. 2012; Mesarich et al. 2018). Our analysis confirmed that *Avr4E*, *Avr5*, and *Avr9* are located in repeat-rich regions, and further showed that they are flanked by some of the longest intergenic regions present in the genome of *C. fulvum* Race 5. In contrast, the *Avr2*, *Avr4*, *Ecp2*, and *Ecp4* effector genes for which selection pressure from cognate resistance genes in tomato led to the emergence of mutated alleles of these effectors with nucleotide substitutions or short INDELs (Stergiopoulos et al. 2007a), are located in repeat-poor, gene-rich regions. Collectively, these results support the hypothesis that complete gene deletions of effector genes in *C. fulvum* are induced by the presence of neighboring repeats and that the location of candidate effectors in the genome of *C. fulvum* could potentially be used to foretell their population genetics and mode of evolution under selection pressure from cognate resistance genes in tomato.

Among the candidate effector genes flanked by repeat-rich regions is *Ecp11-1*, the only non-hypothetical gene in the genome of *C. fulvum* Race 5 with two identical copies. Previous studies have shown that Ecp11-1 triggers a hypersensitive response in various tomato accessions, indicating that it is likely an effector recognized by a cognate resistance protein in tomato (Mesarich et al. 2018). Moreover, a recent study elucidated the crystal structure of Ecp11-1 and showed that it is also recognized by the oilseed rape resistance protein Rlm3 (Lazar et al. 2021). Ecp11-1 is predicted to belong to the so-called LARS (Leptosphaeria avirulence-suppressing) family of effectors, members of which have been detected in several Dothideomycetes (Lazar et al. 2021). *Leptosphaeria maculans* 'brassicae' has an expanded number of candidate LARS effectors (n = 13), the majority of which are found grouped in three regions of the genome, most likely as a result of local duplication events (Lazar et al. 2021). In a similar way, the three *Ecp11-1* copies present in the genome of *C. fulvum* Race 5 are tandemly arranged on Chr5, possibly due to TE activity or genomic instability induced by the repeat-rich neighboring regions. However, *Ecp11-1* copy C is pseudogenized by RIP-like mutations and a TE-like insertion in its coding sequence. This suggests that RIP mutations and TE activity can disrupt duplicated genes, thus preventing copy number variation in *C. fulvum*. Moreover, several polymorphisms were identified between the *Ecp11-1* copy C and copies A and B, which are intact and exhibit identical sequences. This suggests that the *Ecp11-1* copy C was likely derived from an older duplication event, whereas copies A and B were generated from a more recent dupli cation event.

Accessory (a.k.a. dispensable) chromosomes are present in widespread Eukaryotic taxa, including plants, animals, and fungi, and they typically provide no advantage to the host organism (Houben et al. 2014). However, fungal accessory chromosomes receive particular attention because they may harbor genes associated with virulence or host adaptation (Bertazzoni et al. 2018; Witte et al. 2021). For example, the secreted in xylem (SIX) effector genes in *Fusarium oxysporum,* which are pathogenicity factors and drivers of host specificity among *formae speciales* of this species (Houterman et al. 2007; Ma et al. 2010; Rep et al. 2004; Yang et al. 2020), the *AvrLm11* effector gene of *L. maculans* that confers virulence on *Brassica napa* (Balesdent et al. 2013), and a pea pathogenicity cluster of *Fusarium solani* that detoxifies the phytoalexin pisatin and is required for virulence on pea (Temporini and VanEtten 2002), are just a few examples of virulence or pathogenicity factors encoded by genes present in accessory chromosomes. The Capnodiales *Z. tritici* has the largest number of accessory chromosomes (n = 8) identified in fungal species so far (Goodwin et al. 2011) but their role still remains elusive and no clear association with pathogenicity has been established. However, a recent study indicated that their presence provides a small but significant increase in virulence (Stewart et al. 2018). To the best of our knowledge, the presence of accessory chromosomes among the Capnodiales has thus far been demonstrated only for *Zymoseptoria* spp. Here we show that the mini-chromosome Chr14 of *C. fulvum* Race 5 shows presence/absence variation among different isolates of the fungus and is therefore dispensable. Thus, at present, *C. fulvum* is just the second species of Capnodiales in which dispensable chromosomes are detected. It is currently unknown whether the presence of Chr14 in the genome of *C. fulvum* presents with any selective advantage to the fungus, as all the 25 genes present on it encode hypothetical proteins and no candidate effector genes were predicted to reside on this chromosome. However, evidence of gene flow between Chr14 and the core chromosomes was observed, with a duplicated hypothetical gene having one copy in Chr1 and another copy in Chr14 (Table 2.S14). Gene flow between core and accessory compartments of the genome has been reported for the rice pathogen *Magnaporthe oryzae*, in which the effector genes *PWL2* and *BAS1* can be located on the core chromosomes or side-by-side in a dispensable chromosome (Peng et al. 2019). This raises the possibility that Chr14 of *C. fulvum* acts as a reservoir that accelerates the evolution of genes by rapidly accumulating mutations via RIP leakage or other types of mutations induced by TE activity. This is supported by the fact that Chr14 is rich in repetitive DNA and heavily affected by RIP mutations with signs of abundant RIP leakage toward single-copy regions. However, more studies are required to provide insights about the function and importance of Chr14 in *C. fulvum*.

In summary, our work offers important and novel insights into the architecture and organization of the *C. fulvum* genome. Novel findings in this study include the presence of a dispensable mini-chromosome, the organization of genes into gene clusters that are flanked by repeat-rich regions in agreement with the two-speed model of evolution, and the separation between small and large chromosomes based on gene content. The genome of *C. fulvum* Race 5 provides a valuable resource for functional genomic and population genetic studies of this organism. It further highlights potential mechanisms underlying its adaptation to its tomato host by showing, among others, that the repeat-rich regions can serve as a cradle for genomic variability mediated by mutations and duplications induced via TE activity or genome instability. Future studies will focus on deciphering the importance for infections of the dispensable Chr14 and of genes with copy number variation in the genome of *C. fulvum*.

2.4 Materials and Methods

2.4.1 Fungal isolates, nucleic acid extractions, and sequencing

Cladosporium fulvum Race 5 Kim, a race 5 isolate of the fungus that was initially isolated in France in 1979 (Stergiopoulos et al. 2007b), was kindly provided by Emeritus Professor Pierre J. G. M. De Wit, Laboratory of Phytopathology at Wageningen University in the Netherlands. An additional twenty-four isolates of *C. fulvum* (Stergiopoulos et al. 2007b), were kindly provided by Professor Matthieu H. A. J. Joosten from the same laboratory.

High-molecular weight (HMW) genomic DNA from *C. fulvum* Race 5 was isolated according to Jones et al. 2019 (Jones et al. 2019) with some modifications. Specifically, *C. fulvum* Race 5 was grown on potato dextrose agar (PDA), at 22°C for two weeks. Spores were harvested from the PDA plates, and 10^6 spores were inoculated in 100 ml Gamborg's B5 medium and grown at 22°C for one week. The mycelia were filtered by two layers of cheesecloth and freeze-dried for two days. The dried material was ground with mortar and pestle in liquid nitrogen. An amount of 500 mg of starting material was then mixed with 17.5 ml of the lysis buffer, containing 10 kU of RNase A (ThermoFisher Scientific; Catalog #: EN0531), 6.5 ml of buffer A (0.35 M sorbitol in 0.1 M Tris-HCl, pH 9, and 5 mM EDTA, pH 8), 6.5 ml of buffer B (0.2 M Tris-HCl, pH 9, 50 mM EDTA, pH 8, 2 M NaCl, and 2% CTAB), and 2.75 ml of buffer C (5% N-lauroylsarcosine sodium salt). The sample was then incubated at 25°C for 30 mins and inverted every 5 mins. A total of 200 µl of Proteinase K (New England BioLabs Inc.; Catalog #: P8107S) was next added to the sample and incubated at 25 °C for 30 mins, while mixing by inversion every 5 min. After this step, 3.5 ml of 5 M potassium acetate was added to the sample, and the sample was incubated on ice for 5 mins. The sample was then spun at 5,000 g, at 4°C, for 12 mins. The supernatant was transferred to a 50 ml tube containing 17.5 ml of phenol:chloroform:isoamyl alcohol (25:24:1, v/v, Sigma-Aldrich, Catalog #: P3803) and mixed by inversion for 2 mins. The sample was centrifuged at 4,000 g, at 4°C, for 10 mins, and the phenol:chloroform:isoamyl alcohol separation was repeated again. The supernatant was mixed with 1.8 ml 3 M sodium acetate and 18 ml isopropanol, and then incubated at -20°C overnight. The mixture was centrifuged at 10,000 g, 4°C for 30 mins, and the pellet was transferred to a 1.7 ml tube. The pellet was centrifuged at 13,000 g for 5 mins. After removing the supernatant, the pellet was washed with 70% ethanol and then centrifuged at 13,000 g for 5 mins. The ethanol wash step was repeated once, and the pellet was air-dried for 5 mins. Subsequently, the pellet was dissolved in 200 µL of 10 mM Tris, pH8.5. The DNA was further cleaned up by using AMpure XP beads (Beckman, Coulter Inc., Catalog #: A63880) following the manufacturer's instructions. The DNA was quantified using a Qubit™ dsDNA broad range (BR) assay kit (ThermoFisher Scientific, Catalog #: Q32850) and its quality was measured with a Nanodrop ND-1000 (ThermoFisher Scientific) instrument based on the 260/280 and 260/230 ratios.

Library construction and sequencing of highly pure and HMW DNA was outsourced to the DNA Technologies and Expression Analysis Core Laboratory at the UC Davis Genome Center (https://dnatech.genomecenter.ucdavis.edu/). The sample was enriched for HMW fragments prior to library construction by size selection of fragments longer than 20 kb, using the BluePippin pulsed-field gel electrophoresis platform (Sage Science, Beverly, MA). The constructed library was then sequenced using one SMRT Cell 1M v2 on a Sequel Chemistry v2 platform (Pacific Biosciences, Menlo Park, CA) with 10 h of total movie time. DNA extracted from *C. fulvum* Race 5 was also used to generate an Illumina library. In addition, approximately 300 mg of fresh weight of *C. fulvum* Race 5 grown as described above, was used for Hi-C library construction using the Proximo Hi-C Kit (microbial) (Phase Genomics), according to the manufacturer's instructions. Illumina whole-genome sequencing, Hi-C, and RNA-seq libraries (see below) were sequenced on a NovaSeq 6000 instrument (PE150 format) utilizing 1.43%, 2.35%, and 9.78% of a lane, respectively.

Total RNA from *C. fulvum* Race 5 strain was extracted using the Trizol Reagent (Invitrogen, Catalog #: 15596026) according to the manufacturer's instructions. Briefly, 10^6 fungal spores were inoculated in 100 ml Gamborg's B5 medium with vitamins at 200 rpm, at 22°C for six days. The collected mycelia were subsequently inoculated in thirteen induction conditions for an additional 20 hrs. These induction conditions included growth in (i) 100 ml of rich medium (10 g/l yeast extract, 30 g/l glucose), (ii) 100 ml of minimal medium (1 g/l KH_2PO_4, 1 g/l KNO_3, 0.5 g/l $MgSO_4.7H_2O$, 0.5 g/l KCl, 0.5 g/l sucrose, and 0.5 g/l glucose), (iii) 100 ml of Gamborg's B5 medium with vitamins at 4°C, or (iv) 42°C for four days, 100 ml of Gamborg's B5 medium supplemented with (v) 10 mg/l thiamine, (vi) 2 mg/ml sorbitol, (vii) 2 mg/ml maltose, (viii) 2 mg/ml xylose, (ix) 10 mM ammonia sulfate, (x) 5 mM H_2O_2, (xi) 5 mM methanol, (xii) 0.5M glutamine, and (xiii) 100 ml of Gamborg's B5 medium without carbon source. The mycelia were collected by filtering through two layers of cheesecloths and ground into fine powders using a mortar and pestle, and liquid nitrogen. 100 mg of fine powder was then mixed with 1 ml Trizol™ reagent by vortexing and RNA was extracted according to the manufacturer's instructions. The RNA was quantified using a Qubit™ RNA broad range (BR) assay kit (ThermoFisher Scientific, Catalog #: Q10210) and its quality was measured with a Nanodrop™ ND-1000 (ThermoFisher Scientific) instrument based on the 260/280 and 260/230 ratios. Finally, the RNA samples were pooled in equimolar amounts into a single sample prior to Illumina library construction, which was outsourced to the DNA Technologies and Expression Analysis Core of the UC Davis Genome Center. Samples were sequenced (PE150 format) on a NovaSeq 6000 instrument (PE150 format) as described above.

2.4.2 Genome assembly

The genome of *C. fulvum* Race 5 was assembled with Canu v1.8 (Koren et al. 2017) with parameters *genomeSize=70m*, *corOutCoverage=60*, *minReadLength=5000*, *minOverlapLength=3000*, *corMinCoverage=5*, *corMhapSensitivity=normal*, and *correctedErrorRate=0.03*. Bacterial contigs were identified using the *sendsketch.sh* script from BBMap v38 (Bushnell 2014), and the contig containing the mitochondrial genome was identified by querying the mitochondrial genome of *Z. tritici* (Torriani et al. 2008) with BLASTn. Contigs were polished with Arrow v2.3.3 (https://github.com/PacificBiosciences/pbbioconda) based on PacBio reads mapped with pbmm2 v1.0.0

(https://github.com/PacificBiosciences/pbmm2). To further polish the contigs, Illumina reads were obtained and trimmed with fastp v0.20.1 (Chen et al. 2018). Trimmed reads were mapped with BWA-MEM v0.7.17-r1188 (Li and Durbin 2009), PCR duplicates were marked with samblaster v0.1.24 (Faust and Hall 2014), and polishing was carried out with Pilon v1.23 (Walker et al. 2014). Assembled contigs missing a telomere in one of their ends were extend up to 207 bp until the telomeric repeat was reached. To do so, Illumina reads mapping to the last or first 250 bp of contigs' ends that were missing telomeres, were extracted with SAMtools and the script *filterbyname.sh* from BBMap v38. The extracted read pairs were then assembled with SPAdes v3.15.3 (Bankevich et al. 2012) with k-mer values of 33, 55, 77, and 111, and the assembled fragments, which contained telomeric repeats, were merged manually with the respective contigs. The Illumina reads were also used to estimate the genome size with the *kmercountexact.sh* script from BBMap v38 using a k-mer value of 31. To predict the chromosomes of *C. fulvum* Race 5, sequenced Hi-C reads were mapped to the assembled contigs with BWA-MEM v0.7.17-r1188 with parameters *-5, -S,* and *-P* to allow mapping of each read end individually. Mapped reads were processed with samblaster v0.1.24 (Faust and Hall 2014) to mark PCR duplicates, and then with SAMtools v1.9 (Li et al. 2009) to remove mapped reads with mate unmapped, not primary or supplementary alignments (SAM flag = 2316). The scripts *makeAgpFromFasta.py* and *agp2assembly.py* (https://github.com/phasegenomics/juicebox_scripts) were then used to create an *assembly* file. Links were generated with matlock (https://github.com/phasegenomics/matlock) with parameter *bam2juicer*. A *hic* file was then produced from the *assembly* file and the links with the script *run-assembly-visualizer.sh* from the 3D-DNA package (Dudchenko et al. 2017). The Hi-C heat map was visualized and exported with Juicebox v1.11.08 (Durand et al. 2016).

2.4.3 Genome assembly evaluation

To estimate the integrity of the assembly and possible misassemblies, the trimmed Illumina and raw PacBio reads were mapped to the assembly with BWA-MEM v0.7.17-r1188 with default parameters (for Illumina) and parameters *-M* and *-x pacbio* (for PacBio). PCR duplicates were marked with samblaster v0.1.24 (Faust and Hall 2014). The number of mapped reads and properly paired Illumina reads were determined with *flagstat* from SAMtools v1.9 (Li et al. 2009). Sniffles v1.0.12 (Sedlazeck et al. 2018) was used to predict structural variants based on mapped PacBio reads with default settings. PacBio and Illumina read coverage was examined with IGV v2.6.1 (Robinson et al. 2011) at locations of predicted structural variations. Collapsed regions were identified based on the genome-wide coverage of PacBio and Illumina reads obtained with mosdepth v0.3.2 (Pedersen and Quinlan 2018), using a sliding window of 30 kb.

2.4.4 Repetitive DNA annotation

A custom *de novo* repetitive DNA library was obtained with RepeatModeler v1.0.11 using the *ncbi* engine and RepeatModeler v2.0.2 with the parameter *-LTRStruct* enabled to run the LTR structural discovery pipeline (Flynn et al. 2020). The produced consensus repeat library was queried with InterProScan v5.32-71.0 to search for conserved domains not related to transposons that could have been called as repetitive DNA. Repeats were then masked with RepeatMasker v4.0.7 using the consensus library produced by RepeatModeler and with parameters adjusted for higher sensitivity (-s), to output alignments (*-a*), and repeat coordinates in GFF format (*-gff*). The custom repetitive DNA library and repeat coordinates are available at https://doi.org/10.5281/zenodo.6380765. The custom repeat library and the repeat alignments produced by RepeatMasker were used to estimate repeat divergence with the *parseRM.pl* script from the Parsing-RepeatMasker-Outputs package (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs) with parameters *--land 50,1*, *--parse*, *--fa*, and *--nrem*. Genomic regions affected by RIP were identified with RIPper (Van Wyk et al. 2019), using a 1 kb sliding window and a step size of 500 bp. Windows with substrate index value (CpA + TpG)/(ApC + GpT) ≤ 0.75, product index value (TpA/ApT) ≥ 1.1, and composite index value (TpA/ApT) – ([CpA + TpG]/[ApC + GpT]) ≥ 0.01 were considered to be affected by RIP (Van Wyk et al. 2019). The percentage of masked bases covered by RIPped windows was used to estimate the level of RIP mutations in repetitive regions. Windows considered as RIPped were queried with BLASTn against the genome using e-value < 1E-20, identity > 50%, and query coverage > 20%. Based on these cutoff values, RIPped windows with a single BLASTn hit were considered single-copy, and therefore used as evidence of RIP leakage.

2.4.5 Gene prediction

RNA-seq reads were processed with fastp v0.20.1 (Chen et al. 2018) to trim the adapters and low-quality sequences. Reads were then mapped to the *C. fulvum* Race 5 genome with HISAT2 v2.2.0 (Kim et al. 2015) using a maximum intron length of 3,000 bp and the option *--dta* enabled to report alignments tailored for transcriptome reconstruction. Full length transcripts were then assembled with Stringtie v2.1.1 (Pertea et al. 2015). Genes were predicted with the Maker pipeline v2.31.10 (Cantarel et al. 2008). Initially, the assembled transcripts of *C. fulvum* Race 5 and protein sequences from *Z. tritici* isolate IPO323 (GCF_000219625.1) and *Cercospora beticola* isolate 09-40 (GCF_002742065.1) were used by Maker to produce gene models in order to train the *ab initio* gene predictors Augustus v3.2.3 (Stanke et al. 2006) and SNAP v2013-11-29 (Korf 2004). The script *maker2zff* that is incorporated in the SNAP software was used with parameters *-c 1 -o 1 -x 0.1* to extract 3,925 high-confidence gene models, which were then used to train again Augustus and SNAP. After parameter optimization, Augustus reported a sensitivity and a specificity at the nucleotide level of 0.968 and 0.836, respectively based on a testing data set that consisted of 200 genes. To further assist Maker predictions, gene models were also obtained with GeMoMa v1.6.3 (Keilwagen et al. 2019). GeMoMa mapped the gene annotations of *Z. tritici* IPO323 (GCF_000219625.1), *C. beticola* 09-40, and *C. fulvum* 0WU (JGI) with TBLASTn to the assembly of *C. fulvum* Race 5, and used the mapped RNA-seq reads to infer gene models with accurate exon-intron structure. The gene models produced with GeMoMa were filtered using GAF to keep only one isoform per gene (parameter *m=1*). The script *bam2hints* that is incorporated in the Braker software v2.1.5 (Hoff et al. 2019) was used to extract intron hints from the mapped RNA-seq reads, using a minimum and a maximum intron length of 20 and 2,000 bp, respectively. These intron hints and the gene models produced by GeMoMa using *C. beticola* as reference were used to train GeneMark v4.57 (Ter-Hovhannisyan et al. 2008) with parameters --*fungus* --*training* --*soft_mask auto*. Finally, the pre-identified repeats and *gff* format along with all lines of gene evidence, i.e. assembled transcripts of *C. fulvum* Race 5, protein sequences from *Z. tritici* IPO323 (GCF_000219625.1) and *C. beticola* 09-40 (GCF_002742065.1), GeMoMa gene models, trained predictors Augustus, SNAP, and GeneMark, were provided to Maker to select the best gene models for *C. fulvum* Race 5. Splice sites were extracted from mapped RNA-seq reads with RegTools v0.5.2 (Feng et al. 2018) with minimum intron size of 20 bp, maximum intron size of 3,000 bp, and minimum anchor length of 8 bp. The splice sites were annotated with RegTools and genes with splice sites fully supported (i.e. known donor-acceptor [DA]) by at least five reads were used to estimate exon-intron prediction accuracy.

2.4.6 Gene annotation

Gene annotation completeness was estimated with BUSCO (Benchmarking Universal Single-Copy Orthologs) v5.2.1 (Simão et al. 2015) using hmmsearch v3.1 and the database Dothideomycetes_db10 2020-08-05 as reference. Genes encoding key enzymes for secondary metabolism were identified with antiSMASH v6.0.1 (Medema et al. 2011). Genes encoding carbohydrate-active enzymes (CAZymes) were identified and classified with dbCAN2 meta server (Zhang et al. 2018), using the HMM database v9. Genes encoding proteases and transporters were identified based on homology searches performed with BLASTp (e-value < 1E-10) against the MEROPS database v12 (Rawlings et al. 2014) and the transporter classification database (TCDB; 2021-06-20) (Saier Jr et al. 2014), respectively. Proteases and transporters were classified based on the most homologous sequence according to BLASTp. Secreted proteins were identified with SignalP v5 (Armenteros et al. 2019) and transmembrane domains were identified with TMHMM v2 (Krogh et al. 2001). GPI-anchored proteins were identified with PredGPI (Pierleoni et al. 2008), using PFrate < 0.005 as threshold. Candidate effectors were characterized as secreted proteins and classified as effectors with EffectorP v2 (Sperschneider et al. 2016). We also considered as candidate effectors small secreted proteins shorter than 250 aa, with at least 2% cysteine residues, no transmembrane domain in the mature protein, and no GPI anchor. Previously described candidate effectors from *C. fulvum* 0WU were obtained from NCBI and mapped to the genome of *C. fulvum* Race 5 with minimap2 v2.16 (Li 2018) in splice aware mode (parameter *-x splice*). Missing or inaccurate gene annotation of these candidate effectors in *C. fulvum* Race 5 were manually curated when necessary, based on mapped RNA-seq reads of *C. fulvum* Race 5. Differences in gene content among the chromosomes were analyzed with a principal component analysis performed with the *prcomp* function (parameter *scale = TRUE*) and visualized with the *biplot* function within R v4.2.1.

2.4.7 Compartmentalization analysis

The intergenic regions were obtained using the script *complement* from BEDtools v2.29.0 (Quinlan and Hall 2010) to obtain genomic space not covered by genes. Subsequently, the script *closest* from BEDtools v2.29.0 was used to assign up- and downstream intergenic regions to each gene. Heat maps of the intergenic regions were obtained using the *geom_hex* function from the R package ggplot2 v3.3.3 (Wickham 2016) with a bin size of 50 within R v3.5.1. Clusters of genes based on intergenic sizes were obtained with the script *cluster* from BEDtools v2.29.0 by varying the maximum distance parameter (*-d*). Enrichment of PFAM domains of genes in gene-sparse regions was carried out with the *enricher* function from the R package clusterProfiler v3.18.1 (Yu et al. 2012), using the Benjamini and Hochberg p-value correction method and adjusted p-value < 0.05. Enrichment of specific gene categories within gene-sparse regions was performed with the *phyper* function within R v4.0.3.

2.4.8 Comparative analyses with other genomes

The genome of *C. fulvum* was aligned with the genome of other Dothideomycetes using PROmer from the MUMmer package v4.0 (Kurtz et al. 2004) with default settings. Alignments were then filtered with the *delta-filter* script that is incorporated in MUMmer to only retain the best matches (parameter *-1*). Alignment coordinates were used to make dot plots within R v3.5.1 and a circos plot with circos v0.69-8 (Krzywinski et al. 2009). Assembled scaffolds of *C. fulvum* 0WU were split into contigs with the *splitasm* function from RagTag v2.0.1 (Alonge et al. 2019). The resulting contigs were mapped to the *C. fulvum* Race 5 assembly with minimap2 v2.20 (Li 2018) with parameters *-ax asm10*. The alignment was filtered with SAMtools v1.9 (Li et al. 2009) to remove unmapped contigs and not primary alignment (SAM flag = 260). Regions uncovered by the mapped contigs were determined with the *genomecov* function from BEDtools v2.29.0 (Quinlan and Hall 2010). The nucleotide sequences of genes from *C. fulvum* Race 5 were queried with BLASTn v2.12.0 with e-value < 1E-5 and query coverage of at least 50% against the scaffolds of *C. fulvum* 0WU. Genes with no BLASTn hit were considered missing in the genome of *C. fulvum* 0WU. To obtain evidence of expression of genes missing in the genome assembly of *C. fulvum* 0WU, RNA-seq reads of this same isolate were obtained from NCBI SRA database (SRR1171044, SRR1171045, and SRR1171046) and mapped to the genome of *C. fulvum* Race 5 with HISAT2 v2.2.1 (Kim et al. 2015) with parameter *--max-intronlen 3000*. Reads mapped to genes were counted with *featureCounts* from the Subread package v2.0.1 (Liao et al. 2014), and transcripts per million (TPM) values were then calculated with a custom R script available at http://github.com/alexzaccaron/2021_cfr5_gm/tree/main/gene_expression/scripts.

2.4.9 Detection and confirmation of the mini-chromosome Chr14 in a population of *C. fulvum*

Four pairs of primers were designed to PCR-amplify eight predicted genes in four different regions of Chr14 (Table 2.S17 and Fig 2.S14A). Isolates were grown in PDA media for 10 days at 25°C. Spores and mycelial fragments were harvested from the media surface using sterile blades. DNA was extracted using a simple SDS based extraction (Penouilh-Suzette et al. 2020). The PCR reactions were performed using an Apex Red Mix (APEX Bioresearch Products, USA) following the manufacturer instructions. Cycling conditions consisted of 35 cycles of 30 s at 94°C, 30 s at 55°C, 56°C, or 57°C depending on the primer combination, and 2 min at 72°C. A final 7 min extension step at 72°C completed the reaction. PCR products were visualized in a 1% agarose gel (Fig 2.S14B). To further confirm dispensability of chromosome Chr14, DNA from isolates for which Chr14 was predicted to be present (isolates 2, IMI Argent 358077, and Turk 1a) or absent (isolates IPO 2.4.8.9.11 Polen, IPO 249 France, and 2.5) was pooled in equimolar amounts into two samples. DNA libraries were prepared using the Invitrogen Collibri ES DNA Library Prep Kit for Illumina Systems (Thermo Fisher Scientific) according to the manufacturer's instructions (protocol MAN001845). Libraries were multiplexed using unique dual indexes and sequenced at the UC Davis Genome Center on an Illumina NovaSeq 6000 instrument (PE150 format). Reads were trimmed with fastp v0.23.1 (Chen et al. 2018) and mapped to the genome assembly of *C. fulvum* Race 5 with BWA-MEM v0.7.17 (Li and Durbin 2009). Read depth across chromosome Chr14 was determined with *mosdepth* v0.3.3 (Pedersen and Quinlan 2018).

2.5 Data availability

This whole-genome project has been deposited at NCBI BioProject under the accession PRJNA565804. PacBio reads have been deposited at the NCBI Sequence Read Archive (SRA) under the accession SRR16292145. Illumina whole-genome sequencing, Hi-C, and RNA-seq reads have been deposited at the NCBI SRA under the accessions SRR16292144, SRR16292147, and SRR16292146, respectively. Pooled whole-genome sequencing of three isolates containing Chr14 and three isolates absent of Chr14 were deposited at the NCBI SRA under the accessions SRR18210015 and SRR18210014, respectively. The *C. fulvum* Race 5 chromosomes have been deposited at DDBJ/ENA/GenBank under the accessions CP090163 through CP090176. Scripts and code snippets utilized in this study were deposited in a public GitHub repository available at https://github.com/alexzaccaron/2021_cfr5_gm.

Author contributions

Conceptualization: AZZ, IS; **Data Curation:** AZZ, AS; **Formal Analysis:** AZZ; **Funding Acquisition:** IS; **Investigation:** AZZ; **Methodology:** AZZ, LHS, AS, IS; **Project Administration:** IS; **Software:** AZZ; **Supervision:** IS; **Validation:** AZZ; **Visualization:** AZZ, AS, IS; **Writing – Original Draft Preparation:** AZZ, LHC, AS, IS; **Writing – Review & Editing:** AZZ, LHC, AS, IS.

Funding information

This work was supported by the National Science Foundation (NSF) Award number 1557995 (IS, LHC, AS, AZZ) and by the USDA-NIFA Hatch Project CA-D-PPA-2185-H (IS). AZZ received additional funding from University of California, Davis, through the Dean's Distinguished Graduate Fellowship (DDGF) program.

Acknowledgments

We kindly thank Emeritus Prof. Pierre J.G.M De Wit for providing us with *C. fulvum* isolate Race 5 Kim. Prof. Matthieu H. A. J. Joosten and Christiaan Schol are acknowledged for providing us with additional isolates of *C. fulvum*. Adam Taranto is acknowledged for providing assistance during early stages of this work.

2.6 References

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., and Schatz, M. C. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20:1–17

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420–423

Balesdent, M.-H., Attard, A., Kühn, M., and Rouxel, T. 2002. New avirulence genes in the phytopathogenic fungus Leptosphaeria maculans. Phytopathology. 92:1122–1133

Balesdent, M.-H., Fudal, I., Ollivier, B., Bally, P., Grandaubert, J., Eber, F., Chèvre, A.-M., Leflon, M., and Rouxel, T. 2013. The dispensable chromosome of *Leptosphaeria maculans* shelters an effector gene conferring avirulence towards *Brassica rapa*. New Phytol. 198:887–898

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477

Bertazzoni, S., Williams, A. H., Jones, D. A., Syme, R. A., Tan, K.-C., and Hane, J. K. 2018. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic Fungi. Mol. Plant. Microbe Interact. 31:779–788

Bushnell, B. 2014. *BBMap: a fast, accurate, splice-aware aligner*. Lawrence Berkeley National Lab (LBNL), Berkeley, CA.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., and Yandell, M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188–196

Castanera, R., Lopez-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Perez, G., Pisabarro, A. G., Grigoriev, I. V., and others. 2016. Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. PLoS Genet. 12:e1006108

Chang, T.-C., Salvucci, A., Crous, P. W., and Stergiopoulos, I. 2016. Comparative genomics of the Sigatoka disease complex on banana suggests a link between parallel evolutionary changes in *Pseudocercospora fijiensis* and *Pseudocercospora eumusae* and increased virulence on the banana host. PLoS Genet. 12:e1005904

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890

Clutterbuck, A. J. 2011. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. Fungal Genet. Biol. 48:306–326

Collemare, J., Griffiths, S., Iida, Y., Karimi Jashni, M., Battaglia, E., Cox, R. J., and de Wit, P. J. 2014. Secondary metabolism and biotrophic lifestyle in the tomato pathogen *Cladosporium fulvum*. PLoS One. 9:e85877

Covo, S. 2020. Genomic instability in fungal plant pathogens. Genes. 11:421

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

van Dam, P., Fokkens, L., Ayukawa, Y., van der Gragt, M., Ter Horst, A., Brankovics, B., Houterman, P. M., Arie, T., and Rep, M. 2017. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. Sci. Rep. 7:1–15

De Wit, P. J., Joosten, M. H., Thomma, B. H., and Stergiopoulos, I. 2009. Gene for gene models and beyond: the *Cladosporium fulvum*-Tomato pathosystem. Pages 135–156 in: Plant relationships, Springer.

De Wit, P. J., Van Der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., Bahkali, A. H., Beenen, H. G., Chettri, P., Cox, M. P., and others. 2012. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. PLoS Genet. 8:e1003088

Depotter, J. R., Shi-Kunne, X., Missonnier, H., Liu, T., Faino, L., van den Berg, G. C., Wood, T. A., Zhang, B., Jacques, A., Seidl, M. F., and others. 2019. Dynamic virulence-related regions of the plant pathogenic fungus *Verticillium dahliae* display enhanced sequence conservation. Mol. Ecol. 28:3482–3495

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

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., and others. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 356:92–95

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., and Aiden, E. L. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3:99–101

Dutheil, J. Y., Mannhaupt, G., Schweizer, G., MK Sieber, C., Münsterkötter, M., Güldener, U., Schirawski, J., and Kahmann, R. 2016. A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. Genome Biol. Evol. 8:681–704

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., and others. 2009. Real-time DNA sequencing from single polymerase molecules. Science. 323:133–138

Faust, G. G., and Hall, I. M. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 30:2503–2505

Feng, Y.-Y., Ramu, A., Cotto, K. C., Skidmore, Z. L., Kunisaki, J., Conrad, D. F., Lin, Y., Chapman, W., Uppaluri, R., Govindan, R., and others. 2018. RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. BioRxiv. :436634

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., and Smit, A. F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. 117:9451–9457

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., and 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

Frantzeskakis, L., Kusch, S., and Panstruga, R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. Mol. Plant Pathol. 20:3–7

Gan, P., Hiroyama, R., Tsushima, A., Masuda, S., Shibata, A., Ueno, A., Kumakura, N., Narusaka, M., Hoat, T. X., Narusaka, Y., and others. 2021. Telomeres and a repeat-rich chromosome encode effector gene clusters in plant pathogenic *Colletotrichum* fungi. Environ. Microbiol.

Goodwin, S. B., Ben M'Barek, S., Dhillon, B., Wittenberg, A. H., Crane, C. F., Hane, J. K., Foster, A. J., Van der Lee, T. A., Grimwood, J., Aerts, A., and others. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7:e1002070

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., and others. 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. BMC Genomics. 15:1–27

Guo, X., Zhang, R., Li, Y., Wang, Z., Ishchuk, O. P., Ahmad, K. M., Wee, J., Piskur, J., Shapiro, J. A., and Gu, Z. 2020. Understand the genomic diversity and evolution of fungal pathogen *Candida glabrata* by genome-wide analysis of genetic variations. Methods. 176:82–90

Hane, J. K., Rouxel, T., Howlett, B. J., Kema, G. H., Goodwin, S. B., and Oliver, R. P. 2011. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biol. 12:1–16

Haridas, S., Albert, R., Binder, M., Bloem, J., LaButti, K., Salamov, A., Andreopoulos, B., Baker, S., Barry, K., Bills, G., and others. 2020. 101 Dothideomycetes genomes: a test case for predicting lifestyles and emergence of pathogens. Stud. Mycol. 96:141–153

Hoff, K., Lomsadze, A., Borodovsky, M., and Stanke, M. 2019. Whole-genome annotation with BRAKER. Methods Mol. Biol. Clifton NJ. 1962:65

Houben, A., Banaei-Moghaddam, A. M., Klemme, S., and Timmis, J. N. 2014. Evolution and biology of supernumerary B chromosomes. Cell. Mol. Life Sci. 71:467–478

Houterman, P. M., Speijer, D., Dekker, H. L., de Koster, C. G., Cornelissen, B. J., and Rep, M. 2007. The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. Mol. Plant Pathol. 8:215–221

Iida, Y., van 't Hof, P., Beenen, H., Mesarich, C., Kubota, M., Stergiopoulos, I., Mehrabi, R., Notsu, A., Fujiwara, K., Bahkali, A., and others. 2015. Novel mutations detected in avirulence genes overcoming tomato Cf resistance genes in isolates of a Japanese population of *Cladosporium fulvum*. PloS One. 10:e0123271

Jones, A., Nagar, R., Sharp, A., and Schwessinger, B. 2019. High-molecular weight DNA extraction from challenging fungi using CTAB and gel purification. protocols.io.

Joosten, M. H., and De Wit, P. J. 1999. The tomato–*Cladosporium fulvum* interaction: a versatile experimental system to study plant-pathogen Interactions. Annu. Rev. Phytopathol. 37:335–367

van Kan, J. A., Van den Ackerveken, G., and De Wit, P. 1991. Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. Mol Plant-Microbe Interact. 4:52–59

Keilwagen, J., Hartung, F., and Grau, J. 2019. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol. Biol. Clifton NJ. 1962:161–177

Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 12:357–360

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736

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

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

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. L. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:1–9

Lazar, N., Mesarich, C. H., Petit-Houdenot, Y., Talbi, N., de la Sierra-Gallay, I. L., Zelie, E., Blondeau, K., Gracy, J., Ollivier, B., Blaise, F., and others. 2021. A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins. bioRxiv. :2020–12

Levan, A., Fredga, K., and Sandberg, A. A. 1964. Nomenclature for centromeric position on chromosomes. Hereditas. 52:201–220

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

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25:1754–1760

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

Li, J., Fokkens, L., van Dam, P., and Rep, M. 2020. Related mobile pathogenicity chromosomes in *Fusarium oxysporum* determine host range on cucurbits. Mol. Plant Pathol. 21:761–776

Liao, Y., Smyth, G. K., and Shi, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 30:923–930

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., and others. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. science. 326:289–293

Lofgren, L. A., Uehling, J. K., Branco, S., Bruns, T. D., Martin, F., and Kennedy, P. G. 2019. Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. Mol. Ecol. 28:721–730

Lorrain, C., Oggenfuss, U., Croll, D., Duplessis, S., and Stukenbrock, E. 2021. Transposable elements in fungi: coevolution with the host genome shapes, genome architecture, plasticity and adaptation. Pages 142–155 in: Encyclopedia of Mycology, Ó. Zaragoza and A. Casadevall, eds. Elsevier, Oxford.

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., and others. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 464:367–373

McDonald, M. C., Taranto, A. P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., Milgate, A., and Solomon, P. S. 2019. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. MBio. 10:e01515-19

Medema, M. H., Blin, K., Cimermancic, P., De Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39:W339–W346

Mesarich, C. H., Griffiths, S. A., van der Burgt, A., Ökmen, B., Beenen, H. G., Etalo, D. W., Joosten, M. H., and de Wit, P. J. 2014. Transcriptome sequencing uncovers the *Avr5* avirulence gene of the tomato leaf mold pathogen *Cladosporium fulvum*. Mol. Plant. Microbe Interact. 27:846–857

Mesarich, C. H., Ökmen, B., Rovenich, H., Griffiths, S. A., Wang, C., Karimi Jashni, M., Mihajlovski, A., Collemare, J., Hunziker, L., Deng, C. H., and others. 2018. Specific hypersensitive response–associated recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato. Mol. Plant. Microbe Interact. 31:145–162

Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sánchez-García, M., Morin, E., Andreopoulos, B., Barry, K. W., Bonito, G., and others. 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nat. Commun. 11:1–17

Möller, M., Habig, M., Freitag, M., and Stukenbrock, E. H. 2018. Extraordinary genome instability and widespread chromosome rearrangements during vegetative growth. Genetics. 210:517–529

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

Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., and Ginalski, K. 2019. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci. Rep. 9:4307

Navarrete, F., Grujic, N., Stirnberg, A., Saado, I., Aleksza, D., Gallei, M., Adi, H., Alcântara, A., Khan, M., Bindics, J., and others. 2021. The Pleiades are a cluster of fungal effectors that inhibit host defenses. PLoS Pathog. 17:e1009641

Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F. E., Singh, N. K., Abraham, L. N., Karisto, P., Vonlanthen, T., Mundt, C. C., McDonald, B. A., and others. 2021. A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. bioRxiv. :2020–02

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., Condon, B. J., Copeland, A. C., Dhillon, B., Glaser, F., and others. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 8:e1003037

Pedersen, B. S., and Quinlan, A. R. 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 34:867–868

Peng, Z., Oliveira-Garcia, E., Lin, G., Hu, Y., Dalby, M., Migeon, P., Tang, H., Farman, M., Cook, D., White, F. F., and others. 2019. Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. PLoS Genet. 15:e1008272

Penouilh-Suzette, C., Fourré, S., Besnard, G., Godiard, L., and Pecrix, Y. 2020. A simple method for high molecular-weight genomic DNA extraction suitable for long-read sequencing from spores of an obligate biotroph oomycete. J. Microbiol. Methods. 178:106054

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290–295

Pierleoni, A., Martelli, P. L., and Casadio, R. 2008. PredGPI: a GPI-anchor predictor. BMC Bioinformatics. 9:392

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., and Croll, D. 2017. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. Front. Plant Sci. 8:119

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

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

Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42:D503–D509

Rep, M., Van Der Does, H. C., Meijer, M., Van Wijk, R., Houterman, P. M., Dekker, H. L., De Koster, C. G., and Cornelissen, B. J. 2004. A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for *I-3*-mediated resistance in tomato. Mol. Microbiol. 53:1373–1383

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. 2011. Integrative genomics viewer. Nat. Biotechnol. 29:24–26

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., Van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., and others. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:1–10

Saier Jr, M. H., Reddy, V. S., Tamang, D. G., and Västermark, Å. 2014. The transporter classification database. Nucleic Acids Res. 42:D251–D258

Salzberg, S. L., and Yorke, J. A. 2005. Beware of mis-assembled genomes. Bioinformatics. 21:4320–4321

Schrader, L., and Schmitz, J. 2019. The impact of transposable elements in adaptive evolution. Mol. Ecol. 28:1537–1549

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Von Haeseler, A., and Schatz, M. C. 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods. 15:461–468

Selker, E. U. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu. Rev. Genet. 24:579–613

Selker, E. U., and Garrett, P. W. 1988. DNA sequence duplications trigger gene inactivation in *Neurospora crassa*. Proc. Natl. Acad. Sci. 85:6870–6874

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

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743–761

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439

Stergiopoulos, I., De Kock, M. J., Lindhout, P., and De Wit, P. J. 2007a. Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. Mol. Plant. Microbe Interact. 20:1271–1283

Stergiopoulos, I., Groenewald, M., Staats, M., Lindhout, P., Crous, P. W., and De Wit, P. J. 2007b. Mating-type genes and the genetic structure of a world-wide collection of the tomato pathogen *Cladosporium fulvum*. Fungal Genet. Biol. 44:415–429

Stergiopoulos, I., and de Wit, P. J. 2009. Fungal effector proteins. Annu. Rev. Phytopathol. 47:233–263

Stewart, E. L., Croll, D., Lendenmann, M. H., Sanchez-Vallet, A., Hartmann, F. E., Palma-Guerrero, J., Ma, X., and McDonald, B. A. 2018. Quantitative trait locus mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen *Zymoseptoria tritici*. Mol. Plant Pathol. 19:201–216

Temporini, E. D., and VanEtten, H. D. 2002. Distribution of the pea pathogenicity (*PEP*) genes in the fungus *Nectria haematococca* mating population VI. Curr. Genet. 41:107–114

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18:1979–1990

Thomma, B. P., Van Esse, H. P., Crous, P. W., and de Wit, P. J. 2005. *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. Mol. Plant Pathol. 6:379–393

Torres, D. E., Oggenfuss, U., Croll, D., and Seidl, M. F. 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. Fungal Biol. Rev.

Torriani, S. F. F., Goodwin, S. B., Kema, G. H. J., Pangilinan, J. L., and McDonald, B. A. 2008. Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. Fungal Genet. Biol. 45:628–637

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

Tsuge, T., Harimoto, Y., Akimitsu, K., Ohtani, K., Kodama, M., Akagi, Y., Egusa, M., Yamamoto, M., and Otani, H. 2013. Host-selective toxins produced by the plant pathogenic fungus *Alternaria alternata*. FEMS Microbiol. Rev. 37:44–66

Tsushima, A., Gan, P., Kumakura, N., Narusaka, M., Takano, Y., Narusaka, Y., and Shirasu, K. 2019. Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. Genome Biol. Evol. 11:1487–1500

Van den Ackerveken, G. F., Van Kan, J. A., and De Wit, P. J. 1992. Molecular analysis of the avirulence gene *avr9* of the fungal tomato pathogen *Cladosporium fulvum* fully supports the gene-for-gene hypothesis. Plant J. 2:359–366

Van Wyk, S., Harrison, C. H., Wingfield, B. D., De Vos, L., van Der Merwe, N. A., and Steenkamp, E. T. 2019. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. PeerJ. 7:e7447

Van Wyk, S., Wingfield, B. D., De Vos, L., Van Der Merwe, N. A., and Steenkamp, E. T. 2021. Genome-wide analyses of Repeat-Induced Point mutations in the ascomycota. Front. Microbiol. 11:3625

Varoquaux, N., Liachko, I., Ay, F., Burton, J. N., Shendure, J., Dunham, M. J., Vert, J.-P., and Noble, W. S. 2015. Accurate identification of centromere locations in yeast genomes using Hi-C. Nucleic Acids Res. 43:5331–5339

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963

Wang, M., Fu, H., Shen, X.-X., Ruan, R., Rokas, A., and Li, H. 2019. Genomic features and evolution of the conditionally dispensable chromosome in the tangerine pathotype of *Alternaria alternata*. Mol. Plant Pathol. 20:1425–1438

Westerink, N., Brandwagt, B. F., De Wit, P. J., and Joosten, M. H. 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

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wijayawardene, N. N., Hyde, K. D., Rajeshkumar, K. C., Hawksworth, D. L., Madrid, H., Kirk, P. M., Braun, U., Singh, R. V., Crous, P. W., Kukwa, M., and others. 2017. Notes for genera: Ascomycota. Fungal Divers. 86:1–594

de Wit, P. J. 2016. *Cladosporium fulvum* effectors: weapons in the arms race with tomato. Annu. Rev. Phytopathol. 54:1–23

Witte, T. E., Villeneuve, N., Boddy, C. N., and Overy, D. P. 2021. Accessory chromosome-acquired secondary metabolism in plant pathogenic fungi: the evolution of biotrophs into host-specific pathogens. Front. Microbiol. 12

Wu, J., Kou, Y., Bao, J., Li, Y., Tang, M., Zhu, X., Ponaya, A., Xiao, G., Li, J., Li, C., and others. 2015. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector *AvrPi9* that triggers *Pi9*-mediated blast resistance in rice. New Phytol. 206:1463–1475

Yang, H., Yu, H., and Ma, L.-J. 2020. Accessory chromosomes in *Fusarium oxysporum*. Phytopathology. 110:1488–1496

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J. Integr. Biol. 16:284–287

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46:W95–W101

2.7 Supplementary materials

2.7.1 Supplementary figures

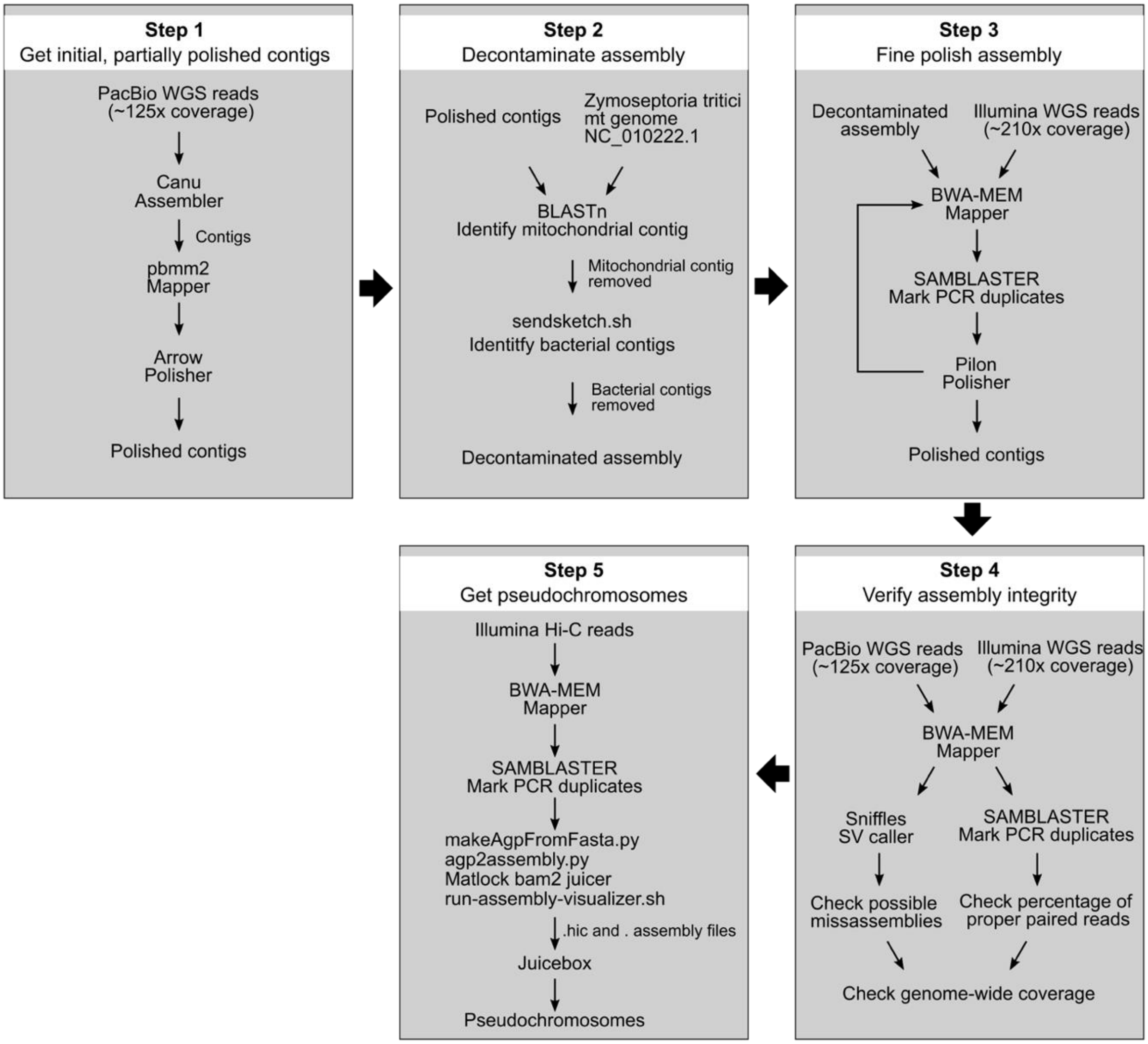

Step 1
Get initial, partially polished contigs
PacBio WGS reads (~125x coverage)
Canu Assembler
Contigs
pbmm2 Mapper
Arrow Polisher
Polished contigs
Step 2
Decontaminate assembly
Polished contigs
Zymoseptoria tritici mt genome NC_010222.1
BLASTn
Identify mitochondrial contig
Mitochondrial contig removed
sendsketch.sh
Identitfy bacterial contigs
Bacterial contigs removed
Decontaminated assembly
Step 3
Fine polish assembly
Decontaminated assembly
Illumina WGS reads (~210x coverage)
BWA-MEM Mapper
SAMBLASTER Mark PCR duplicates
Pilon Polisher
Polished contigs
Step 4
Verify assembly integrity
PacBio WGS reads (~125x coverage)
Illumina WGS reads (~210x coverage)
BWA-MEM Mapper
Sniffles SV caller
SAMBLASTER Mark PCR duplicates
Check possible missassemblies
Check percentage of proper paired reads
Check genome-wide coverage
Step 5
Get pseudochromosomes
Illumina Hi-C reads
BWA-MEM Mapper
SAMBLASTER Mark PCR duplicates
makeAgpFromFasta.py
agp2assembly.py
Matlock bam2 juicer
run-assembly-visualizer.sh
.hic and . assembly files
Juicebox
Pseudochromosomes

Figure 2.S1: Workflow used to assemble the genome of *Cladosporium fulvum* Race 5. The workflow has five major steps. First, contigs are assembled with Canu based on PacBio reads, followed by an initial round of polishing performed with Arrow. This round of polishing performed 7,347 changes, most of which (n = 7,020) were short INDELs of at most 13 bp. Assembly decontamination is performed in step 2. Specifically, from the 43 assembled contigs, 24 of them with a total size of 12.9 Mb matched bacterial genomes. These 24 contigs were then removed along with another four contigs of a total size of 186.4 kb as they were either contained within other contigs or they were formed from a single PacBio read. Another contig of 179.1 kb in size containing the mitochondrial genome of *C. fulvum* was also removed in this step. In step 3, the decontaminated assembly is fine polished with Illumina reads. This step was performed twice.

In the first round, 692 changes were performed, whereas only four changes were performed in the second round. In step 4, potential misassemblies are identified by mapping the PacBio reads to the contigs and calling structural variants with Sniffles. Collapsed regions in the assembly are identified by mapping the PacBio and Illumina reads to the contigs, and verifying regions with abnormally high coverage. Finally, in step 5 Hi-C reads are mapped to the contigs and pseudochromosomes are identified.

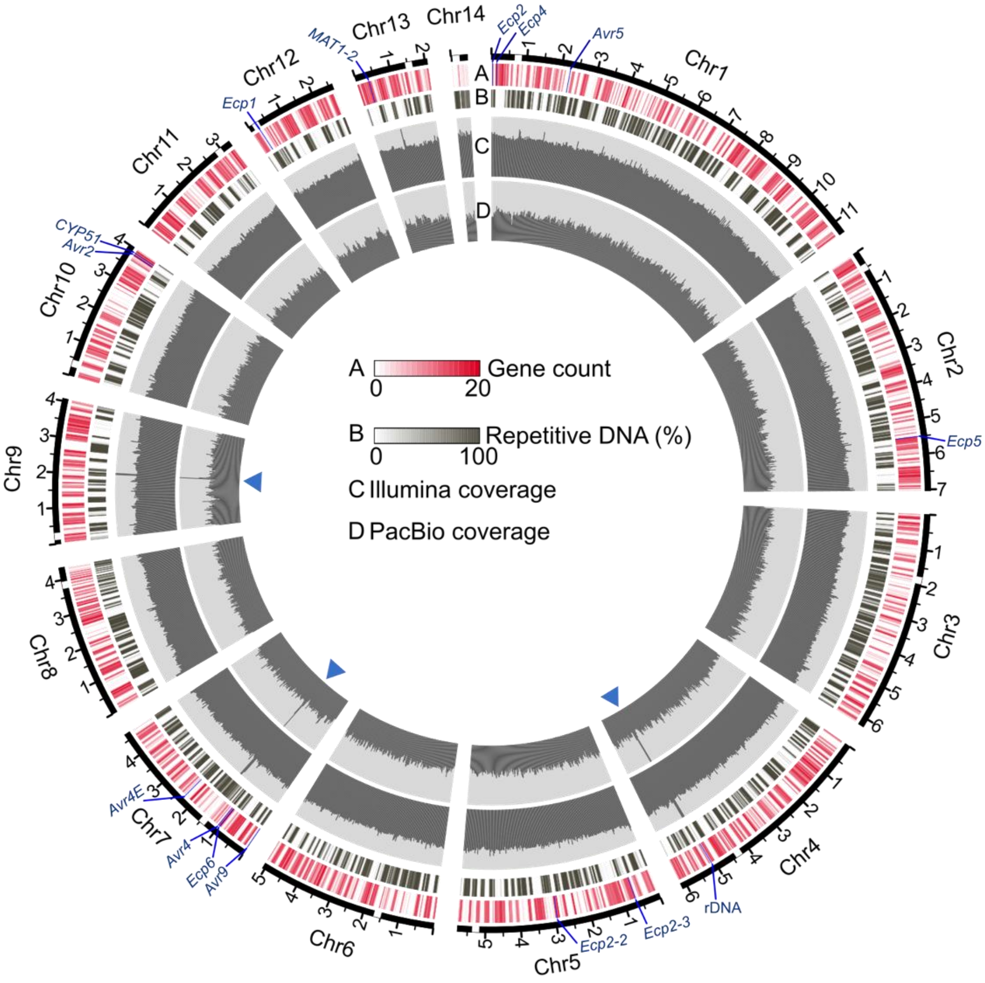


Figure 2.S2: Two putative misassembled regions weakly supported by raw sequencing data in the genome assembly of *Cladosporium fulvum* Race 5. Coverage of PacBio and Illumina reads are show for both regions. For PacBio, reads with supplementary alignments are shown at the top and supplementary alignments of the same read are linked (links are not visible because supplementary alignments of the same read have overlapping mapping coordinates). Each of these PacBio reads shown at the top in both regions have exactly one supplementary alignment mapped to the same location with opposite orientation as the primary alignment. These PacBio reads are likely the reason that Sniffles predicted an inverted tandem duplication in both regions (approximate location indicated with square boxes). However, most of the PacBio reads as well as the Illumina reads do no support the presence of inverted tandem duplications or

other type of misassembly, because most PacBio reads span the entire regions and almost all Illumina reads mapped to these two regions are properly paired and uniquely mapped (mapping quality = 60).

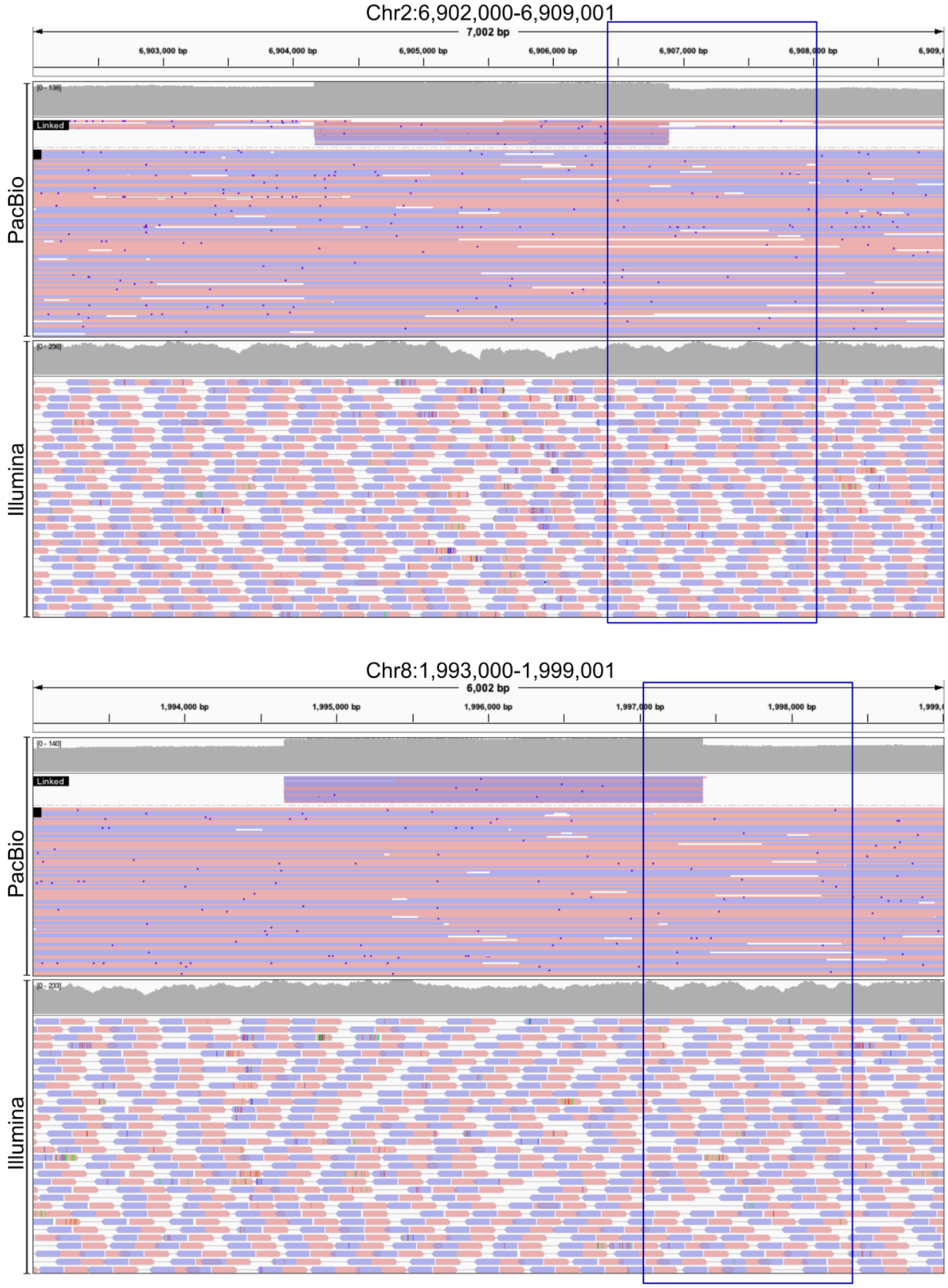


Figure 2.S3: Integrity of the assembly of *Cladosporium fulvum* Race 5. Two genome locations where Sniffles predicted inverted tandem duplications (approximate location with square boxes) based on the PacBio reads, suggesting possible misassemblies. The PacBio and Illumina reads that mapped to these regions are shown in grey, with soft clipped regions indicated in color-coded base pairs (i.e., adenine in

green, cytosine in blue, guanine in brown, and thymine in red). Some PacBio reads were only partially mapped to these regions but the mapped Illumina reads do not suggest problems with the assembly in these regions.

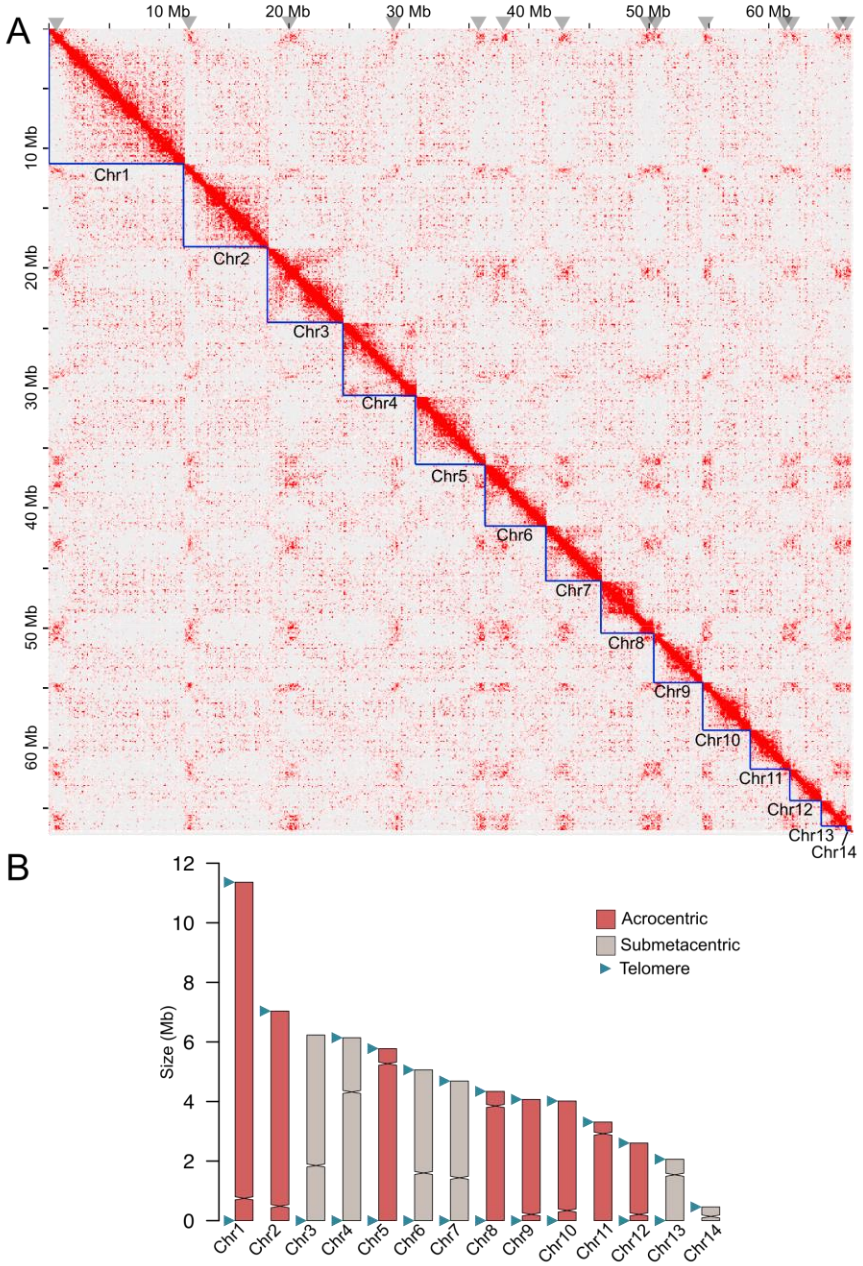


Figure 2.S4: Chromosomes of *Cladosporium fulvum* Race 5 identified based on an analysis of Hi-C data. (A) Heat map showing *all-versus-all* interaction frequency across the genome. The predicted 14 chromosomes are highlighted with blue lines. Putative centromeric regions are indicated with triangles at the top of the figure. In fungi, centromeric regions typically have high inter-chromosomal interaction frequency. The heat map was visualized and exported with Juicebox v1.11.08 at resolution (i.e. bin size) of 100 kb. (B) Classification of the 14 chromosomes of *C. fulvum* Race 5 into submetacentric or acrocentric,

based on the putative location of the centromeres. Submetacentric chromosomes have a ratio of long arm and short arm sizes between 1 and 3, and acrocentric chromosomes have a ratio of long arm and short arm sizes greater than 7. Presence of telomeric repeats at the chromosomes' immediate ends is indicated by triangles.

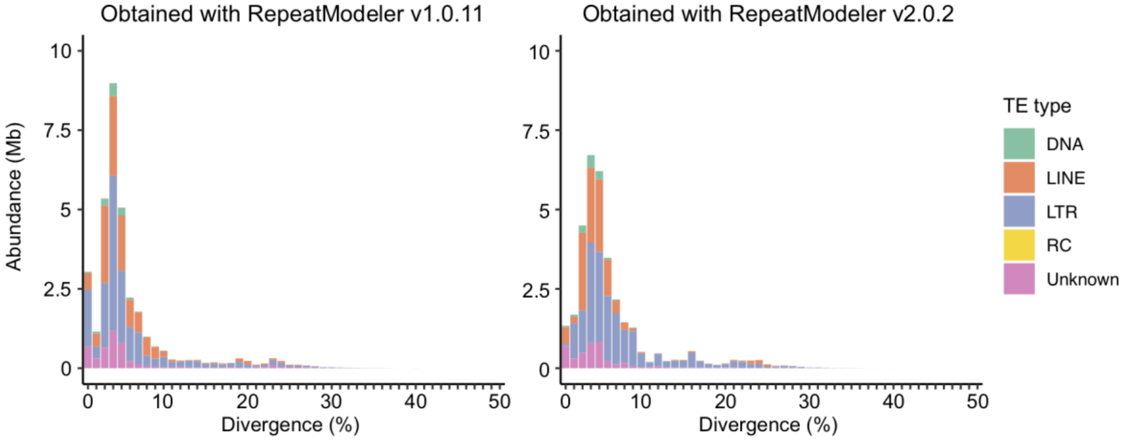


Figure 2.S5: RepeatModeler v1 and RepeatModeler v2 produce similar results using the Cladosporium fulvum Race 5 genome data as input, and support an overall low repeat divergence in this pathogen. Bar plot showing the number of bases (y-axis) covered by predicted transposable elements (TEs) of different (sub)classes, i.e., DNA transposons (DNA), long interspersed nuclear elements (LINE), long terminal repeats (LTR), rolling-circles (RC), and unclassified TEs. The x-axis shows the divergence of repeats from the consensus sequences. The figure shows that the genome of *C. fulvum* Race 5 is abundant in repeats with an overall low divergence.

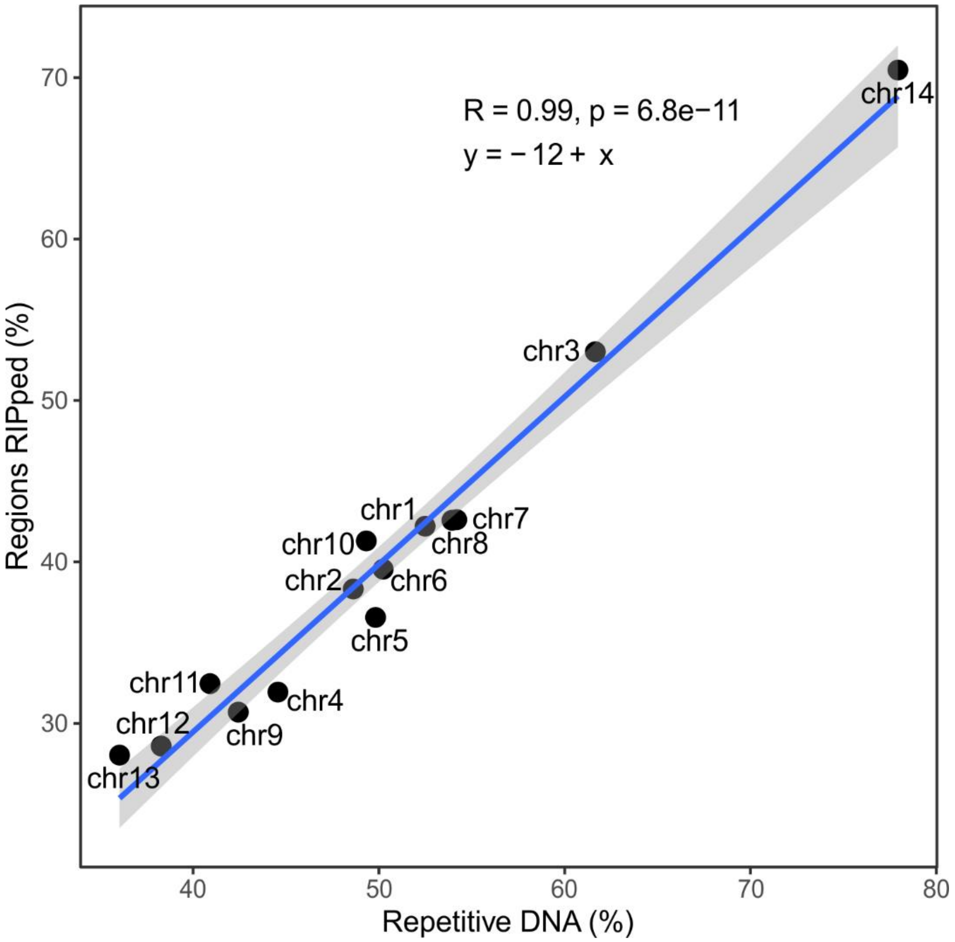


Figure 2.S6: Correlation between repetitive DNA content and percentage of regions affected by Repeat-Induced Point (RIP) mutations in the chromosomes of *Cladosporium fulvum* Race 5. The scatter plot shows a positive correlation between repetitive DNA content and the percentage of regions affected by RIP. A regression line is shown in blue and was determined with the *geom_smooth* function from the R package ggplot2, utilizing the lm method. Dark areas represent confidence intervals (95%). Correlation coefficient, p-value and the equation of the regression line are shown at the top of the plot.

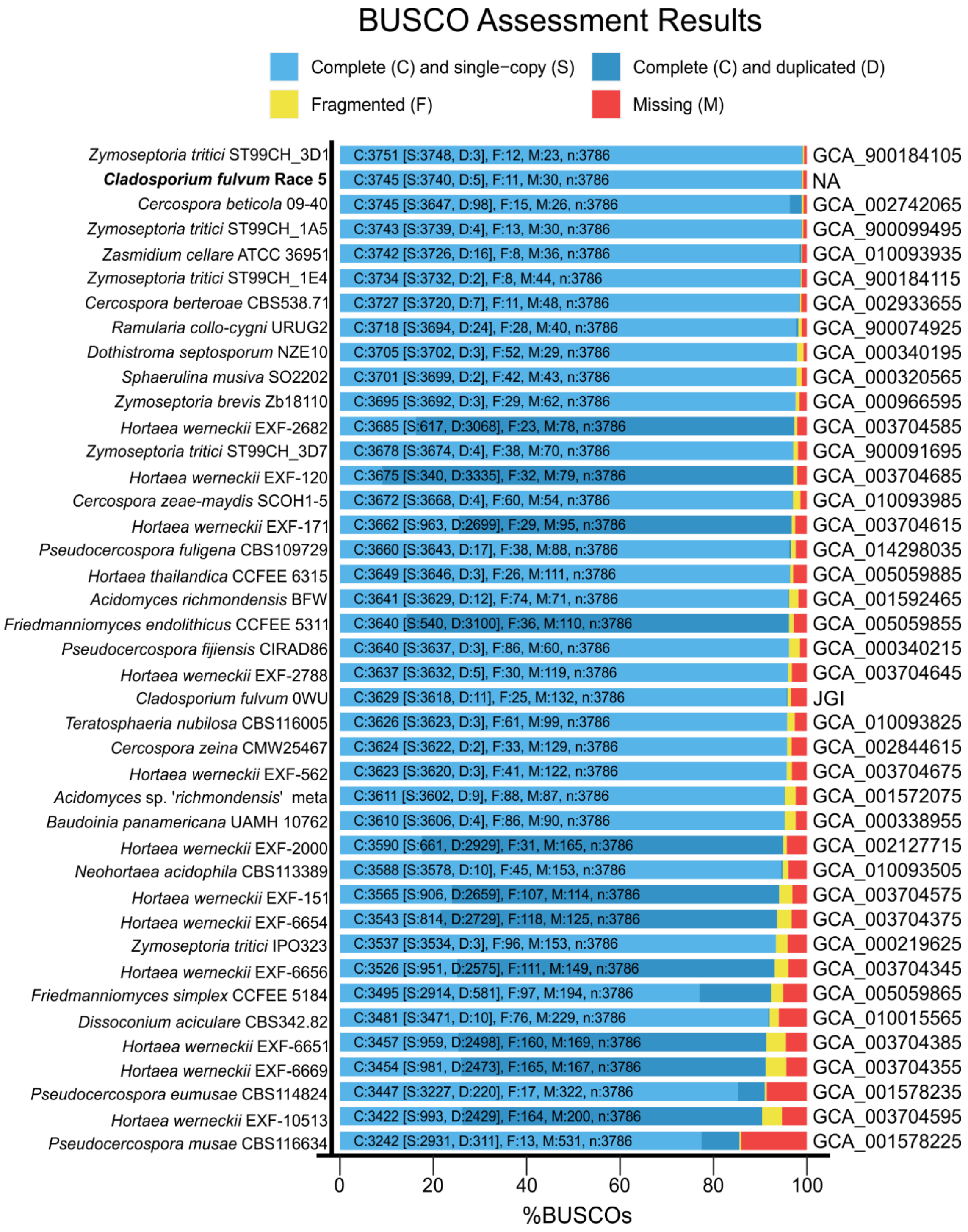


Figure 2.S7: Comparison of BUSCO completeness of *Cladosporium fulvum* Race 5 and other annotated genomes of Capnodiales. Genomes were ordered by estimated completeness (i.e., complete single copy plus complete duplicated BUSCOs). Genome accession numbers are shown on the right-hand side, except for *C. fulvum* Race 5 and *C. fulvum* 0WU (obtained from JGI MycoCosm). Values were obtained with BUSCO v5.2.1 in protein mode, using the Dothideomycetes_odb10 (2020-08-05) data base containing 3,786 BUSCOs as reference. Genomes were obtained from NCBI (2021-07-28).

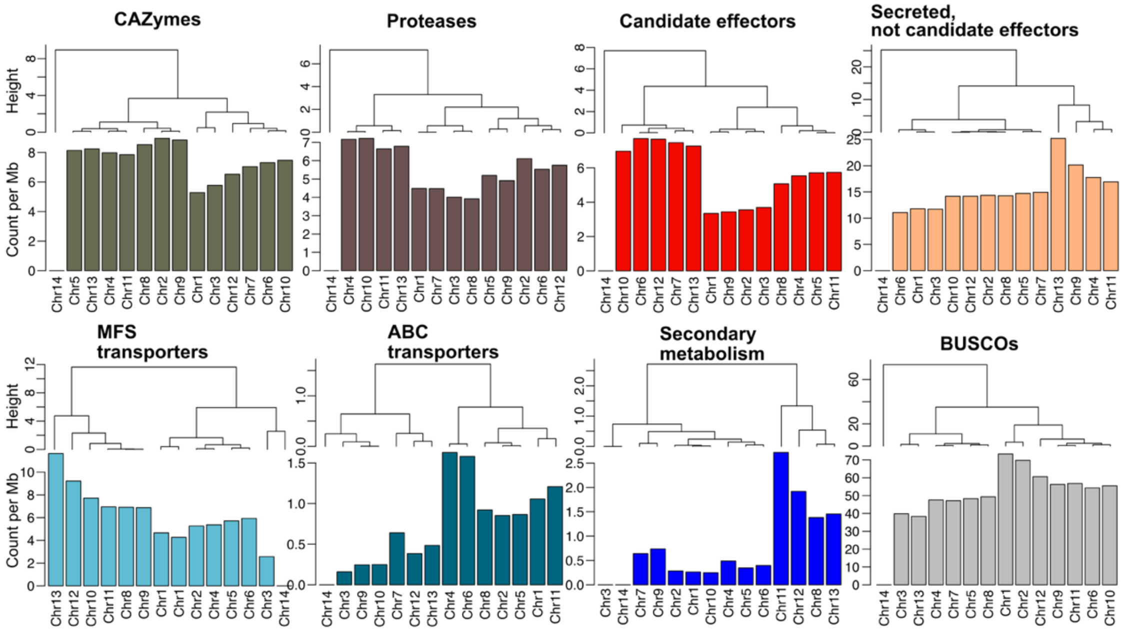


Figure 2.S8: Distribution of various gene categories within the chromosomes of *Cladosporium fulvum* Race 5. Bar plots showing hierarchical clustering of chromosomes based on gene densities of specific gene categories. Hierarchical clustering was performed using the *complete* method based on the Euclidean distances.

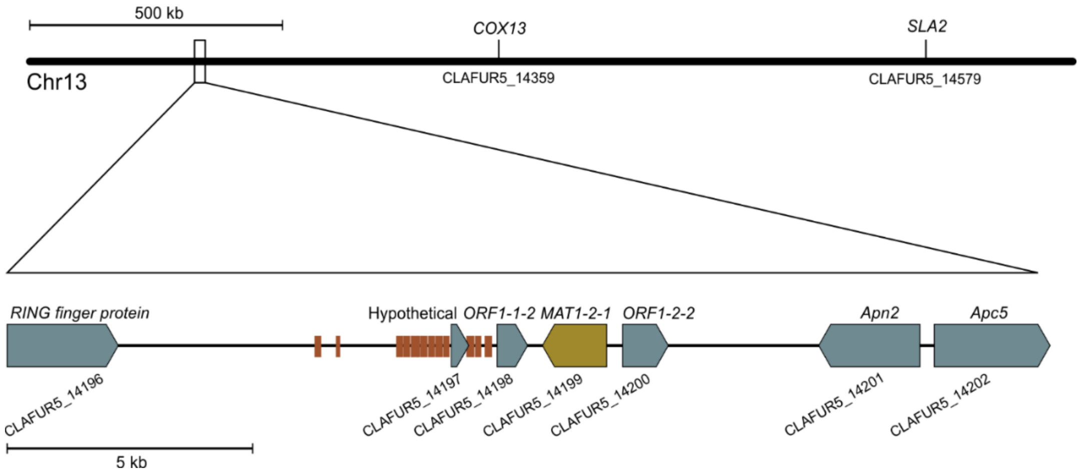
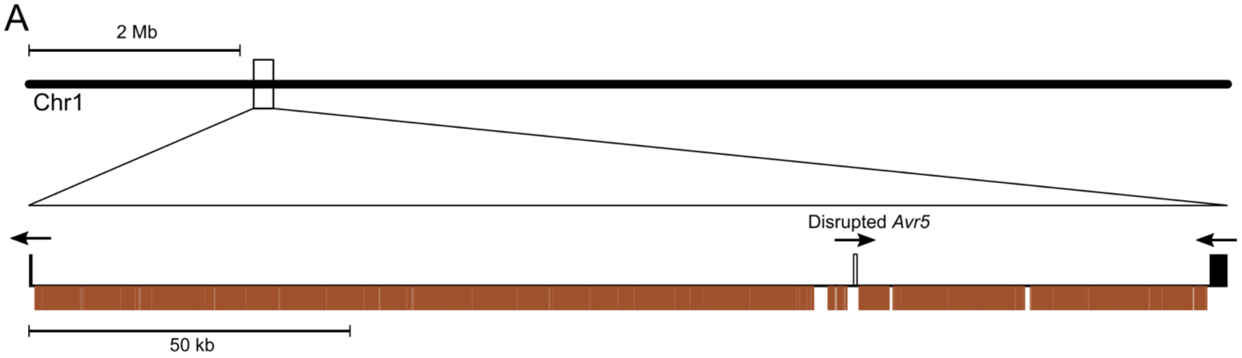


Figure 2.S9: Organization of the *MAT1-2* mating locus present in *Cladosporium fulvum* Race 5. The figure shows the location of the *MAT1-2-1* gene in Chr13. Two genes, named *ORF1-1-2* and *ORF1-2-2,* encoding hypothetical proteins flank *MAT1-2-1*. Three genes, i.e. *Apn2, SLA2,* and *COX13*, are typically located near *MAT* genes in Ascomycetes. However, only *Apn2* was near *MAT1-2-1* in *C. fulvum*. The location of *SLA2* and *COX13* in Chr13 are indicated with vertical lines. Other genes surrounding the MAT locus in *C. fulvum* encode a putative RING finger protein, a hypothetical protein, and the putative subunit 5 of the anaphase-promoting complex (Apc5). Repetitive regions are indicated with rectangles.



B

```
Chr1 : ATGAAG--TCTCCTATCGTAATCACAATTCTGGCTACTGCACTCGGGGCACTTGGTAGCTACGACGCC  >>>>     Intron 1    >>>>  CTACCTATTAACTGT 2258118
       ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||++           51 bp         ++|||||||||||||||
Avr5 : ATGAAGTCTCTCCTATCGTAATCACAATTCTGGCTACTGCACTCGGGGCACTTGGTAGCTACGACGCCgt.........................agCTACCTATTAACTGT 81


Chr1 : AGGGACACGACCAACTACTGTTTTAACGGTAACGGACGTCACGAAGTGTG  >>>>     Intron 2    >>>>  CTCATACTGCAACCAGGCCAAAGAAGAGCCCCT 2258269
       ||||||||||||||||||||||||||||||||||||||||||||||||||++           68 bp         ++|||||||||||||||||||||||||||||||||
Avr5 : AGGGACACGACCAACTACTGTTTTAACGGTAACGGACGTCACGAAGTGTGgt.........................agCTCATACTGCAACCAGGCCAAAGAAGAGCCCCT 164


Chr1 : TAAACTAGGCCGTCGAGGAGGCCAGCGTGATTGCGGTGTAGCAGGGAGCCAATGTAACGACGTAGACCATCAGCAATGC  >>>>     Intron 3    >>>>  GATG 2258400
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||++           48 bp         ++||||
Avr5 : TAAACTAGGCCGTCGAGGAGGCCAGCGTGATTGCGGTGTAGCAGGGAGCCAATGTAACGACGTAGACCATCAGCAATGCgt.........................agGATG 247


Chr1 : CCAGATGTTGCTCAAAAATAGGCTCGCCCACGTTTTATG  >>>>     Intron 4    >>>>  GAGTGCGATGCCCGTATCCGTACTAG 2258530
       |||||||||||||||||||||||||||||||||||||||++           65 bp         ++||||||||||||||||||||||||||
Avr5 : CCAGATGTTGCTCAAAAATAGGCTCGCCCACGTTTTATGgt.........................agGAGTGCGATGCCCGTATCCGTACTAG 312
```

Figure 2.S10: Location of *Avr5* in *Cladosporium fulvum* Race 5. (A) The disrupted avirulence gene *Avr5* is located on chromosome Chr1, next to long, repeat-rich intergenic regions. Repeats are represented as rectangles below the baseline, and genes are indicated with rectangles above the base line, with arrows indicating their transcriptional orientation. (B) Coding sequence alignment of the disrupted *Avr5* from *C. fulvum* Race 5 and *Avr5* from *C. fulvum* isolate 0WU (KJ452245.1).

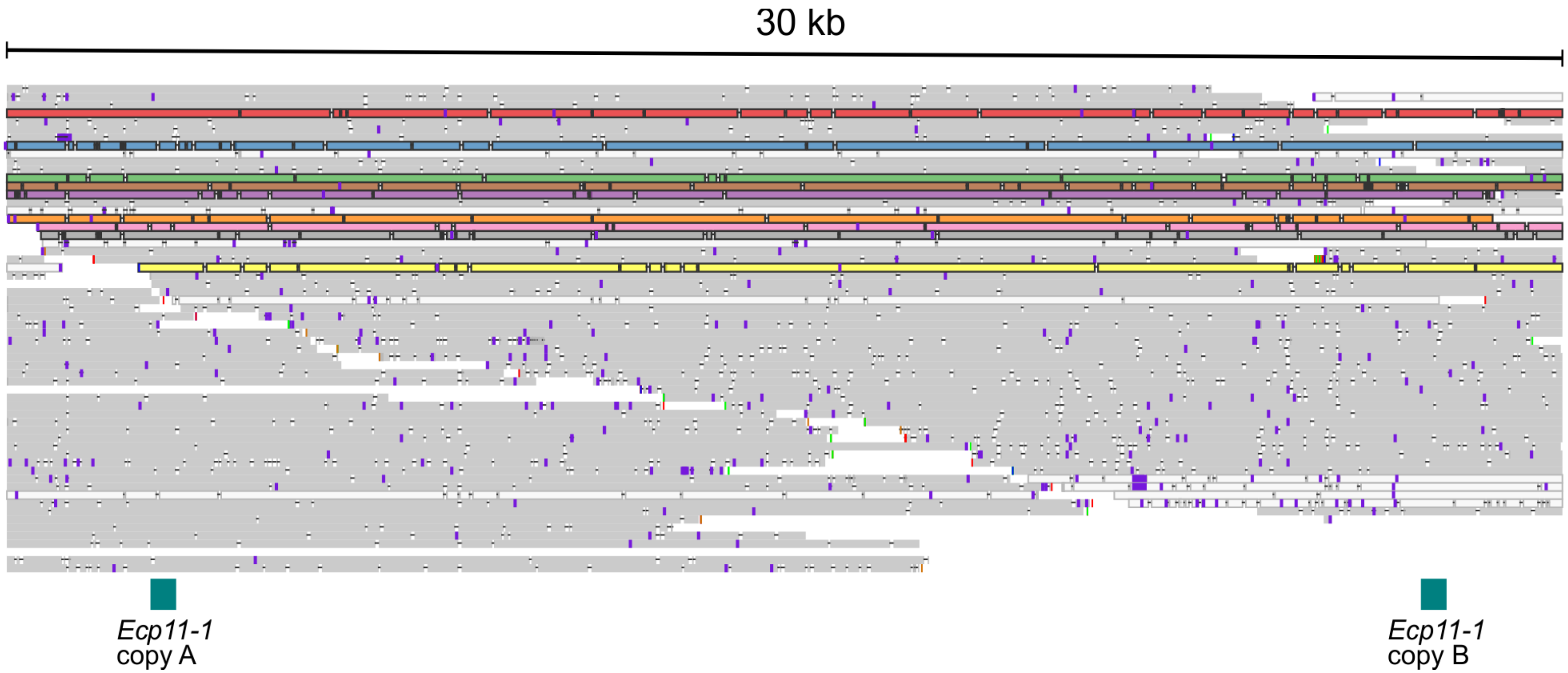


Figure 2.S11: The region in chromosome 5 of *Cladosporium fulvum* Race 5 containing two identical copies of the candidate effector gene *Ecp11-1*. The two gene copies (copy A and copy B) are separated by an intergenic region of 24,419 bp and both have the same transcriptional orientation. The figure shows PacBio reads mapped to the locus. Coverage ranges from 45x to 59x. The nine long reads highligh ted span the entire sequences of both *Ecp11-1* copies. These nine reads were uniquely mapped to the genome of *C. fulvum* race 5 (mapping quality = 60). The figure was visualized and exported with the Integrative Genomics Viewer (IGV) v2.6.1.

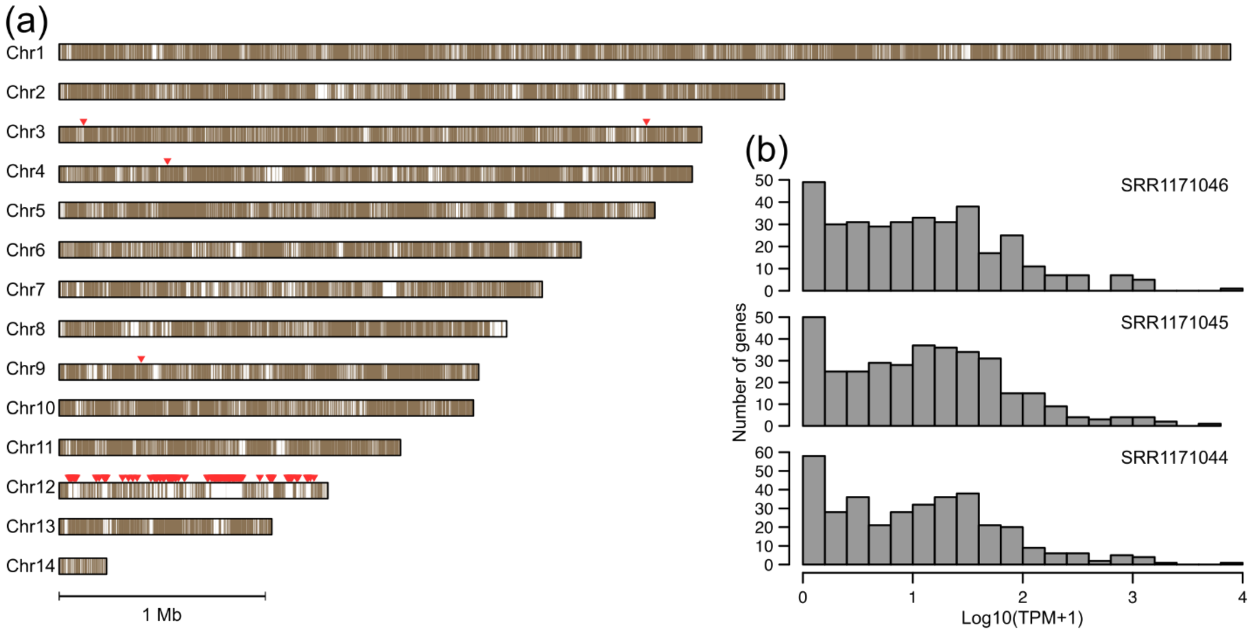


Figure 2.S12: Comparison of the genome assemblies of *Cladosporium fulvum* isolates Race 5 and 0WU. (A) The 14 assembled chromosomes of *C. fulvum* Race 5 are shown with filled regions indicating regions covered by *C. fulvum* 0WU contigs after mapping them with minimap2. Blank regions correspond to missing portions of the chromosomes from isolate Race 5 in the assembly of isolate 0WU. Triangles indicate 352 genes present in the assembly of *C. fulvum* Race 5 but missing in the assembly of *C. fulvum* 0WU. The figure shows that nearly half of Chr12 is not present in the assembly of *C. fulvum* 0WU and that 348 out of these 352 missing genes are in this chromosome. (B) Histogram showing the expression values in transcripts per million (TPM) of the 352 genes missing in the assembly of *C. fulvum* 0WU, based on three different RNA-seq data sets corresponding to three different conditions of isolate 0WU grown *in vitro* (SRR1171046, SRR1171045, and SRR1171044). The figure shows that most of the genes that are absent in the assembly of *C. fulvum* 0WU have clear evidence of expression based on the RNA-seq data of this isolate, indicating that these genes were misassembled in the previous reference genome of *C. fulvum* 0WU.

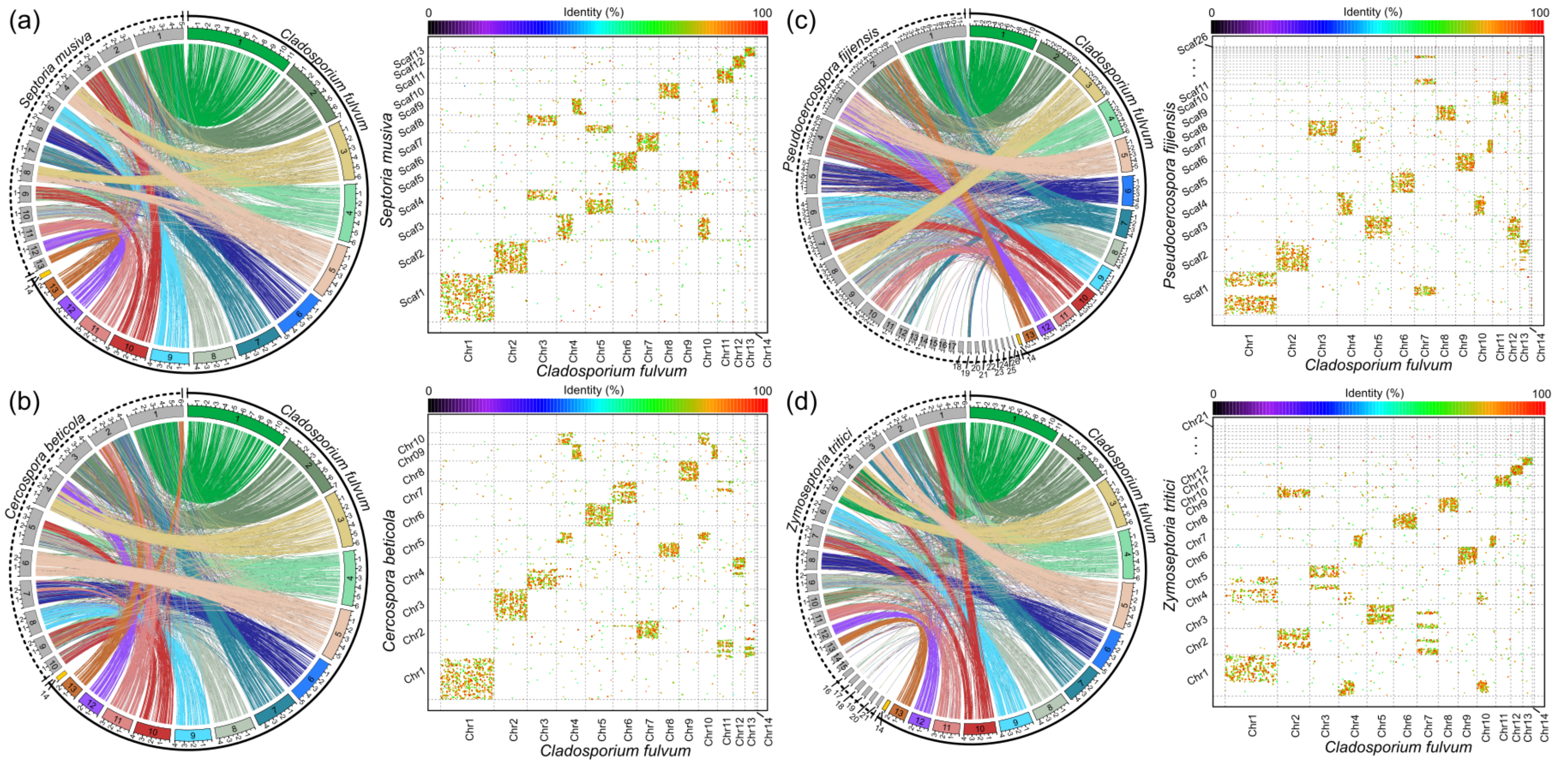


Figure 2.S13: Ribbon plots and dot plots showing the patterns of mesosynteny observed between *Cladosporium fulvum* Race 5 and other plant pathogenic species of Capnodiales. Genomes were aligned at the amino acid level with PROmer. Before alignment, scaffolds shorter than 100 kb were removed, and the remaining scaffolds were sorted by size in decreasing order. The 14 chromosomes of *C. fulvum* Race 5 are shown on the x-axis, and the chromosomes or scaffolds of the other spescies are shown on the y-axis. RefSeq accession numbers of the genomes of *Septoria musiva* isolate SO2202, *Cercospora beticola* isolate 09-40, *Pseudocercospora fijiensis* isolate CIRAD86, and *Zymoseptoria tritici* isolate IPO323 used for the alignments are GCF_000320565.1, GCF_002742065.1, GCF_000340215.1, and GCF_000219625.1, respectively.

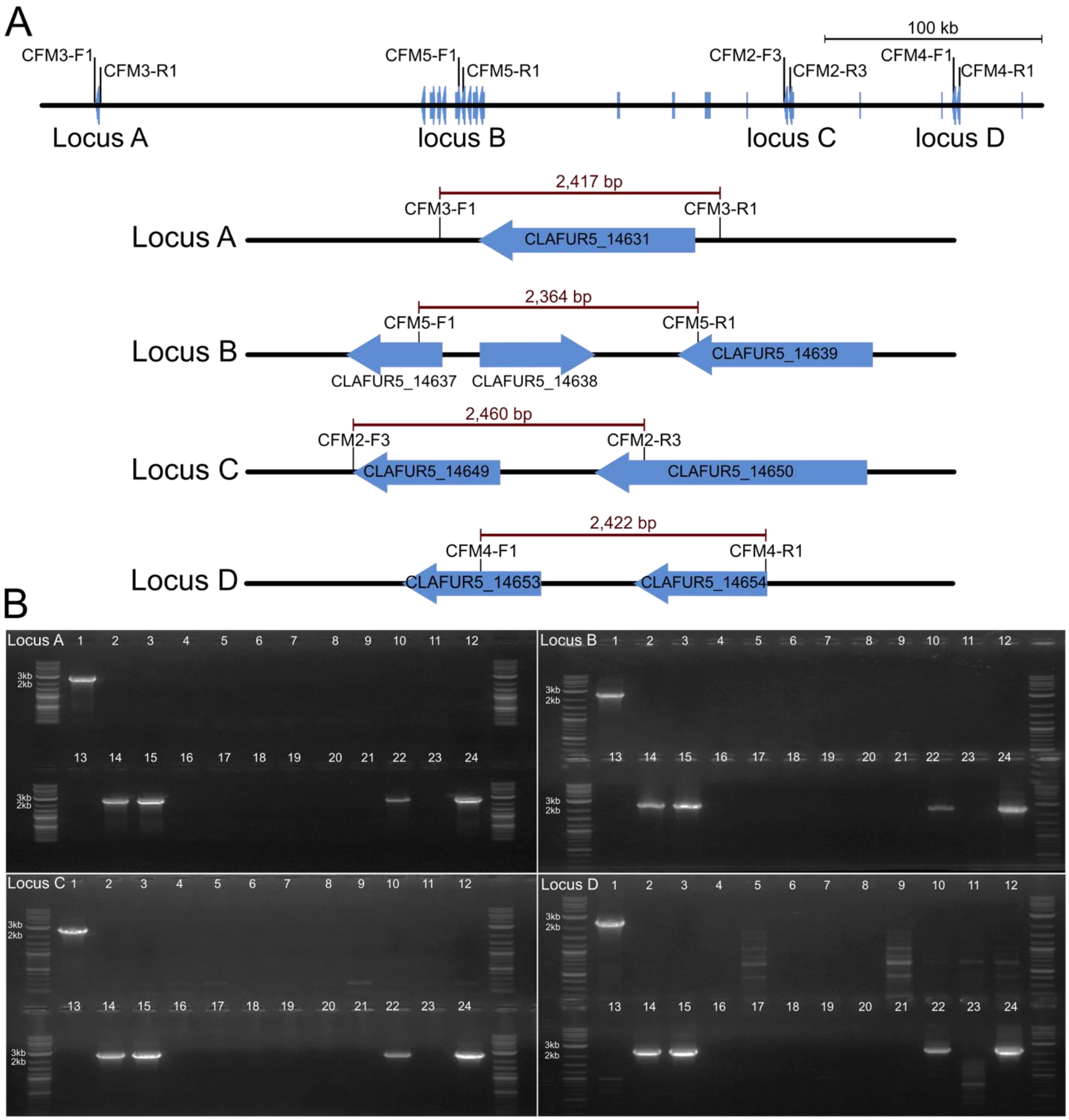


Figure 2.S14: Presence/absence variation of the mini-chromosome 14 (Chr14) in 24 isolates of *Cladosporium fulvum*. (A) Location of the primers used to determine the presence or absence of Chr14 among the 24 *C. fulvum* isolates examined in this study (Table 2.4). The figure shows an overview of the entire Chr14 as well as zoomed-in regions where primers are located. The regions (i.e. loci) were named A, B, C, and D. Predicted genes are represented as arrows. Primer sequences are shown in Table 2.S17. (B) Gel electrophoresis analysis of polymerase chain reaction (PCR) amplicons obtained using the primer combinations in regions a, b, c, and d. The figure shows that only five isolates, i.e. isolates 1, 14, 15, 22, and 14, had the chromosome Chr14. Ladder is a 1kb Plus (New England Biolabs).

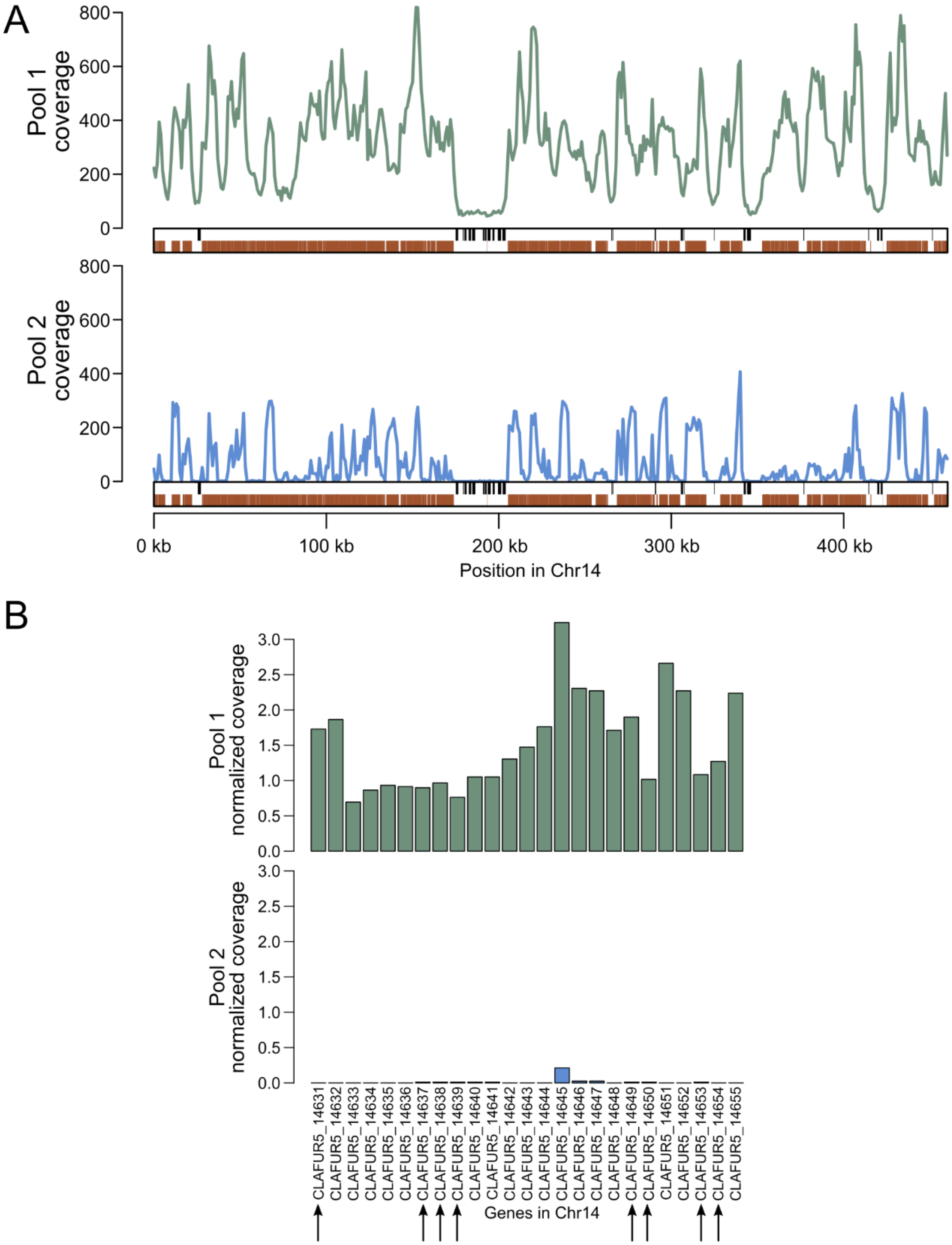


Figure 2.S15: Pooled whole-genome sequencing confirms that the mini-chromosome Chr14 of *Cladosporium fulvum* Race 5 is dispensable. *Cladosporium fulvum* isolates 2, IMI Argent 358077, and Turk 1a, for which Chr14 was predicted to be present, were grouped in pool 1, whereas isolates IPO 2.4.8.9.11 Polen, IPO 249 France, and 2.5, for which Chr14 was predicted to be absent, were pooled in pool 2. (A) Read depth across chromosome Chr14 for both pools. Predicted genes are represented with black rectangles and repetitive regions are represented with brown rectangles. (B) Bar plot showing the median

coverage of all 25 genes predicted in chromosome Chr14 for pool 1 and pool 2. Coverage was normalized by the median coverage of 100 BUSCO genes (59x for pool1 and 80x for pool2). Arrows indicate PCR-amplified genes used verify the presence of Chr14 in a collection of 24 *C. fulvum* isolates (Fig 2.S14). The figure shows that all predicted genes in Chr14 exhibits practically no coverage for the isolates in pool 2. The only exception was the gene CLAFUR5_14645, which is duplicated in *C. fulvum* Race 5 with two identical copies, one in Chr14 and the other in Chr1. In contrast, all 25 genes in Chr14 exhibited high coverage levels for isolates in pool 1.

2.7.2 Supplementary tables

Table 2.S1: Contigs assembled with Canu using sequenced PacBio reads of *Cladosporium fulvum* Race 5. Assembled bacterial contigs contained high GC content, and were discarded from the final assembly. Other smaller contigs were also discarded, as they matched the mitochondrial genome of *C. fulvum*, were contained within other contigs, or were assembled from a single PacBio read. Discarded contigs are highlighted. This table is shown in the next page.

Contig ID	Size (bp)	GC content (%)	No. reads used to form contig	Note
tig00000003	11359455	48.85	17616	Keep
tig00000006	7034911	49.18	10959	Keep
tig00004032	6232441	47.22	9546	Keep
tig00000008	6140979	49.91	9980	Keep
tig00004031	6025800	66.66	3775	Bacterial, discard
tig00000001	5821070	65.96	8323	Bacterial, discard
tig00004034	5761526	49.22	8865	Keep
tig00000014	5061379	48.88	7900	Keep
tig00000019	4686178	48.33	7499	Keep
tig00000011	4339671	48.48	6773	Keep
tig00000022	4070318	49.83	6505	Keep
tig00000025	4017557	48.77	6157	Keep
tig00000028	3310984	49.59	5283	Keep
tig00000031	2606401	50.21	4103	Keep
tig00000033	2062990	50.12	3302	Keep
tig00000035	647739	62.07	1106	Bacterial, discard
tig00000037	459899	45.64	609	Keep
tig00000039	179077	29.55	1706	Mitochondrial genome, discard
tig00000017	66995	52.96	23	Similar to 28S rDNA. Contained within tig00000008 (Chr4), discard
tig00004033	60865	48.38	1	Assembled from single read, discard
tig00000024	58531	45.46	56	Contained within tig00000022, discard
tig00004035	45736	53.24	1	Assembled from single read, discard
tig00004030	38108	62.07	1	Bacterial, discard
tig00000042	36804	66.6	5	Bacterial, discard
tig00000120	32778	68.32	5	Bacterial, discard
tig00000044	32603	67.42	6	Bacterial, discard
tig00000114	31263	66.67	4	Bacterial, discard
tig00000041	30563	60.15	6	Bacterial, discard

tig00000077	23405	69.99	3	Bacterial, discard
tig00000052	20732	69.7	3	Bacterial, discard
tig00000070	20447	62.2	7	Bacterial, discard
tig00000049	20427	68.62	3	Bacterial, discard
tig00000081	19825	66.75	5	Bacterial, discard
tig00000053	17317	68.73	3	Bacterial, discard
tig00000061	15563	68.14	8	Bacterial, discard
tig00000091	12012	67.99	6	Bacterial, discard
tig00000098	10415	69.15	4	Bacterial, discard
tig00000074	10057	69.27	7	Bacterial, discard
tig00000111	9648	68.36	4	Bacterial, discard
tig00000069	9610	64.08	5	Bacterial, discard
tig00000102	7803	68.54	3	Bacterial, discard
tig00000116	7762	65.43	4	Bacterial, discard
tig00000092	7288	65	5	Bacterial, discard

Table 2.S2: Summary of estimated abundance of transposable elements in the genome of *Cladosporium fulvum* Race 5. Custom repeat libraries were generated with RepeatModeler v1 and v2 and then used to mask the genome with RepeatMasker. The output of RepeatMasker was parsed with the script parseRM.pl (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs), which counted the number of masked bases as reported in this table. In order to more accurately estimate the percentage of repeats of each class or family of transposable elements, bases masked twice (i.e. overlapping repeats) were not considered.

	RepeatModeler v1.0.11			RepeatModeler v2.0.2		
Total repeats						
nt_total_in_genome	**nt_masked-minus-double**	**%_masked**	**nt_masked_double**	**nt_masked-minus-double**	**%_masked**	**nt_masked_double**
67169167	33429091	49.76850614	2569413	33452881	49.80392417	3140188

	RepeatModeler v1.0.11			RepeatModeler v2.0.2		
Repeats by class						
class	**nt_masked-minus-double**	**%_masked**	**nt_masked_double**	**nt_masked-minus-double**	**%_masked**	**nt_masked_double**
LINE	10952603	16.30599796	553321	10326061	15.37321581	490953
RC	9963	0.014832698	3	0	0	0
Unknown	4725911	7.035833867	442302	4172845	6.212441193	256728
DNA	1196546	1.781391751	32951	1194814	1.778813187	83929
LTR	16544068	24.63044986	1328276	17759161	26.43945398	1674038

		RepeatModeler v1.0.11			RepeatModeler v2.0.2		
Repeats by family							
class	**family**	**nt_masked-minus-double**	**%_masked**	**nt_masked_double**	**nt_masked-minus-double**	**%_masked**	**nt_masked_double**
DNA	MULE-MuDR	32905	0.048988251	2552	33315	0.04959865	4721
DNA	TcMar-Tc1	89536	0.133299256	3024	93451	0.139127823	2083
RC	Helitron	9963	0.014832698	3	0	0	0
DNA	CMC-EnSpm	20398	0.0303681	1634	0	0	0
LINE	Tad1	10952603	16.30599796	553321	10326061	15.37321581	490953
DNA	DNA	32747	0.048753024	2458	3687	0.005489126	37
LTR	Gypsy	10340053	15.3940468	888644	12200352	18.16361963	1006993
Unknown	Unknown	4725911	7.035833867	442302	4172845	6.212441193	256728
DNA	TcMar-Fot1	926003	1.378613196	20998	930240	1.38492115	76423
LTR	Copia	6204015	9.236403066	439632	4676697	6.962565131	597784
DNA	hAT-Restless	94957	0.141369923	2285	134121	0.199676438	665
LTR	Unknown	0	0	0	882112	1.313269227	69261

Table 2.S3: Statistics of Repeat-Induced Point (RIP) mutations in the chromosomes of *Cladosporium fulvum* Race 5. The chromosomes were analyzed using a 1 kb sliding window and a step size of 500 bp. Windows were considered RIPped if the substrate index value (CpA + TpG)/(ApC + GpT) was ≤ 0.75, product index value (TpA/ApT) was ≥ 1.1, and composite index value (TpA/ApT) – ([CpA + TpG]/[ApC + GpT]) was ≥ 0.01. Single-copy windows were considered as RIPed windows when they had no secondary BLASTn hit against the *C. fulvum* Race 5 genome (e-value < 1E-20, identity > 50%, and query coverage > 20%). The last three columns show (i) the estimated percentage of RIPped regions determined as the total percentage to RIPped windows, (ii) the estimated percentage of repetitive regions RIPped determined as the percentage of repeat-masked bases that overlap RIPped windows, and (iii) the estimated percentage of single-copy regions RIPped determined as the percentage of unmasked bases within single-copy RIPped windows.

Chromosome	Total windows	Ripped windows	Base pairs in ripped windows	Single-copy windows ripped	Base pairs in single copy ripped windows	Masked bases covered by ripped windows	Perc. total ripped	Perc. repeat ripped	Perc. single-copy ripped
Chr1	22725	9593	4988790	17	16500	4866644	42.21342	81.603	0.3056413
Chr2	14072	5393	2811000	28	22500	2724172	38.32433	79.63401	0.6223982
Chr3	12466	6609	3426000	40	32500	3279782	53.0162	85.35406	1.3596597
Chr4	12283	3923	2074000	16	12000	2001862	31.93845	73.14614	0.3524736
Chr5	11555	4225	2213500	15	11500	2142544	36.56426	74.43739	0.3966683
Chr6	10124	4004	2092772	31	23272	2002385	39.54959	78.75868	0.9237336
Chr7	9374	3995	2097295	42	31000	1992979	42.61788	78.47044	1.4438671
Chr8	8682	3697	1931606	7	6500	1864952	42.58235	79.66476	0.325064
Chr9	8141	2499	1310492	3	2500	1286350	30.69647	74.47853	0.1066849
Chr10	8036	3319	1726737	44	32737	1656674	41.30164	83.60487	1.6077616
Chr11	6623	2150	1126222	17	12722	1085641	32.46263	80.15175	0.650163
Chr12	5214	1491	785583	5	4500	752252	28.59609	75.38774	0.2797222
Chr13	4127	1157	605147	7	5500	584590	28.03489	78.61979	0.416799
Chr14	921	649	339779	34	24500	296932	70.46688	82.77081	24.1286599

Table 2.S4: Genes in the genome of Cladosporium fulvum Race 5 encoding key enzymes for secondary metabolism. These key enzymes are classified into non-ribosomal peptide synthetases (NRPS), type 1 polyketide synthases (T1PKS), and terpene synthases (Terpene). NRPS-like represent fragments of NRPS genes.

Gene	SM key enzyme type	Chromosome	Start	End
CLAFUR5_03847	NRPS	Chr2	6813405	6815531
CLAFUR5_04416	NRPS	Chr4	1954012	1961643
CLAFUR5_07436	NRPS	Chr6	3939882	3948075
CLAFUR5_10739	NRPS	Chr7	4487178	4501969
CLAFUR5_10777	NRPS	Chr8	97553	102082
CLAFUR5_11006	NRPS	Chr8	1189252	1194770
CLAFUR5_11318	NRPS	Chr8	2781270	2784524
CLAFUR5_08791	NRPS	Chr9	15649	30345
CLAFUR5_09052	NRPS	Chr9	1044867	1048418
CLAFUR5_12064	NRPS	Chr10	1822777	1823751
CLAFUR5_12662	NRPS	Chr11	487132	490677
CLAFUR5_12895	NRPS	Chr11	1291719	1300463
CLAFUR5_12913	NRPS	Chr11	1475292	1489509
CLAFUR5_13278	NRPS	Chr11	3023997	3029972
CLAFUR5_13418	NRPS	Chr12	140493	146903
CLAFUR5_02708	NRPS-like	Chr2	1536435	1539600
CLAFUR5_04428	NRPS-like	Chr4	2177494	2182757
CLAFUR5_04894	NRPS-like	Chr4	4293503	4296608
CLAFUR5_07484	NRPS-like	Chr6	4109767	4112823
CLAFUR5_09844	NRPS-like	Chr7	138507	142073
CLAFUR5_09474	NRPS-like	Chr9	3037671	3040749
CLAFUR5_12787	NRPS-like	Chr11	788002	791250
CLAFUR5_13164	NRPS-like	Chr11	2658682	2661623
CLAFUR5_13788	NRPS-like	Chr12	1618731	1621862
CLAFUR5_13962	NRPS-like	Chr12	2241276	2244338
CLAFUR5_14350	NRPS-like	Chr13	852968	856885
CLAFUR5_06232	PKS-NRPS hybrid	Chr5	3898442	3910634
CLAFUR5_00278	T1PKS	Chr1	1252087	1256478
CLAFUR5_06161	T1PKS	Chr5	3500997	3508032
CLAFUR5_10765	T1PKS	Chr8	62927	71161
CLAFUR5_10780	T1PKS	Chr8	105680	112961
CLAFUR5_11407	T1PKS	Chr8	3064219	3071038
CLAFUR5_12824	T1PKS	Chr11	1138506	1145063
CLAFUR5_12905	T1PKS	Chr11	1407056	1412575
CLAFUR5_13273	T1PKS	Chr11	2980353	2988770
CLAFUR5_13909	T1PKS	Chr12	2084657	2090304
CLAFUR5_13961	T1PKS	Chr12	2234102	2240730
CLAFUR5_14164	T1PKS	Chr13	251167	258429
CLAFUR5_00877	Terpene	Chr1	4307265	4308887
CLAFUR5_02085	Terpene	Chr1	10411834	10413359
CLAFUR5_10460	Terpene	Chr7	3134219	3134988
CLAFUR5_14251	Terpene	Chr13	569930	571917

Table 2.S5: Genes in the genome of Cladosporium fulvum Race 5 encoding predicted carbohydrate-active enzymes (CAZymes). CAZymes are classified into six major classes, i.e., auxiliary activity (AA), carbohydrate-binding module (CBM), carbohydrate esterase (CE), glycoside hydrolase (GH), glycosyltransferase (GT), and polysaccharide lyase (PL). Predicted secreted proteins are indicated in the third column. This table is available at https://zenodo.org/records/11211529.

Table 2.S6: Genes in the genome of Cladosporium fulvum Race 5 encoding predicted proteases. Proteases are classified into aspartic (A), cysteine (C), metallo (M), asparagine (N), serine (S), threonine (T), and inhibitory (I) proteases. Proteases were identified based on BLASTp searches (e-value < 1E-10) against the MEROPS database. Predicted secreted proteins are indicated in the third column. This table is available at https://zenodo.org/records/11211529.

Table 2.S7: Genes in the genome of Cladosporium fulvum Race 5 encoding putative transporters. Transporters were identified based on BLASTp searches (e-value < 1E-10) against the Transporter Classification Database (TCDB). This table is available at https://zenodo.org/records/11211529.

Table 2.S8: Secreted proteins and candidate effectors in the genome of Cladosporium fulvum Race 5. The table shows all proteins containing a predicted signal peptide (SP), their size (amino acids), prediction of EffectorP, number of transmembrane (TM) domains in the mature protein, predicted GPI-anchor (PFrate ≤ 0.005 means probable GPI-anchor), number and percentage of cysteines residues, name of homologous candidate effector previously described, and classification of the proteins into predicted effector (SSP) or non-effector (noSSP). This table is available at https://zenodo.org/records/11211529.

Table 2.S9: Genes located in subtelomeric regions in the genome of *Cladosporium fulvum* Race 5. This table is available at https://zenodo.org/records/11211529.

Table 2.S10: Over- and under-representation of different gene categories in subtelomeric regions in the genome of *Cladosporium fulvum* Race 5. Gene densities (count per Mb) for the whole genome and within subtelomeric regions (i.e., within 25 kb of telomeric repeats) are shown. There was a total of 143 genes presented within subtelomeric regions. The columns 4 to 7 show the number of genes from the specific category located in subtelomeric regions, number of genes in subtelomeric regions that do not belong to the specific gene category, total number of genes from the specific category in the genome, and total number of genes in the genome that do not belong to the specific category. These numbers were used in the phyper function within R to perform hypergeometric tests for over- and under-representation. Resulting p-values are shown in the last columns.

Gene category	Density for the whole genome (count per Mb)	Density in subtelomeric regions (count per Mb)	Number in subtelomeric regions from category	Number in subtelomeric regions not from category	Total number of genes from category	Total number of genes not from category	Under-representation p-value	Over-representation p-value
BUSCOs	55.68	22.86	16	127	3740	10950	1.70E-05	0.9999941
Candidate effectors	5.14	14.29	10	133	345	14354	0.9994443	0.002016587
Secreted, not candidate effectors	14.50	20.00	14	129	974	13716	0.9479977	0.09227297
ABC transporters	0.85	0.00	0	143	57	14633	0.5719771	1
CAZymes	7.27	2.86	2	141	488	14202	0.1415134	0.953552
MFS transporters	5.69	1.43	1	142	382	14308	0.1101271	0.9773221
Secondary metabolism enzymes	0.63	1.43	1	142	42	14648	0.9370843	0.337298
Proteases	5.34	4.29	3	140	359	14331	0.5359565	0.6829426

Table 2.S11: Summary of gene clustering in the genome of *Cladosporium fulvum* Race 5. Genes were clustered based on different maximum threshold distances of 1 to 20 kb. For each threshold distance, the table shows the total number of gene clusters and the number of clusters with only one gene or more than one gene. Other fields present average numbers for the clusters identified, including average cluster size, average number of genes, and average size and repeat content of intergenic regions inside out outside clusters.

Threshold distance (kb)	No. of clusters	No. of single-gene clusters	No. of multi-gene clusters	Mean cluster size (bp)	No. of genes in cluster	Mean repeat content in intergenic regions (%)		Mean size of intergenic regions (bp)	
						Outside clusters	Inside clusters	Outside clusters	Inside clusters
1	4726	1689	3037	5248.6	3.1	15.1	1.1	8937.5	461.6
2	1872	362	1510	15369.9	7.8	33.0	1.6	20358.4	668.6
3	1066	159	907	28810.0	13.8	53.2	1.9	33756.5	771.5
4	756	102	654	42028.5	19.4	70.4	2.1	45967.5	830.6
5	649	83	566	49681.1	22.6	80.0	2.1	52677.9	857.7
6	593	73	520	54884.8	24.8	84.9	2.2	57037.6	875.9
7	555	64	491	59082.4	26.5	88.1	2.3	60417.8	890.8
8	531	61	470	62094.0	27.7	89.7	2.4	62746.0	902.1
9	518	59	459	63864.8	28.4	90.9	2.4	64072.4	909.1
10	509	58	451	65159.3	28.9	90.9	2.5	65014.1	914.4
11	500	54	446	66522.8	29.4	91.5	2.5	65967.1	920.6
12	495	52	443	67310.3	29.7	91.5	2.6	66502.7	924.3
13	491	52	439	67957.7	29.9	91.6	2.6	66933.1	927.4
14	485	50	435	68964.9	30.3	91.8	2.6	67576.2	932.7
15	477	46	431	70363.8	30.8	92.0	2.7	68441.8	940.4
16	473	44	429	71089.9	31.1	92.1	2.7	68876.7	944.5
17	466	42	424	72404.9	31.5	92.2	2.7	69641.2	952.1
18	458	41	417	73973.2	32.1	92.5	2.8	70527.1	961.3
19	456	41	415	74376.6	32.2	92.6	2.8	70750.5	963.7
20	449	39	410	75839.7	32.7	92.6	2.8	71525.8	972.8

Table 2.S12: Predicted genes in the genome of *Cladosporium fulvum* Race 5 that overlap with masked repetitive regions. The table shows genes that overlap more than 25% of their sequences with regions of the genome masked with RepeatMasker based on repeat libraries produced with RepeatModeler v1.0.11 (RM1) and RepeatModeler v2.0.2 (RM2). The percentages of the gene regions that overlap with masked regions are shown, as well as conserved PFAM domains found in the genes. The table shows that the genes that overlap with masked regions typically contain conserved domains commonly found in transposable elements. This table is available at https://zenodo.org/records/11211529.

Table 2.S13: Statistical analysis to test enrichment of specific gene categories within gene-sparse regions of the genome of *Cladosporium fulvum* Race 5. A total of 990 genes with up- or downstream intergenic size of at least 8 kb were considered in genes-sparse regions. The columns 2 to 5 show the number of genes from the specific category located in gene-sparse regions, number of genes in gene-sparse regions that do not belong to the specific gene category, total number of genes from the specific category in the genome, and total number of genes in the genome that do not belong to the specific category. These numbers were used in the phyper function within R to perform a hypergeometric t est. Resulting p-values are shown in the last column.

Gene category	Number in gene-sparse regions from category	Number in gene-sparse regions not from category	Total number of genes from category	Total number of genes not from category	P-value
Candidate effectors	66	924	345	14345	6.59E-15
Secreted proteins	142	848	1320	13370	7.61E-09
Secreted, not candidate effectors	76	914	974	13716	0.09788024
CAZymes	35	955	488	14202	0.3750396
Secreted CAZymes	20	970	229	14461	0.1407404
Proteases	20	970	359	14331	0.8417786
Secreted proteases	9	981	77	14613	0.0734955
Key enzymes for secondary metabolism	7	983	42	14648	0.0209943
Transporters	155	835	2287	12403	0.4829155
ABC transporters	7	983	57	14633	0.08714286
MFS transporters	30	960	382	14308	0.2152802
BUSCOs	187	803	3740	10950	0.9999998

Table 2.S14: Putative recently duplicated genes in *Cladosporium fulvum* Race 5. Gene clusters identified with cd-hit-test by grouping coding sequences with at least 90% nucleotide identity. Representative sequences of the clusters are indicated with an asterisk. This table is available at https://zenodo.org/records/11211529.

Table 2.S15: Genes from the genome of *Cladosporium fulvum* Race 5 and their corresponding ortholog in the genome of *C. fulvum* strain 0WU (JGI). This table is available at https://zenodo.org/records/11211529.

Table 2.S16: Genes present in the genome of *Cladosporium fulvum* Race 5 but missing the assembly of *C. fulvum* isolate 0WU. Location of the genes in the genome and functional descriptions are shown. Genes encoding universal single-copy orthologs conserved in species of Capnodiales are indicated with the respective BUSCO ID. Genes encoding putative transporters are indicated with the respective transporter family. Genes encoding Carbohydrate-active enzymes (CAZymes) and proteases are indicated with the respective CAZyme and protease family. Genes encoding secreted proteins or candidate effectors are also shown. Expression values in transcripts per million are shown in the last three columns for three different data sets of *C. fulvum* 0WU grown in vitro. Most missing genes have evidence of expression, indicating that they were misassembled in the genome of isolate 0WU. This table is available at https://zenodo.org/records/11211529.

Table 2.S17: Primers designed to capture genes located in the dispensable chromosome 14 (Chr14) of *Cladosporium fulvum* Race 5. A map with the location of these primers is shown in Fig 2.S14A.

Primer	Sequence (5'->3')	Region	Region start in Chr14	Region end in Chr14	Region size
CFM3-F1	TGGATACTGCACAGCATCTGTC	Locus A	25119	27535	2417
CFM3-R1	GTAATGCGGCATGTTGTCTCTTG				
CFM5-F1	AGTTGGTGGAAGAGACGACAAG	Locus B	191562	193925	2364
CFM5-R1	AGCAGGGAATTGTTGGAGCAG				
CFM2-F3	AAGACCTAGTAGACCGAGGATC	Locus C	342165	344624	2460
CFM2-R3	AGCAGTGTGTACCAACTTCGAG				
CFM4-F1	TATCACAAGGCGAGACTCGAAC	Locus D	420358	422779	2422
CFM4-R1	AACATGACCCACTACCTCACTC				

Chapter 3

Analysis of five near-complete genome assemblies of the tomato pathogen *Cladosporium fulvum* uncovers additional accessory chromosomes and structural variations induced by transposable elements effecting the loss of avirulence genes

Alex Z. Zaccaron
Ioannis Stergiopoulos

Author contributions

Conceptualization: AZ, IS; Formal analyses: AZ; Investigation: AZ; Methodology: AZ; Project administration: IS; Software: AZ; Visualization: AZ; Writing – original draft: AZ; Writing – review & editing: AZ, IS.

This chapter was published in: *BMC Biology*, January 2024, volume 22, number 1 (https://doi.org/10.1186/s12915-024-01818-z).

Abstract

Fungal plant pathogens have dynamic genomes that allow them to rapidly adapt to adverse conditions and overcome host resistance. One way by which this dynamic genome plasticity is expressed is through effector gene loss, which enables plant pathogens to overcome recognition by cognate resistance genes in the host. However, the exact nature of these loses remains elusive in many fungi. This includes the tomato pathogen *Cladosporium fulvum*, which is the first fungal plant pathogen from which avirulence (*Avr*) genes were ever cloned and in which loss of *Avr* genes is often reported as a means of overcoming recognition by cognate tomato *Cf* resistance genes. A recent near-complete reference genome assembly of *C. fulvum* isolate Race 5 revealed a compartmentalized genome architecture and the presence of an accessory chromosome, thereby creating a basis for studying genome plasticity in fungal plant pathogens and its impact on avirulence genes. Here, we obtained near-complete genome assemblies of four additional *C. fulvum* isolates. The genome assemblies had similar sizes (66.96 Mb to 67.78 Mb), number of predicted genes (14,895 to 14,981), and estimated completeness (98.8% to 98.9%). Comparative analysis that included the genome of isolate Race 5 revealed high levels of synteny and colinearity, which extended to the density and distribution of repetitive elements and of repeat-induced point (RIP) mutations across homologous chromosomes. Nonetheless, structural variations, likely mediated by transposable elements and effecting the deletion of the avirulence genes *Avr4E*, *Avr5*, and *Avr9,* were also identified. The isolates further shared a core set of 13 chromosomes, but two accessory chromosomes were identified as well. Accessory chromosomes were significantly smaller in size, and one carried pseudogenized copies of two effector genes. Whole-genome alignments further revealed genomic islands of near-zero nucleotide diversity interspersed with islands of high nucleotide diversity that co-localized with repeat-rich regions. These regions were likely generated by RIP, which generally asymmetrically affected the genome of *C. fulvum*. Our results reveal new evolutionary aspects of the *C. fulvum* genome and provide new insights on the importance of genomic structural variations in overcoming host resistance in fungal plant pathogens.

3.1 Introduction

Fungal plant pathogens have a remarkable capacity to evolve rapidly in order to adapt to adverse conditions and overcome host resistance, which poses challenges to the establishment of sustainable strategies for crop protection. The rapid adaptation of fungal pathogens to unfavorable environments is predominantly orchestrated by their genome plasticity, including changes in their genome size, organization, and chromosome number (Möller and Stukenbrock 2017). Genome plasticity is in terms facilitated by the proliferation of transposable elements (TEs), which can comprise up to 90% of the genomic content in some fungal plant pathogens (Gupta et al. 2023; Tobias et al. 2021; Mat Razali et al. 2019; Seidl and Thomma 2017). The presence or mobilization of TEs is often further associated with adaptive genomic changes, such as gene deletion (Chuma et al. 2011), gene duplication (Pedersen et al. 2012), and horizontal gene transfer (McDonald et al. 2019) that can accelerate genome evolution and create opportunities for overcoming stressful environments. TEs may also trigger single nucleotide polymorphisms (SNPs) through repeat-induced point (RIP) mutations. RIP is a premeiotic mechanism present in fungal genomes that acts in defense against the deleterious effects of TE proliferation (Clutterbuck 2011; Selker 1990, 2002). RIP induces transition nucleotide substitutions (C-to-T or the complement G-to-A) in duplicated genomic sequences, with a strong bias toward CpA-to-TpA (or the complement TpG-to-TpA) dinucleotides (Hane and Oliver 2008; Selker 1990). In fungi, RIP can occur at every cycle of sexual reproduction, thus resulting in high rates of mutation that can be from the highest among non-viral organisms (Wang et al. 2020).

Although TEs can, in principle, proliferate almost randomly in a genome, they are often unevenly distributed within fungal genomes, thereby resulting in genomic regions with clustered TEs (Faino et al. 2016; Muszewska et al. 2019; Torres et al. 2021; Zaccaron et al. 2022). This architecture is likely instigated by purifying selection against the deleterious effects of TE insertion into more sensitive regions of the genome, such as gene-dense regions (Torres et al. 2020). The uneven distribution of TEs compartmentalizes fungal genomes into a bipartite architecture composed of TE-rich, gene-sparse regions and TE-poor, gene-rich regions (Dong et al. 2015; Frantzeskakis et al. 2019; Raffaele and Kamoun 2012). This bipartite genome architecture is referred to as the "two-speed genome" model of evolution (Croll and McDonald 2012; Raffaele et al. 2010) and is often encountered in plant pathogens, as it enables them to overcome host immunity through the fast evolution of their effector and other pathogenicity-related genes (Wacker et al. 2023; Wang et al. 2017).

Another component of fungal genomes that contributes to their plasticity and compartmentalization is accessory chromosomes. Also known as "dispensable" or "B chromosomes", accessory chromosomes are richer in TEs than core chromosomes and are present in only some individuals of a species (Bertazzoni et al. 2018). Accessory chromosomes are also typically small (< 2 Mb), are not required for basic growth, and exhibit non-Mendelian segregation ratios (Covert 1998). In fungi, accessory chromosomes have been reported in many species (Coleman et al. 2009; He et al. 1998; Syme et al. 2018; Talbot et al. 1993; Wang et al. 2003; Zaccaron et al. 2022), including the wheat pathogen *Zymoseptoria tritici* which has eight accessory chromosomes, the largest number reported thus far for fungi (Badet et al. 2020; Goodwin et al. 2011). Even though fungal accessory chromosomes can carry virulence-associated genes (Ma et al. 2010) and genes involved in the biosynthesis of host-selective toxins (Witte et al. 2021), their function in most fungal species remains elusive (Bertazzoni et al. 2018), which makes their persistence within fungal populations intriguing.

Cladosporium fulvum (Dothideomycetes; Ascomycota; synonyms *Passalora fulva*, *Fulvia fulva*), is a fungal plant pathogen that causes tomato leaf mold (Thomma et al. 2005). While the disease is currently of notable concern only in certain regions of the world, *C. fulvum* has been used extensively as a model species for studying molecular plant-pathogen interactions (Mesarich et al. 2018; de Wit 2016). To date, at least 12 effector genes, namely *Avr2, Avr4, Avr4E, Avr5, Avr9, Ecp1, Ecp2, Ecp2-2, Ecp2-3, Ecp4, Ecp5,* and *Ecp6*, have been cloned from this pathogen and are shown to be avirulence determinants in tomato accessions with matching *Cf* resistance genes (de Wit 2016). Of these, *Avr9* was the first fungal avirulence (*Avr*) effector gene to ever be cloned from fungal plant pathogens (van Kan et al. 1991) but its intrinsic function still remains elusive. In tomato, Avr9 is recognized by the cognate Cf-9 resistance protein (Jones et al. 1994) but isolates of the fungus have emerged that can overcome Cf-9-mediated resistance through loss of *Avr9*, the only mechanism reported in isolates that overcome Cf-9 (Stergiopoulos et al. 2007a). Although complete or partial deletion of avirulence effector genes is a common strategy among plant pathogens for overcoming recognition by cognate resistance proteins (Fouché et al. 2018; Latorre et al. 2020; Plissonneau et al. 2017; Stergiopoulos et al. 2007a; Stukenbrock and McDonald 2009), the mechanisms that promote these deletions are often still unknown (Fouché et al. 2018; Plissonneau et al. 2017).

The first reference genome for *C. fulvum* isolate Race 0WU was released in 2012 (De Wit et al. 2012). However, the assembly was highly fragmented because repetitive regions were not properly assembled. Since then, efforts were made to unravel the genome organization of this pathogen (Mesarich et al. 2023), and a new chromosome-scale reference genome for *C. fulvum* isolate Race 5 was recently obtained (Zaccaron et al. 2022). The new assembly revealed many features of the *C. fulvum* genome that were hidden by the former highly fragmented assembly, including the presence of 13 core and 1 accessory chromosome, and a 'checkerboard' genome architecture composed of gene-dense and TE-poor regions interspersed with gene-sparse and TE-rich regions. It also showed that nearly 40% of the genome is affected by RIP mutations, making it one of the fungal species impacted the most by RIP, and laid the foundation to perform chromosome-scale comparative analyses (Zaccaron et al. 2022) .

In this study, we obtained near-complete genome assemblies for four additional isolates of *C. fulvum* that were collected during the 1970s, 1980s, or 1990s from the Netherlands or Poland and, together with the genome of isolate Race 5 that was collected in the Netherlands in 1979 (Zaccaron et al. 2022), performed chromosome-level comparative analyses among these five genomes. Our findings provide novel insights on the impact of repetitive DNA, RIP, and SVs on effector genes and genome evolution of a fungal plant pathogen.

3.2 Results

3.2.1 Long-read sequencing of four *C. fulvum* isolates yielded near-complete genome assemblies

Whole-genome sequencing libraries for *C. fulvum* isolates Race 0WU (Netherlands, 1997) (De Wit et al. 2012), Race 4 (Netherlands, 1971) (Boukema 1981), Race 2.4.5.9.11 IPO (Netherlands, 1980s) (Boukema 1981; Lindhout et al. 1989), and Race 2.4.9.11 (Poland, 1980s) (Lindhout et al. 1989) were multiplexed into one single SMRT cell and sequenced with the PacBio HiFi technology (Wenger et al. 2019). The SMRT cell yielded a total of 1,978,275 HiFi high-quality reads with an average size of 10,472 bp (Fig 3.S1). After demultiplexing, between 272,759 and 1,031,973 reads per isolate were obtained with an estimated genome coverage of 35x to 167x (Table 3.S1). The reads were next assembled with Canu (Koren et al. 2017) into representative genomes containing 15 to 18 contigs and ranging from 66.96 Mb to 67.78 Mb in size (Table 3.1). These genome assemblies are similar in size to the 67.17 Mb genome assembly obtained previously for isolate Race 5 (Zaccaron et al. 2022). Using the 14 chromosomes of isolate Race 5 as reference, the genome assemblies of the other four isolates could be further translated into 13 to 15 chromosomes (Fig 3.1A). Nearly all assembled chromosomes had the canonical telomeric repeat (TTAGGG)n at both ends, except of Chr13 of isolate Race 4, and Chr5 and Chr12 of isolate Race 2.4.9.11, which were missing telomeric repeats at one chromosome end. However, these three chromosomes had similar sizes compared to their homologous complete chromosomes in other isolates (Table 3.S2). The genomes of isolates Race 0WU and Race 4 had two unplaced small contigs shorter than 60 kb, whereas the assemblies of the other isolates had no unplaced contigs. Finally, closed circular contigs of 86.6 kb to 86.8 kb in size were assembled for all four isolates that represented their mitochondrial genomes (Table 3.1). Collectively, these results indicate that the four *C. fulvum* genomes obtained are nearly complete.

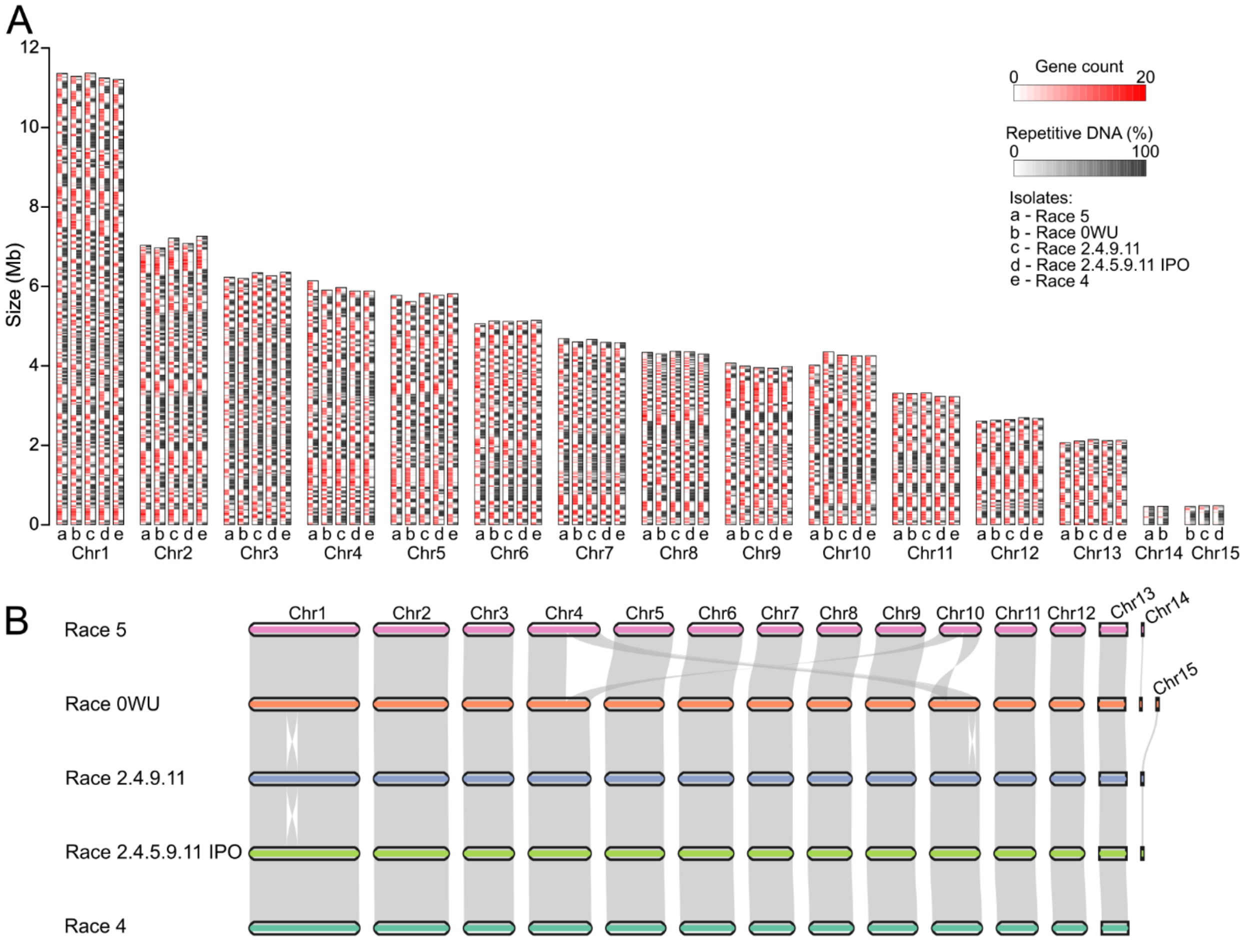


Figure 3.13: Chromosome-scale genome assemblies of five *Cladosporium fulvum* isolates. (A) Comparison of the size, gene content, and repetitive DNA content among matching chromosomes of *C. fulvum* isolates Race 5, Race 0WU, Race 2.4.9.11, Race 2.4.5.9.11 IPO, and Race 4. Matching chromosomes from different isolates are grouped and depicted as rectangles composed of two tracks representing the gene density (in red) and repetitive DNA content (in black), using a sliding window of 30 kb. (B) Pairwise synteny among the five *C. fulvum* genomes. Ribbons connect syntenic regions of the chromosomes. The figure shows a reciprocal translocation between Chr4 and Chr10 in isolate Race 5, an inversion in Chr10 in isolate Race 0WU, and another inversion in Chr1 in isolate Race 2.4.9.11.

Table 3.6: Genome assembly statistics of five *Cladosporium fulvum* isolates.

	Race 5[a]	Race 0WU	Race 4	Race 2.4.5.9.11 IPO	Race 2.4.9.11
Assembly size (bp)	67,169,167	67,489,499	66,963,592	67,158,745	67,784,504
GC (%)	48.9	48.9	49.0	48.9	48.9
Number of chromosomes	14	15	13	14	14
Number of unplaced contigs	0	2	2	0	0
Contig N50 (Mb)	5.77	5.62	5.77	5.78	5.82
Contig L50	5	5	5	5	5
mtDNA size (bp)	NA	86,642	86,835	86,771	86,731
Repetitive DNA (%)	50.25	50.15	50.00	49.92	50.47
Short tandem repeats	0.45	0.46	0.45	0.46	0.46
Dispersed repeats	49.80	49.69	49.55	49.46	50.01

[a] (Zaccaron et al. 2022).

3.2.2 The genomic landscape of repeats is conserved among the isolates of *C. fulvum*

A *de novo* annotation of repeats in the genomes of the four *C. fulvum* isolates showed that they shared a similar content in repetitive DNA, ranging from 49.9% (33.5 Mb) of the genomic content in isolate Race 2.4.5.9.11 IPO to 50.5% (34.2, Mb) in isolate Race 2.4.9.11 (Table 3.1). These values are in agreement with the 49.7% (33.4 Mb) of repetitive DNA content reported previously in the genome of isolate Race 5 (Zaccaron et al. 2022). The majority of the repeats in the genomes of the isolates were dispersed repeats, which accounted for 99.08% (33.2 Mb) to 99.1% (33.4 Mb) of the repetitive content in the four isolates, whereas short tandem repeats accounted for only 0.90% (0.30 Mb) to 0.92% (0.31 Mb) of the repetitive content. Further annotation of the TEs in the genomes of the four *C. fulvum* isolates produced similar results and showed again only small differences among them (Fig 3.S2 and Table 3.S3). As for isolate Race 5, the majority of TEs in the genomes of the four *C. fulvum* isolates were retrotransposons, which accounted for a minimum of 88.5% (30.2 Mb) of the repetitive content in isolate Race 2.4.9.11 to a maximum of 90.3% (30.2 Mb) of the repetitive content in isolate Race 0WU. In contrast, DNA transposons and unclassified repeats ranged from 3.5% to 3.6%, and from 5.5% to 7.4% of the repetitive content in the four isolates, respectively. Among retrotransposons, the most common families were the LTR Ty3/mdg-4 family, ranging from 36.5% to 38.3% of the repetitive content in the four isolates, the LINE Tad1 family (29.2 % to 31.0%), and the LTR Copia family (18.0% to 20.0%). When compared to the TE content of isolate Race 5, this isolate had less

Copia retrotransposons (13.9% of the repetitive content) and more unclassified TEs (12.5% of the repetitive content) compared to the other four isolates (Fig 3.S2 and Table 3.S3). However, considering that the genome of isolate Race 5 was assembled using PacBio's error prone contiguous long reads (CLR) whereas the genomes of the other four isolates were assembled using PacBio's HiFi reads, the small differences in TE content could be perhaps explained by the effect of the sequencing technology on the accuracy of assembling repetitive DNA.

3.2.3 RIP density and distribution patterns are also conserved among the isolates of *C. fulvum*

It was previously shown that *C. fulvum* exhibits one of the highest rates of RIP among fungi, with 39.2% of the genome of isolate Race 5 affected by RIP mutations (Zaccaron et al. 2022). Genome-wide RIP analyses using a sliding-window approach bolstered these results by showing that between 40.7% and 41.0% of the genomic content of isolates Race 0WU, Race 4, Race 2.4.9.11, and Race 2.4.5.9.11 IPO was affected by RIP mutations. As expected, between 95.3% to 96.5% of the RIPed regions in the genomes of the four isolates co-localized with repetitive DNA, and no major differences in RIP density and distribution patterns were observed among homologous sets of chromosomes in the five isolates (Fig 3.S3 and Table 3.S4). Among the core chromosomes, RIP levels were highest in Chr3, ranging from 52.7% to 53.3% in the five isolates, followed by Chr12 (28.6%-30.0%) and Chr13 (28.0%-30.0%) (Table 3.S4). When present, the two accessory chromosomes displayed even higher levels of RIP, ranging between 56.7% to 58.3% in Chr15, and 70.4% to 70.7% for Chr14. Accessory chromosomes also showed higher abundance of RIP leakage toward non-repetitive regions, ranging from 2.82% in Chr15 to 3.69% in Chr14. In contrast, RIP leakage in all core chromosomes, was estimated to be less than 0.05%. Genome-wide RIP analyses also revealed the presence of many large RIP-Affected Regions (LRARs) longer than 4 kb in size, with LRARs numbers ranging from 1492 in isolate Race 4 to 1536 in isolate Race 0WU. Moreover, the average size of LRARs ranged from 16766 bp in isolate Race 0WU to 16991 bp in isolate Race 2.4.9.11, and their average GC content was 42.5%. Finally, given that the isolates exhibited similar patterns of RIP across their chromosomes, they also displayed a similar bimodal distribution in GC content with major peaks at approximately 54% and a minor peak at approximately 42% (Fig 3.S4). Collectively, the above results indicate that isolates of *C. fulvum* exhibit limited intraspecific diversity in terms of their genomic landscape of repeats and of RIP patterns, possibly because the fungus reproduces mainly asexually in nature (Stergiopoulos et al. 2007b).

3.2.4 A pangenome analysis of the five *C. fulvum* genomes indicates a stable gene content with a low number of accessory genes

The genomes of isolates Race 0WU, Race 4, Race 2.4.9.11, and Race 2.4.5.9.11 IPO were annotated using a combination of *ab initio* gene predictions and available gene models from *C. fulvum* isolates Race 5 (Zaccaron et al. 2022) and Race 0WU (De Wit et al. 2012). We also refined the gene annotation of *C. fulvum* Race 5 and removed 69 transposon-like gene models while adding 372 new gene models, which increased the number of genes in this isolate from 14690 to 14993 (Table 3.S5).

The total number of predicted genes was similar among the five isolates, ranging from 14,895 genes in isolate Race 4 to 14,993 genes in isolate Race 5. A BUSCO-based assessment of the quality and completeness of the genes annotations (Simão et al. 2015) in the five *C. fulvum* genomes showed that they were 98.8% and 98.9% complete and that less than 1% of the genes were missing in any of the isolates (Table 3.2). Further functional annotations showed that the five isolates shared a similar number of protein-coding genes in different functional categories (Table 3.S6), including categories with relevance to fungal plant pathogens such as CAZymes (519 to 525 genes) (Fig 3.S5 and Table 3.S7), proteases (357 to 362 genes) (Fig 3.S6 and Table 3.S8), cytochrome P450s (133 to 134 genes) (Fig 3.S7A and Table 3.S9), putative transporters (2277 to 2293 genes) (Fig 3.S7B and Table 3.S10), key enzymes for the biosynthesis of secondary metabolites (SMs) (41 to 42 genes) (Fig 3.S7C and Table 3.S11), secreted proteins (SPs) (1404 to 1425 genes) (Table 3.S12), and candidate effectors (427 to 440 genes) (Table 3.S13). Included among the candidate effectors are the previously characterized *Avr2, Avr4, Avr4E, Avr5, Avr9, Ecp1, Ecp2, Ecp2-2, Ecp2-3, Ecp4, Ecp5,* and *Ecp6* effector genes as well as the additional 67 candidate effectors previously described as extracellular proteins (Mesarich et al. 2014). Similar results were obtained when the protein-coding genes from the five *C. fulvum* isolates were assigned functions based on annotations against the major categories and subcategories of gene ontology (GO, 8047 to 8079 genes) (Fig 3.S8A and Table 3.S14) and the eukaryotic orthologous groups (KOG, 8867 to 8079 genes) (Fig 3.S8B and Table 3.S14).

To further construct a gene-based pangenome for *C. fulvum*, the 74,756 genes that were predicted among the five isolates were organized into hierarchical orthogroups (HOGs) with OrthoFinder. A total of 15,041 HOGs were obtained, which included 99.8% to 99.9% of all predicted genes from each isolate. Nearly all (*n* = 14,962; 99.4%) of these HOGs contained at most one gene per isolate, corresponding to one-to-one orthologs. Surprisingly, all five isolates shared 14,714 HOGs, corresponding to 98.3% to 98.8% of all their genes (Fig 3.2A). This indicated that less than 2% of the *C. fulvum* genes were accessory genes. From the 326 HOGs containing accessory genes, 57 contained genes assigned to different functional categories. These include HOGs containing genes encoding CAZymes (*n* = 8), proteases (*n* = 4), cytochrome P450s (*n* = 2), transporters (*n* = 25), key enzymes for biosynthesis of SMs (*n* = 1), secreted proteins (*n* = 22), and candidate effectors (*n* = 12) (Table 3.S15). However, no significant functional gene category was enriched among the accessory HOGs (hypergeometric test *p*-value < 0.05).

To investigate the extent to which the sizes of the pan- and core genomes of *C. fulvum* changed as a function of the number of genomes analyzed, the five genomes were sampled into subsets of sizes between one and five, and the number of core and accessory HOGs was used as a proxy for the size of pan- and core genome. The size of the pangenome stabilized at 15,043 genes, and although the size of the core genome continued to decrease after including all five isolates, it trended toward stabilizing rapidly (Fig 3.2B). These results suggest that *C. fulvum* has a stable gene complement, and that the inclusion of more genomes will likely not increase considerably the number of novel genes. To better support this hypothesis, the inclusion of more isolates that better represent the genetic diversity of *C. fulvum* is need.

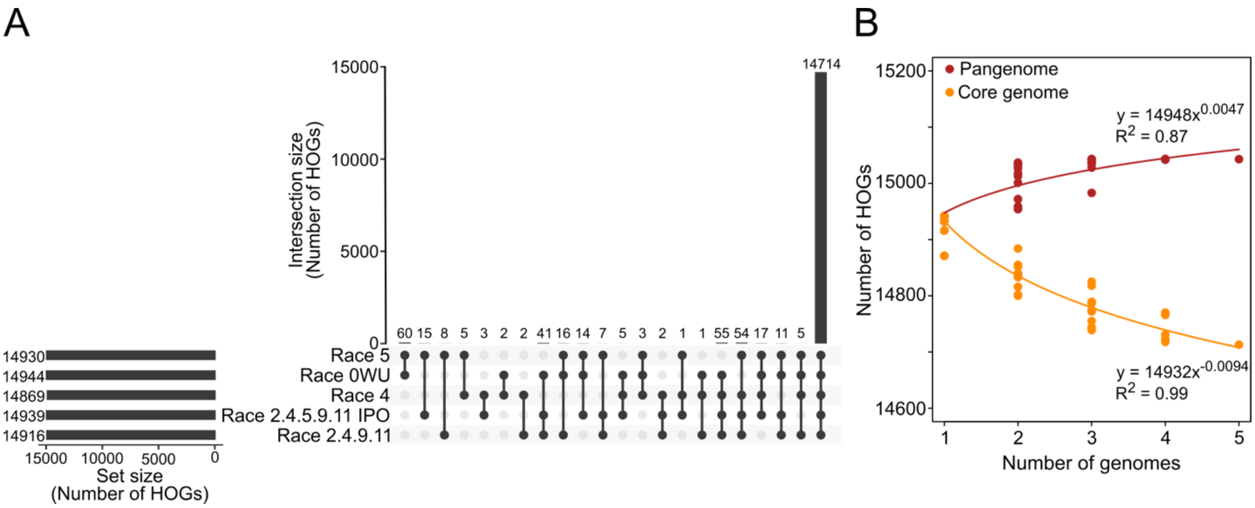


Figure 3.14: *Cladosporium fulvum* has a low number of accessory genes. (A) Upset plot showing the number of hierarchical orthogroups (HOGs) containing genes from one or more isolates. The figure shows that almost all HOGs are shared by all isolates. (B) Scatterplot showing the estimated sizes of pan - and core genome of *C. fulvum*. The five genomes were sampled in all possible combinations of size *x*, with 1≤ *x* ≤5. Points represent the number of all HOGs (pangenome) and HOGs containing genes from all sampled genomes (core genome). The curves were fitted by linear regressions of the log-transformed median values of the pan- and core genome. The figure shows that the pangenome size grows only slowly as more genomes are included, suggesting that the five sequenced genomes already capture most of the gene space in *C. fulvum*.

Table 3.7: Gene prediction statistics for five *Cladosporium fulvum* isolates. BUSCO completeness was estimated based on the Dothideomycetes dataset (*n*= 3786 genes).

	Race 5[a]	Race 0WU	Race 4	Race 2.4.5.9.11 IPO	Race 2.4.9.11
Number of genes	14,993	14,981	14,895	14,944	14,943
Average gene length (bp)	1,359	1,354	1,355	1,356	1,356
Average exon length (bp)	594	593	594	594	594
Average intron length (bp)	79	79	78	79	79
Average protein length (aa)	422	421	421	421	421
BUSCO completeness					
Complete	98.9	98.9	98.8	98.8	98.8
Complete single	98.8	98.7	98.6	98.6	98.5
Complete duplicated	0.1	0.2	0.2	0.2	0.3
Complete fragmented	0.3	0.3	0.3	0.3	0.3
Missing	0.8	0.8	0.9	0.9	0.9

[a] (Zaccaron et al. 2022).

3.2.5 The five genomes of *C. fulvum* exhibit chromosome-scale conservation of synteny and collinearity with few large-scale chromosomal structural variations

A synteny analysis among the five *C. fulvum* genomes indicated that homologous chromosomes shared one-to-one macrosynteny and a large degree of microsynteny and collinearity as well, as both the order and orientation of the genes on them were fairly conserved in the sequenced isolates. Indeed, based on the order of the genes on chromosomes, the number of synteny blocks between any pair of genomes ranged from 15 to 19 and contained between 98.8% to 99.8% of the all the genes in the genomes (Fig 3.S9A and Table 3.S16). Moreover, whole-genome alignments that consider synteny breaks that do not affect the gene order revealed a total of 373 to 1437 of synteny blocks that covered between 92.7% to 98.5% of the genomes (Fig 3.S9B, Fig 3.S10 and Table 3.S16).

Overall, only three large-scale chromosomal structural variations (SVs) were observed among the genomes of the five isolates. These SVs corresponded to a reciprocal translocation between chromosomes Chr4 and Chr10 of isolate Race 5 (Fig 3.1B) and two large inversions in Chr1 of isolate Race 2.4.9.11 and in Chr10 of isolate Race 0WU, respectively. Mapping of the PacBio reads to the junctions of the reciprocal translocation in isolate Race 5 and the two large inversions in isolates Race 0WU and Race 2.4.9.11 supported their presence and indicated that they were not caused by misassemblies (Fig 3.S11). Further analysis of the SVs indicated that the synteny breaks of the reciprocal translocation between Chr4 and Chr10 of isolate Race 5 were localized in repeat-rich regions (Fig 3.S12A), thus raising the possibility that the exchange of the chromosome arms was facilitated by the repeats. In addition, although no genes were disrupted by the synteny break points in Chr4, there were two genes flanking the break points. The genes encoded a hypothetical secreted protein of unknown function (CLAFUR5_04694) and the candidate effector Ecp46 (CLAFUR5_12163), indicating that the chromosome arms exchange disrupted the intergenic region of a putative virulence-associated gene. Generally, reciprocal translocations are rarely reported in fungi, but a case has been described in the pine tree pathogen *Dothistroma septosporum*, a close relative of *C. fulvum*

(Bradshaw et al. 2019), in which the translocation occurred between chromosomes Chr5 and Chr13 that are homologous to the *C. fulvum* Chr3 and Chr12, respectively (Fig 3.S13). When considering the other two large-scale SVs, the largest of the two was 1.2 Mb long present in Chr1 of isolate Race 2.4.9.11 (Fig 3.1B). The synteny break points of this inversion were also in repeat-rich regions and did not disrupt any protein coding sequences (Fig 3.S12B). In contrast, synteny breaks of the 654 kb inversion in Chr10 of isolate Race 0WU were in repeat-poor regions and physically close (< 500 bp) to the nearest predicted genes (Fig 3.S12C). Interestingly, both synteny breaks of this second largest inversion colocalized with a segment of 7.2 kb that was duplicated in Race 0WU but not in the other four genomes analyzed (Fig 3.S12B). The two copies of this duplicated segment were identical and contained three predicted genes, one encoding a hypothetical protein (CLAFUR0_10547), and two encoding two predicted secreted chloroperoxidases (CLAFUR0_10548 and CLAFUR0_10549). Collectively, these results indicate that large-scale SVs in *C. fulvum* often colocalize with repetitive or duplicated DNA, which could either promote or be caused by these large SVs. They further show that large chromosomal rearrangements do not play a significant role in genome evolution of *C. fulvum* but may occasionally affect its infectivity by impacting virulence-associated genes such as effector-encoding genes.

3.2.6 Loss of the avirulence genes *Avr4E*, *Avr5*, and *Avr9* is due to SVs induced by transposable elements

Effector gene deletion (Westerink et al. 2004; Mesarich et al. 2014) is often reported in *C. fulvum* as a mean to overcome resistance mediated by their cognate resistance genes in tomato, but the mechanisms mediating these deletions remain elusive. We had previously hypothesized that effector gene loss could be a consequence of SVs and the effectors' physical location in the *C. fulvum* genome (Stergiopoulos et al. 2007a). *Avr9,* in particular, whose loss is commonly reported in race 9 strains of the fungus that overcome the Cf-9-mediated resistance in tomato, is situated in a repeat-rich region of the genome that is present at 6.6 kb from the telomeric repeat at the left end of Chr7. This makes *Avr9* putatively prone to deletions

(Zaccaron et al. 2022). To investigate the mechanism that promotes loss of *Avr9*, Chr7 of isolate Race 2.4.9.11 which lacks *Avr9* (Stergiopoulos et al. 2007a) was aligned to Chr7 of isolate Race 0WU which has *Avr9*. The alignment revealed that the first 7.8 kb of Chr7 chromosome in isolate Race 0WU is replaced in isolate Race 2.4.9.11 by a 13.1 kb fragment that is largely composed of interspersed repeats and contains no predicted genes (Fig 3.3A). Homology searches revealed that the 13.1 kb fragment was nearly identical to the first 13.1 kb of Chr2 of the same isolate (Fig 3.3B), suggesting that in isolate Race 2.4.9.11, the first 7.8 kb of Chr7 carrying *Avr9*, was replaced by the first 13.1 kb of Chr2. Interestingly, both copies of the 13.1 kb fragment in Chr2 and Chr7 of isolate Race 2.4.9.11 were flanked on one side by truncated copies of a Ty1/Copia retrotransposon (Fig 3.3B). The consensus of this Ty1/Copia retrotransposon was a 5.6 kb sequence flanked by direct repeats of 240 bp long and contained typical domains found in LTR retrotransposons (Fig 3.3D). These truncated copies of a Ty1/Copia retrotransposon are also present in other isolates, including Race 0WU (Fig 3.3C). Mapping of the HiFi reads from isolate Race 2.4.9.11 to the genome of isolate Race 0WU confirmed the absence of the *Avr9* locus and that its deletion colocalized with the truncated Ty1/Copia copy (Fig 3.S14). Collectively, these results support the hypothesis that deletion of *Avr9* in isolate Race 2.4.9.11 was the result of a nonreciprocal translocation between Chr7 and Chr2, promoted by the presence of truncated copies of a Ty1/Copia retrotransposon.

As for *Avr9*, loss of *Avr4E* and *Avr5* is also commonly reported in race 4E and race 5 isolates of the fungus that overcome the cognate *Cf-4E* and *Cf-5* resistance genes, respectively in tomato. To investigate the mechanisms promoting the deletion of *Avr4E* and *Avr5*, the genomes of Race 0WU which has both genes, Race 2.4.5.9.11 IPO which lacks both genes, and/or Race 2.4.9.11 which lacks *Avr4E*, were aligned. *Avr4E* was located within a 8270 bp segment of Chr7 in isolate Race 0WU that was absent in isolates Race 2.4.5.9.11 IPO and Race 2.4.9.11 (Fig 3.S15). This segment was flanked by two near-identical copies of a putative DNA transposon Tc1/mariner that was similar (47.1% nucleotide identity) to the Tc1/mariner Molly from the wheat fungal pathogen *Stagonospora nodorum* (AJ488502). A similar organization of repetitive

DNA flanking the deletion of the *Avr5* locus in Chr1 was observed in isolate Race 2.4.5.9.11 IPO (Fig 3.S16). In this isolate, we noticed the deletion of a long 91338 bp fragment containing *Avr5* and part of its up- and downstream intergenic regions. The deleted fragment was flanked by two similar (85.9% identity) copies of a putative non-LTR LINE/Tad1 retrotransposon located on the same DNA strand (Fig 3.S16).

Collectively, the above results demonstrate that the deletion of *Avr4E*, *Avr5*, and *Avr9* in the genome of race 4E, race 5, and race 9 isolates of *C. fulvum*, respectively, is due to SVs mediated by the presence of neighboring copies of transposable elements, which possibly serve as templates for nonallelic homologous recombination.

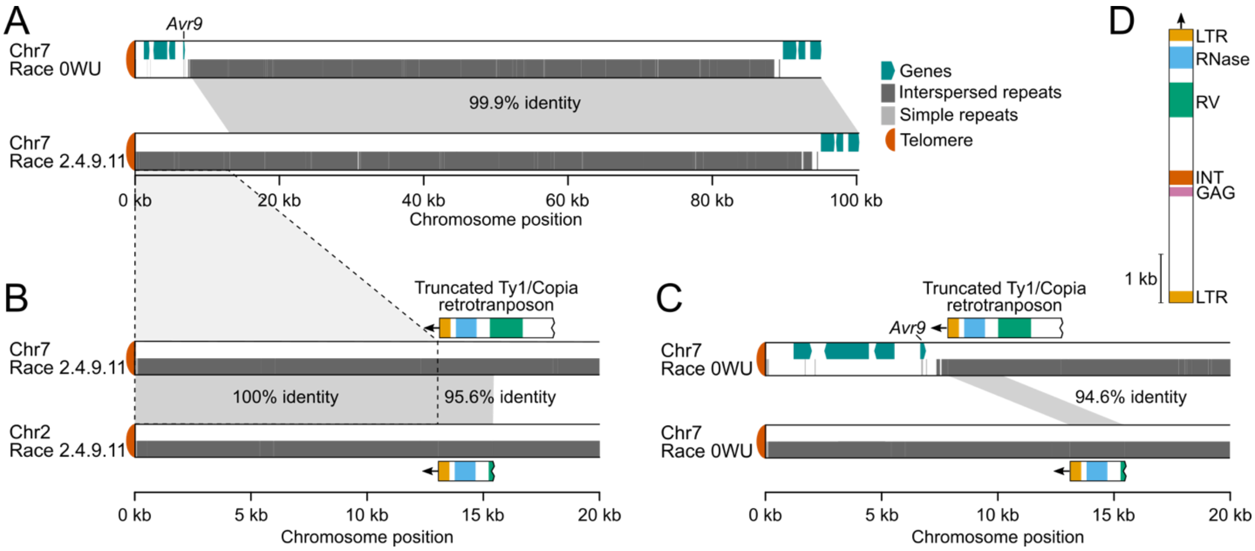


Figure 3.15: A nonreciprocal translocation between Chr7 and Chr2 causes the deletion of the *Avr9* locus. (A) Alignment of the first 100 kb of Chr7 from isolates Race 0WU which has *Avr9* and Race 2.4.9.11 in which *Avr9* is lost. The 15 kb region that harbors *Avr9* in isolate Race 0WU is absent in isolate Race 2.4.9.11. (B) Alignment of the first 20 kb of Chr7 and Chr2 of isolate Race 2.4.9.11. The left-hand side tip of Chr7 of isolate Race 2.4.9.11 is identical to the sequence of the left-hand side tip of Chr2, and both sequences are flanked by truncated copies of a Ty1/Copia retrotransposon. (C) Alignment of the first 20 kb of Chr7 and Chr2 of isolate Race 0WU. Truncated copies of the same Ty1/Copia retrotransposon are present in the first 15 kb of Chr7 and Chr2 of isolate Race 0WU. (D) Representation of the intact Ty1/Copia retrotransposon shown in (B) and (C). LTR: long terminal repeat, GAG: group-specific antigen domain, INT: integrase domain, RV: reverse transcriptase domain, RNase: ribonuclease H domain.

3.2.7 Most SVs in the genome of *C. fulvum* colocalize with TE-rich regions and do not affect genes

The identification of SVs that affected avirulence genes indicated that some SVs can serve *C. fulvum* to overcome host resistance. To search for other genes affected by SVs, we performed pairwise whole genome alignments using isolate Race 0WU as reference. The number of SVs identified varied from 718 in the genome of isolate Race 5 to 843 in the genome of isolate Race 4 (Table 3.S17). From the identified SVs, between 697 and 822 (97% to 98%) were long insertions and deletions, indicating that most SVs in *C. fulvum* corresponded to INDELs. Colocalized INDELs were merged, resulting in a total of 662 insertions and 564 deletions (Fig 3.4), which varied in size from 205 bp to 108,643 bp and averaged 6050 bp in length. To investigate the extent to which these large INDELs colocalized with TEs, their coordinates were compared with masked regions of the genome of isolate Race 0WU. The analysis showed that 593 (89.6%) insertions had their insertion sites located within predicted TEs, and 502 (89.0%) deletions had both their start and end coordinates located within predicted TEs. Moreover, 1184 (96.5%) of the INDELs had more than 95% of their sequences composed of predicted TEs (Fig 3.S17). Collectively, these results indicate that the vast majority of SVs colocalize with TE-rich regions in the genome of *C. fulvum*.

Although INDELs were largely associated with TE-rich regions of the genome, their impact on predicted genes was minimal. From the 1226 INDELs identified, only 31 (2.5%) overlapped with gene coding regions (Fig 3.4), affecting a total of 46 genes (Table 3.S18). Of these, 13 genes were located within deletions, and, as expected, *Avr5* and *Avr4E* were among them, along with genes predicted to encode an alkaline phosphatase, a serine/threonine-protein kinase, a transcription factor, and a hypothetical secreted protein (Table 3.S18). Of the rest 33 genes affected by SVs, 15 were predicted to have been duplicated due to the insertion of duplicated segments, with the largest of these duplicated segments containing a group of nine genes in isolate Race 2.4.9.11 (Fig 3.S18B). Finally, nine of the genes affected by SVs were disrupted by an insertion in their coding sequence, including a gene that encoded a candidate effector in isolate Race 0WU

(CLAFUR0_01596). Notably, the identified insertion in CLAFUR0_01596 is a tandem duplication in isolate Race 2.4.5.9.11 IPO that duplicated a fragment that contained the CLAFUR0_01596 ortholog in this isolate (CLAFUW4_01596) together with a neighboring gene encoding a putative laccase (CLAFUW4_01597) (Fig 3.S18A).

Collectively, these results indicate that most of the SVs present in *C. fulvum* are long INDELs largely composed of TEs, which correspond to presence/absence of TE-rich regions or TE insertion site polymorphisms. Furthermore, a small number of the identified SVs affect predicted genes, thus corroborating with our previous observation that *C. fulvum* has a stable gene complement.

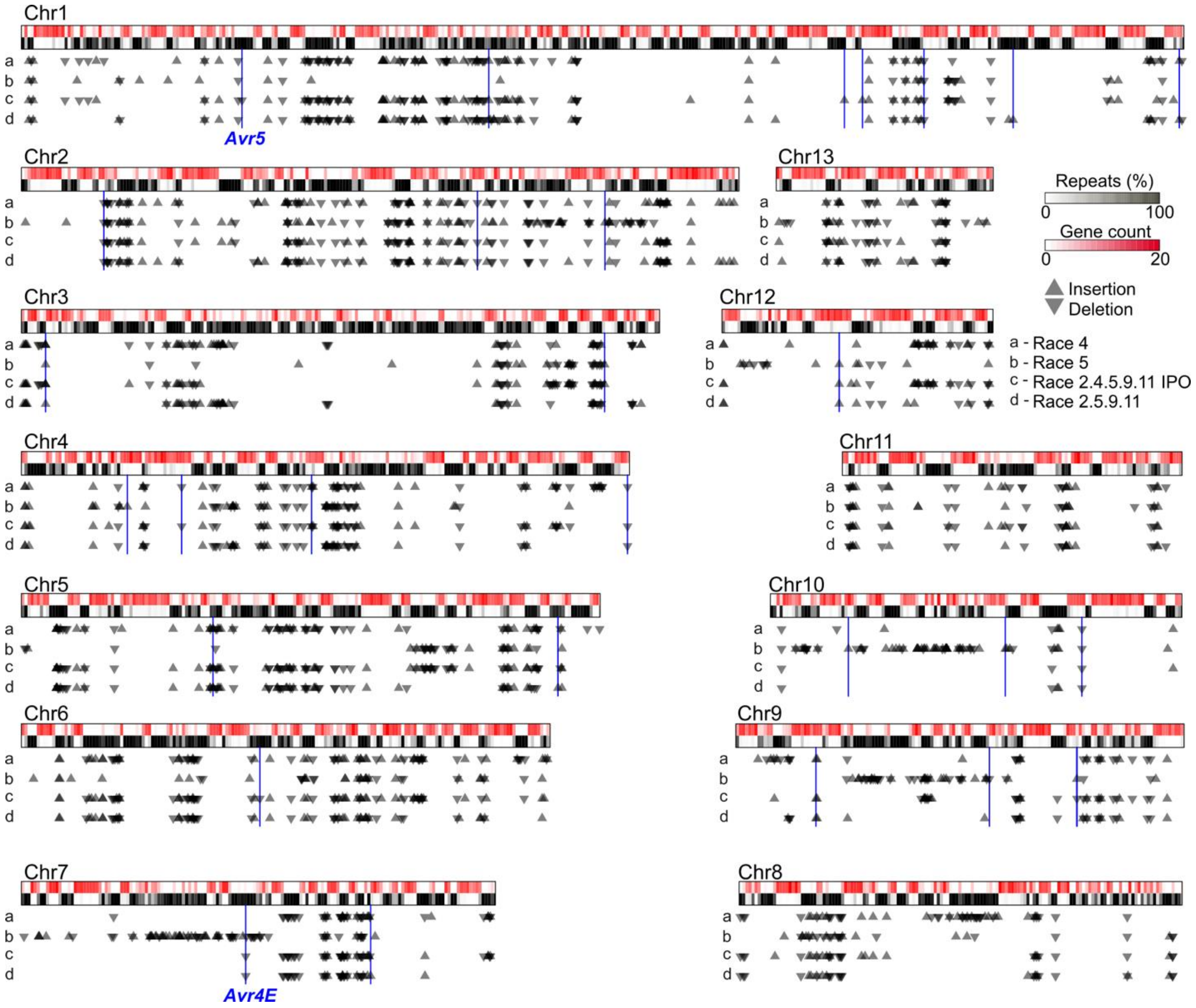


Figure 3.16: SVs in the genome of *Cladosporium fulvum* are mostly located in repeat-rich regions. The figure shows diagrams of the 13 chromosomes of *C. fulvum* isolate Race 0WU as rectangles with two tracks, representing gene content (top track) and repetitive DNA content (bottom track). Location of SVs (i.e. INDELs longer than 200 bp) are shown as upward (insertion) or downward (deletions) triangles for the four isolates compared to isolate Race 0WU. SVs that affect predicted genes are highlighted with vertical blue lines and as shown by the figure, overall SVs tend to not affect protein-coding genes The SVs that resulted in the deletion of *Avr4E* and *Avr5* are labeled.

3.2.8 *Cladosporium fulvum* has at least two accessory chromosomes, one of which carries pseudogenized copies of candidate effector genes

A total of 15 chromosomes were assembled from the five genomes of *C. fulvum*, 13 of which (Chr1-Chr13) were core chromosomes common to all isolates and two represented accessory chromosomes that were selectively present in two (Chr4) and three isolates (Chr15), respectively (Fig 3.1A). Both Chr14 and Chr15 were further differentiated from the core chromosomes by their small size and high repetitive DNA content (Table 3.S2), which are typical features of accessory chromosomes (Bertazzoni et al. 2018; Houben et al. 2014). Pairwise alignments of the two accessory chromosomes showed that, when present, they were highly syntenic among isolates, with aligned segments sharing more than 99.9% of nucleotide identity and a conserved complement of 28 (Chr14) or 40 to 41 (Chr15) genes (Fig 3.S19). However, of the 69 genes present collectively in Chr14 and Chr15 of isolate Race 0WU, 67 encoded hypothetical proteins, 1 (CLAFUR0_14817) encoded a protein with a conserved kinesin motor domain (PF00225), and 1 (CLAFUR0_14809) encoded a secreted protein. Moreover, 52 (75.4%) of the genes had no homolog in the NCBI nr database based on BLASTp searches (e-value < 1E-5) (Table 3.S19). To investigate whether any of the predicted genes in Chr14 and Chr15 were expressed during host infection, public RNA-seq data of *C. fulvum* Race 0WU-*Solanum lycopersicum* cv. Heinz interaction (NCBI accessions SRR1171035, SRR1171040, SRR1171043) (De Wit et al. 2012) was used to quantify gene expression. From the 69 genes in Chr14 and Chr15, 30 had almost no detectable levels of expression (TPM < 3) at any time point (Table 3.S19). In contrast, six genes in Chr14 and four genes in Chr15, all of which encoded hypothetical proteins and had no BLASTp hits in the NCBI nr database, had considerable levels of expression (TPM > 50). These results indicate that most genes in accessory chromosomes are transcriptionally inactive during host infection.

We have previously demonstrated the presence of gene flow between the core and accessory chromosomes of *C. fulvum*, with a case of a gene (CLAFUR5_14645) in isolate Race 5 that had two identical copies, i.e., one in the core Chr1 and one in the accessory Chr14 (Zaccaron et al. 2022). To investigate whether additional genes were shared by core and accessory chromosomes, Chr14 and Chr15 of isolate Race 0WU were hard masked and queried with BLASTn against the 13 core chromosomes (e-value < 1E-10). Of the 98,513 bp that were unmasked in Chr14, only 8,090 bp (8.2%) had BLAST hits (Fig 3.S20). Included was a 934 bp fragment that contained a gene encoding a hypothetical protein with two identical copies in Chr14 (CLAFUR0_14855) and Chr1 (CLAFUR0_00411), respectively and which was homologous to the CLAFUR5_14645 gene previously reported as duplicated in isolate Race 5 (Zaccaron et al. 2022). In a similar way, of the 178,773 bp that were unmasked in Chr15, 45,829 bp (25.6%) had BLAST hits in core chromosomes (Fig 3.S20). Included were two fragments of 7.5 kb and 8.6 kb in size, respectively, which were shared by Chr6 and Chr15 (Fig 3.5A). These fragments exhibited a peculiar arrangement, as in Chr15 they were situated nearly next to each other, whereas in Chr6 they were present at the opposite ends of this chromosome i.e., at 42 kb from the left-end telomere and at 300 bp from the right-end telomere. Moreover, the 8.6 kb fragment was further tandemly duplicated once in Chr15 (Fig 3.5A). Further inspection of the two fragments shared between Chr6 and Chr15 showed that the 7.5 kb long fragment contained two genes of unknown function in Chr6, while the 8.6 kb fragment harbored five genes, of which two encoded hypothetical proteins (CLAFUR0_07628 and CLAFUR0_07629), one encoded a predicted prolyl 4-hydroxylase (CLAFUR0_07630), and two encoded the candidate effectors Ecp13 (CLAFUR0_07631) and CE29 (CLAFUR0_07632), respectively. However, the copies of four of these genes on the 8.6 kb fragment were pseudogenized in Chr15, including the two genes encoding the candidate effectors Ecp13 and CE29. Pseudogenization was caused by the accumulation of C<->T/G<->A nucleotide substitutions in their coding sequences, possibly as a result of RIP (Fig 3.5B and Fig 3.5C). Collectively, these observations suggest the presence of gene flow from a core to an accessory chromosome that involved candidate effectors, followed by pseudogenization of these genes by RIP mutations.

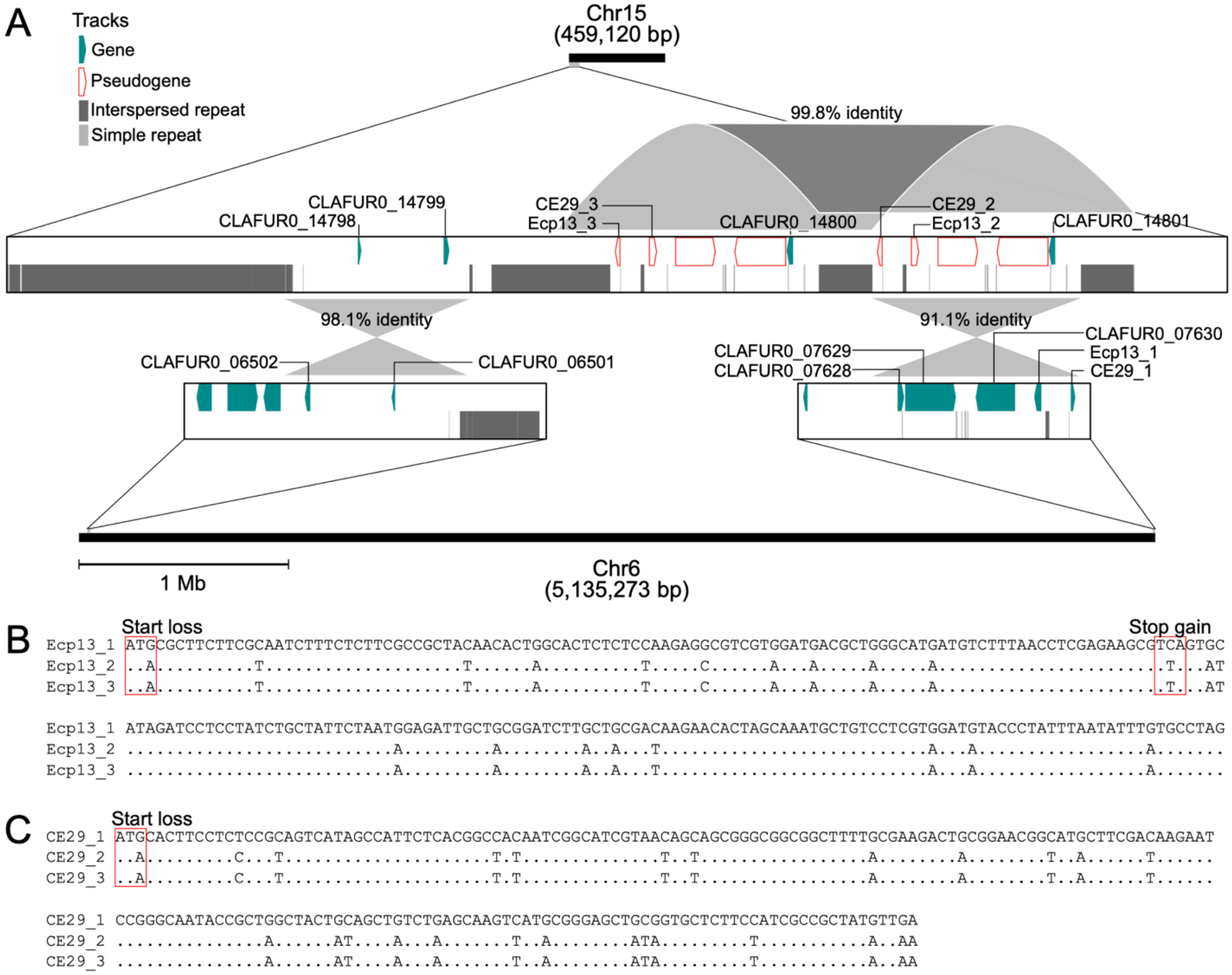

A
Tracks
Gene
Pseudogene
Interspersed repeat
Simple repeat
Chr15
(459,120 bp)
99.8% identity
CLAFUR0_14799
CLAFUR0_14798
CE29_3
Ecp13_3
CLAFUR0_14800
CE29_2
Ecp13_2
CLAFUR0_14801
98.1% identity
91.1% identity
CLAFUR0_06502
CLAFUR0_06501
CLAFUR0_07629
CLAFUR0_07628
CLAFUR0_07630
Ecp13_1
CE29_1
1 Mb
Chr6
(5,135,273 bp)
B
Start loss
Stop gain
Ecp13_1 ATGCGCTTCTTCGCAATCTTTCTCTTCGCCGCTACAACACTGGCACTCTCTCCAAGAGGCGTCGTGGATGACGCTGGGCATGATGTCTTTAACCTCGAGAAGCGTCAGTGC
Ecp13_2 ..A..........T....................T......A..........T.....C......A...A.....A.....A.......................T...AT
Ecp13_3 ..A..........T....................T......A..........T.....C......A...A.....A.....A.......................T...AT
Ecp13_1 ATAGATCCTCCTATCTGCTATTCTAATGGAGATTGCTGCGGATCTTGCTGCGACAAGAACACTAGCAAATGCTGTCCTCGTGGATGTACCCTATTTAATATTTGTGCCTAG
Ecp13_2 ...........................A.........A........A..A...T...........................A...A.................A.......
Ecp13_3 ...........................A.........A........A..A...T...........................A...A.................A.......
C
Start loss
CE29_1 ATGCACTTCCTCTCCGCAGTCATAGCCATTCTCACGGCCACAATCGGCATCGTAACAGCAGCGGGCGGCGGCTTTTGCGAAGACTGCGGAACGGCATGCTTCGACAAGAAT
CE29_2 ..A.........C...T.....................T.T..............T..T.................A........A........T..A......T......
CE29_3 ..A.........C...T.....................T.T..............T..T.................A........A........T..A......T......
CE29_1 CCGGGCAATACCGCTGGCTACTGCAGCTGTCTGAGCAAGTCATGCGGGAGCTGCGGTGCTCTTCCATCGCCGCTATGTTGA
CE29_2 ...............A......AT....A...A.......T..A........ATA.........T...........A..AA
CE29_3 ...............A......AT....A...A.......T..A........ATA.........T...........A..AA

Figure 3.17: Duplicated segments between a core and a dispensable chromosome of *Cladosporium fulvum* isolate Race 0WU. (A) Intra- and interchromosomal duplications within the first 50 kb of Chr15. A tandem duplication of 12.9 kb fragment is shown. This duplication harbors pseudogenized copies of the candidate effector genes *Ecp13* and *CE29,* for which the functional copies are located 12 kb from the right telomere of Chr6. Underscores followed by numbers were used to distinguish copies of *Ecp13* and *CE29*. The figure also shows another 7.5 kb fragment having one copy in Chr15 and another copy at 40 kb from the left telomere of Chr6. (B) and (C) shows the alignments of the coding sequences of *Ecp13* and *CE29* with their pseudogenized copies. Conserved nucleotides are represented by dots. Codons that harbor predicted loss-of-function substitutions are indicated.

3.2.9 Repetitive regions are asymmetrically affected by RIP in the genome of *C. fulvum*

To better understand the genetic and genomic diversity in *C. fulvum*, we performed a whole-genome single-nucleotide polymorphism (SNP) analysis by aligning the genomes of isolates Race 5, Race 4, Race 2.4.9.11, and Race 2.4.5.9.11 IPO on the genome of isolate Race 0WU. A total of 192,279 SNPs were identified, most of which (n=183,160; 95.2%) were in intergenic regions. A total of 8794 SNPs were identified within the 14,714 genes conserved in all five isolates. A phylogenetic tree based on these 8794 SNPs indicated considerable genetic diversity among the five isolates analyzed (Fig 3.S21). Interestingly, 90% of the SNPs (n=173,651) could be organized into 2000 clusters of 17 bp to 93,951 bp in size that accounted for 19% of the genomic content of the 13 core chromosomes (12,659,908 bp). A sliding window analysis along the chromosomes revealed genomic regions with low and high nucleotide diversity, suggesting the presence of SNP hotspots (Fig 3.6). Such contrasting patterns of nucleotide diversity were essentially due to nucleotide transitions, as the average nucleotide diversity of transitions per site (π_{Ts}) ranged from 0 to 0.2, and the average nucleotide diversity of transversions per site (π_{Tv}) ranged from 0 to 0.003. Further mapping of the SNPs on the *C. fulvum* chromosomes revealed that the SNP hotspots co-localized with repetitive regions of the genome that were RIPed (Fig 3.6), suggesting that they were formed by RIP mutations. This was further supported by the observation that transition nucleotide substitutions in the SNP hotspots exhibited the typical dinucleotide bias of RIP mutations (i.e., CpA ↔ TpA) (Fig 3.S22). However, while RIP has been reported to induce only transitions, the SNP hotspots across the *C. fulvum* chromosomes also exhibited elevated π_{Tv} (Fig 3.6 and Fig 3.S23), suggesting that RIP can also induce transversion substitutions or that, next to RIP, another mechanism is promoting SNP hotspots in *C. fulvum*. Alternatively, it might also be that RIPed regions are under relaxed selection, which allows for the faster accumulation of random mutations.

Although SNP hotspots co-localized with RIPed regions, conversely, several long chromosome segments were present that contained RIPed regions with low nucleotide diversity (Fig 3.6). This suggested that repetitive regions of the *C. fulvum* genome were asymmetrically affected by recent RIP mutations. Further analysis showed no evident correlation between the estimated age of transposon families and their nucleotide diversity (Fig 3.S24A), as the estimated diversity of TE families that overlapped with regions of $\pi_{Ts} < 0.005$ were not significantly different from the estimated diversity of TE families that overlapped with regions of $\pi_{Ts} > 0.005$ (Fig 3.S24B). Moreover, there was no evident differences in GC content between TE copies within regions of $\pi_{Ts} > 0.005$ as compared to TE copies within regions of $\pi_{Ts} < 0.005$ (Fig 3.S25).

Taken together, the above observations support the existence of genomic islands that are less likely to accumulate RIP mutations compared to other regions of the genome, and that the occurrence of transition substitutions caused by RIP is associated with higher occurrence of transversion substitutions.

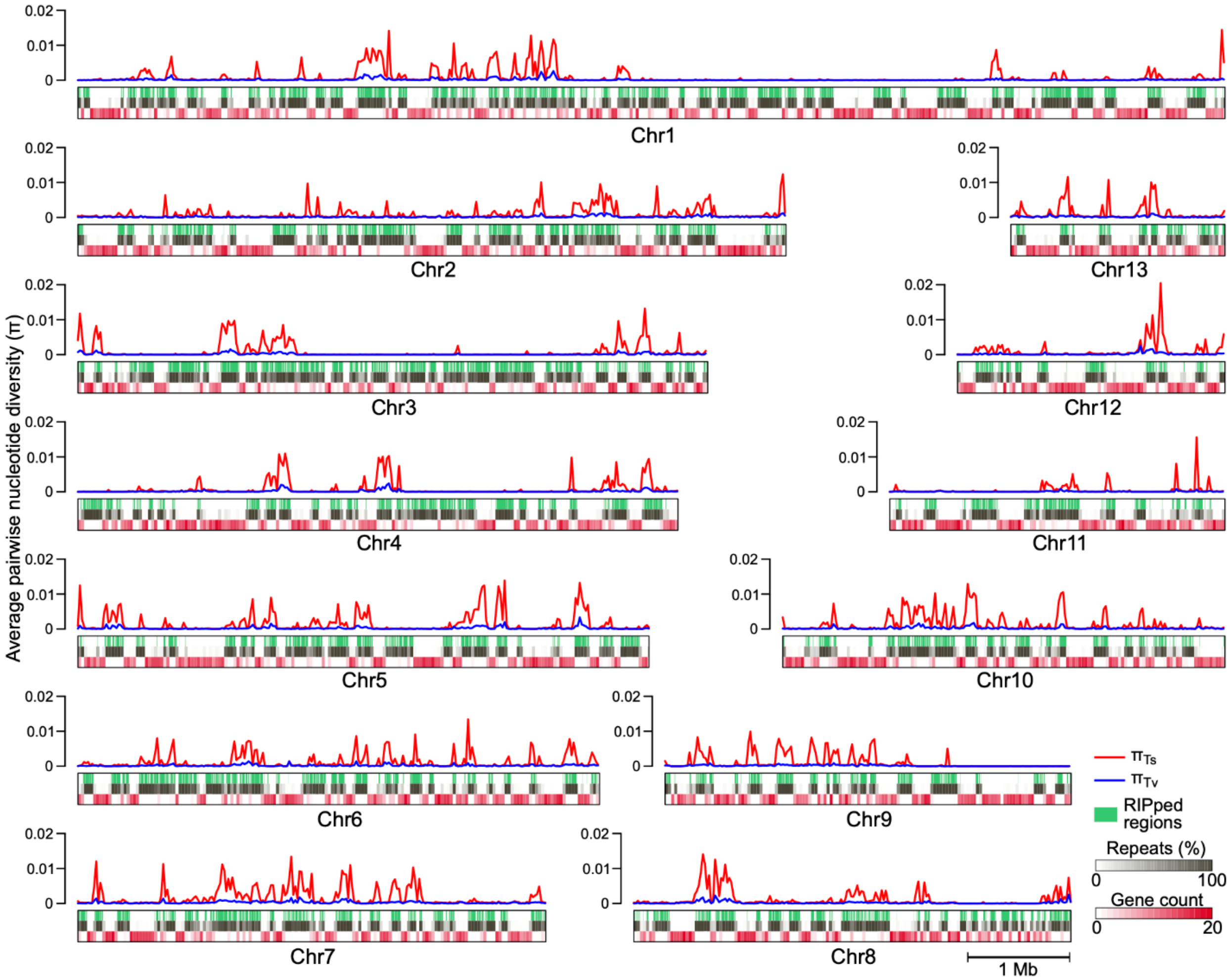


Figure 3.18: Repetitive regions in the genome of *Cladosporium fulvum* are asymmetrically affected by RIP mutations. Shown are diagrams of the 13 core chromosomes of *C. fulvum* isolate Race 0WU as rectangles with three tracks. From top to bottom, tracks indicate regions affected by RIP (RIPed) (green lines), repetitive DNA content (black lines), and gene content (red lines). The lines on top of the tracks represent average nucleotide diversity values calculated using either transitions (π_{Ts}) (red lines) or transversions (π_{Tv}) (blue lines) among the complete genomes of five isolates. The figure shows genomic regions of high variability due to transitions in RIPed regions, as well as islands of RIPed regions with almost no variability. Nucleotide diversity was calculated within 20 kb windows.

3.3 Discussion

The availability of high-quality genome assemblies can significantly advance our understanding of genome plasticity in fungi and its key role in overcoming host resistance in plant pathogens (Hartmann 2022; Schikora-Tamarit and Gabaldón 2022). In this study, we generated high-quality chromosome-level genome assemblies and gene annotations for four isolates of the tomato pathogen *C. fulvum*, thereby increasing the number of *C. fulvum* isolates with near-complete genome assemblies from one (Zaccaron et al. 2022) to five and allowing the in depth study of genomic SVs in this pathogen. Our whole-genome alignments indicated high levels of synteny among the five *C. fulvum* genomes but uncovered a few large-scale chromosomal SVs as well, including a balanced reciprocal translocation between Chr4 and Chr10 in isolate Race 5. Such large interchromosomal translocations are often reported in asexual fungal species (Bradshaw et al. 2019; de Jonge et al. 2013; Olarte et al. 2019; Tsushima et al. 2019) and rarely only in sexually reproducing ones (Demené et al. 2021) since they could result in improper chromosome pairing and nondisjunction during meiosis (Kistler and Miao 1992). Although the functional impact of interchromosomal translocations in fungal genomes remains mostly elusive, they nonetheless have been associated with acquisition of novel gene clusters for SM biosynthesis (Olarte et al. 2019) and adaptation to new hosts by the deletion and recovery of effector-encoding genes (Chuma et al. 2011). We found no evidence that the reciprocal translocation in isolate Race 5 physically disrupted any protein-coding genes, indicating no gain or loss of fitness by sequence diversification. However, it remains unknown whether this reciprocal translocation impacted the expression of genes that were translocated from one chromosome to another.

Fungal plant pathogens typically tolerate many accessory genes that exhibit presence/absence variation among isolates. For instance, in the cereal pathogens *Claviceps purpurea*, *Z. tritici,* and *Pyrenophora tritici-repentis,* 38%, 45%, and 57% of the genes, respectively, are allegedly accessory (Chen et al. 2023; Gourlie et al. 2022; Wyka et al. 2022). These genes contribute to the pathogens' genome plasticity and are believed to be important for adaptation to novel hosts and adverse environmental conditions. In *C. fulvum*, however, less than 2% of the genes were found to be accessory, indicating a highly stable gene complement among isolates of the fungus. This is likely due to the rare recombination events in *C. fulvum*, as the pathogen reproduces almost exclusively asexually (De Wit et al. 2012; Stergiopoulos et al. 2007b). However, lack of sexual reproduction might not solely explain the low number of accessory genes. This is evidenced in the asexual fungal pathogen *Verticillium dahliae*, which is abundant in genomic rearrangements and lineage-specific genes (Faino et al. 2016). Alternatively, it is plausible that the five isolates analyzed in this study may underestimate the population diversity of *C. fulvum* since they all originate from Europe, and that a more extensive sampling that includes isolates from different continents may reveal higher number of accessory genes. An amplified fragment length polymorphism (AFLP)-based multilocus analysis of 67 isolates of *C. fulvum* collected worldwide had shown, for example, that European isolates were significantly genetically differentiated from isolates that were collected in the Americas or Japan (Stergiopoulos et al. 2007b). The same study, however, which included four of the isolates in this study (i.e., Race 5, Race 0WU, Race 4, and Race 2.4.5.9.11 IPO), had also shown that the sequenced isolates represent different haplotypes of the fungus and they are phylogenetically distinct (Stergiopoulos et al. 2007b). Therefore, it is unlikely that the low number of accessory genes is an artifact of sampling or caused in its entirety by the lack of genetic diversity among the five isolates. Finally, the low number of accessory genes in *C. fulvum* contrasts the assumption that many pathogenicity-related genes in this species, such as carbohydrate-degrading enzymes and genes for SM biosynthesis, are not expressed during infection or are pseudogenized (De Wit et al. 2012). Assuming that these genes are inactive and no longer contribute to fitness, it is intriguing why they persist among the core genes of the genome.

Despite the low number of accessory genes, our study revealed that *C. fulvum* has an additional accessory chromosome, next to the one reported previously (Zaccaron et al. 2022). Interestingly, the two accessory chromosomes were both present only in isolate Race 0WU, suggesting that they are regularly gained or lost in isolates of the fungus. The true origin of fungal accessory chromosomes remains largely elusive, but it is widely accepted that they spawn from core chromosomes following major structural changes such as inversions, translocations, and fissions (Croll et al. 2013; Houben et al. 2014). In support of this assumption, it has been shown that accessory chromosomes can accumulate gene fragments from core chromosomes. Such fragments can be associated with diverse functions that enable accessory chromosomes to acquire novel functions and thereby promote their persistence in a population (Ahmad et al. 2020; Martis et al. 2012). In *C. fulvum*, the duplication of a gene of unknown function between a core and an accessory chromosome has been reported, supporting the existence of gene flow between core and accessory chromosomes (Zaccaron et al. 2022). Our current results provided further support for this idea and revealed that the accessory Chr15 of *C. fulvum* carries segments of DNA from subtelomeric regions of the core Chr6, including a fragment with pseudogenized copies of the candidate effectors *Ecp13* and *CE29*. One possibility is that the copies of these two candidate effectors were active when migrated to Chr15, thereby increasing the overall fitness of the pathogen. However, because they were spawn by gene duplications, they were eventually pseudogenized by accumulating RIP mutations. A similar scenario was reported for the candidate effector *Ecp11*, which has three tandem copies in *C. fulvum*, one of which is pseudogenized likely by RIP mutations (Zaccaron et al. 2022). Overall, our findings support the hypothesis that accessory chromosomes of *C. fulvum* could be a reservoir of genes that rapidly accumulate mutations induced by the presence of TEs (Zaccaron et al. 2022).

Repetitive DNA and TEs in fungal genomes are targeted by RIP mutations, which typically materialize between the plasmogamy and karyogamy stages of sexual reproduction (Irelan and Selker 1996). However, although sexual reproduction is thought to be rare in *C. fulvum*, nearly all predicted TEs in its genome exhibit evidence of RIP. Our whole-genome alignments also showed the presence within repetitive regions of islands with high nucleotide diversity, mostly caused by transition substitutions with dinucleotide bias, typical of RIP mutations. Large genomic islands with low or near-zero nucleotide diversity were also present within repetitive regions, but their size makes it unlikely that they were fashioned by typical processes that reduce genetic variation, such as selective sweeps. One possibility is that these highly conserved regions have accumulated considerably less RIP mutations compared to regions of high nucleotide diversity, suggesting that the genome of *C. fulvum* is asymmetrically affected by RIP mutations. This might be the case since the RIP machinery does not mutate all repetitive DNA evenly. For instance, short repeats of less than 400 nucleotides frequently escape RIP mutations (Watters et al. 1999). Also, tandem duplications are much more likely to be affected by RIP compared to interspersed duplications (Selker 2002), while divergent copies of less than 80% nucleotide identity are typically not affected by RIP (Galagan and Selker 2004). Although an attractive hypothesis, we found no evidence that TEs in highly conserved regions of the genome of *C. fulvum* escape RIP due to their short size or high divergence among copies, and thus, the origin of the alternating patterns of high and low nucleotide diversity within repetitive regions remains elusive.

Many fungal pathogens are known for their compartmentalized genome architecture with gene-sparse, TE-rich compartments and gene-dense TE-poor regions. As RIP spillage from the TEs often leads to higher mutation rates in neighboring genes (Rouxel et al. 2011), the placement of genes in TE-rich compartments is thought to facilitate their faster evolution (Faino et al. 2016; Raffaele et al. 2010; Wacker et al. 2023; Wang et al. 2017). Even so, TEs may still accommodate genome evolution by inducing gene loss. This is particularly important for fungal plant pathogens as virulence-associated genes such as effector-encoding genes are often enriched in TE-rich regions (Dong et al. 2015; Raffaele and Kamoun 2012). Indeed, TEs have been associated with the loss of the *Avr-Pita* effector in the rice blast fungus *Magnaporthe oryzae* (Chuma et al. 2011), the *Ave1* effector in *Verticillium dahliae* (Faino et al. 2016), and the candidate effector *Zt_8_609* in *Z. tritici* (Hartmann et al. 2017). Loss of these genes provided an advantage to the pathogens in terms of evading effector-triggered immunity mediated by cognate resistance genes in the host. Similarly, we could show that TEs instigated the loss of the *Av4E*, *Avr5*, and *Avr9* effectors in *C. fulvum* to overcome their matching resistance gene in tomato (De Wit et al. 2012; Van den Ackerveken et al. 1992). The precise mechanism by which TEs induce gene loss is often elusive but it has been connected to nonhomologous recombination (Seidl and Thomma 2014). For instance, upon random double-strand breaks induced by ionizing radiation in *Saccharomyces cerevisiae*, chromosome rearrangements, including a nonreciprocal translocation, emerged by homologous recombination between nonallelic Ty1 retrotransposons (Argueso et al. 2008). The authors of this study suggested that the observed chromosomal aberrations could have occurred during DNA repair via the break-induced replication (BIR) pathway. BIR has been associated with restoration of collapsed replication forks by repairing double-strand DNA breaks through invasion into a homologous template (Malkova and Ira 2013; McEachern and Haber 2006). Nonreciprocal translocations can occur via BIR when resection at a double-strand break exposes TEs that allow recombination with other homologous TEs located at ectopic positions (Argueso et al. 2008; VanHulle et al. 2007). We found that loss of the *Avr9* locus in *C. fulvum* isolate Race 2.4.9.11 is due to a nonreciprocal translocation between Chr7 and Chr2, possibly mediated by BIR while using the Ty1/Copia copies as substrate for strand invasion. The location of *Avr9* and of the Ty1/Copia copies in close proximity to the telomeres was likely the key contributing factor to this nonreciprocal translocation and the deletion of *Avr9* in isolates under selection pressure by the tomato *Cf-9* resistance gene. Similarly, we revealed that the borders of the deleted segments carrying *Avr4E* and *Avr5* colocalized with homologous copies of TEs that likely served as template for nonallelic homologous recombination, thus resulting in the deletion of the effectors *Avr4E* and *Avr5*. These findings highlight the importance of TEs and of the genome organization for the evolution of fungal pathogens.

By obtaining four additional near-complete genome assemblies of the tomato pathogen *C. fulvum* and comparing five of them in total, in this study, we provided new insights on the role of repetitive DNA, RIP, and SVs in the evolution of this fungal plant pathogen. Notably, the presence of a Ty1/Copia retroelement likely served as a substrate for a nonreciprocal translocation that resulted in the deletion of the effector gene *Avr9*. Moreover, although nearly all TEs in the genome of *C. fulvum* had footprints of RIP mutations, recent RIP mutations that were variable among isolates appeared to have given rise to genomic islands of high nucleotide variability that increased allelic diversity in nearby genes. Our study also provides evidence of effector gene flow between core and accessory chromosomes that support the hypothesis that accessory chromosomes can gain new functions by acquiring sequences from core chromosomes. Finally, the genomes presented herein are of high value for future comparative genomic analyses and functional studies.

3.4 Materials and methods

3.4.1 Nucleic acid extraction and sequencing

High-molecular weight (HMW) genomic DNA from *C. fulvum* isolates Race 0WU, Race 4, Race 2.4.5.9.11 IPO, and Race 2.4.9.11 was obtained essentially following the protocol of Jones et al. 2019 (Jones et al. 2019). PacBio libraries were multiplexed and sequenced using the HiFi protocol on a Sequel II instrument and one SMRT Cell 8M. Libraries were prepared and sequenced at the DNA Technologies & Expression Analysis Core Lab of the UC Davis Genome Center.

3.4.2 Genome assembly

Quality of the sequenced PacBio HiFi reads of *C. fulvum* isolates Race 0WU, Race 4, Race 2.4.5.9.11 IPO, and Race 2.4.9.11 was assessed with FastQC v0.12.1 (Andrews 2010). Reads were then assembled with Canu v2.2 (Koren et al. 2017) using parameter -pacbio-hifi and genomeSize=70m. Assembled contigs were identified as chromosomes and properly oriented by pairwise alignments performed with NUCmer from the MUMmer package v4 (Marçais et al. 2018) using the 14 chromosomes of *C. fulvum* Race 5 as reference (Zaccaron et al. 2022). Contigs representing the mitochondrial genomes were identified by querying the mitochondrial genes of the fungal pathogen *Erysiphe necator* (Zaccaron et al. 2021) with BLASTn (e-value < 1E-10).

3.4.3 Repetitive DNA annotation

Repetitive DNA was annotated *de novo* for each genome. Specifically, repeat libraries of interspersed repeats were obtained with RepeatModeler v2.0.2 (Flynn et al. 2020) using the parameter *-LTRStruct*. Short tandem repeats were identified with the Tandem Repeats Finder v4.09.1 (Benson 1999). The interspersed repeat libraries were used by RepeatMasker v4.1.2 in sensitive mode (parameter -s) to mask the genomes. Alignments produced by RepeatMasker were used by the script *parseRM.pl* (Kapusta 2023) with parameters *--land 50,1*, *--parse* and *--nrem* to estimate content of repetitive DNA from different classes and families. The script *parseRM.pl* was also used to estimate average divergence of repeat families, which were then used to estimate repeat divergence based on a 20-kb sliding window as described in Zaccaron et al. 2023 (Zaccaron et al. 2023). Genomic regions predicted to be affected by repeat-induced point (RIP) mutations were identified with RIPper (Van Wyk et al. 2019) with default parameters. Specifically, the genomes were analyzed using a 1 kb sliding windows with step size of 500 bp. Windows with composite index (TpA/ApT) - ((CpA + TpG)/(ApC + GpT)) > 0.01, product index TpA/ApT > 1.1, and substrate index (CpA + TpG)/(ApC + GpT) < 0.75 were considered RIPed. RIPed windows were queried with BLASTn (e-value < 1e-20, identity > 50%, query coverage > 20%) against the genome assemblies, and those with a single hit were considered as evidence of RIP leakage toward single-copy regions.

3.4.4 Gene prediction

Predicted gene models of *C. fulvum* isolates Race 5 (Zaccaron et al. 2022) and Race 0WU (De Wit et al. 2012) were mapped to the genomes of isolates Race 0WU, Race 4, Race 2.4.5.9.11 IPO, and Race 2.4.9.11 with liftoff v1.6.3 (Shumate and Salzberg 2020). A round of *ab initio* predictions was performed with Augustus v3.3.3 (Hoff and Stanke 2019) trained to predict the genes of *C. fulvum* Race 5 (Zaccaron et al. 2022). Mapped gene models with more than 50% overlap with interspersed repeats, detected with the script *coverage* from BEDtools v2.30.0 (Quinlan and Hall 2010), were removed. The remaining mapped gene models were analyzed interactively using the script *overlap* from BEDtools in the following approach. First, because the gene annotation of isolate Race 5 (Zaccaron et al. 2022) is overall better compared to the annotation of isolate Race 0WU (De Wit et al. 2012), all mapped gene models from Race 5 were retained. Next, mapped gene models from Race 0WU that did not overlap with mapped gene models from Race 5 were added. Similarly, gene models predicted by Augustus that did not overlap with mapped gene models from Race 5 and Race 0WU were added.

To predict additional genes that could be important for pathogenicity, public RNA-seq data of isolate Race 0WU growing *in vitro* (SRR1171044), and from infections of tomato (cv. Heinz) at 4 dpi (SRR1171035), 8 dpi (SRR1171040), and 12 dpi (SRR1171043), were obtained from NCBI (De Wit et al. 2012). Reads were mapped to the genome of Race 0WU with STAR v2.7.10a (Dobin et al. 2013) with a mapping rate of 94.5%, 0.7%, 3.3%, and 16.8%, respectively. The mapped reads were merged with BAMtools v1.9, and 14,401 transcripts were reconstructed with Stringtie v2.2.1 (Pertea et al. 2015). The nucleotide sequences of the assembled transcripts were obtained with gffread v0.12.7 (Pertea 2023; Pertea and Pertea 2020), and 99,699 open reading frames (ORFs) were predicted with ORFfinder v0.4.3 with minimum ORF size of 180 bp and starting with ATG only. These ORFs were mapped back to the reference genome of Race 0WU with GMAP v2021.08.25 (Wu and Watanabe 2005) to obtain a gff file with their coordinates. BEDtools was used to identify ORFs overlapping with interspersed repeats and already predicted genes in isolate Race 0WU. From the 99,699 ORFs, 80,842 were removed as they overlapped with repeats or existing gene models. From the remaining 18,859 ORFs, 114 had a signal peptide predicted with SignalP6 (Teufel et al. 2022) and further confirmed with DeepLoc v2 (Thumuluri et al. 2022). These 114 ORFs were added as new gene models in the annotation of isolate Race 0WU. Finally, the gene models of isolate Race 0WU were mapped to the genomes of isolates Race 5, Race 4, Race 2.4.9.11, and Race 2.4.5.9.11 IPO with liftoff, and mapped genes that did not overlap with existent genes were added. Gene completeness was estimated with BUSCO v5.4.4 (Simão et al. 2015) in protein mode using the Dothideomycetes_db10 2020-08-05 as reference.

3.4.5 Functional annotation of genes

Genes encoding candidate effectors were predicted as described in (Zaccaron et al. 2022). Briefly, secreted proteins were identified with Signalp5 (Armenteros et al. 2019) and were further classified as effectors with EffectorP v2 (Sperschneider et al. 2016). Specifically, proteins that were shorter than 250 aa, had at least 2% of cysteine residues, no transmembrane domains according to DeepTMHMM (Hallgren et al. 2022) in the mature protein, and no GPI anchors according to PredGPI (Pierleoni et al. 2008), were considered as candidate effectors. GO terms were assigned to genes with the PANNZER2 web server (Törönen et al. 2018), using a positive predictive value of at least 0.4. Genes were assigned to KOG categories using eggNOG-mapper v2.1.9 (Cantalapiedra et al. 2021). Genes encoding CAZymes were predicted with the dbCAN2 meta server (Zhang et al. 2018) using HMMdb v11 and the default threshold values for HMMER (e-value < 1e-15, coverage > 0.35), DIAMOND (e-value < 1e-102), and HMMER (e-value < 1e-15, coverage > 0.35). CAZymes from families previously described to contain PCWDEs (Hage and Rosso 2021) were considered as PCWDEs. Genes encoding proteases were predicted by querying the proteins with BLASTp (e-value < 1E-10) against the MEROPS database v12 (Rawlings et al. 2014). Genes encoding transporters were identified by querying the proteins with BLASTp (e-value < 1E-10) against the transporter classification database v2021-06-20 (Busch and Saier 2002). Genes encoding cytochrome P450s were identified by querying the predicted proteins with the script hmmsearch from HMMER v3.3.2 (e-value< 1E-3) using the HMM model for cytochrome P450 (PF00067) obtained from the PFAM website. Cytochrome P450s were classified based on BLASTp searches (e-value < 1E-10; identity > 40%; query coverage> 40%) against the Dr. Nelson's database of curated fungal cytochrome P450s. Genes encoding key enzymes for secondary metabolism were identified with antiSMASH v7 (Medema et al. 2011).

3.4.6 Identification and visualization of SVs

To detect large-scale SVs, synteny plots of assembled chromosomes were generated based on pairwise gene homology searches implemented in the MCscan pipeline (Tang et al. 2008) within the JCVI utilities libraries (Tang 2023). Confirmation of the chromosomal variations was obtained by mapping the PacBio reads to the genomes using minimap2 v2.24 (Li 2018) with parameters *-ax map-pb*, and visualizing the borders of the SVs in IGV v2.16.2 (Robinson et al. 2011). Dot plots based on pairwise whole-genome alignments were generated with NUCmer from the MUMmer package v4 (Marçais et al. 2018). To detect small-scale SVs, pairwise whole-genome alignments were generated with minimap2 v2.24 (Li 2018) with parameters *-a -x asm5 --cs -r2k*. The alignments were then parsed by SVIM-asm v1.0.3 (Heller and Vingron 2020) with parameters *haploid --min_sv_size 50 --max_sv_size 100000*. Insertions and deletions were extracted and then merged with SURVIVOR v1.0.7 (Jeffares et al. 2017) with parameters adjusted to use maximum distance between breaking points of 100 bp, to take the type and orientation of SVs into account, and minimum SV size of 200 bp. Repetitive DNA content of INDELs was estimated by extracting the INDEL sequences from the output of SURVIVOR and masking them with RepeatMasker using the repetitive DNA library of *C. fulvum* isolate 0WU. Genes overlapping with SVs were identified using the script *overlap* of BEDtools v2.30 (Quinlan and Hall 2010). Plots showing the impact of SVs on genes were generated by extracting homologous regions between two or more genomes, then aligning them using NUCmer (Marçais et al. 2018) while keeping only the best match of each aligned block, and using R v4.3.1 to plot the aligned blocks, genes, and repetitive DNA. To detect duplications between core and accessory chromosomes, Chr14 and Chr15 from isolate Race 0WU were hard masked using the output of RepeatMasker and the *maskfasta* script from BEDtools v2.30 (Quinlan and Hall 2010). The hard-masked sequences were then queried with BLASTn (e-value = 1E-10) against the core chromosomes of isolates Race 5, Race 4, Race 2.4.9.11, and Race 2.4.5.9.11 IPO. The script *intersect* from BEDtools was used to detect genes from core chromosomes that overlapped with BLASTn hits. Genes that overlapped with BLASTn hits were considered duplicated between core and accessory chromosomes.

3.4.7 Gene-based pangenome

Predicted genes were organized into hierarchical orthogroups with OrthoFinder v2.5.3 (Emms and Kelly 2015). Number of shared HOGs were counted and visualized with an UpSet plot (Lex et al. 2014) using the R package UpSetR v1.4.0 (Conway et al. 2017). HOGs containing genes from all five isolates analyzed were considered as the core pangenome. Genes from HOGs not shared by all isolates were considered as accessory genes. The core and pangenome curves were obtained using the linear model function *lm* within R to obtain linear least squares fit of the $\log_e$-transformed sizes of core and pangenome sizes in response to the $\log_e$-transformed sizes of the number of genome combinations.

3.4.8 Gene expression

RNA-seq reads from an isolate Race 0WU-*Solanum lycopersicum* cv. Heinz interaction at 4 dpi (SRR1171035), 8 dpi (SRR1171040), and 12 dpi (SRR1171043) (De Wit et al. 2012), and from isolate Race 0WU grown in potato-dextrose broth (SRR1171044) (De Wit et al. 2012), were mapped to the genome assembly of isolate Race 0WU as described above. Number of paired-end reads mapped to the genes was counted with featureCounts from the subread package v2.0.1 (Liao et al. 2014). Transcripts per million (TPM) values were estimated with a custom R script (Zaccaron et al. 2022).

3.4.9 Nucleotide diversity across chromosomes

The nucleotide diversity across the chromosomes was calculated based on pairwise whole-genome alignments using isolate Race 0WU as reference. Specifically, the genomes of isolates Race 4, Race 5, Race 2.4.9.11, and Race 2.4.5.9.11 IPO were aligned with the genome of isolate Race 0WU using NUCmer and parameters *--maxmatch*, *-c 100*, *-b 500*, and *-l 50*. Alignments were filtered with *delta-filter* with parameters *-m*, *-i 90*, and *-l 100*, and then converted to tabular format with *show-coords* with parameters -*THrd*. The filtered alignments were used by SyRI v1.6.3 (Goel et al. 2019) to identify polymorphisms. SNPs were extracted with the script *vcfasm* that comes with SyRI and then merged into a single VCF file using

BCFtools v1.16 (Danecek et al. 2021). The VCF file was further converted to a genotype matrix using a custom Unix command. The genotype matrix was split into transitions and transversions using the custom R script *split_tstv.R*. The custom R script *calculate_window_pi.R* was then used to calculate the average nucleotide diversity per site of transitions and transversions using a 20 kb sliding window. Dinucleotide bias in regions of high nucleotide diversity was observed by extracting the nucleotides flanking the point mutation and obtaining a sequence logos using WebLogo (Crooks et al. 2004). A phylogenetic tree of the isolates was obtained by selecting SNPs within 14,713 genes present in all 5 isolates using the script *intersect* from BEDtools v2.30.0 (Quinlan and Hall 2010). SNPs were converted to a *fasta* file using the script *phylo* from vcfkit v0.2.9 (Andersen Lab 2023), and a tree was generated with RAxML v8.2.12 (Stamatakis 2014) with parameters *-m ASC_GTRGAMMA*, and *--asc-corr=lewis*. Pairwise number of segregating sites were obtained with the script *snp-dists* v0.8.2 (Seemann 2023).

3.5 Data availability

The genome assemblies of *C. fulvum* isolates Race 2.4.5.9.11 IPO and Race 2.4.9.11, which have no unplaced contigs, have been deposited at NCBI under accessions CP121173-CP121187 and CP120815-CP120829, respectively. The genome assemblies of *C. fulvum* isolates Race 0WU and Race 4, which have unplaced contigs, have been deposited at NCBI under accessions JARNMG010000000 and JARJJH010000000, respectively. The SRA accessions for the PacBio HiFi reads for isolates Race 0WU, Race 4, Race 2.4.5.9.11 IPO, and Race 2.4.9.11 are SRR24302839, SRR23862434, SRR24303573, SRR24303582, respectively. Scripts and code snippets used to generate the results are available at https://github.com/alexzaccaron/2023_cfulv_pangen/. Supplementary files that include *vcf* files of the structural variations and SNPs, repetitive DNA libraries and annotation, hierarchical orthogroups, expression values of all *C. fulvum* Race 0WU genes during interaction with *Solanum lycopersicum* cv. Heinz, and RIP indices values across the genomes are available at Zenodo (https://zenodo.org/doi/10.5281/zenodo.10019509).

Authors' Contributions

A.Z. and I.S. conceived and supervised the project. A.Z performed genome assemblies, gene annotation, and comparative genomics analyses. A. Z. and I.S. wrote and revised the manuscript. All authors read and approved the final manuscript.

Abbreviations

Avr: Avirulence; BIR: Break-induced replication; Chr: Chromosome; HOG: Hierarchical orthogroup; INDEL: Insertion/deletion; RIP: Repeat-induced point; SNP: Single-nucleotide polymorphism; SV: Structural variation; TE: Transposable element; TPM: transcripts per million

Acknowledgements

We thank Professor Matthieu H. A. J. Joosten and Christiaan Schol for providing us with isolates of *C. fulvum*. We are also grateful to Anastasios Samaras for isolating DNA and providing assistance during the sequencing process.

Funding

This work was supported by the National Science Foundation (NSF) Award number 1557995 (IS, AZZ) and by the USDA-NIFA Hatch Project CA-D-PPA-2185-H (IS).

3.6 References

Ahmad, S. F., Jehangir, M., Cardoso, A. L., Wolf, I. R., Margarido, V. P., Cabral-de-Mello, D. C., O'Neill, R., Valente, G. T., and Martins, C. 2020. B chromosomes of multiple species have intense evolutionary dynamics and accumulated genes related to important biological processes. BMC Genomics. 21:1–25

Andersen Lab. 2023. VCF-kit. Available at https://github.com/AndersenLab/VCF-kit.

Andrews, S. 2010. *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Argueso, J. L., Westmoreland, J., Mieczkowski, P. A., Gawel, M., Petes, T. D., and Resnick, M. A. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. Proc. Natl. Acad. Sci. 105:11845–11850

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420–423

Badet, T., Oggenfuss, U., Abraham, L., McDonald, B. A., and Croll, D. 2020. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. BMC Biol. 18:12

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580

Bertazzoni, S., Williams, A. H., Jones, D. A., Syme, R. A., Tan, K.-C., and Hane, J. K. 2018. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic Fungi. Mol. Plant. Microbe Interact. 31:779–788

Boukema, I. 1981. Races of *Cladosporium fulvum* Cke.(*Fulvia fulva*) and genes for resistance in the tomato (Lycopersicon Mill.). Pages 287--292 in: Genetics and breeding of tomato: proceedings of the meeting of the Eucarpia Tomato Working Group, Avignon-France, May 18-21, 1981, Versailles, France: Institut national de la recherche agronomique, 1981.

Bradshaw, R. E., Sim, A. D., Chettri, P., Dupont, P.-Y., Guo, Y., Hunziker, L., McDougal, R. L., Van der Nest, A., Fourie, A., Wheeler, D., and others. 2019. Global population genomics of the forest pathogen *Dothistroma septosporum* reveal chromosome duplications in high dothistromin-producing strains. Mol. Plant Pathol. 20:784–799

Busch, W., and Saier, M. H. 2002. The transporter classification (TC) system, 2002. Crit. Rev. Biochem. Mol. Biol. 37:287–337

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. 38:5825–5829

Chen, H., King, R., Smith, D., Bayon, C., Ashfield, T., Torriani, S., Kanyuka, K., Hammond-Kosack, K., Bieri, S., and Rudd, J. 2023. Combined pangenomics and transcriptomics reveals core and redundant virulence processes in a rapidly evolving fungal plant pathogen. BMC Biol. 21:24

Chuma, I., Isobe, C., Hotta, Y., Ibaragi, K., Futamata, N., Kusaba, M., Yoshida, K., Terauchi, R., Fujita, Y., Nakayashiki, H., and others. 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

Clutterbuck, A. J. 2011. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. Fungal Genet. Biol. 48:306–326

Coleman, J. J., Rounsley, S. D., Rodriguez-Carres, M., Kuo, A., Wasmann, C. C., Grimwood, J., Schmutz, J., Taga, M., White, G. J., Zhou, S., and others. 2009. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. PLoS Genet. 5:e1000618

Conway, J. R., Lex, A., and Gehlenborg, N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 33:2938–2940

Covert, S. F. 1998. Supernumerary chromosomes in filamentous fungi. Curr. Genet. 33:311–319

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

Croll, D., Zala, M., and 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

Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. 2004. WebLogo: A Sequence Logo Generator. Genome Res. 14:1188–1190

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10

De Wit, P. J., Van Der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., Bahkali, A. H., Beenen, H. G., Chettri, P., Cox, M. P., and others. 2012. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. PLoS Genet. 8:e1003088

Demené, A., Laurent, B., Cros-Arteil, S., Boury, C., and Dutech, C. 2021. Chromosomal rearrangements but no change of genes and transposable elements repertoires in an invasive forest-pathogenic fungus. bioRxiv. :2021–03

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21

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

Emms, D. M., and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:1–14

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

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., and Smit, A. F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. 117:9451–9457

Fouché, S., Plissonneau, C., and Croll, D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr. Opin. Microbiol. 46:34–42

Frantzeskakis, L., Kusch, S., and Panstruga, R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. Mol. Plant Pathol. 20:3–7

Galagan, J. E., and Selker, E. U. 2004. RIP: the evolutionary cost of genome defense. TRENDS Genet. 20:417–423

Goel, M., Sun, H., Jiao, W.-B., and Schneeberger, K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20:1–13

Goodwin, S. B., Ben M'Barek, S., Dhillon, B., Wittenberg, A. H., Crane, C. F., Hane, J. K., Foster, A. J., Van der Lee, T. A., Grimwood, J., Aerts, A., and others. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7:e1002070

Gourlie, R., McDonald, M., Hafez, M., Ortega-Polo, R., Low, K. E., Abbott, D. W., Strelkov, S. E., Daayf, F., and Aboukhaddour, R. 2022. The pangenome of the wheat pathogen *Pyrenophora tritici-repentis* reveals novel transposons associated with necrotrophic effectors ToxA and ToxB. BMC Biol. 20:239

Gupta, Y. K., Marcelino-Guimarães, F. C., Lorrain, C., Farmer, A., Haridas, S., Ferreira, E. G. C., Lopes-Caitar, V. S., Oliveira, L. S., Morin, E., Widdison, S., and others. 2023. Major proliferation of transposable elements shaped the genome of the soybean rust pathogen *Phakopsora pachyrhizi*. Nat. Commun. 14:1–16

Hage, H., and Rosso, M.-N. 2021. Evolution of fungal carbohydrate-active enzyme portfolios and adaptation to plant cell-wall polymers. J. Fungi. 7:185

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Almagro Armenteros, J. J., Marcatili, P., Nielsen, H., Krogh, A., and Winther, O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. BioRxiv. :2022–04

Hane, J. K., and Oliver, R. P. 2008. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics. 9:1–12

Hartmann, F. E. 2022. Using structural variants to understand the ecological and evolutionary dynamics of fungal plant pathogens. New Phytol. 234:43–49

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

He, C., Rusu, A. G., Poplawski, A. M., Irwin, J. A., and Manners, J. M. 1998. Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*. Genetics. 150:1459–1466

Heller, D., and Vingron, M. 2020. SVIM-asm: structural variant detection from haploid and diploid genome assemblies. Bioinformatics. 36:5519–5521

Hoff, K. J., and Stanke, M. 2019. Predicting genes in single genomes with AUGUSTUS. Curr. Protoc. Bioinforma. 65:e57

Houben, A., Banaei-Moghaddam, A. M., Klemme, S., and Timmis, J. N. 2014. Evolution and biology of supernumerary B chromosomes. Cell. Mol. Life Sci. 71:467–478

Irelan, J. T., and Selker, E. U. 1996. Gene silencing in filamentous fungi: RIP, MIP and quelling. J. Genet. 75:313–324

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

Jones, A., Nagar, R., Sharp, A., and Schwessinger, B. 2019. High-molecular weight DNA extraction from challenging fungi using CTAB and gel purification. protocols.io.

Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J., and Jones, J. D. 1994. Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science. 266:789–793

de Jonge, R., Bolton, M. D., Kombrink, A., van den Berg, G. C., Yadeta, K. A., and Thomma, B. P. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23:1271–1282

van Kan, J. A., Van den Ackerveken, G., and De Wit, P. 1991. Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. Mol Plant-Microbe Interact. 4:52–59

Kapusta, A. 2023. Parsing-RepeatMasker-Outputs. Available at https://github.com/4ureliek/Parsing-RepeatMasker-Outputs.

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

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736

Latorre, S. M., Reyes-Avila, C. S., Malmgren, A., Win, J., Kamoun, S., and Burbano, H. A. 2020. Differential loss of effector genes in three recently expanded pandemic clonal lineages of the rice blast fungus. BMC Biol. 18:1–15

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. 2014. UpSet: visualization of intersecting sets. IEEE Trans. Vis. Comput. Graph. 20:1983–1992

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

Liao, Y., Smyth, G. K., and Shi, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 30:923–930

Lindhout, P., Korta, W., Cislik, M., Vos, I., and Gerlagh, T. 1989. Further identification of races of *Cladosporium fulvum* (*Fulvia fulva*) on tomato originating from the Netherlands France and Poland. Neth. J. Plant Pathol. 95:143–148

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., and others. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 464:367–373

Malkova, A., and Ira, G. 2013. Break-induced replication: functions and molecular mechanism. Curr. Opin. Genet. Dev. 23:271–279

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. 2018. MUMmer4: A fast and versatile genome alignment system. PLoS Comput. Biol. 14:e1005944

Martis, M. M., Klemme, S., Banaei-Moghaddam, A. M., Blattner, F. R., Macas, J., Schmutzer, T., Scholz, U., Gundlach, H., Wicker, T., Šimková, H., and others. 2012. Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. Proc. Natl. Acad. Sci. 109:13343–13346

Mat Razali, N., Cheah, B. H., and Nadarajah, K. 2019. Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. Int. J. Mol. Sci. 20:3597

McDonald, M. C., Taranto, A. P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., Milgate, A., and Solomon, P. S. 2019. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. MBio. 10:e01515-19

McEachern, M. J., and Haber, J. E. 2006. Break-induced replication and recombinational telomere elongation in yeast. Annu Rev Biochem. 75:111–135

Medema, M. H., Blin, K., Cimermancic, P., De Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39:W339–W346

Mesarich, C. H., Barnes, I., Bradley, E. L., de la Rosa, S., de Wit, P. J. G. M., Guo, Y., Griffiths, S. A., Hamelin, R. C., Joosten, M. H. A. J., Lu, M., McCarthy, H. M., Schol, C. R., Stergiopoulos, I., Tarallo, M., Zaccaron, A. Z., and Bradshaw, R. E. 2023. Beyond the genomes of *Fulvia fulva* (syn. *Cladosporium fulvum*) and *Dothistroma septosporum*: New insights into how these fungal pathogens interact with their host plants. Mol. Plant Pathol. 24:474–494

Mesarich, C. H., Griffiths, S. A., van der Burgt, A., Ökmen, B., Beenen, H. G., Etalo, D. W., Joosten, M. H., and de Wit, P. J. 2014. Transcriptome sequencing uncovers the *Avr5* avirulence gene of the tomato leaf mold pathogen *Cladosporium fulvum*. Mol. Plant. Microbe Interact. 27:846–857

Mesarich, C. H., Ökmen, B., Rovenich, H., Griffiths, S. A., Wang, C., Karimi Jashni, M., Mihajlovski, A., Collemare, J., Hunziker, L., Deng, C. H., and others. 2018. Specific hypersensitive response–associated recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato. Mol. Plant. Microbe Interact. 31:145–162

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

Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., and Ginalski, K. 2019. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci. Rep. 9:4307

Olarte, R. A., Menke, J., Zhang, Y., Sullivan, S., Slot, J. C., Huang, Y., Badalamenti, J. P., Quandt, A. C., Spatafora, J. W., and Bushley, K. E. 2019. Chromosome rearrangements shape the diversification of secondary metabolism in the cyclosporin producing fungus *Tolypocladium inflatum*. BMC Genomics. 20:1–23

Pedersen, C., van Themaat, E. V. L., McGuffin, L. J., Abbott, J. C., Burgis, T. A., Barton, G., Bindschedler, L. V., Lu, X., Maekawa, T., Weßling, R., and others. 2012. Structure and evolution of barley powdery mildew effector candidates. BMC Genomics. 13:694

Pertea, G. 2023. GffRead. Available at https://github.com/gpertea/gffread.

Pertea, G., and Pertea, M. 2020. GFF Utilities: GffRead and GffCompare [version 1; peer review: 3 approved]. F1000Research. 9

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290–295

Pierleoni, A., Martelli, P. L., and Casadio, R. 2008. PredGPI: a GPI-anchor predictor. BMC Bioinformatics. 9:392

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., and Croll, D. 2017. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. Front. Plant Sci. 8:119

Quinlan, A. R., and 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., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., and others. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science. 330:1540–1543

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

Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42:D503–D509

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. 2011. Integrative genomics viewer. Nat. Biotechnol. 29:24–26

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., Van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., and others. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:1–10

Schikora-Tamarit, M. À., and Gabaldón, T. 2022. Using genomics to understand the mechanisms of virulence and drug resistance in fungal pathogens. Biochem. Soc. Trans. 50:1259–1268

Seemann, T. 2023. snp-dists. Available at https://github.com/tseemann/snp-dists.

Seidl, M. F., and Thomma, B. P. 2017. Transposable elements direct the coevolution between plants and microbes. Trends Genet. 33:842–851

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

Selker, E. U. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu. Rev. Genet. 24:579–613

Selker, E. U. 2002. Repeat-induced gene silencing in fungi. Adv. Genet. 46:439–450

Shumate, A., and Salzberg, S. L. 2020. Liftoff: accurate mapping of gene annotations. Bioinformatics.

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

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743–761

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

Stergiopoulos, I., De Kock, M. J., Lindhout, P., and De Wit, P. J. 2007a. Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. Mol. Plant. Microbe Interact. 20:1271–1283

Stergiopoulos, I., Groenewald, M., Staats, M., Lindhout, P., Crous, P. W., and De Wit, P. J. 2007b. Mating-type genes and the genetic structure of a world-wide collection of the tomato pathogen *Cladosporium fulvum*. Fungal Genet. Biol. 44:415–429

Stukenbrock, E. H., and McDonald, B. A. 2009. Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. Mol. Plant. Microbe Interact. 22:371–380

Syme, R. A., Tan, K.-C., Rybak, K., Friesen, T. L., McDonald, B. A., Oliver, R. P., and Hane, J. K. 2018. Pan-*Parastagonospora* comparative genome analysis—effector prediction and genome evolution. Genome Biol. Evol. 10:2443–2457

Talbot, N. J., Salch, Y. P., Ma, M., and Hamer, J. E. 1993. Karyotypic variation within clonal lineages of the rice blast fungus, *Magnaporthe grisea*. Appl. Environ. Microbiol. 59:585–593

Tang, H. 2023. JCVI utility libraries. Available at https://github.com/tanghaibao/jcvi.

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. 2008. Synteny and collinearity in plant genomes. Science. 320:486–488

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 40:1023–1025

Thomma, B. P., Van Esse, H. P., Crous, P. W., and de Wit, P. J. 2005. *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. Mol. Plant Pathol. 6:379–393

Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H., and Winther, O. 2022. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. Nucleic Acids Res. 50:W228–W234

Tobias, P. A., Schwessinger, B., Deng, C. H., Wu, C., Dong, C., Sperschneider, J., Jones, A., Luo, Z., Zhang, P., Sandhu, K., and others. 2021. *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. G3. 11:jkaa015

Törönen, P., Medlar, A., and Holm, L. 2018. PANNZER2: a rapid functional annotation web server. Nucleic Acids Res. 46:W84–W88

Torres, D. E., Oggenfuss, U., Croll, D., and Seidl, M. F. 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. Fungal Biol. Rev.

Torres, D. E., Thomma, B. P., and Seidl, M. F. 2021. Transposable elements contribute to genome dynamics and gene expression variation in the fungal plant pathogen *Verticillium dahliae*. Genome Biol. Evol. 13:evab135

Tsushima, A., Gan, P., Kumakura, N., Narusaka, M., Takano, Y., Narusaka, Y., and Shirasu, K. 2019. Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. Genome Biol. Evol. 11:1487–1500

Van den Ackerveken, G. F., Van Kan, J. A., and De Wit, P. J. 1992. Molecular analysis of the avirulence gene *avr9* of the fungal tomato pathogen *Cladosporium fulvum* fully supports the gene-for-gene hypothesis. Plant J. 2:359–366

Van Wyk, S., Harrison, C. H., Wingfield, B. D., De Vos, L., van Der Merwe, N. A., and Steenkamp, E. T. 2019. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. PeerJ. 7:e7447

VanHulle, K., Lemoine, F. J., Narayanan, V., Downing, B., Hull, K., McCullough, C., Bellinger, M., Lobachev, K., Petes, T. D., and Malkova, A. 2007. Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. Mol. Cell. Biol. 27:2601–2614

Wacker, T., Helmstetter, N., Wilson, D., Fisher, M. C., Studholme, D. J., and Farrer, R. A. 2023. Two-speed genome evolution drives pathogenicity in fungal pathogens of animals. Proc. Natl. Acad. Sci. 120:e2212633120

Wang, C., Skrobek, A., and Butt, T. M. 2003. Concurrence of losing a chromosome and the ability to produce destruxins in a mutant of *Metarhizium anisopliae*. FEMS Microbiol. Lett. 226:373–378

Wang, L., Sun, Y., Sun, X., Yu, L., Xue, L., He, Z., Huang, J., Tian, D., Hurst, L. D., and Yang, S. 2020. Repeat-induced point mutation in *Neurospora crassa* causes the highest known mutation rate and mutational burden of any cellular life. Genome Biol. 21:1–23

Wang, Q., Jiang, C., Wang, C., Chen, C., Xu, J.-R., and Liu, H. 2017. Characterization of the two-speed subgenomes of *Fusarium graminearum* reveals the fast-speed subgenome specialized for adaption and infection. Front. Plant Sci. 8:140

Watters, M. K., Randall, T. A., Margolin, B. S., Selker, E. U., and Stadler, D. R. 1999. Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. Genetics. 153:705–714

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N. D., and others. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. 37:1155–1162

Westerink, N., Brandwagt, B. F., De Wit, P. J., and Joosten, M. H. 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

de Wit, P. J. 2016. *Cladosporium fulvum* effectors: weapons in the arms race with tomato. Annu. Rev. Phytopathol. 54:1–23

Witte, T. E., Villeneuve, N., Boddy, C. N., and Overy, D. P. 2021. Accessory chromosome-acquired secondary metabolism in plant pathogenic fungi: the evolution of biotrophs into host-specific pathogens. Front. Microbiol. 12

Wu, T. D., and Watanabe, C. K. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 21:1859–1875

Wyka, S., Mondo, S., Liu, M., Nalam, V., and Broders, K. 2022. A large accessory genome and high recombination rates may influence global distribution and broad host range of the fungal plant pathogen *Claviceps purpurea*. PloS One. 17:e0263496

Zaccaron, A. Z., Chen, L.-H., Samaras, A., and Stergiopoulos, I. 2022. A chromosome-scale genome assembly of the tomato pathogen *Cladosporium fulvum* reveals a compartmentalized genome architecture and the presence of a dispensable chromosome. Microb. Genomics. 8:000819

Zaccaron, A. Z., De Souza, J. T., and Stergiopoulos, I. 2021. The mitochondrial genome of the grape powdery mildew pathogen *Erysiphe necator* is intron rich and exhibits a distinct gene organization. Sci. Rep. 11:13924

Zaccaron, A. Z., Neill, T., Corcoran, J., Mahaffee, W. F., and Stergiopoulos, I. 2023. A chromosome-scale genome assembly of the grape powdery mildew pathogen *Erysiphe necator* reveals its genomic architecture and previously unknown features of its biology. Mbio. :e00645-23

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46:W95–W101

3.8 Supplementary materials

3.8.1 Supplementary figures

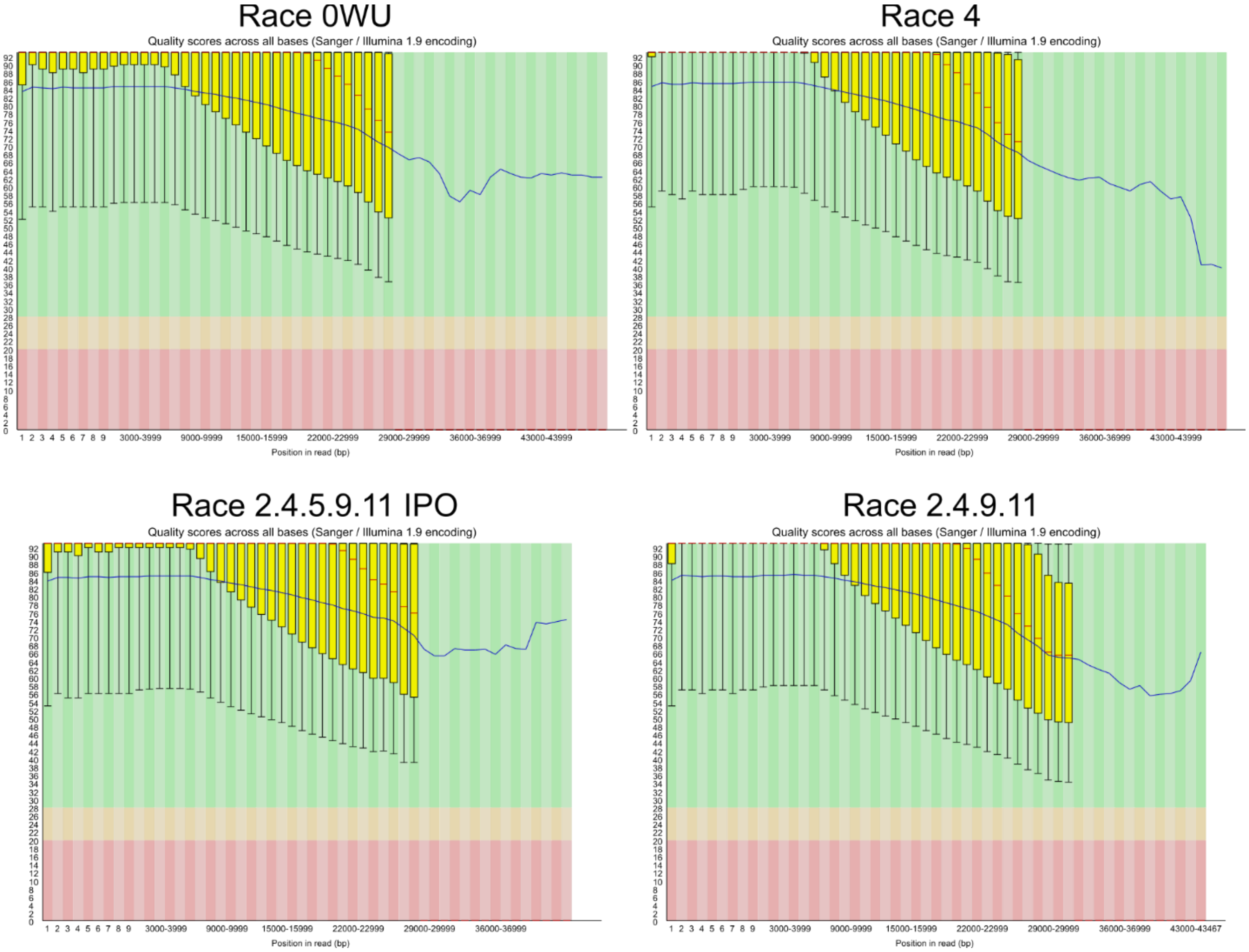


Figure 3.S1: Quality of the sequenced PacBio HiFi reads of five *Cladosporium fulvum* isolates. The figure shows box plots representing the distribution of quality scores (Y-axis) along the HiFi reads (X-axis) of *C. fulvum* isolates Race 0WU, Race 4, Race 2.4.5.9.11 IPO, and Race 2.4.9.11 . Blue lines indicate average quality scores. Plots were generated with FastQC.

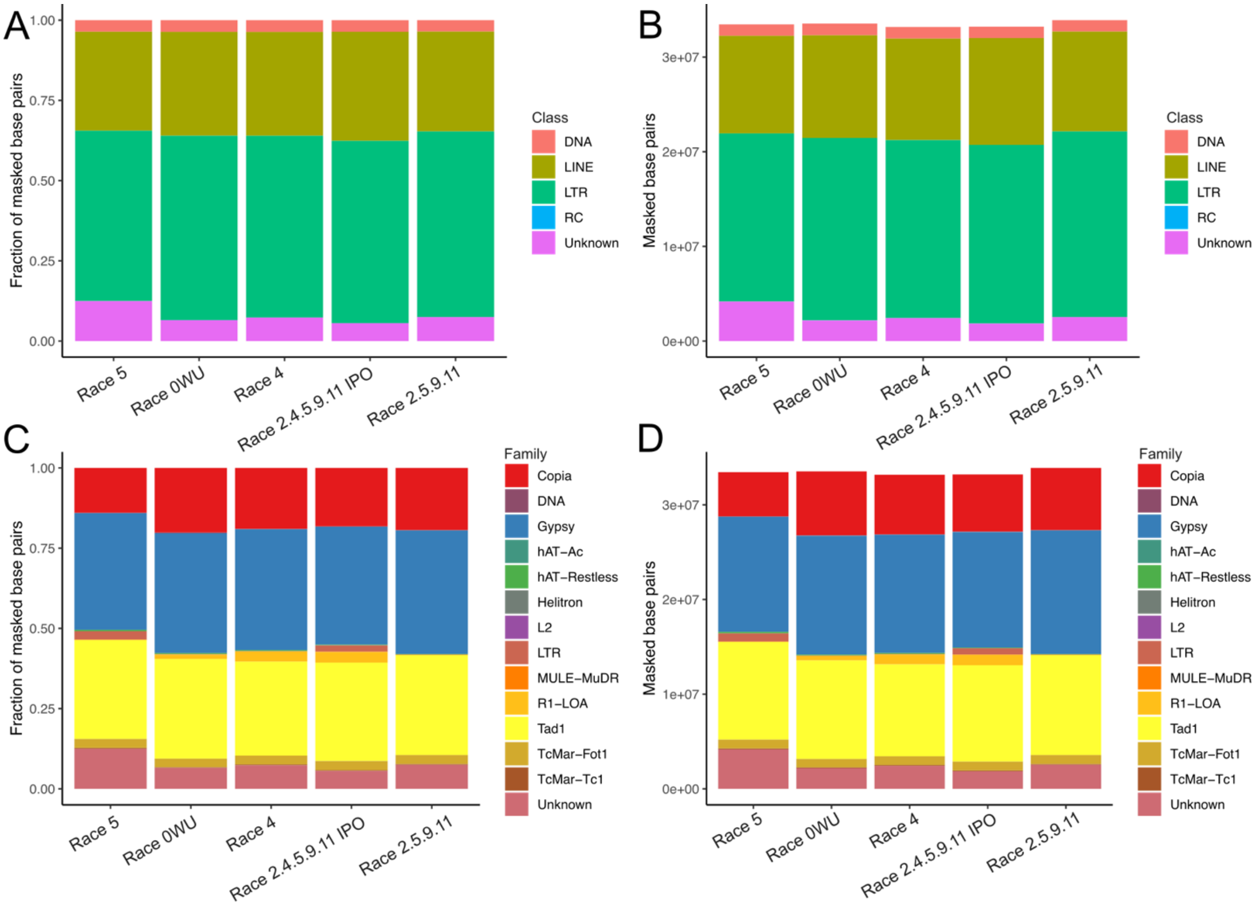


Figure 3.S2: The genomes of five *Cladosporium fulvum* isolates have similar complements of predicted transposable elements (TEs). The bar plots show the overall abundance of TEs in the genomes of *C. fulvum*. (A) The percentage and the total number of masked bases, organized into different TE classes. (B) The percentage and the total number of masked bases, organized into different TE (sub)families.

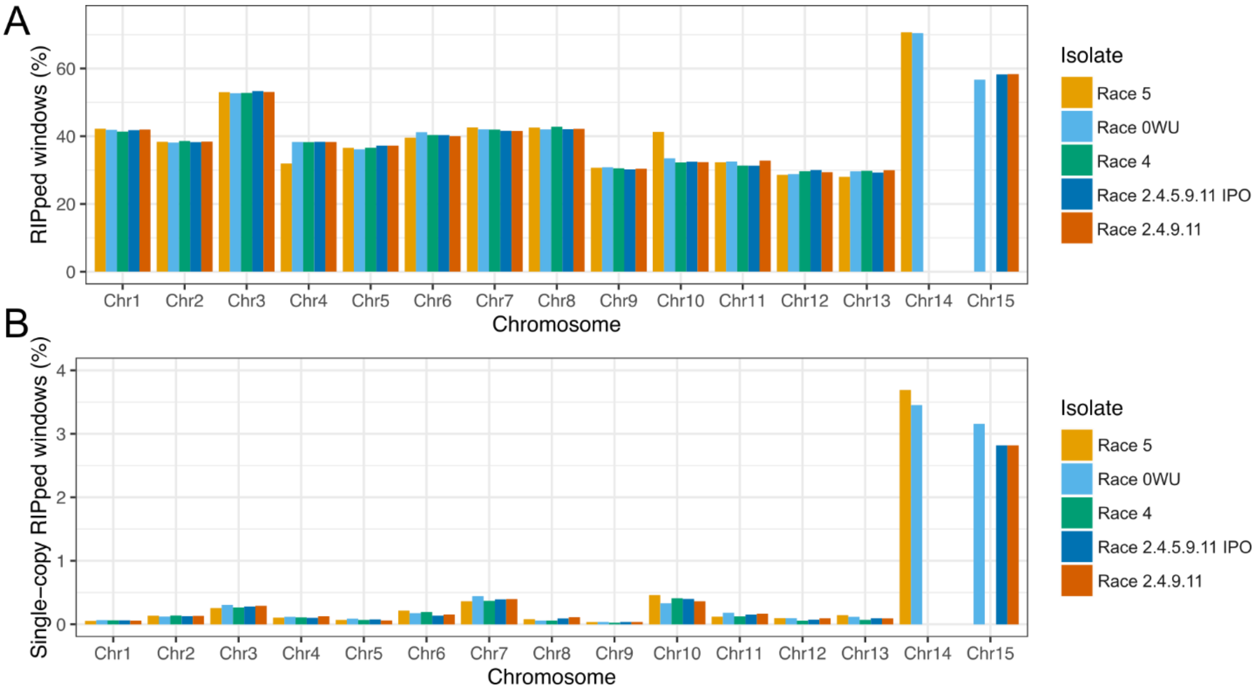


Figure 3.S3: The chromosomes of five *Cladosporium fulvum* isolates are heavily affected by Repeat-Induced Point (RIP) mutations. The figure shows percentages of scanned sliding windows with evidence of RIP (i.e., RIP indices substrate < 0.75, product > 1.1, and composite > 0) for homologous chromosomes. (A) Percentage of total windows with evidence of RIP. (B) Percentage of single-copy windows (BLASTn e-value < 1e-20, identity > 50%, query coverage > 20%) with evidence of RIP. The dispensable chromosomes, Chr14 and Chr15, exhibit higher percentage of sequences affected by RIP and exhibit higher percentage of single-copy regions affected by RIP, which indicates more frequent RIP slippage compared to core chromosomes. The RIP indices scan analysis was performed with a sliding window of 1 kb and a step size of 500 bp.

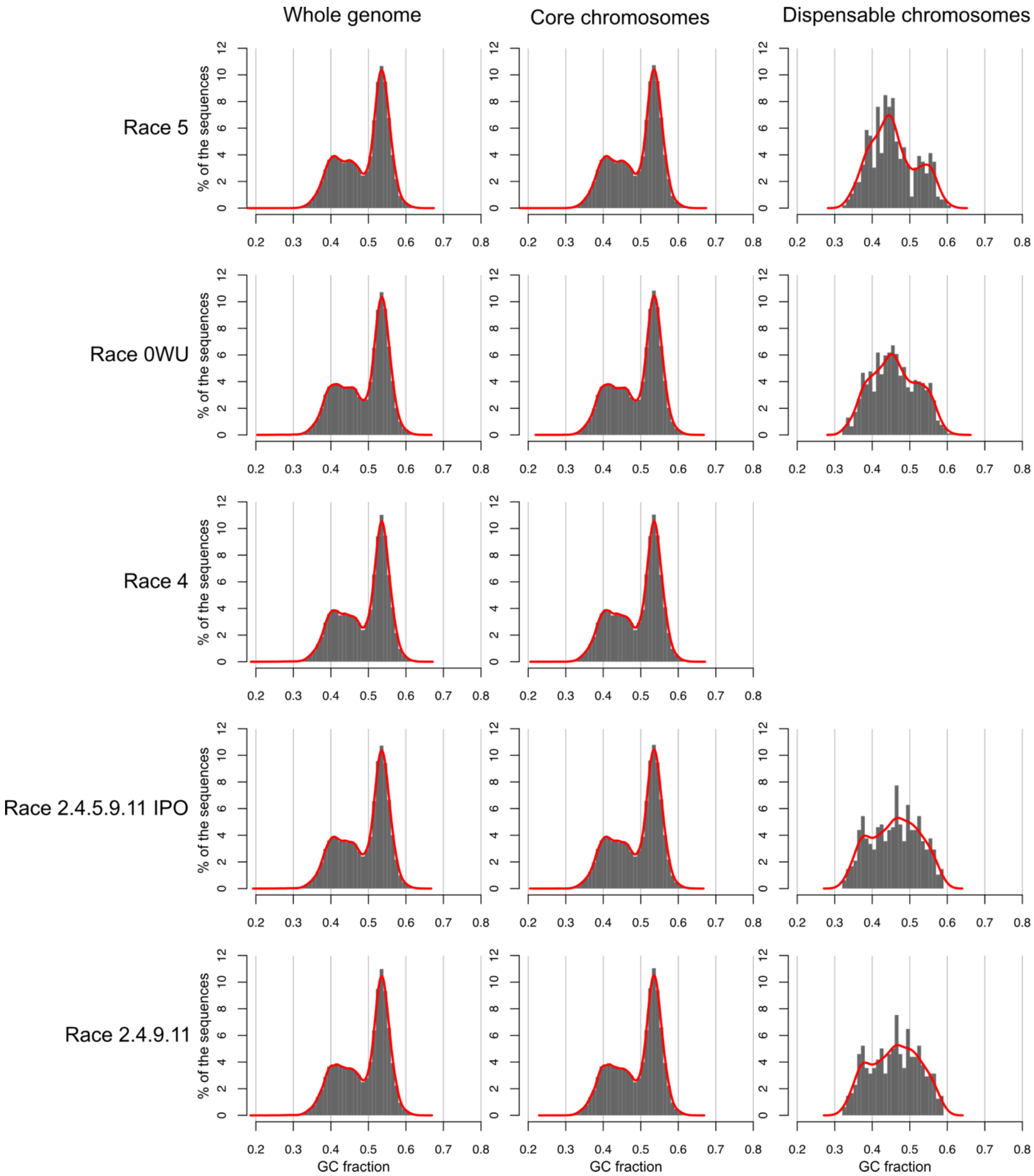


Figure 3.S4: Bimodal GC content distribution of five *Cladosporium fulvum* genomes. The histograms show the percentages (Y-axis) of the whole genome, core chromosomes (Chr1-to-Chr13), and dispensable chromosomes (Chr14 and/or Chr15) containing the indicated GC fractions (X-axis). The five genomes show a bimodal distribution of the whole genome and core chromosomes with peaks at 0.42 GC (42% GC) and 0.54 GC (54% GC). In contrast, bimodal distribution of the GC content in dispensable chromosomes is less evident. The GC content was calculated using a sliding window of 1 kb. Lines represent estimated distributions based on the kernel density of the histograms. GC distribution of dispensable chromosomes for isolate Race 4 is not shown because this isolate has no dispensable chromosomes.

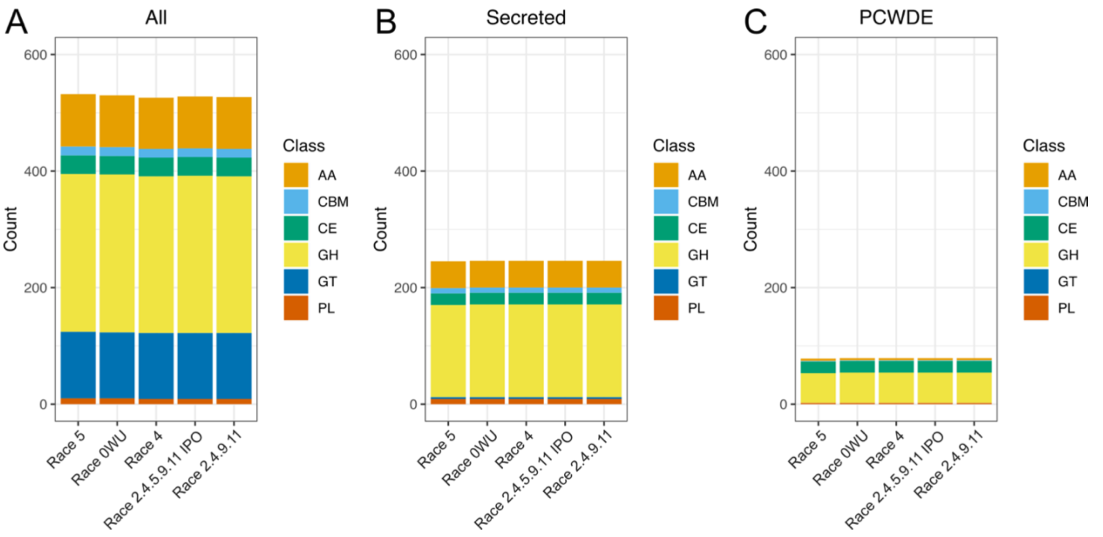


Figure 3.S5: Number of genes encoding carbohydrate-active enzymes (CAZymes) in five *Cladosporium fulvum* genomes. The bar plots show number of CAZyme modules for (A) all genes encoding CAZymes, (B) genes encoding secreted CAZymes, and (C) genes encoding CAZymes associated with plant cell wall degrading enzymes (PCWDEs). CAZymes were classified based on the presence of auxiliary activity (AA) modules, glycoside hydrolase (GH) modules, glycosyl transferase (GT) modules, polysaccharide lyase (PL) modules carbohydrate esterase (CE) modules, and carbohydrate-binding module (CBM) modules.

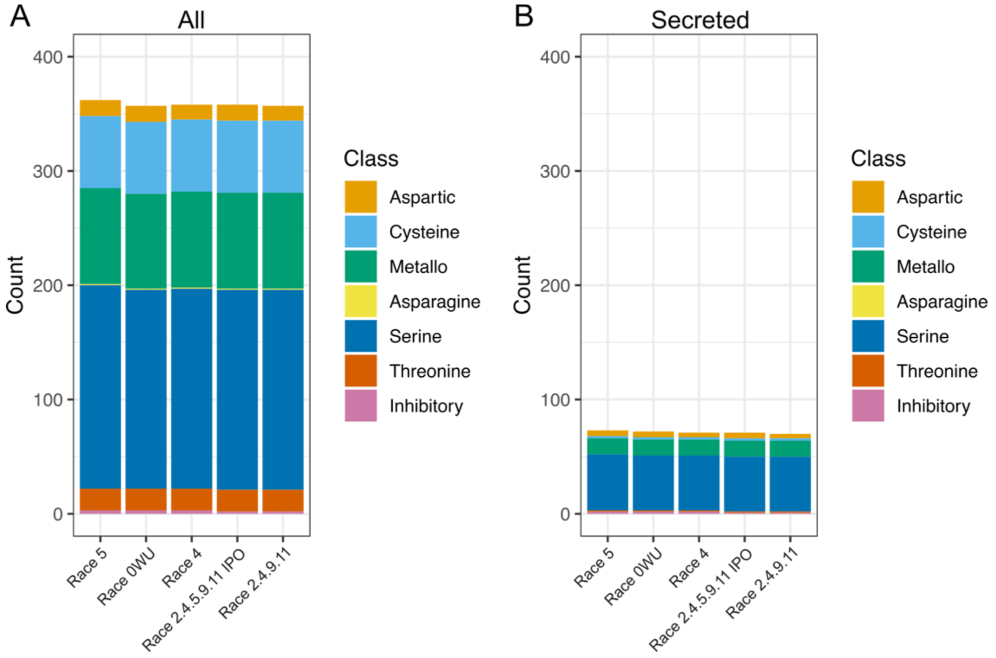


Figure 3.S6: Number of genes encoding proteases in five *Cladosporium fulvum* genomes. The bar plots show the number of (A) all genes encoding proteases and (B) genes encoding secreted proteases from the seven main classes of proteases.

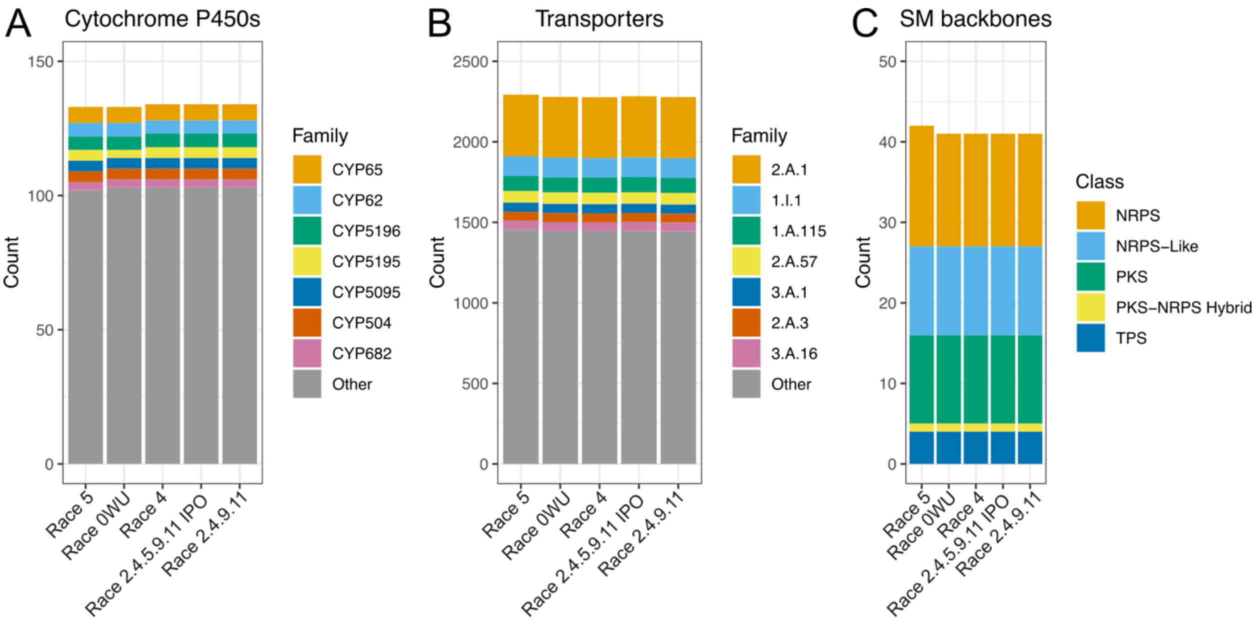


Figure 3.S7: Number of genes encoding cytochrome P450s, transporters, and key enzymes for secondary metabolite biosynthesis (SM) in five *Cladosporium fulvum* genomes. The bar plots show the number of genes encoding (A) cytochrome P450s, (B) transporters, and (C) key enzymes for SM biosynthesis. Cytochrome P450s are classified based on homology searches (BLASTn 40% identity, 50% query coverage, e-value< 1E-10) against Dr. Nelson's database of curated fungal cytochrome P450s (https://drnelson.uthsc.edu/P450seqs.dbs.html). The seven most abundant classes are indicated. Transporters were classified based on homology searches (e-value< 1E-10) against the Transporter Classification Databse (TCDB; https://tcdb.org). The seven most abundant families are indicated. These include the major facilitator (MFS) superfamily (2.A.1), the nuclear pore complex (NPC) family (1.I.1), the pore-forming NADPH-dependent 1-acyldihydroxyacetone phosphate reductase (Ayr1) family Ayr1 (1.A.115), the equilibrative nucleoside transporter (ENT) family ENT (2.A.57), The ATP-binding cassette (ABC) superfamily ABC (3.A.1), the amino acid-polyamine-organocation (APC) family APC (2.A.3), and the endoplasmic reticular retrotranslocon (ER-RT) family ER-RT (3.A.16). Key enzymes for the biosynthesis of SMs are classified into non-ribosomal peptide synthetases (NRPS), NRPS-Like, type 1 polyketide synthases (PKS), PKS-NRPS hybrid, and terpene synthases (TPS).

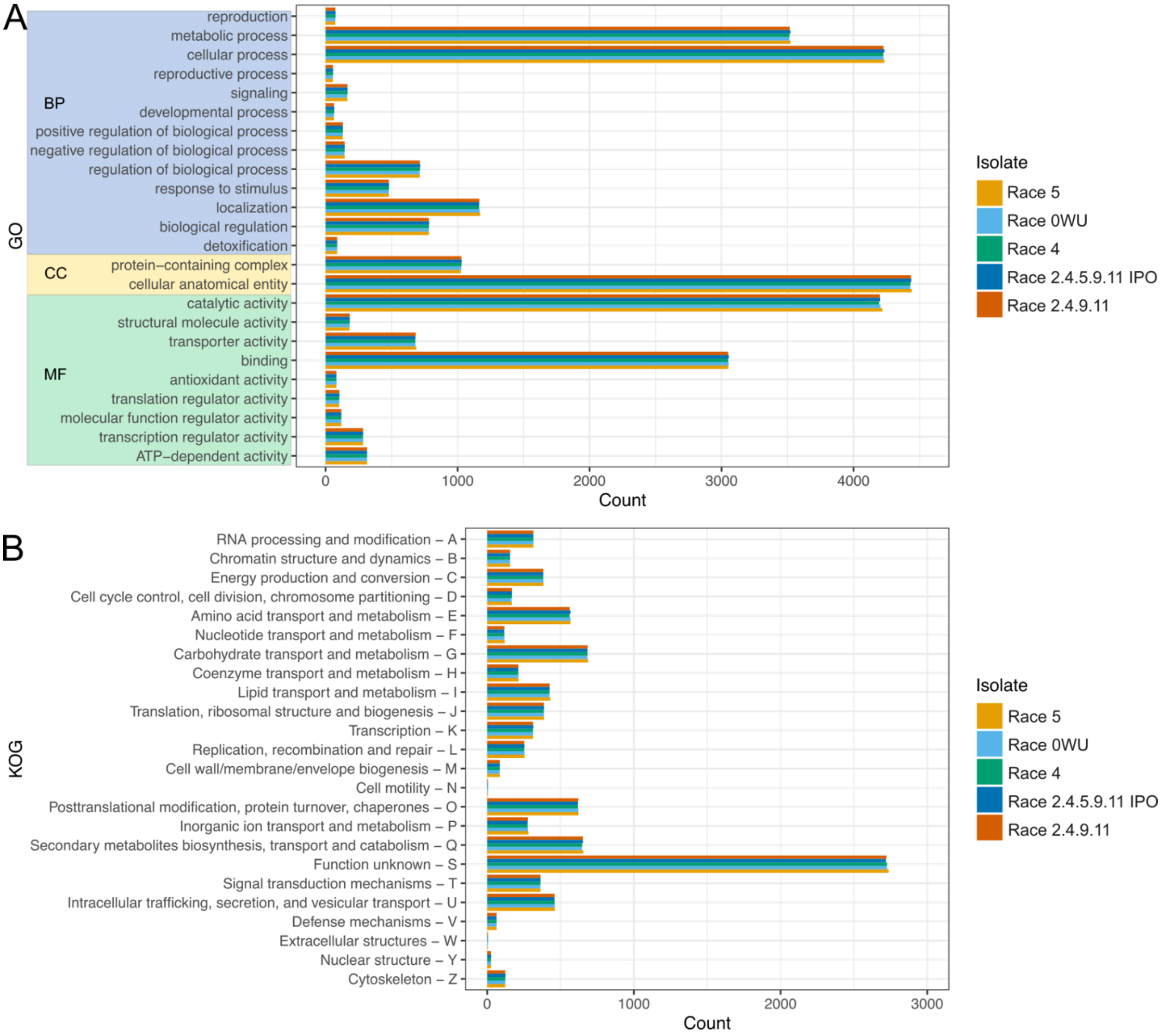


Figure 3.S8: Number of genes in five *Cladosporium fulvum* genomes assigned to different Gene Ontology (GO) terms and EuKaryotic Ortholog Group (KOG) categories. (A) Number of genes from main GO terms from the Biological Process (B), Cellular Component (CC), and Molecular Function (MF) classes. (B) Number of genes assigned to 24 main KOG classes.

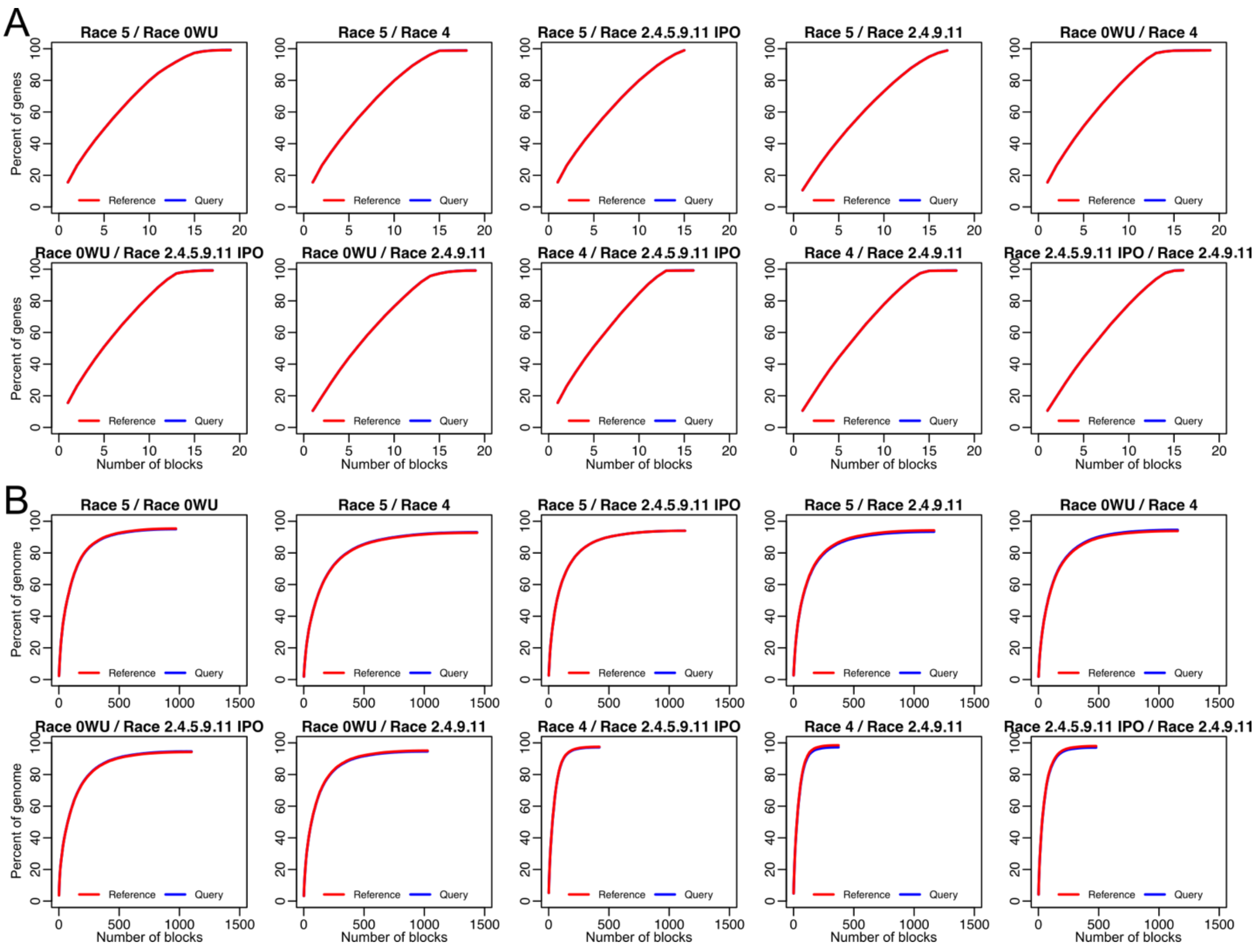


Figure 3.S9: Overall number of pairwise synteny blocks in pairwise alignments of five *Cladosporium fulvum* genomes. The figure shows line plots that represent the minimum number of syntenic blocks between reference and query (shown at the top of each plot in the format reference / query) to cover the percentage of genes (A) and the whole genome (B), of the reference and query. Plots in (A) represent syntenic blocks at the gene level, i.e., the syntenic blocks were obtained by matching orthologous genes between reference and query. Plots in (B) represent syntenic blocks at the DNA level, i.e., the genome of the reference and query were aligned with NUCmer and the best one-to-one aligned blocks were considered.

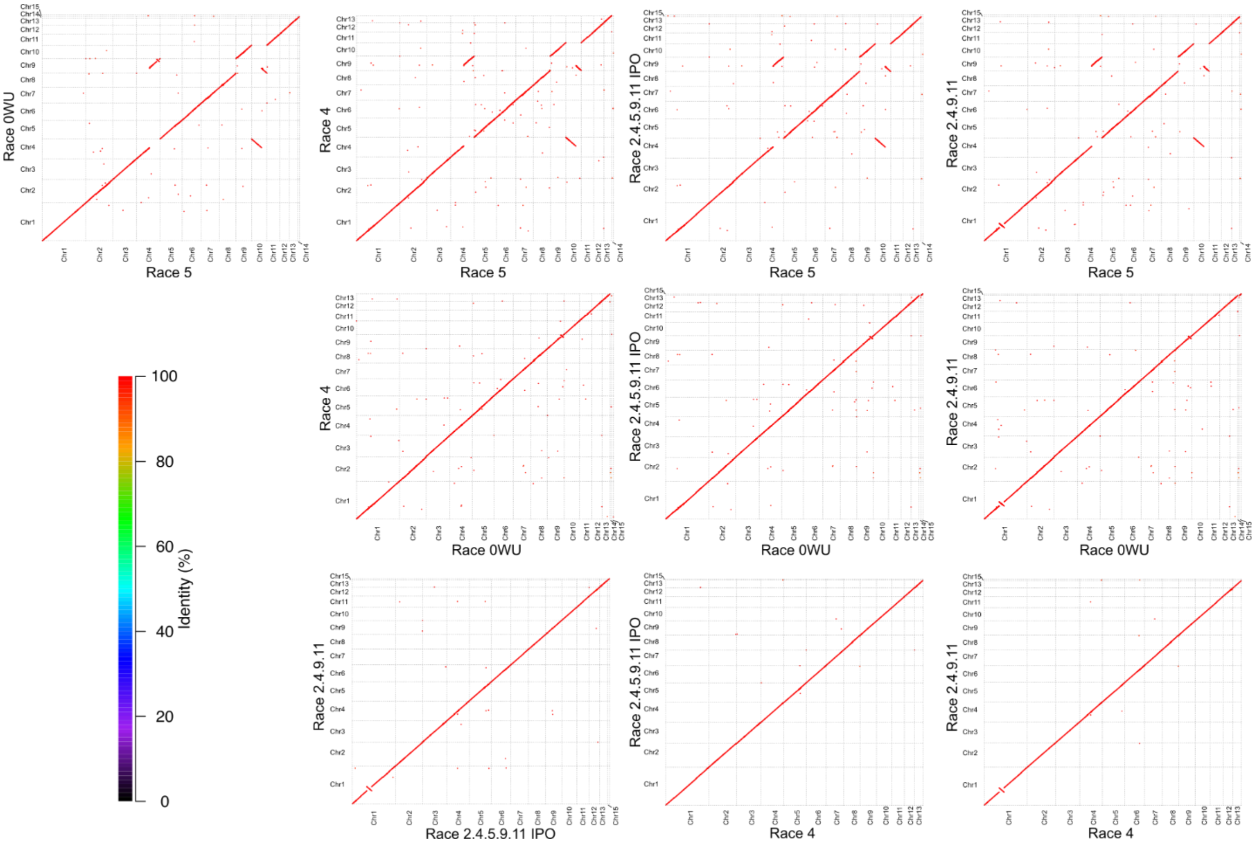


Figure 3.S10: Alignment dot plots showing pairwise syntenic regions among *Cladosporium fulvum* genomes. The figure shows that the five genomes of *C. fulvum* have long syntenic regions. The dot plots also show a reciprocal translocation between Chr4 and Chr10 in isolate Race 5, an inversion in Chr10 in isolate Race 0WU, and another inversion in Chr1 in isolate Race 2.4.9.11. Alignments were generated with NUCmer at the DNA level.

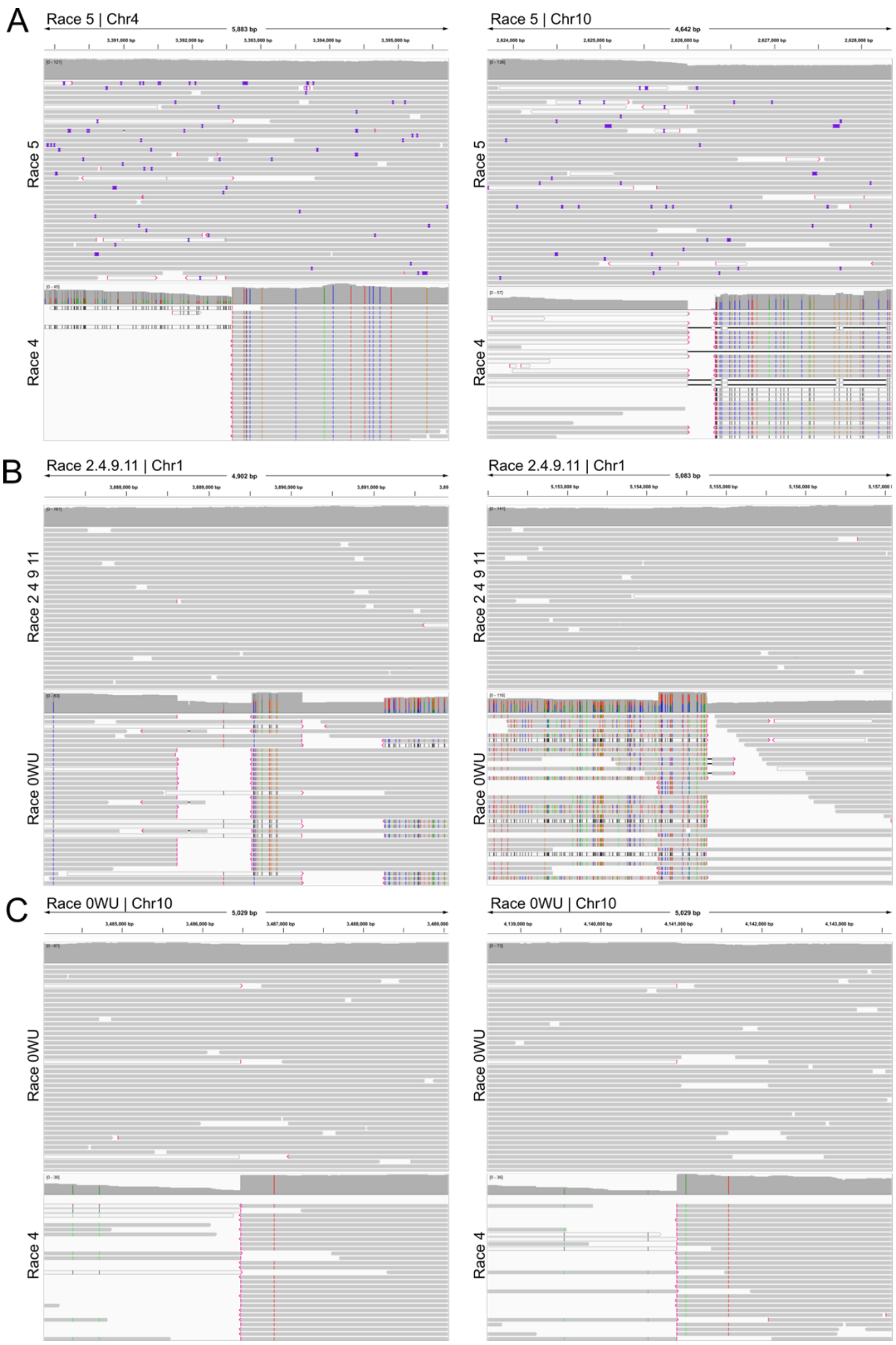


Figure 3.S11: No evidence of misassembly at the borders of large-scale structural variations in isolates of *Cladosporium fulvum*. The figure shows the location of the left-end and right-end side borders of the (A) reciprocal translocation between Chr4 and Ch10 of isolate Race 5, (B) large inversion in Chr1 of isolate Race 2.4.9.11, and (C) large inversion in Chr1 of isolate 0WU. PacBio reads mapped to each border are shown. Reads from the same isolate harboring the large-scale structural variation shows no evidence of misassembly, whereas reads from another isolate not harboring the large-scale structural variation are hard clipped around the structural variation borders.

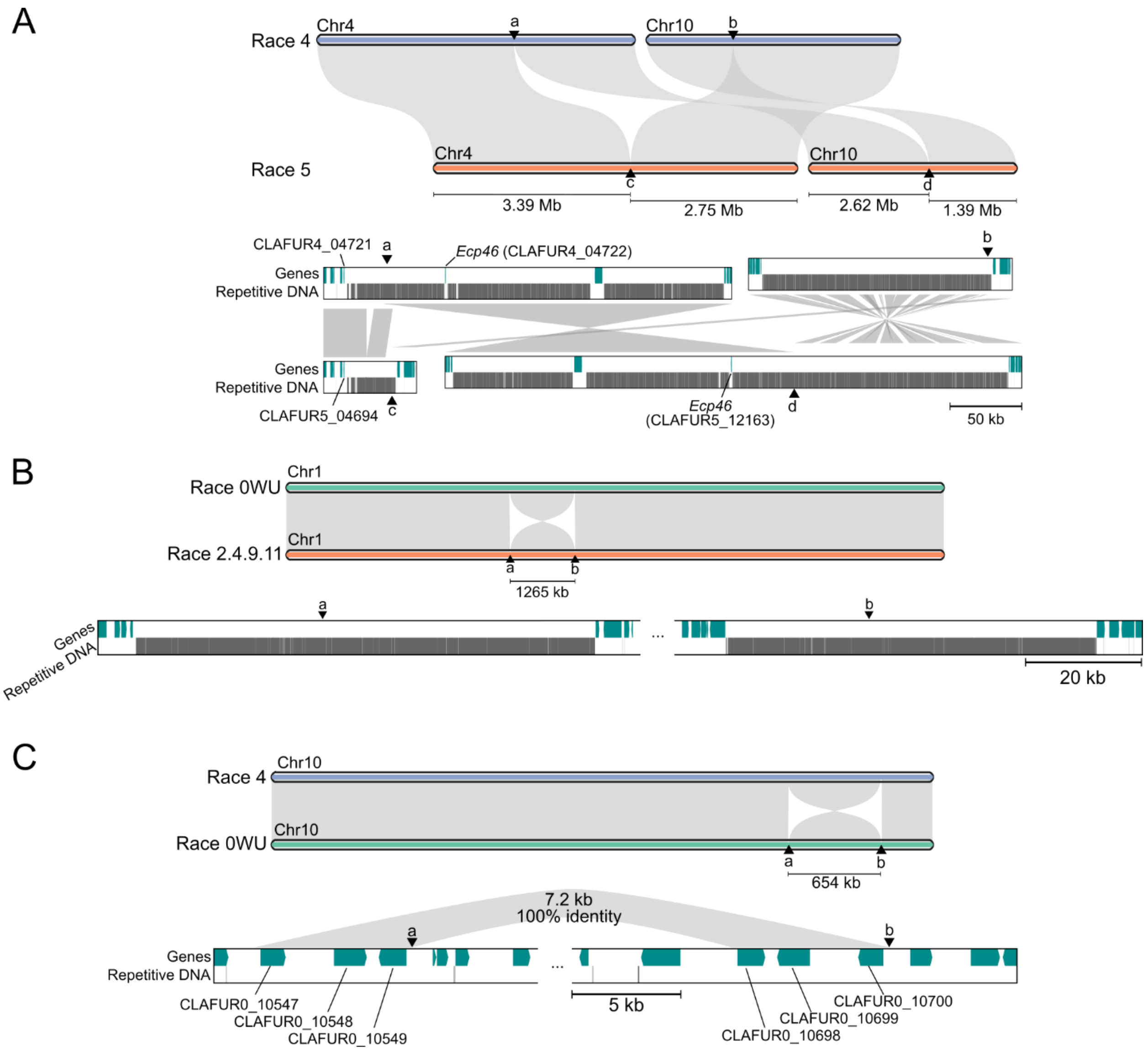


Figure 3.S12: Three large-scale chromosomal structural variations were identified among the five isolates of *Cladosporium fulvum*. Chromosomes are represented as thick solid lines with ribbons connecting homologous regions. Synteny break points are pointed with triangles. Zoomed -in locations of the synteny break points are shown below the synteny plots. (A) The balanced reciprocal translocation between Chr4 and Chr10 in isolate Race 5. This translocation affected the intergenic region of the candidate effector gene *Ecp46*. (B) the inversion present in Chr1 of isolate Race 2.4.9.11. (C) The inversion present in Chr10 of isolate Race 0WU. The break points of this inversion colocalize with a duplicated segment of 7.2 kb containing three genes.

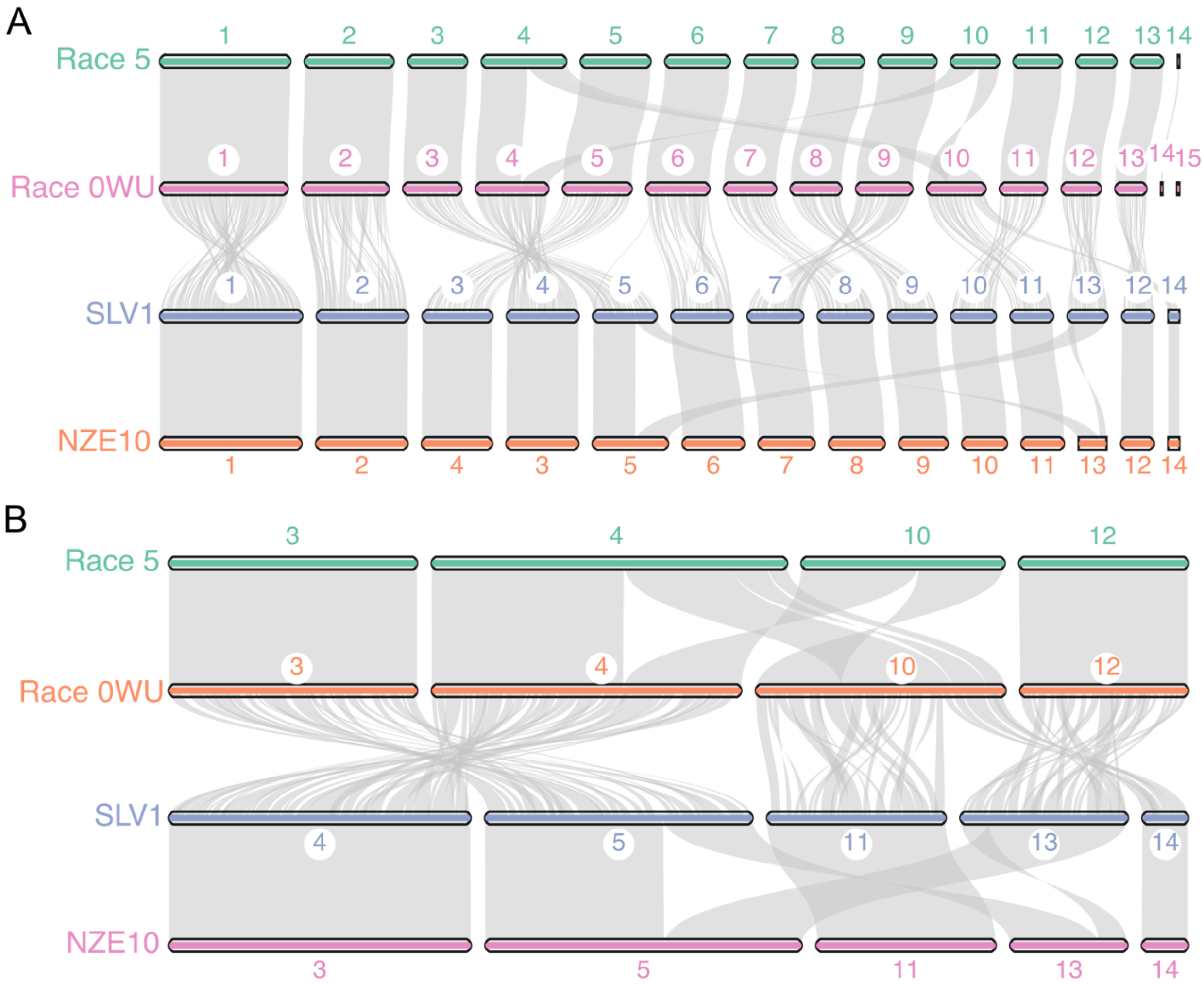


Figure 3.S13: Comparison of reciprocal translocation events in *Cladosporium fulvum* and the pine tree pathogen *Dothistroma septosporum*. (A) Plot showing pairwise synteny between *C. fulvum* isolates Race 5 and Race 0WU, and *D. septosporum* isolates SLV1 and NZE10. In *C. fulvum*, a reciprocal translocation occurs between Chr4 and Chr10. In *D. septosporum*, a reciprocal translocation occurs between Chr5 and Chr13. (B) Detailed synteny plot showing only the chromosomes that have reciprocal translocations. The Chr4 and Chr10 of *C. fulvum* that exhibit a reciprocal translocation are not homologs of the Chr5 and Chr13 of *D. septosporum* that also exhibit a reciprocal translocation.

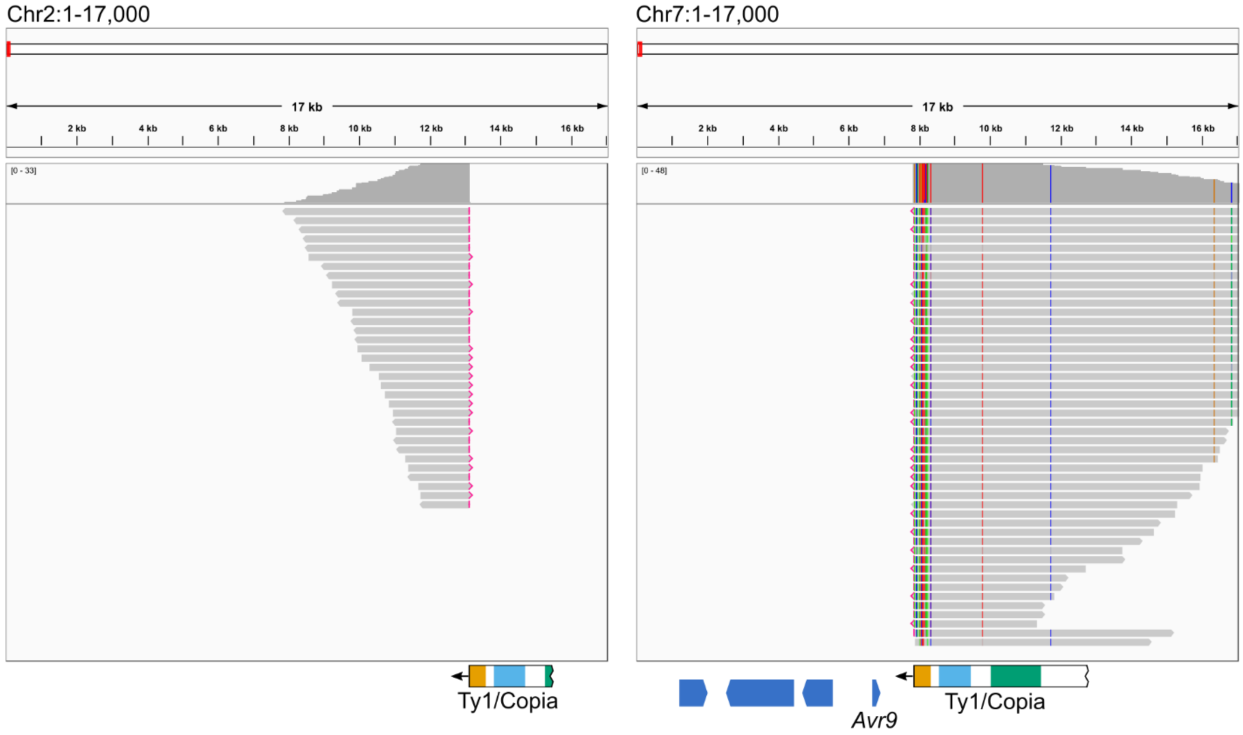
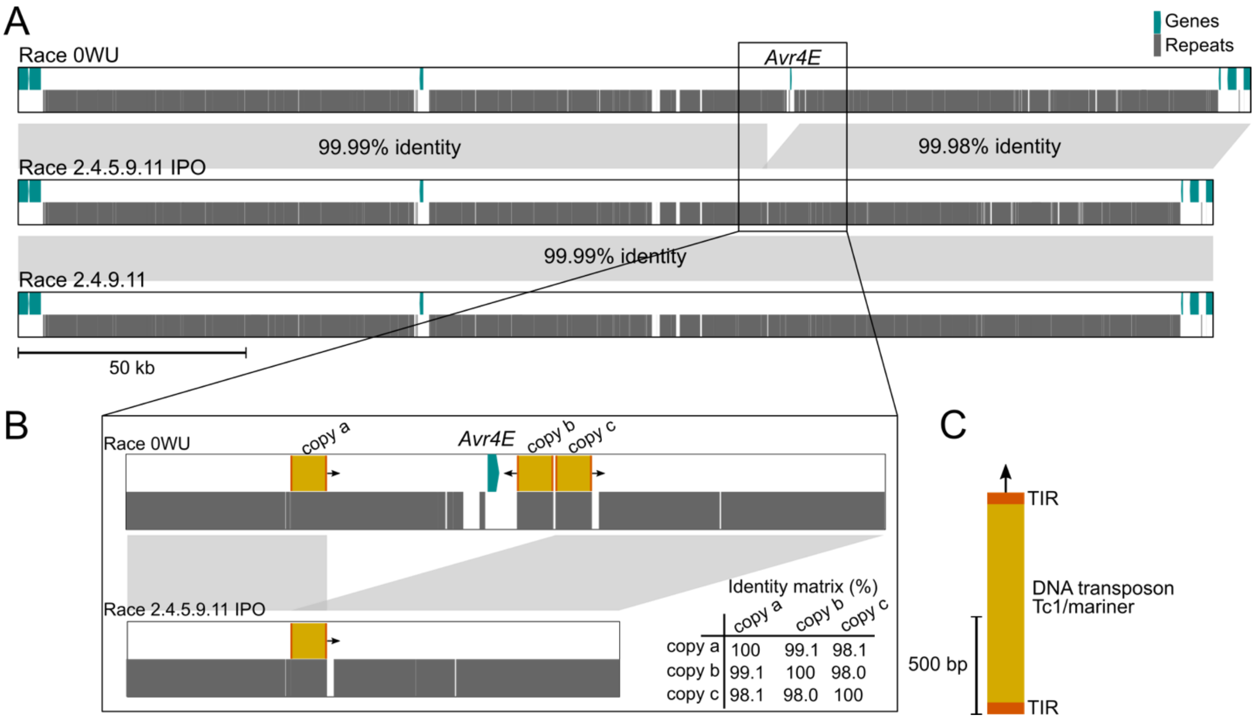


Figure 3.S14: PacBio HiFi reads mapped to the *Avr9* locus of *Cladosporium fulvum* support a non-reciprocal translocation. The figure shows IGV snapshots of HiFi reads from isolate Race 2.4.9.11 as they map on Chr2 and Chr7 of isolate Race 0WU. Reads from isolate Race 2.4.9.11 mapping on Chr7 of isolate Race 0WU extend until the 3'-end of a truncated copy of a Ty1/Copia retrotransposon. Then, instead of continuing to map over the *Avr9* locus, the reads split and a portion of them maps to the immediate downstream region of another truncated copy of the same Ty1/Copia retrotransposon present in Chr2 of isolate Race 0WU. Images were exported from IGV v2.11.4 and edited in Inkscape v1.0.2.



	copy a	copy b	copy c
copy a	100	99.1	98.1
copy b	99.1	100	98.0
copy c	98.1	98.0	100

Figure 3.S15: The deletion of *Avr4E* in *Cladosporium fulvum* likely requires neighboring copies of a Tc1/mariner DNA transposon. (A) Homologous regions of Chr7 of isolates Race 0WU, Race 2.4.5.9.11 IPO, and Race 2.4.9.11. Regions are represented with two tracks indicating the locations of predicted genes (top track) and of repetitive DNA (bottom track). The location of *Avr4E* is indicated. Parise syntenic blocks are indicated with ribbons. (B) Zoomed-in region of the deleted *Avr4E* locus. The deleted region contains the whore *Avr4E* sequence, and it is flanked by near-intact copies of a Tc1/mariner DNA transposon in the same predicted transcriptional orientation (copies a and b). Copy c has the inverse orientation. A fragment of 8270 bp is deleted in both isolates Race 2.4.5.9.11 IPO and Race 2.4.9.11. This fragmented contained *Avr4E*, two copies of the Tc1/mariner DNA transposon, and other predicted transposable elements. The nucleotide identity matrix of the Tc1/mariner copies obtained from a multiple sequence alignment is shown. (C) Structure of the Tc1/mariner DNA transposon flanking the *Avr4E* deletion. The consensus sequence is 1138 bp and is flanked by terminal inverted repeats (TIR) of 60 bp. No conserved domain was present within the Tc1/mariner DNA transposon sequence.

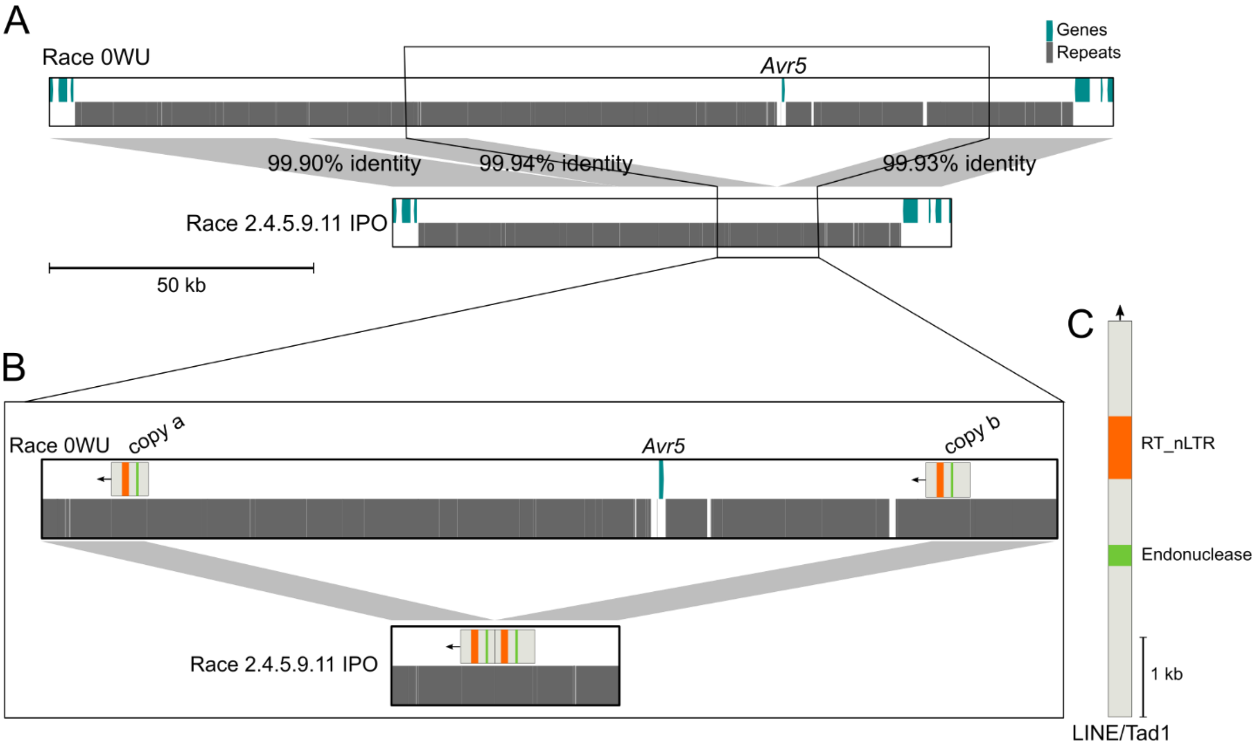


Figure 3.S16: The deletion of *Avr5* in *Cladosporium fulvum* likely requires neighboring copies of a LINE/Tad1 non-LTR retrotransposon. (A) Homologous regions of chromosome Chr1 of isolates Race 0WU and Race 2.4.5.9.11 IPO. Regions are represented with two tracks indicating the locations of predicted genes and repetitive DNA. The location of *Avr5* is indicated. Pairwise syntenic blocks are indicated with ribbons. (B) Zoomed-in region of the deleted *Avr5* locus. A region of 91338 bp containing *Avr5* and neighboring repetitive DNA is deleted in isolate Race 2.4.5.9.11 IPO. This deleted region is flanked by two copies (a and b) of a predicted LINE/Tad1 non-LTR retrotransposon. (C) Consensus sequence of the non-LTR retrotransposon LINE/Tad1 flanking the deleted *Avr5* locus. The sequence has 4970 bp and contains an endonuclease-reverse transcriptase domain (PF14529) and a reverse transcriptase domain of non-long terminal repeat retrotransposons (RT_nLTR; cd01650). The LINE/Tad1 copies a and b have 85.9% nucleotide identity based on a pairwise local alignment.

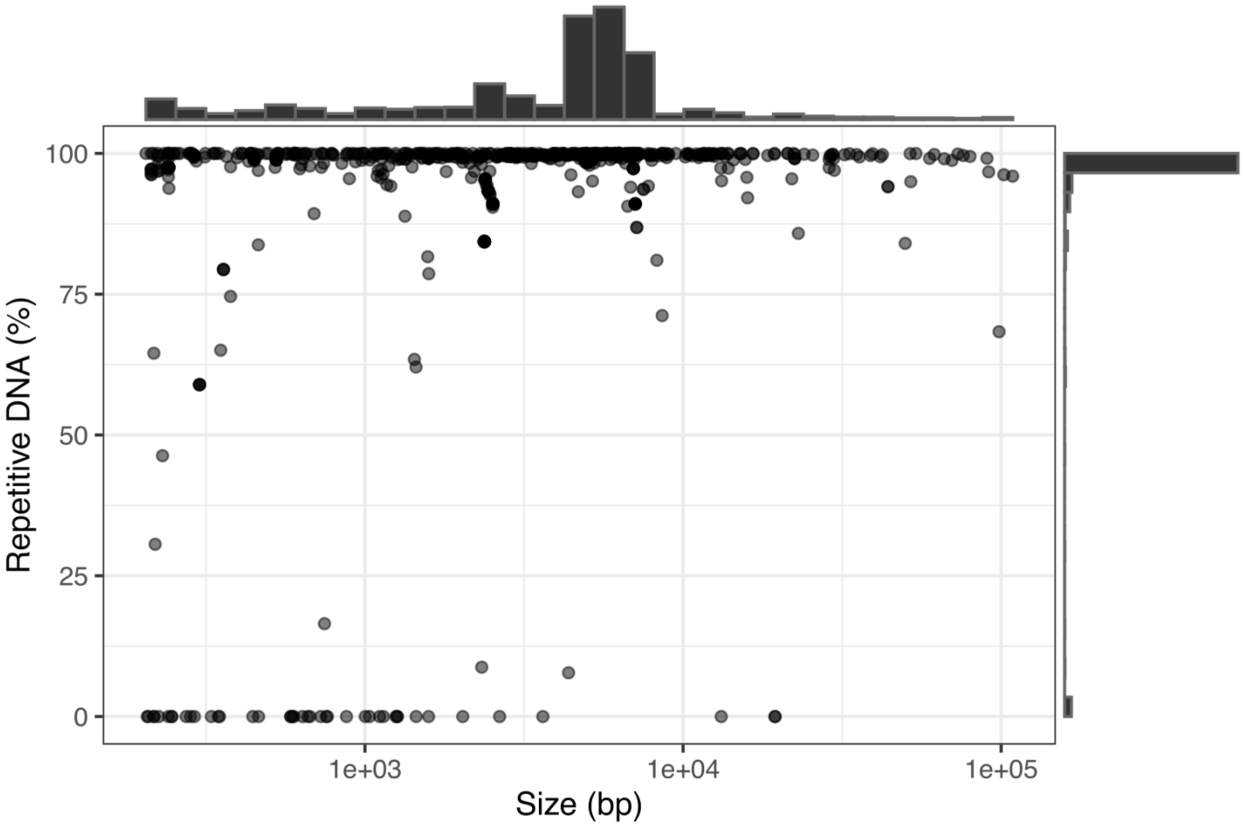


Figure 3.S17: Most long INDELs in the genome of *Cladosporium fulvum* are composed of repetitive DNA. Scatter plot showing 1226 INDELs as points. The size and predicted transposable element content of the INDELs is show on the X-axis and on the Y-axis, respectively. Marginal histograms are shown for both axes.

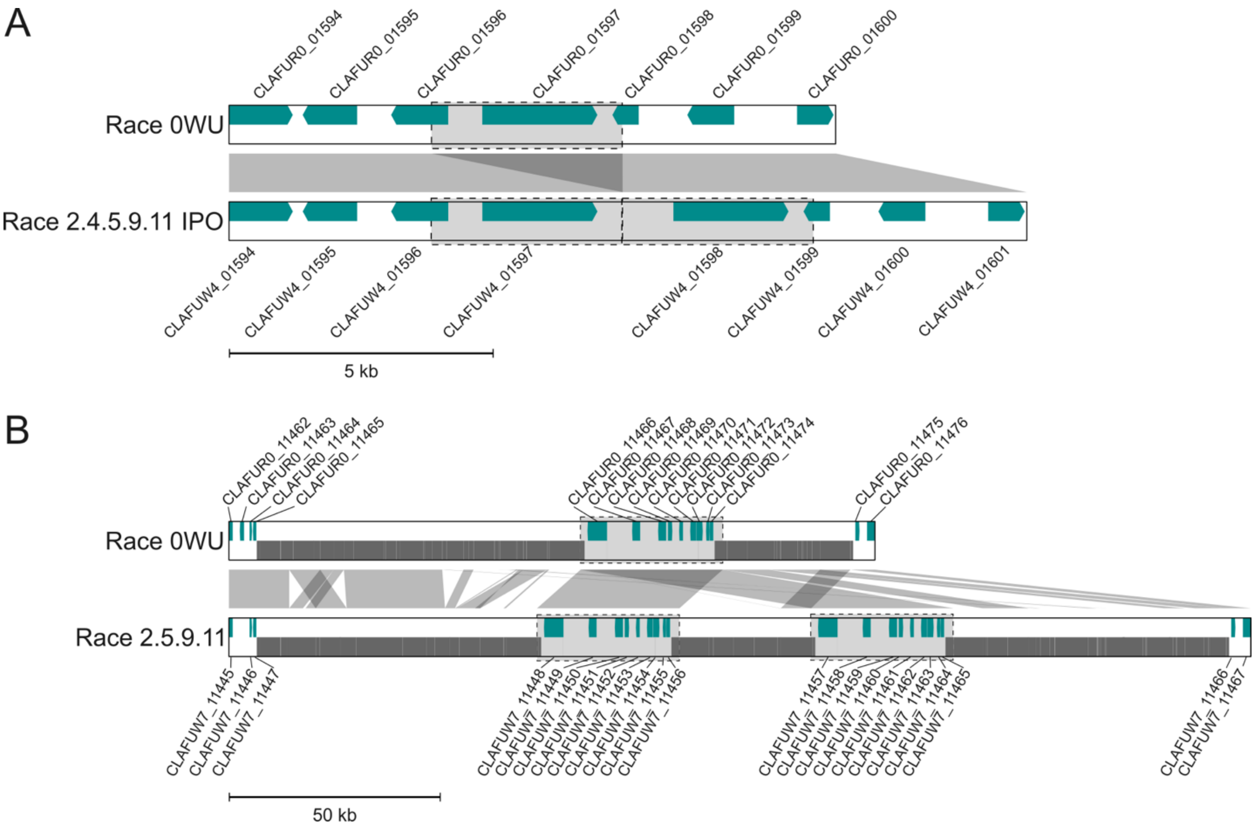


Figure 3.S18: Cases of tandem gene duplications in the genome of *Cladosporium fulvum*. (A) Tandem duplication of a 3.6 kb fragment that resulted in the duplication of the gene CLAFUR0_01597 encoding a putative laccase in isolate Race 2.4.5.9.11 IPO. The duplication also partially duplicated the neighboring genes CLAFUR0_01596, encoding a candidate effector, and CLAFUR0_01598, encoding a hypothetical protein. (B) Tandem duplication of a 33.8 kb fragment that resulted in the duplication of nine genes in isolate Race 2.4.9.11. Regions of the genome of *C. fulvum* Regions are represented with two tracks indicating the locations of predicted genes (upper track) and repetitive DNA (bottom track). Duplicated segments are highlighted with rectangles.

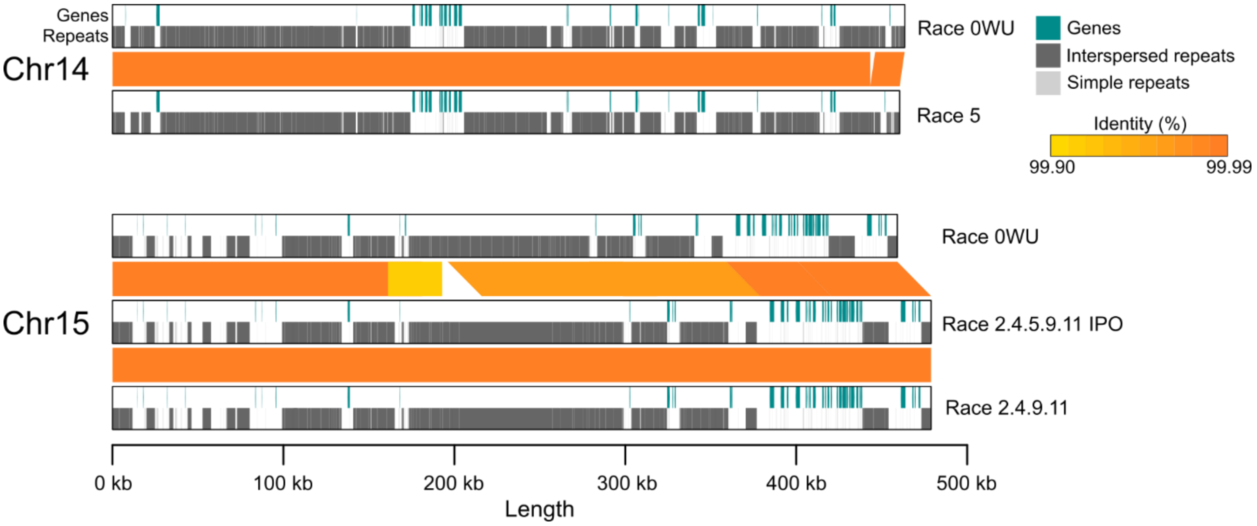


Figure 3.S19: Matching dispensable chromosomes present in different isolates of *Cladosporium fulvum* exhibit high nucleotide identity. Gene tracks and repeat tracks are shown for the two dispensable chromosomes, Chr14 and Chr15, of *C. fulvum*. Ribbons connect homologous regions and the respective percentages of nucleotide identity are shown as a heatmap. Homologous regions were identified by pairwise alignments using NUCmer of the MUMmer package

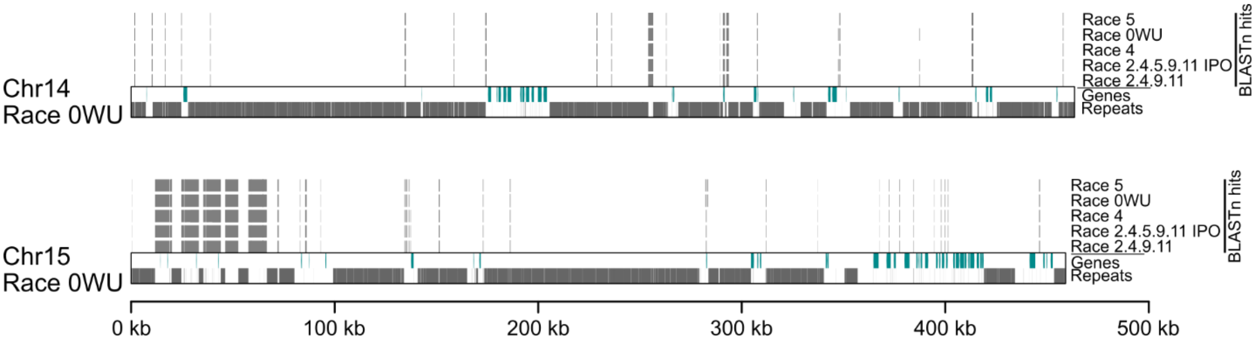
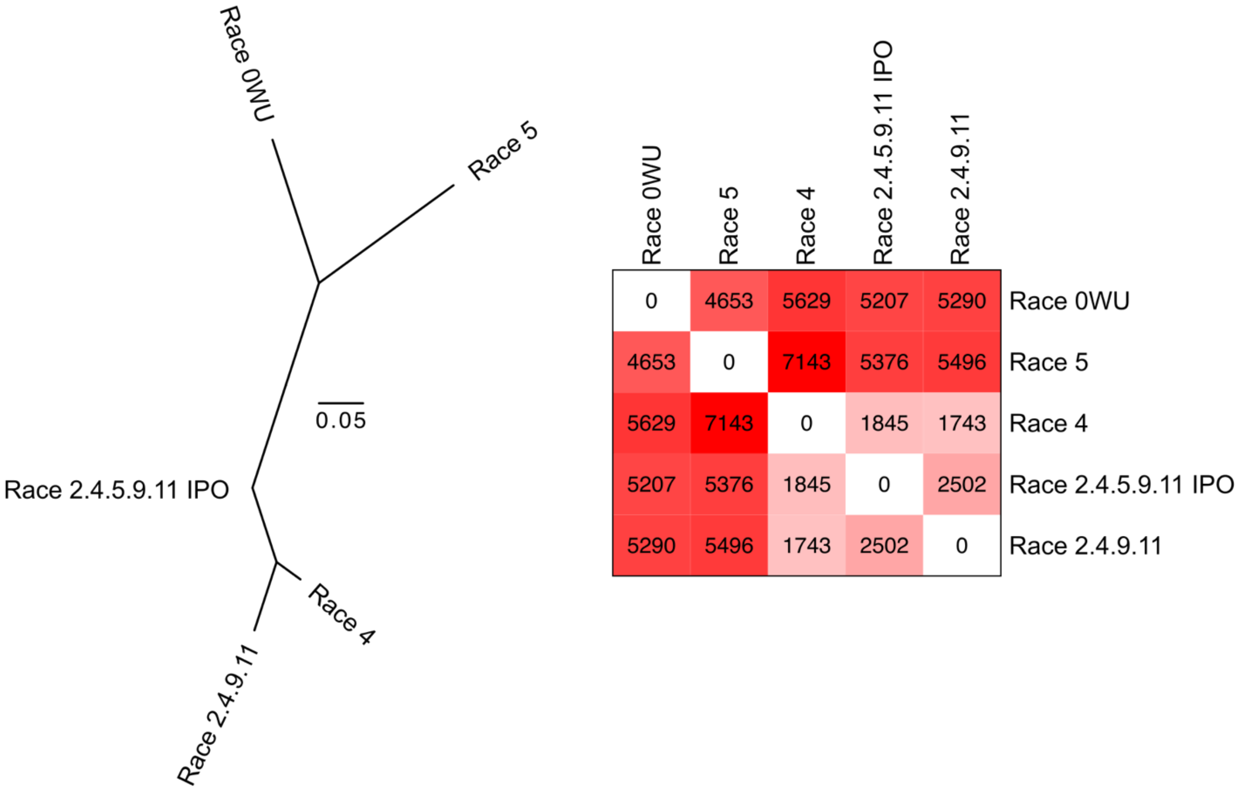


Figure 3.S20: The left end of the dispensable Chr15 of *Cladosporium fulvum* is composed of segments from core chromosomes. The figure shows the two dispensable chromosomes, Chr14 and Chr15, of isolate Race 0WU and two tracks that represent predicted genes (green track) and repetitive DNA (grey track). Grey rectangles above the tracks represent unmasked regions that have BLASTn hits (e-value< 1E-10) against core chromosomes of isolates Race 0WU, Race 5, Race 4, Race 2.4.5.9.11 IPO, and Race 2.4.9.11.



	Race 0WU	Race 5	Race 4	Race 2.4.5.9.11 IPO	Race 2.4.9.11
Race 0WU	0	4653	5629	5207	5290
Race 5	4653	0	7143	5376	5496
Race 4	5629	7143	0	1845	1743
Race 2.4.5.9.11 IPO	5207	5376	1845	0	2502
Race 2.4.9.11	5290	5496	1743	2502	0

Figure 3.S21: Phylogeny of the sequenced *Cladosporium fulvum* isolates. The figure shows an unrooted maximum likelihood phylogenetic tree of five *C. fulvum* isolates based on 8794 SNPs identified within 14,713 genes conserved in all five isolates. The heatmap shows the total number of pairwise segregating sites among the 8794 SNPs between isolates.

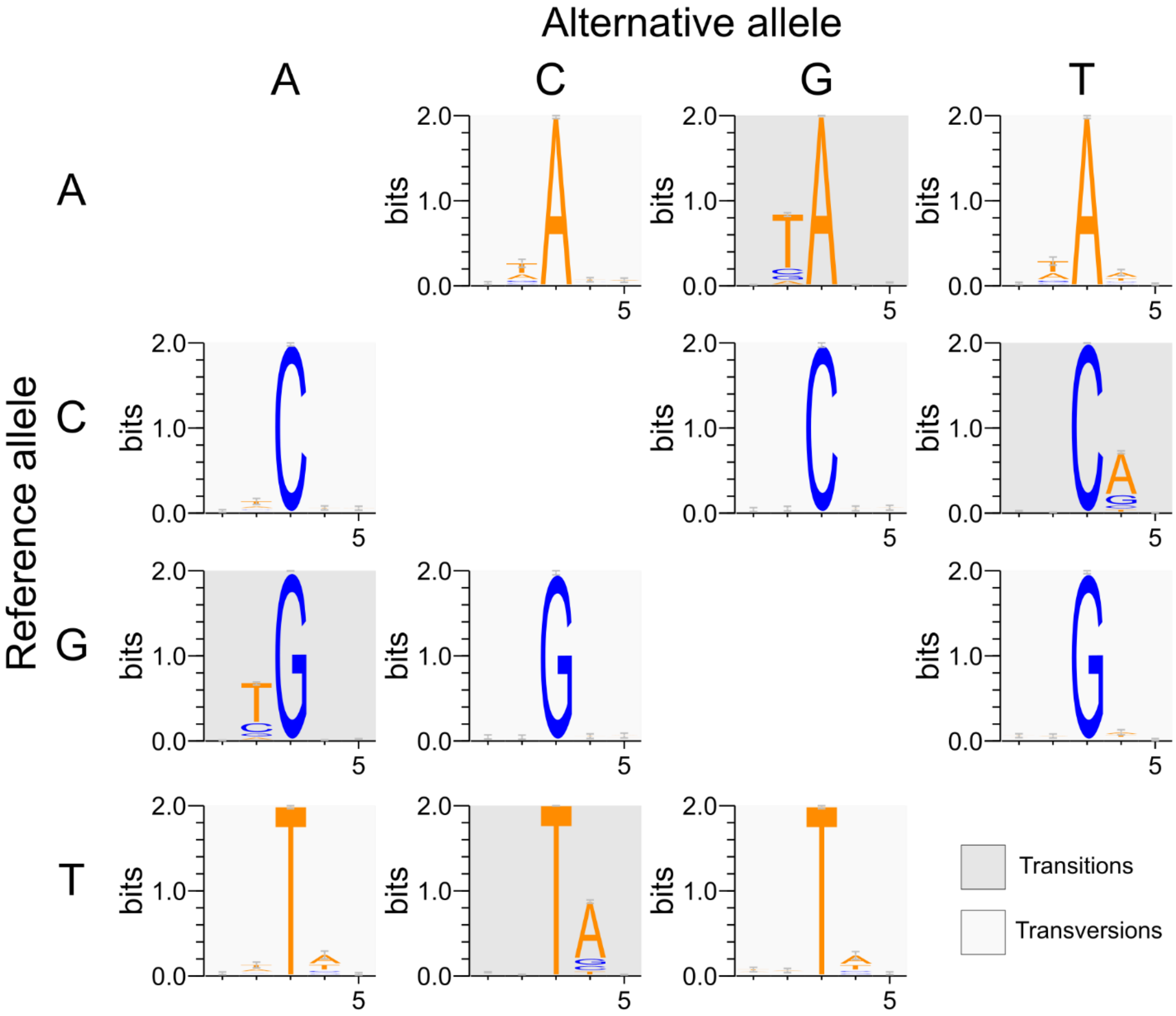


Figure 3.S22: Dinucleotide bias in regions of high nucleotide diversity in the genome of *Cladosporium*. The figure shows sequence logos for the nucleotides flanking the 12 possible types of nucleotide substitutions located in regions of high nucleotide diversity (π> 0.005) in the genome. The dinucleotide bias CpA ↔ TpA (or the complement TpG ↔ TpA) typical of RIP mutations can be observed.

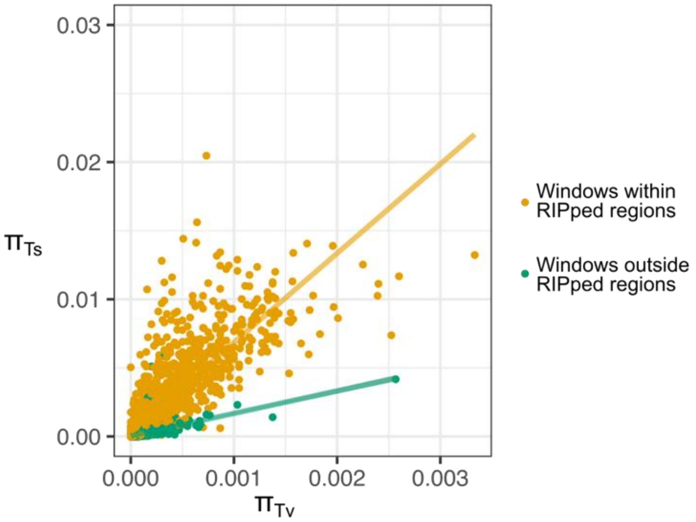


Figure 3.S23: Positive correlation between nucleotide diversity of transitions and of transversions in the genome of *Cladosporium fulvum*. In the scatter plot, each point represents a 20 kb window of the genome and indicates its overall nucleotide diversity of transitions (π_{Ts}) and of transversions (π_{Tv}). Windows with at least 10% overlap with regions predicted to be affected by RIP were considered within RIPped regions. Windows with less than 10% overlap with RIPped regions were considered outside RIPped regions. Trend lines were obtained with the 'lm' function using the formula $\pi_{Ts} \sim \pi_{Tv}$.

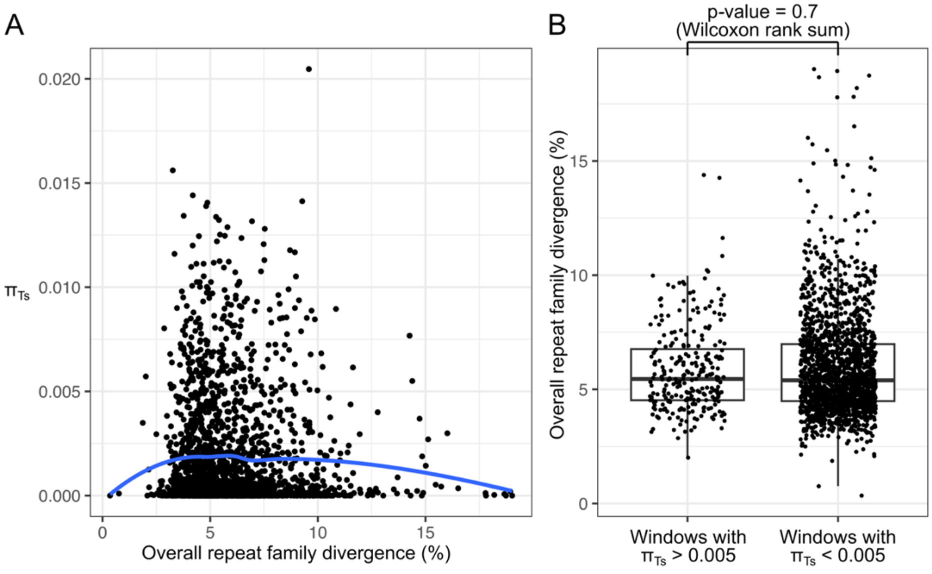


Figure 3.S24: No differences of repeat family divergence in regions of high and low nucleotide diversity in the *Cladosporium fulvum* genome. Points in (A) and (B) represent 20 kb windows of the genome. The scatter plot in (A) shows no clear correlation between repeat family divergence and transition nucleotide diversity. The line represents the trend of the points using the 'loess' smoothing method. The boxplot in (B) shows that there is no significant difference in divergence of repeat families located in windows with high nucleotide diversity *versus* low nucleotide diversity.

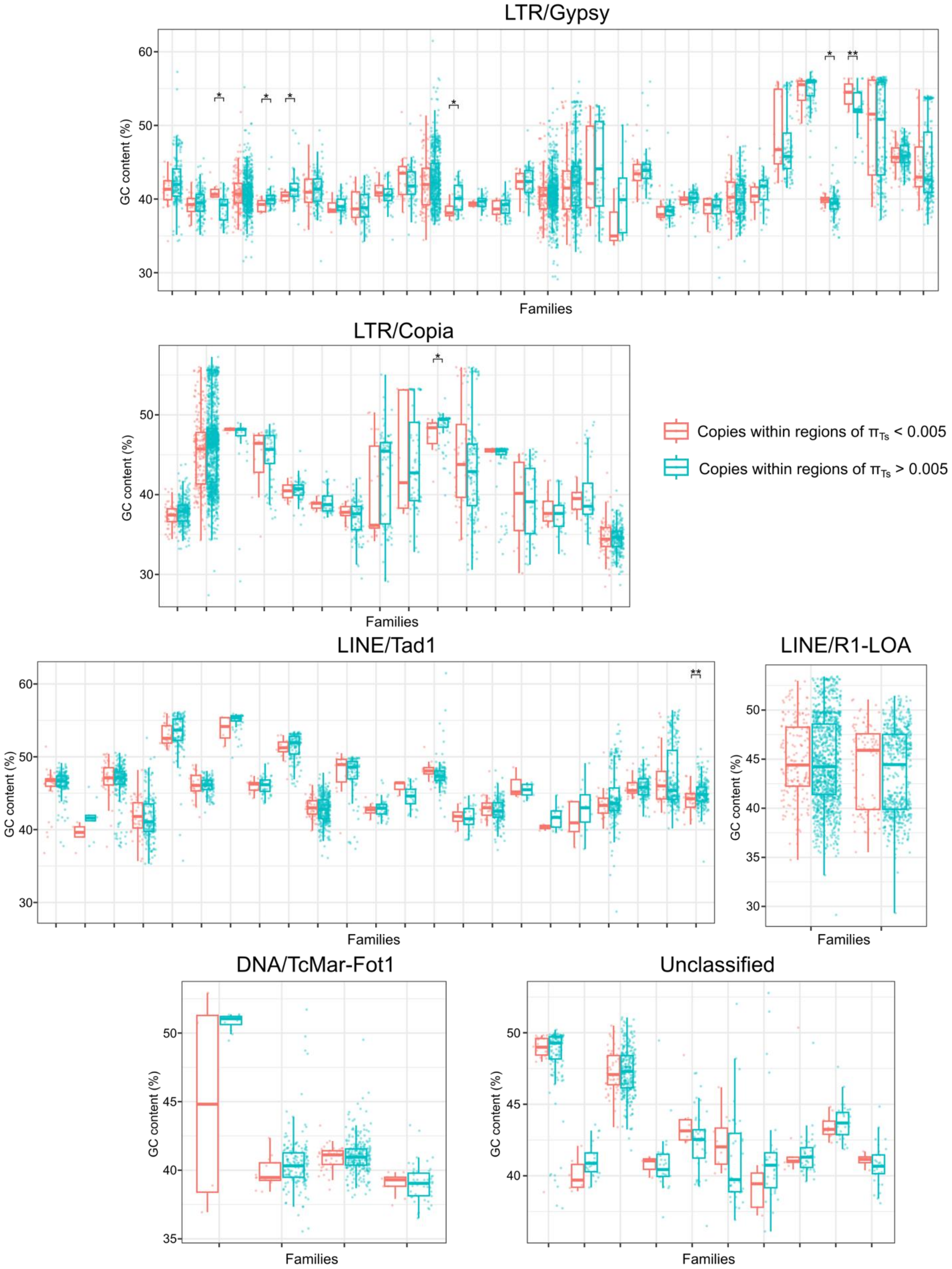

LTR/Gypsy
GC content (%)
60
50
40
30
Families
*
*
*
*
*
**
LTR/Copia
GC content (%)
50
40
30
Families
*
Copies within regions of π_{Ts} < 0.005
Copies within regions of π_{Ts} > 0.005
LINE/Tad1
GC content (%)
60
50
40
30
Families
**
LINE/R1-LOA
GC content (%)
50
45
40
35
30
Families
DNA/TcMar-Fot1
GC content (%)
50
45
40
35
Families
Unclassified
GC content (%)
50
45
40
Families

Figure 3.S25: No differences in GC content of transposable elements copies within regions of low and high diversity of transitions in the genome of *Cladosporium fulvum*. The boxplots show the GC content of predicted TE copies from different TE families organized into five sub(classes) and unclassified. Each point represents a TE copy. For all families, copies were divided into two groups, those that are entirely within regions of the genome with low nucleotide diversity of transitions (π_{Ts}; π_{Ts}< 0.005) and within high π_{Ts} (π_{Ts}> 0.005). Using the Mann-Whitney U test, no statistically significant difference of GC content between copies within low π_{Ts} and within high π_{Ts} was observed for 80 out of the 88 TE families analyzed. Only TE families which have at least three copies within low π_{Ts} and high π_{Ts} regions are shown. Moreover, only TE copies longer than 500 bp were considered. * *p*-value< 0.05, ** *p*-value< 0.01.

3.8.2 Supplementary tables

Table 3.S1: *Cladosporium fulvum* isolates for which near-complete genome assemblies were obtained using PacBio HiFi sequencing technology. Race, origin, year of collection, and mating time of the isolates were obtained from Stergiopoulos et al. (2007a). Isolate Race 4 is also known as isolate 4(2).

Isolate	Race	Country of origin	Year of collection	Mating type	Total number of HiFi reads	Median read size (bp)	Average read size (bp)	Estimated genome coverage	SRA accession
Race 0WU	4E	Netherlands	1997	MAT1-2	411,162	10,540	11,303.8	66x	SRR24302839
Race 4	4.4E	Netherlands	1971	MAT1-1	272,759	8,590	9,584.1	35x	SRR23862434
Race 2.4.5.9.11 IPO	2.4.4E.5.9.11	Netherlands	1980s	MAT1-2	262,381	9,960	10,854.5	37x	SRR24303573
Race 2.4.9.11	2.4.4E.9.11	Poland	1980s	MAT1-1	1,031,973	9,460	10,279.3	167x	SRR24303582

Table 3.S2: Comparison of the assembled chromosomes of five *Cladosporium fulvum* isolates. Repetitive DNA percentage corresponds to the percentage of all bases masked in the genome based on a de novo repeat annotation performed with RepeatModeler. This table is available at https://zenodo.org/records/11211529.

Table 3.S3: Summary of estimated abundance of transposable elements in the genomes of five *Cladosporium fulvum* isolates. Custom repeat libraries were generated with RepeatModeler v2 and then used to mask the genome with RepeatMasker. The output of RepeatMasker was parsed with the script parseRM.pl (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs), which counted the number of masked bases as reported in this table. In order to more accurately estimate the percentage of repeats of each class or family of transposable elements, bases masked twice (i.e. overlapping repeats) were not considered. This table is available at https://zenodo.org/records/11211529.

Table 3.S4: Summary of regions of the five *Cladosporium fulvum* genomes affected by Repeat-Induced Point (RIP) mutations. The table shows statistics for each chromosome and for the overall genomes of *C. fulvum*. These statistics correspond to the total number of windows scanned, the number and percentage of RIP-affected (i.e., RIPped) windows and single-copy RIP-affected windows, total number of Large RIP-Affected Regions (LRARs), total size of LRARs, average and median sizes of LRARs, longest LRARs and the average GC content within LRARs. The table also shows the number and percentage of RIP-affected windows that overlap with interspersed repetitive regions of the genome. Statistics were obtained using a sliding window of 1 kb and step size of 500 bp. Single-copy windows were considered as the windows with a single BLASTn hit against the genome using e-value< 1e-20, identity> 50%, query coverage> 20%. Windows with the RIP indices substrate, product, and composite were considered RIP affected. LRARs are formed by at least seven consecutive RIP-affected windows. This table is available at https://zenodo.org/records/11211529.

Table 3.S5: Update of the gene annotations of *Cladosporium fulvum* isolate Race 5. Transposon-like genes were removed, whereas new gene models were added. This table is available at https://zenodo.org/records/11211529.

Table 3.S6: Summary of genes from different functional categories and select sub-categories in five *Cladosporium fulvum* genomes. Emphasis is given to functional categories and sub-categories that are of relevance to fungal plant pathogens.

Functional gene category	Race 5	Race 0WU	Race 4	Race 2.4.5.9.11 IPO	Race 2.4.9.11
BUSCOs	1697	1694	1694	1693	1693
CAzymes	525	523	519	520	521
auxiliary activity (AA)	90	89	88	89	89
glycoside hydrolase (GH)	271	271	269	269	270
glycosyl transferase (GT)	114	113	113	113	113
polysaccharide lyase (PL)	10	10	9	9	9
carbohydrate esterase (CE)	32	32	32	32	32
carbohydrate-binding module (CBM)	15	15	15	15	15
plant cell wall degrading enzymes (PCWDE)	129	130	129	129	129
Cytochrome P450 monooxygenases (CYPs)	133	133	134	134	134
CYP51	1	1	1	1	1
CYP53	1	1	1	1	1
CYP61	1	1	1	1	1
Peptidases and Proteases (Total)	362	357	358	357	358
aspartic (A) proteases	14	14	13	13	14
cysteine (C) proteases	63	63	63	63	63
metallo (M) proteases	84	83	84	84	84
serine (S) proteases	1	1	1	1	1
threonine (T) proteases	178	174	175	175	175
mixed type proteases	19	19	19	19	19
inhibitory (I) proteases	3	3	3	2	2
Secondary metabolism enzymes (Total)	42	41	41	41	41
polyketide sythase (PKS)	11	11	11	11	11
non-ribosomal peptide synthetase (NRPS)	15	14	14	14	14
NRPS-Like	11	11	11	11	11
PKS-NRPS hybrid	1	1	1	1	1
terpene synthase (TPS)	4	4	4	4	4
Transporters	2293	2279	2277	2278	2283
ABC transporters	57	57	57	57	57
MFS transporters	383	378	380	380	380
Secreted proteins (Total)	1417	1425	1408	1404	1410
Secreted proteins, not candidate effectors	985	986	977	978	981
Candidate effectors	432	439	431	426	429

Table 3.S7: Genes encoding carbohydrate-active enzymes (CAZymes) in five *Cladosporium fulvum* genomes. The table shows the genes organized into hierarchical orthogroups (HOGs) and their location in the *C. fulvum* genomes. CAZymes were classified based on the presence of auxiliary activity (AA) modules, glycoside hydrolase (GH) modules, glycosyl transferase (GT) modules, polysaccharide lyase (PL) modules carbohydrate esterase (CE) modules, and carbohydrate-binding module (CBM) modules. CAZymes modules associated with plant cell wall degrading enzymes (PCWDE) and indicated, as well as orthogroups containing genes encoding secreted proteins. Orthogroups containing multiple sequences from the same isolate were manually split based on sequence homology so that all orthogroups include 1-to-1 orthologs. This table is available at https://zenodo.org/records/11211529.

Table 3.S8: Genes encoding proteases in five *Cladosporium fulvum* genomes. The table shows the genes organized into hierarchical orthogroups (HOGs) and their location in the *C. fulvum* genomes. Protease are classified into serine (S), metallo (M), aspartic (A), cysteine (C), threonine (T), inhibitory (I), Asparagine (N), and further family IDs are shown based on the most homologous sequence (BLASTp; e-value< 1e-10) in the MEROPS database. Orthogrous containing multiple sequences from the same isolate were manually split based on sequence homology so that all orthogroups include 1-to-1 orthologs. This table is available at https://zenodo.org/records/11211529.

Table 3.S9: Genes encoding cytochromes P450 in five *Cladosporium fulvum* genomes. The table shows the genes organized into hierarchical orthogroups HOGs) and their location in the *C. fulvum* genomes. These cytochrome P450s are classified based on homology searches (BLASTn 40% identity, 50% query coverage, e-value< 1E-10) against Dr. Nelson's database of curated fungal cytochrome P450s (https://drnelson.uthsc.edu/P450seqs.dbs.html). Orthogrous containing multiple sequences from the same isolate were manually split based on sequence homology so that all orthogroups include 1-to-1 orthologs. This table is available at https://zenodo.org/records/11211529.

Table 3.S10: Genes encoding ABC and MFS transporters in five *Cladosporium fulvum* genomes. The table shows the genes organized into hierarchical orthogroups (HOGs) and their location in the *C. fulvum* genomes. The table shows the genes organized into orthogroups and their location in the C. fulvum genomes. These transporters are classified based on homology searches (e-value< 1E-10) against the TCDB database. Orthogrous containing multiple sequences from the same isolate were manually split based on sequence homology so that all orthogroups include 1-to-1 orthologs. This table is available at https://zenodo.org/records/11211529.

Table 3.S11: Genes encoding key secondary metabolite enzymes in five *Cladosporium fulvum* genomes. The table shows the genes organized into hierarchical orthogroups (HOGs) and their location in the *C. fulvum* genomes. These key enzymes are classified into non-ribosomal peptide synthetases (NRPS), type 1 polyketide synthases (T1PKS), and terpene synthases (Terpene). NRPS-like represent fragments of NRPS genes. Orthogrous containing multiple sequences from the same isolate were manually split based on sequence homology so that all orthogroups include 1-to-1 orthologs. This table is available at https://zenodo.org/records/11211529.

Table 3.S12: Genes encoding secreted proteins in *Cladosporium fulvum* genomes. The table shows the genes organized into hierarchical orthogroups (HOGs) and their location in the *C. fulvum* genomes. Orthogrous continaining multiple sequences from the same isolate were manually split based on sequence homology so that all orthogroups include 1-to-1 orthologs. This table is available at https://zenodo.org/records/11211529.

Table 3.S13: Genes encoding candidate effector proteins in five *Cladosporium fulvum* genomes. The table shows the genes organized into hierarhical orthogroups (HOGs) and their location in the *C. fulvum* genomes. Names of candidate effector genes previously described are shown in the first column. Orthogrous containing multiple sequences from the same isolate were manually split based on sequence homology so that all orthogroups include 1-to-1 orthologs. This table is available at https://zenodo.org/records/11211529.

Table 3.S14: Number of genes in five *Cladosporium fulvum* genomes assigned to different Gene Ontology (GO) terms and EuKaryotic Ortholog Group (KOG) categories. Subcategories from the second level of the GO hierarchy are shown. This table is available at https://zenodo.org/records/11211529.

Table 3.S15: Accessory genes from five *Cladosporium fulvum* genomes. The table shows the accessory genes organized into hierarchical orthogroups (HOGs). HOGs containing genes from ten different functional categories are indicated in columns seven to sixteen by their respective category classification. A summary table is also shown containing over-representation p-values of HOGs containing accessory genes from the different functional classes. This table is available at https://zenodo.org/records/11211529.

Table 3.S16: Summary of pairwise synteny blocks in pairwise alignments of five *Cladosporium fulvum* genomes. The table shows statistics for syntenic blocks at the gene and genome level. At the gene level, the table shows the number of syntenic blocks, number of genes from these syntenic blocks, number of genes from the longest syntenic block, percentage of the genes from reference and query assigned to synteni c blocks, and the minimum number of syntenic blocks that contain 50% to 90% of all genes from the reference genome. The table also shows syntenic blocks, size of the longest syntenic block, percent of the reference and query genomes in syntenic blocks, and the minimum number of syntenic blocks that contain 50% to 90% of the reference genome, based on whole-genome alignments.

		Race 5 / Race 0WU	**Race 5 / Race 4**	**Race 5 / Race 2.4.5.9.11 IPO**	**Race 5 / Race 2.4.9.11**	**Race 0WU / Race 4**	**Race 0WU / Race 2.4.5.9.11 IPO**	**Race 0WU / Race 2.4.9.11**	**Race 4 / Race 2.4.5.9.11 IPO**	**Race 4 / Race 2.4.9.11**	**Race 2.4.5.9.11 IPO / Race 2.4.9.11**
Gene level	Synteny blocks	19	18	15	17	19	17	19	16	18	16
	No. genes within blocks	14862	14819	14837	14823	14837	14865	14859	14867	14856	14885
	Longest block (genes)	2336	2335	2339	1586	2333	2330	1581	2337	1588	1589
	Coverage ref	99.1	98.8	99	98.9	99	99.2	99.2	99.8	99.7	99.6
	Coverage qry	99.2	99.5	99.3	99.2	99.6	99.5	99.4	99.5	99.4	99.6
	Number of blocks to cover:										
	50% of the genes	6	6	6	7	5	5	6	5	6	6
	60% of the genes	7	7	7	8	7	7	8	7	8	8
	70% of the genes	9	9	9	10	8	8	9	8	9	9
	80% of the genes	11	11	11	12	10	10	11	10	11	11
	90% of the genes	13	13	13	14	12	12	13	11	13	13
Genome level	Synteny blocks	971	1437	1131	1166	1155	1100	1027	421	373	479
	Longest block (Mb)	1.56	1.26	1.75	1.79	1.23	2.47	2.22	3.48	3.27	2.85
	Percent of the reference genome in blocks	95.4	92.7	94.0	94.3	93.8	94.2	95.1	97.5	98.5	98.0
	Percent of the query genome in blocks	95.0	93.1	94.0	93.4	94.5	94.6	94.6	97.2	97.2	97.1
	Number of blocks to cover:										
	50% of the genome	68	104	69	70	81	75	64	32	28	31
	60% of the genome	100	154	103	106	118	112	96	44	39	44
	70% of the genome	143	229	156	157	173	165	140	59	52	61
	80% of the genome	211	364	248	249	272	254	216	80	71	85
	90% of the genome	379	787	495	496	526	481	398	122	106	135

Table 3.S17: Number and type of structural variations (SVs) identified in four *Cladosporium fulvum* genomes. The table shows number of SVs identified based on whole-genome pairwise alignments of *C. fulvum* isolates Race 5, Race 4, Race 2.4.5.9.11 IPO, and Race 25 9 11 using isolate Race 0WU as reference.

	Race 5	**Race 4**	**Race 2.4.5.9.11 IPO**	**Race 2 5 9 11**
Translocation breakend	12	12	6	6
Deletions	321	416	396	354
Insertions	376	406	393	364
Inversions	9	9	7	6
Total SVs	718	843	802	730

Table 3.S18: Structural variations (INDELs) affecting genes in the genome of *Cladosporium fulvum*. The table shows SVs that overlap with predicted genes of *C. fulvum*. SVs are either insertions or deletions. Location and size of SVs are shown, as well as the genes affected by them. Further details of the impact of the SVs on the genes are shown in the last column. This includes insertions of duplicated segments that result in duplication of genes. This table is available at https://zenodo.org/records/11211529.

Table 3.S19: Homologs and expression of predicted genes from the dispensable chromosomes Chr14 and Chr15 of *Cladosporium fulvum*. The table shows the genes from Chr14 and Chr15 of *C. fulvum* isolate Race 0WU, and description of the ten most homologous sequences in NCBI nr database based on BLASTp searches (e-value < 1E-5). The table also shows transcripts per million (TPM) values for the genes based on RNA-seq data of *C. fulvum* 0WU-Solanum lycopersicum cv. Heinz interaction at 4 days post inoculation (dpi) (SRR1171035), 8 dpi (SRR1171040), and 12 dpi (SRR1171043), as well as *C. fulvum* 0WU grown in potato-dextrose broth (SRR1171044). This table is available at https://zenodo.org/records/11211529.

Chapter 4

Transcriptome analysis of two isolates of the tomato pathogen *Cladosporium fulvum* uncovers genome-wide patterns of alternative splicing events during a host infection cycle

Alex Z. Zaccaron
Li-Hung Chen
Ioannis Stergiopoulos

Leading author statement

I contributed to the conceptualization of this chapter, performed all analyses, generated all figures, contributed to the investigation and interpretation, wrote the original draft, and contributed to the review and editing of the final draft.

Abstract

Alternative splicing (AS) is a key element of eukaryotic gene expression that increases transcript and proteome diversity in cells, thereby altering their physiological responses to external stimuli and adaptation to stresses. While AS is widely present and studied in higher eukaryotes, its frequency, conservation within a species, and functional consequence for virulence are still largely unexplored in fungi. In this study, we sequenced the transcriptomes and systematically analyzed the AS events taking place in *Cladosporium fulvum* isolates Race 5 and Race 4 during nearly a complete infection cycle of their tomato host. In both isolates, approximately 40% of protein-coding genes were predicted to be AS, which is at the high end compared to other fungi. Nearly 59% of the AS genes were predicted to lead to multiple protein isoforms, which collectively led to a 31% increase in the proteome diversity of *C. fulvum*. The remaining 41% of AS genes had splicing events in their 5' or 3' UTRs, suggesting an effect on gene expression and protein translation. Analysis of AS types showed that intron retention was most prevalent amongst AS genes of *C. fulvum,* which were further enriched in genes encoding major facilitator superfamily transporters, sugar transporters, transcription factors, and cytochrome P450 enzymes, but not candidate effectors. Mapping of the AS genes in the genome of *C. fulvum* showed that they were mostly located in repeat-rich core chromosomes and exhibited significantly longer 5’ intergenic regions that were richer in repetitive DNA compared to non-AS genes. Finally, 246 genes from isolate Race 5 and 103 genes from isolate Race 4 were identified that produced transcript isoforms with evidence of differential usage across the timepoints of the infection process. Amongst these were a few effector-encoding genes, suggesting that infection stage-dependent modulation of AS in virulence-associated genes could possibly finetune infections on the host.

4.1 Introduction

Alternative splicing (AS) is a molecular process that occurs during splicing of transcribed pre-mRNA, through which diverse mature mRNA molecules are produced by single genes by fluctuations in the usage of intron splice sites (Kornblihtt et al. 2013). There are five main types of AS events that can occur when the spliceosome interacts with different splice sites of the pre-mRNA, including exon skipping, mutually exclusive exons, intron retention, and alternative 3' and 5' splice sites (Xing and Lee 2006). A key outcome of AS is that it increases diversity of proteins in cells, but it may further lead to additional aberrations in a cell's proteome, including the production of protein isoforms with loss, gain and reshuffling of functional domains, altered patterns of cellular localization, stability, enzymatic activity, or post-translational modifications. Notable examples include membrane-bound and secreted protein isoforms that differ from their cytoplasmic counterparts by the presence of alternatively spliced (AS) N-terminal trans-membrane domains and signal peptides (SP). Moreover, next to its effect on the proteome, AS may further affect many aspects of RNA metabolism, including mRNA translation efficiency and degradation, thereby modulating gene expression at the post-transcriptional level. Aberrant mRNA splicing, for instance, can generate isoforms that are often the targets for nonsense-mediated mRNA decay (NMD), an mRNA surveillance system that detects and degrades prematurely terminated transcripts generated by errors in mRNA processing. AS variants can also regulate the abundance of functional transcripts via a mechanism known as regulated unproductive splicing and translation (RUST). Taken together, it is thus not surprising that a strong positive correlation exists between frequency of AS and functional complexity in cells that affects many of their cellular, metabolic, and physiological processes, including responses to environmental stressors and other stimuli (Chen et al. 2014a). In humans, for instance, it is estimated that 95% of all genes undergo AS at different developmental stages of tissues (Nilsen and Graveley 2010; Wang et al. 2015), thereby often leading to hereditary diseases and cancers, and increasing phenotypic variability in populations.

Next to its ephemeral effects on cell functionality, in eukaryotes, AS is also an important substrate for adaptive evolution, especially at short timescales (Singh and Ahi 2022). An important observation supporting this assumption is that AS take places more readily in cells than changes in gene expression (Barbosa-Morais et al. 2012; Merkin et al. 2012), thereby allowing for rapider responses to environmental stimuli in which a quick response would be beneficial or even crucial to organismal survival, such as abiotic and biotic stresses. In plants, for example, rapid induction of AS has been shown to hav e a significant functional impact in processes such as photosynthesis, flowering, abiotic stress -related responses such as drought, cold and heat, and defense against microbial attacks, including viruses bacteria and fungi (Lin et al. 2009; Calixto et al. 2018; Laloum et al. 2018; Mastrangelo et al. 2012; Leviatan et al. 2013; Qin et al. 2007; Staal and Dixelius 2008; Michael Weaver et al. 2006; Costanzo and Jia 2009; Dinesh -Kumar and Baker 2000). In most cases, AS can be spatially and developmentally regulated, and it is frequently associated with environmental stresses and a response to pathogen infections (Dinesh-Kumar and Baker 2000; Wang et al. 2015; Laloum et al. 2018).

While the biological significance of AS in higher eukaryotes such as plants and mammals is well documented, AS in fungi is relatively understudied. Studies have shown that the occurrence of AS varies considerably among fungal species, and its frequency is typically low compared to other eukaryotes (Fang et al. 2020; Grützmann et al. 2014). In the baker's yeast *Saccharomyces cerevisiae*, for instance, less than 1% of the genes undergo AS (Muzafar et al. 2021), probably due to the fact that only 5% of the genes in *S. cerevisiae* contain introns that can be differentially spliced (Hooks et al. 2014). In contrast, the percentage of AS genes in the biocontrol fungus *Trichoderma longibrachiatum* is 48.9%, which is one of the highest reported in fungi (Xie et al. 2015). Irrespective of the variations in the percentage of AS genes among different fungal species, IR was the major type of AS in fungi, although all forms of AS can occur. As IR typically results in frameshifts and the introduction of premature stop codons in the produced mRNAs that are subsequently targeted for NMD, in most cases it remains unknown whether the bulk of AS in fungi is functional or simply the result of incomplete splicing and splicing errors. Nonetheless, evidence from mostly individual gene studies suggest that AS can have a significant impact on several cellular and physiological functions in fungi, including growth and development (Hoppins et al. 2007), response to environmental stressors such as nutrients, pH, and extracellular nitrogen and phosphate concentration (Leal et al. 2009; Trevisan et al. 2011; Zhang and Miyake 2009), histone deubiquitination (Hossain et al. 2011), gene expression (Shaul 2017; Preker et al. 2002), secretion of enzymes (Gehrmann et al. 2016), subcellular localization of proteins (Freitag et al. 2012; Strijbis et al. 2012), and others (Fang et al. 2020; Muzafar et al. 2021).

An increasing body of evidence also suggests that AS in fungi can further affect virulence on hosts and resistance to antifungal treatments (Sieber et al. 2018; Grützmann et al. 2014; Jeon et al. 2022; Lopes et al. 2022; Muzafar et al. 2021; Fang et al. 2020). For instance, in *Sclerotinia sclerotiorum* that causes white mold disease, the landscape of AS changes in according to the host that the pathogen is interacting with (Ibrahim et al. 2021), indicating that AS could be important for host adaption. In the rice blast fungus *Magnaporthe oryzae*, the frequency of AS appears to increase during interaction with the host, with many AS events putatively being upregulated during host infection (Jeon et al. 2022). One AS gene in particular of *M. oryzae*, *MoPTEN*, has been further shown to generate two protein isoforms during infection, both of which are required for appressorium development and growth of invasive hyphae in plant cells (Wang et al. 2021). A genome-wide investigation of AS in human fungal pathogens revealed that AS is not only species-specific, but also is more frequent during stress conditions, and that AS likely play important regulatory roles to promote infection (Sieber et al. 2018). Collectively, these studies indicate that AS can play an important role in modulating virulence of pathogenic fungi on their hosts and adaptation to stressful conditions.

Cladosporium fulvum (syn. *Passalora fulva*, *Fulvia fulva*) is a biotrophic fungal pathogen member of the Ascomycetes (Dothideomycetes; Capnodiales) that is the causative agent of tomato leaf mold (Thomma et al. 2005). Over the last 40 years, this fungus has been a valuable model for the study of plant-microbe interactions (De Wit et al. 2009; de Wit 2016). Recent advances to obtain a high-quality reference genome of *C. fulvum* have been successfully deployed (De Wit et al. 2012; Mesarich et al. 2023; Zaccaron et al. 2022; Zaccaron and Stergiopoulos 2024). Notably, the genomes of *C. fulvum* isolates Race 5, Race 0WU, Race 4, Race 2.4.5.9.11 IPO, and Race 2.5.9.11 were assembled at the chromosome-level, three of which telomere-to-telomere (Zaccaron et al. 2022; Zaccaron and Stergiopoulos 2024). Comparative genomics analyses revealed that the genome of *C. fulvum* exhibits a 'checkerboard' pattern composed of gene-dense regions interspersed with repeat-rich regions, with an enrichment of genes encoding candidate effectors located in repeat-rich regions, in accordance with the 'two-speed genome' model of evolution (Dong et al. 2015). However, despite significant progress in understanding genomic aspects of *C. fulvum*, its transcriptome and particularly AS profile during infections remains largely unexplored.

In this study, we systematically studied the frequency and changes in AS events occurring in two *C. fulvum* isolates, i.e. Race 5 and Race 4, during a complete fungal infection cycle of the tomato host. Our studies revealed a high frequency and dynamic landscape of AS events taking place in the two isolates during host infections that is particularly affecting genes likely to be involved in the transport of nutrients, regulation of gene expression, and monooxygenase activity, thereby suggesting a role for AS in modulating *C. fulvum* infections on its tomato host.

4.2 Results

4.2.1 RNA sequencing of *C. fulvum* isolates Race 5 and Race 4 during compatible interactions with tomato

The transcriptomes of *C. fulvum* isolates Race 5 and Race 4 were sequenced at high-depth during interaction with *Solanum lycopersicum* cv. Moneymaker at seven timepoints, i.e., 2, 4, 6, 8, 10, 12, and 14 dpi, and from three independent infections (i.e. biological replicates), generating between 65 M and 221 M RNA-seq reads per sample (Table 4.S1). As fungal biomass during the early stages of the infection is typically very low, we sequenced the transcriptomes from the 2, 4, 6, and 8 dpi at higher depth compared to 10, 12, and 14 dpi (Table 4.S1), thus increasing sensitivity to capture transcript expression at early timepoints. Before read mapping, reads were assigned to either the *C. fulvum* isolate Race 5 (Zaccaron et al, 2022) or the tomato genome (Hosmani et al. 2019), using the alignment-free *k*-mer method implemented in the *seal.sh* script from the BBMap package (Table 4.S2). Reads matching the tomato genome were removed, and the remaining reads were mapped to the *C. fulvum* genome. As expected, at early timepoints, the percentage of reads mapping to the *C. fulvum* genome was very low (Table 4.1). Specifically, at 2 dpi and 4 dpi, between 135,953 (0.08%) and 437,311 (0.20%) of the reads from *C. fulvum* Race 5, and between 409,880 (0.23%) and 1,852,149 (0.92%) of the reads from *C. fulvum* Race 4 mapped to the *C. fulvum* genome. However, the percentage of reads that successfully mapped to the *C. fulvum* genome significantly increased at later timepoints, reaching between 27.1 M (32.55%) and 37.6 M (43.87%) for isolate Race 5, and between 12.7 M (19.5%) and 45.7 M (57.64%) reads for isolate Race 4 at 12 dpi and 14 dpi, respectively (Table 4.1).

Table 4.8: The number of RNA-seq reads that mapped to the genome of *Cladosporium fulvum* increases over time during infection of tomato. The table shows numbers and percentages of RNA-seq reads obtained from three independent infections (i.e. biological replicates) and seven timepoints during the interaction between the *C. fulvum* isolates Race 5 and Race 4 and tomato that mapped to the genomes of *C. fulvum* isolates Race 5 and Race 4.

Statistic	Isolate	Timepoint	Rep. 1	Rep. 2	Rep. 3
Number of reads	Race 5	2 dpi	135,953	150,535	195,887
	Race 5	4 dpi	231,260	313,138	437,311
	Race 5	6 dpi	3,270,510	3,036,751	5,898,146
	Race 5	8 dpi	23,392,396	16,071,949	19,383,926
	Race 5	10 dpi	22,863,783	60,588,111	17,654,682
	Race 5	12 dpi	27,109,200	34,630,359	28,289,494
	Race 5	14 dpi	27,894,580	30,191,286	37,633,905
	Race 4	2 dpi	409,880	1,217,740	1,237,896
	Race 4	4 dpi	736,856	976,981	1,852,149
	Race 4	6 dpi	712,610	2,970,591	1,449,788
	Race 4	8 dpi	2,013,657	17,251,769	2,229,050
	Race 4	10 dpi	10,724,443	23,799,801	2,530,422
	Race 4	12 dpi	23,492,901	25,559,655	12,719,704
	Race 4	14 dpi	28,462,052	45,780,891	26,803,887
Percentage of reads	Race 5	2 dpi	0.08	0.10	0.10
	Race 5	4 dpi	0.16	0.18	0.20
	Race 5	6 dpi	2.22	1.91	2.67
	Race 5	8 dpi	12.08	8.96	8.95
	Race 5	10 dpi	29.57	32.10	17.53
	Race 5	12 dpi	32.55	36.42	34.16
	Race 5	14 dpi	40.58	35.07	43.87
	Race 4	2 dpi	0.23	0.60	0.62
	Race 4	4 dpi	0.34	0.51	0.92
	Race 4	6 dpi	0.45	1.55	0.95
	Race 4	8 dpi	1.48	10.86	1.59
	Race 4	10 dpi	13.49	26.02	2.56
	Race 4	12 dpi	29.28	36.13	19.52
	Race 4	14 dpi	35.52	57.64	38.11

4.2.2 Isolates Race 5 and Race 4 of *C. fulvum* share nearly all their intron splice sites

Intron-containing genes may undergo alternative splicing (AS), thereby generating different RNA isoforms. A total 9,356 and 9,278 intron-containing genes were recently reported in the genomes of isolates Race 5 and Race 4, respectively (Zaccaron and Stergiopoulos 2024) (Table 4.S3). As introns are spliced by spliceosomes at conserved sequences, a splice site analysis was first performed in order to determine the extent to which intron splice sites were conserved between orthologous genes in the two isolates. A previous study that used high-depth RNA-seq data of *C. fulvum* isolate Race 5 grown in thirteen induction conditions, revealed that 7,843 (83.8 %) of the intron-containing genes had at least one of their introns supported by RNA-seq reads and 6,935 (74.1 %) genes had all their introns supported (Zaccaron et al. 2022), thereby largely confirming the accuracy of the predicted exon-intron annotations in this isolate. To investigate whether predicted introns of *C. fulvum* isolates Race 5 and Race 4 were also well supported by RNA-seq data during interaction with tomato, the RNA-seq reads obtained in this study were mapped to the genomes of *C. fulvum* Race 5 and Race 4. Ensuing splice site analysis revealed that from the 17,029 and 16,842 introns predicted in the genes of isolates Race 5 and Race 4, 13,639 (80.1%) and 12,718 (75.5%) could be confirmed with at least five RNA-seq reads, respectively. Moreover, from the 9,356 and 9,278 intron-containing genes in isolates Race 5 and Race 4, 7,769 (83.0%) and 7,284 (78.5%) had at least one of their introns confirmed, and 5,309 (56.7%) and 4,453 (48.0%) had all their introns confirmed, respectively. Thus, the predicted exon-intron structures of the gene annotations in *C. fulvum* isolate Race 4 were also well-supported by the RNA-seq data.

To further investigate whether orthologous genes in isolates Race 5 and Race 4 had similar number and size of introns, a total of 14,747 one-to-one ortholog gene pairs were analyzed. These ortholog pairs were obtained with OrthoFinder (Emms and Kelly 2015), and included 98.4% and 99.0% of all predicted genes in isolates Race 5 (*n*= 14,993) and Race 4 (*n*= 14,895), respectively. From the 14,747 ortholog gene pairs, 14,739 (99.9%) pairs had the same number of introns in both orthologous genes (Fig 4.1A) and 14,648 (99.3%) pairs had the same total size of intronic sequences (Fig 4.1B), thereby indicating a high conservation in the number and size of introns between the two isolates.

Next, to investigate whether orthologous genes between the two isolates shared the same start and end coordinates of their introns, the gene annotation of isolate Race 4 was mapped to the genome of Race 5 using Liftoff, thereby resulting in a new annotation for isolate Race 4 but with its gene coordinates now modeled on the genome of Race 5. This enabled intron coordinates to be compared between pairs of ortholog genes, which revealed that from the 14,747 one-to-one gene orthologs between isolate Race 5 and Race 4, 14,729 (99.9%) had the same number of introns with the same start and end coordinates and only

18 ortholog pairs had different number of introns or intron coordinates. These results indicated that splicing sites and intron coordinates are highly conserved between isolates Race 5 and Race 4.

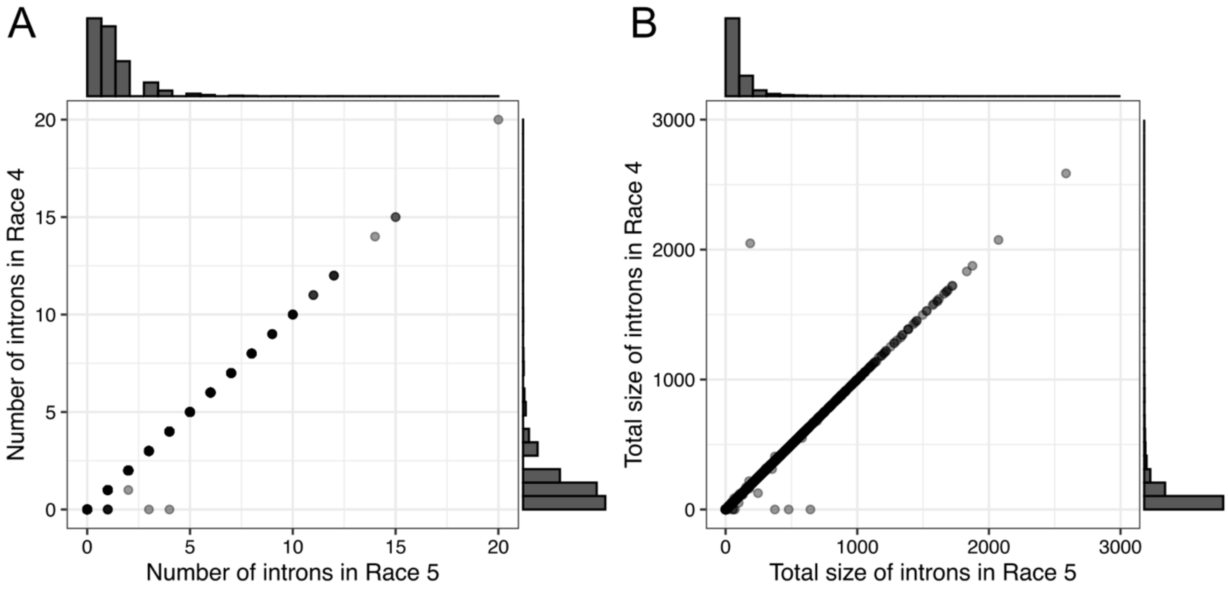


Figure 4.1: Number and size of introns that are conserved among pairs of orthologous genes in *Cladosporium fulvum* isolates Race 5 and Race 4. (A) Number of introns predicted in orthologous genes in isolates Race 5 and Race 4. (B) Total size of introns in base pairs (bp) between orthologous genes. Both scatter plots were generated by comparing a total of 14,747 one-to-one orthologous gene pairs from the genome annotations of isolates Race 5 (Zaccaron et al. 2022) and Race 4 (Zaccaron and Stergiopoulos 2024).

4.2.3 Transcriptome profiling of *C. fulvum* during host infections reveals significant transcript heterogeneity among isolates and different infections

The RNA-seq reads that were obtained in this study for the two isolates were used to perform reference-based transcriptome assemblies, using the genome of isolate Race 5 (Zaccaron et al. 2022) as reference. However, preliminary attempts indicated the presence of chimeric transcripts in the data, which possibly resulted from overlapping untranslated regions of physically close neighboring genes (Fig 4.S1). Therefore, transcripts were instead assembled in a gene-by-gene strategy, in which reads that mapped to the genes were first extracted and then used to assemble the unique transcript isoforms that originated from each individual gene.

Using the reads that mapped to the genomes of isolates Race 5 and Race 4, a total of 612,075 transcripts could be assembled across all timepoints (Table 4.S4). However, many of the transcripts were assembled multiple times among the samples, indicating that they were redundant. To obtain a set of unique transcripts, all the 612,075 assembled transcripts were clustered such that identical or fully contained sequences (i.e. a sequence within a sequence) were clustered together, and the longest sequence from each cluster was chosen as the representative. In doing so, a total of 57,148 clusters were generated, each of which represented a putative unique transcript isoform.

Of the 57,148 transcript isoforms, 43,824 transcript isoforms originated from genes of isolate Race 5, and 41,173 transcript isoforms originated from genes of isolate Race 4, indicating that both strains have similar numbers of uniquely assembled transcript isoforms. However, an overall low number of 27,849 (49%) transcript isoforms were common to both isolates, whereas 15,975 (28%) and 13,324 (23%) of them were unique to isolate Race 5 and isolate Race 4, respectively (Fig 4.S2A), indicating substantial heterogeneity in the transcript isoforms originating from orthologous genes in different isolates. A similar trend was also seen when examining the extent to which individual genes generated the same pool of transcript isoforms in different infections, in which case only 44% of the transcript isoforms from isolate Race 5 and 42% of the transcript isoforms from isolate Race 4 were present in all biological replicates of the experiment. Likewise, when examining the conservation of transcript isoforms in the individual timepoints of the interaction, then again a similar low level of reproducibility was seen in all three biological replicates, with only 35% to 56% of the transcript isoforms in isolate Race 5 being conserved in all three biological replicates for any given timepoint, and 39% to 52% of the transcript isoforms being conserved in isolate Race 4 (Fig 4.S3). Collectively, these results reveal that a significant level of transcript heterogeneity exists among biological samples, and although its source of variation is not known, it may suggest an abundance of noisy splicing in the data.

Upon further analysis, we observed that most of the putative isolate-specific transcript isoforms were assembled only in one sample, i.e. in one biological replicate of one timepoint of the interaction. Such one-off transcript isoforms accounted for 12,688 (79.4%) of the 15,975 transcript isoforms present in isolate Race 5, and 11,325 (85.0%) of the 13,324 transcript isoforms present in isolate Race 4, and likely represented spurious or random transcriptional events that were therefore removed. As a result, the combined number of unique transcript isoforms for the two isolates reduced considerably from 57,148 to 33,135, whereas within each isolate the number of unique transcript isoforms reduced from 43,824 to 31,136 in isolate Race 5 and from 41,173 to 29,848 in isolate Race 4 (Fig 4.S4A). As expected, the percentage of transcript isoforms shared by both isolates now increased from 49% to 84% and so were the overall percentages of transcript isoforms supported by all three replicates when considering all transcripts (Fig 4.S4B) and at each timepoint (Fig 4.S5). Collectively, these results highlight the difficulty of reproducing transcriptome assemblies and identifying recurrent AS events among fungal isolates and multiple biological replicates.

4.2.4 Genes that are alternatively spliced (AS) in *C. fulvum* during tomato infections are enriched in transporters, transcription factors (TFs), and cytochrome P450s

To further increase the accuracy of our AS analysis, transcript isoforms were filtered such that isoforms that shared all their intron coordinates (or lack of them) were considered as duplicates, and thus only the longest of them was kept. Following this filtering step, the number of unique transcript isoforms was reduced from 31,136 to 26,818 in isolate Race 5, and from 29,848 to 26,397 in isolate Race 4. Also, genes associated with more than one transcript isoform were now essentially considered as alternatively spliced, which included 6,034 (40.3%) genes from isolate Race 5 and 6,069 (40.5%) genes from isolate Race 4 (Table 4.S5).

Among the AS genes, 5,611 genes (i.e. 92.3% of the AS genes in isolate Race 5 and 92.5% of the AS genes in isolate Race 4) were common to both isolates, with 4,111 (73.3%) of them further generating the same number of transcript isoforms in each isolate (Table 4.S6). Most also of the AS genes in isolates Race 5 (*n*= 4,590; 76.1%) and Race 4 (*n*= 4,669; 76.9%) encoded just 2 or 3 transcript isoforms, and only a small number of AS genes from isolates Race 5 (*n*= 55; 0.9%) and Race 4 (*n*= 45; 0.7%) encoded for larger numbers of 10 or more transcript isoforms (Fig 4.2A). This indicates that, although many genes are predicted to undergo AS, overall, they generate a low number of transcript isoforms. Indeed, an analysis of the AS events revealed a total of 10,673 and 9,432 AS events in isolates Race 5 and Race 4, with an average of 1.77 and 1.55 AS events per AS gene in the two isolates, respectively. Further classification of AS events based on their type showed similar patterns for isolates Race 5 and Race 4, with intron retention (IR) being the most frequent AS type, and accounting for 71.1% and 70.7% of all classified AS events in isolates Race 5 and Race 4. respectively (Fig 4.2B and 4.2C, Table 4.S7). Also, in both isolates, 96% of all AS events were classified as either intron retention or alternative 3'/5' splice site, indicating a strong preference for these AS events over skipped exon, mutually exclusive exons, and alternative first/last exon, which collectively account for 4% of all AS events in both isolates (Fig 4.2B and 4.2C, Table 4.S7). Overall, these data indicate that isolates Race 5 and Race 4 have similar sets of genes undergoing AS during host infections.

We next investigated whether AS in pathogen genes during host infection affected some gene categories more than others, by performing a functional enrichment analysis based on conserved PFAM domains, gene ontology (GO) terms, and functional gene categories. Among the 6,034 and 6,069 AS genes in isolates Race 5 and Race 4, respectively significant overrepresentation (adjusted p-value< 0.01) of PFAM domains was observed for sugar-like transporters of the major facilitator superfamily (MFS), TFs of the fungal Zn(2)-cys(6) family, and cytochrome P450 enzymes (Fig 4.3A and 4.3B). Accordingly, among all the AS genes in both isolates, the most significantly enriched biological function GO terms were transmembrane transport (GO:0055085), regulation of transcription by RNA polymerase II (GO:0006357), carbohydrate transport (GO:0008643), and oxidoreductase activity (GO:0016491) (Fig 4.3C, Table 4.S8). Finally, based on hypergeometric tests, AS genes predicted to encode transporters (p-value< 1e-30), secreted proteins (p-value< 1e-8), cytochrome P450 enzymes (p-value< 1e-6), carbohydrate-active enzymes (CAZymes; p-value< 1e-6), and proteases (p-value< 0.01) were significant overrepresented among all AS genes in both isolates (Table 4.2). In contrast, no significant enrichment or depletion of genes encoding candidate effectors was detected among the pool of AS genes in isolate Race5 and/or Race 4. Collectively, these results indicate that AS in *C. fulvum* during host infections occurs more frequently in genes likely to be involved in the transport of sugars or other carbohydrates, regulation of genes, and monooxygenase activity, but less frequently in genes encoding proteins that are directly involved in modulation of host-immunity, such as effectors. Another possibility is that transcripts from these gene categories have high rates of turnover, making splicing intermediates (apparent IR isoforms) more likely to be detected and their proportion elevated relative to the mature RNAs.

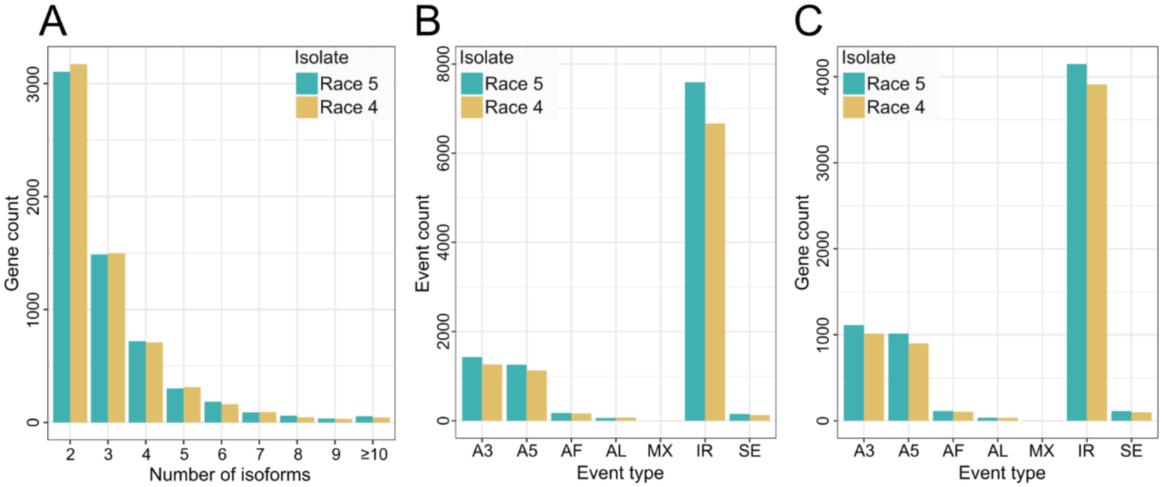


Figure 4.2: Genes that are alternatively spliced (AS) in *Cladosporium fulvum* isolates Race 5 and Race 4 during tomato infections, generally produce low numbers of transcript isoforms and undergo mostly intron retention events. (A) Bar chart showing the number of AS genes from isolates Race 5 and Race 4 (y-axis), producing 2 or more transcript isoforms (x-axis). (B) Bar chart showing the number of AS events in isolates Race 5 and Race 4 (y-axis), classified in one of the major types of AS (x-axis). (C) Bar chart showing the number of AS genes in isolates Race 5 and Race 4 (y-axis), classified in one of the major types of AS (x-axis). In (B) and (C), types of AS events shown are alternative 5'/3' splice sites (A5/A3), alternative first/last exons (AF/AL), mutually exclusive exons (MX), intron retained (IR), and skipping exon (SE).

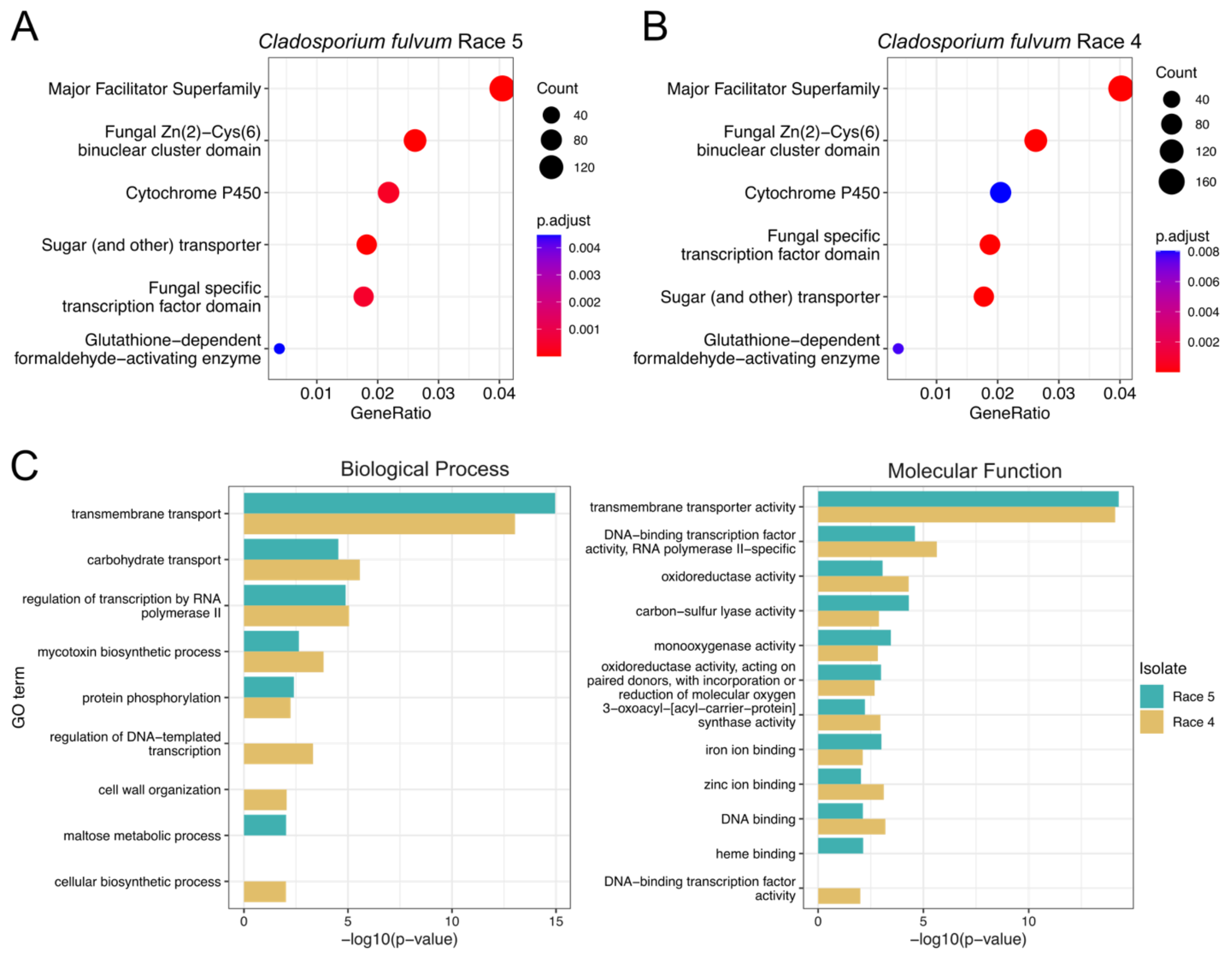


Figure 4.3: Genes that are alternatively spliced (AS) in *Cladosporium fulvum* isolates Race 5 and Race 4 during tomato infections, are enriched for MFS and sugar transporters, transcription factors, and cytochrome P450s. (A and B) Dot plots showing the conserved PFAM domains that are significantly enriched among the AS genes in isolates Race 5 and Race 4. The size of the dots corresponds to the number of AS genes containing the respective PFAM domain. The x-axis shows the proportion to which the AS genes containing the respective PFAM domain contribute to all AS genes containing a conserved PFAM domain. Dots are color-coded based on enrichment p-values adjusted using the Benjamini–Hochberg method. (C) Bar charts showing p-values of gene ontology (GO) terms from the classes of biological process and molecular function that are enriched among the genes predicted to undergo AS in isolates Race 5 and Race 4. The x-axis indicates the enrichment p-values in negative log scale.

Table 4.9: Genes that are alternatively spliced (AS) in *Cladosporium fulvum* isolates Race 5 and Race 4 during tomato infections are enriched for different functional categories. The table shows the total number of genes and of AS genes in *C. fulvum* isolates Race 5 and Race 4 that are classified in the different gene functional categories. Enrichment p-values were obtained with hypergeometric tests.

	Isolate Race 5			Isolate Race 4	
Functional category	**Total in the genome**	**Count AS genes**	**p-value**	**Count AS genes**	**p-value**
Transporters	2293	1186	2.53E-33	1189	4.80E-34
Secreted proteins	1416	677	1.20E-09	673	4.63E-09
Cytochrome P450	133	86	1.11E-08	83	2.03E-07
CAZymes	525	270	1.07E-07	268	2.75E-07
Proteases	362	167	0.013	171	0.004
SM backbones	42	22	0.076	23	0.041
Candidate effectors	432	182	0.233	180	0.298

4.2.5 Genes that are AS in *C. fulvum* during tomato infections are more abundant in repeat-rich chromosomes and exhibit longer upstream intergenic regions

We have previously determined that, as in several other fungal plant pathogens (Dong et al. 2015), the genome of *C. fulvum* shows a compartmentalized architecture composed of gene-dense/repeat-poor regions interspersed with gene-sparse/repeat-rich regions (Zaccaron et al. 2022; Zaccaron and Stergiopoulos 2024). We therefore examined whether the frequency and type of AS events was affected by the localization of genes in gene-rich or gene-poor regions. An inspection of the distribution of AS genes on the different chromosomes using the genome assembly of isolate Race 5 as reference (Zaccaron et al. 2022), showed that the number of AS genes varied among the core chromosomes, ranging between 33.6% and 33.9% of the genes in Chr4 of isolates Race 4 and Race 5, respectively to 45.7% and 45.7% of the genes in Chr1 of isolates Race 4 and Race 5, respectively (Fig 4.4A, Fig 4.S6, Table 4.S9). Interestingly, in the dispensable Chr14, which is present in isolate Race 5 but absent in isolate Race 4, only 25.0% of the genes were AS. Overall, a somewhat positive correlation was observed between abundance of AS genes and repetitive DNA content in the chromosomes, except for the dispensable Ch14 (Fig 4.4B). For instance, 43.8% and 44.5% of the genes in Chr3, which has the highest repetitive DNA content among core chromosomes (62%), were AS during host infections, whereas only 36.9% of the genes in Chr13, which has the lowest repetitive DNA content (36%), were AS. These results indicate that AS occurs in higher frequency in genes located in repeat-rich regions of the genome, and that dispensable chromosomes are less likely to carry AS genes.

Next, we investigated whether AS is also more prevalent in genes located in gene-poor regions that are typically characterized by long intergenic regions and high amounts of repetitive DNA. To do so, the distribution of intergenic sizes was compared between AS and non-AS genes. In isolate Race 5, the upstream intergenic regions of AS genes (mean= 3,911 bp, median= 759 bp) were significantly longer (p-value< 2E-12) compared to the upstream intergenic regions of genes with no evidence of AS (mean= 3,290 bp, median= 684 bp) (Fig 4.4C). Similarly, in isolate Race 4, the upstream intergenic regions of AS genes (mean= 3,873 bp, median= 762 bp) were significantly longer (p-value< 3E-13) compared to the upstream intergenic regions of genes with no evidence of AS (mean= 3,313 bp, median= 682 bp) (Fig 4.4C). In contrast, no significant difference was observed when comparing the size of downstream intergenic regions of AS genes and genes with no evidence of AS in both isolates (Fig 4.4C). These results indicate that, in *C. fulvum*, the upstream intergenic regions of AS genes, which include promoter and other *cis*-regulatory regions of genes, are significantly longer compared to genes with no evidence of AS. Because long intergenic regions are almost always associated with high repetitive DNA content in *C. fulvum* (Zaccaron et al. 2022), the amount of repetitive DNA was investigated. Indeed, the average repetitive DNA content of the upstream intergenic regions of genes with evidence of AS was significantly (p-value< 0.01) larger in both isolates compared to genes with no evidence of AS (Fig 4.S7). However, no significant difference in repetitive DNA content was observed for the downstream intergenic regions comparing genes with and without evidence of AS (Fig 4.S7).

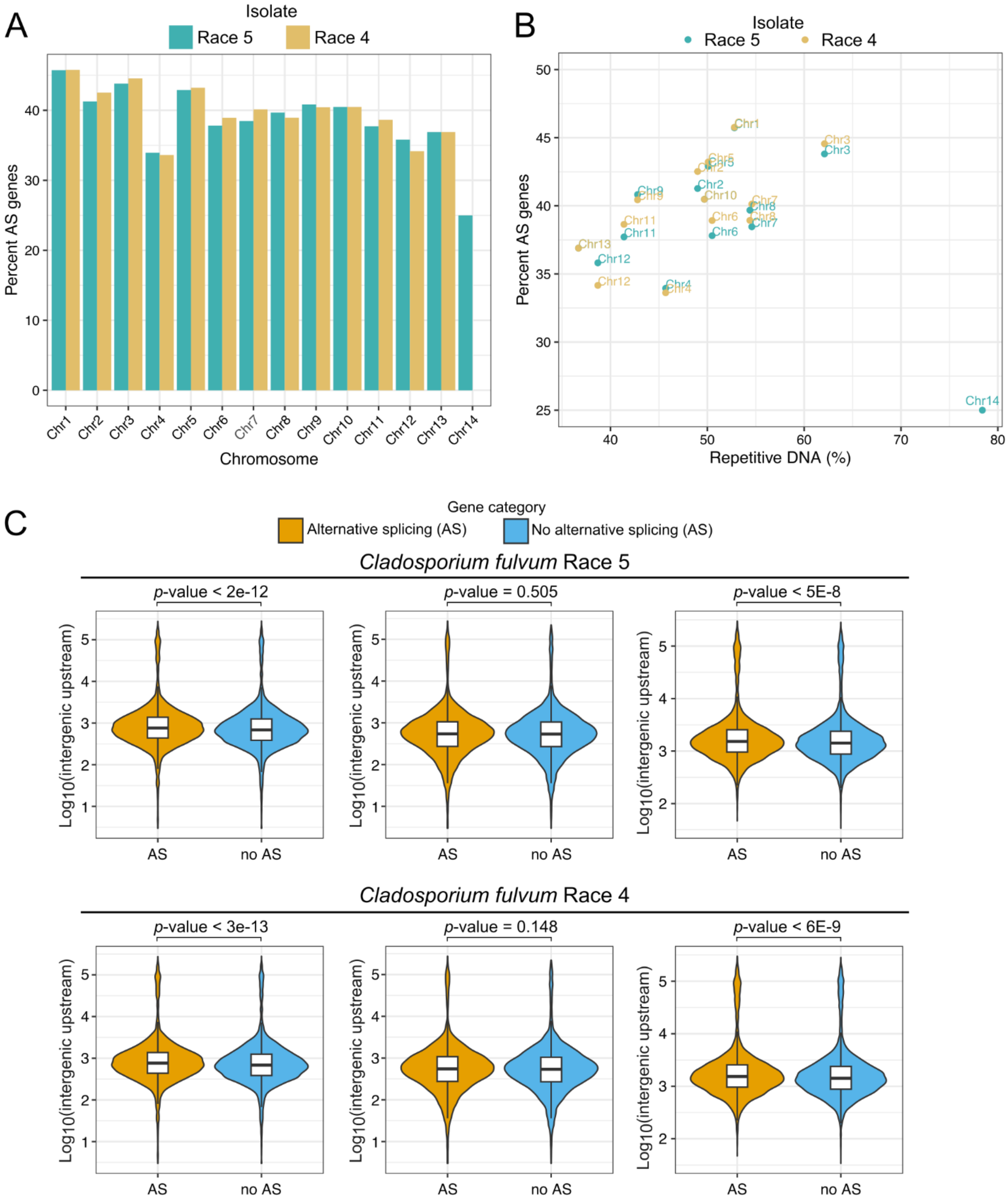


Figure 4.4: Genes that are alternatively spliced (AS) in *Cladosporium fulvum* isolates Race 5 and Race 4 during tomato infections, are preferentially located in repeat-rich chromosomes and exhibit longer upstream intergenic regions compared to non-AS genes. (A) Bar chart showing the percentage of AS genes (y-axis) present in each of the chromosomes (x-axis) of the reference genome of isolate Race 5. (B) Scatter plot showing that chromosomes with overall higher repeat content typically contain higher

percentages of AS genes, except for the dispensable chromosome Chr14 that is absent in isolate Race 4. Each point represents a chromosome, and points are color-coded to distinguish isolates Race 5 and Race 4. (C) Violin plots showing the distribution of intergenic sizes of genes predicted or not to undergo AS. Intergenic up- or downstream, and the total up- plus downstream intergenic sizes. P-values were obtained with the Wilcoxon rank sum test.

4.2.6 Genes that are AS in *C. fulvum* during tomato infections may increase proteome diversity

AS has the potential to increase protein diversity in cells when affecting coding sequences, as opposed to AS events in 3' or 5' untranslated regions (UTRs). To investigate the extent to which AS theoretically increased proteome diversity in *C. fulvum*, ORFs in the assembled transcript isoforms of isolates Race 5 and Race 4 were predicted. To do so, the transcript isoforms from isolate Race 4 that were assembled using the reference genome of isolate Race 5, were mapped to the genome of isolate Race 4 in order to obtain their nucleotide sequences.

A total of 26,632 and 26,108 ORFs could be predicted from the 26,818 and 26,397 transcript isoforms that were assembled from isolates Race 5 and Race 4, respectively. The sequences of the translated ORFs were subsequently organized into clusters with cd-hit such that identical or fully contained sequences were grouped together and each cluster represented a unique protein isoform. By doing so, a total of 19,757 and 19,551 protein isoforms were identified in isolates Race 5 and Race 4, respectively. These numbers are 31% higher than the predicted number of protein-encoding genes in the genomes of isolates Race 5 (*n*= 14,993) and Race 4 (n= 14,895) (Zaccaron and Stergiopoulos 2024), suggesting that AS could theoretically increase proteome diversity in *C. fulvum* during host infections.

From the 6,034 and 6,069 genes with evidence of AS in isolate Race 5 and Race 4, respectively 3,545 (58.7%) and 3,554 (58.5%) genes were predicted to produce distinct protein isoforms (Table 4.S10), indicating that the rest ~41% of the AS genes in *C. fulvum* (i.e. 2,489 (41.3%) genes in isolate Race 5 and 2,515 (41.5%) genes in isolate Race 4) experienced splicing events in non-coding sequences, such as 3' or 5' UTRs. When considering this data in view of all whole genome, then only between 23.6% to 23.9% of the total protein-coding genes in *C. fulvum* contributed through AS to proteome diversity. A functional enrichment analysis indicated that genes encoding transporters (p-value< 4E-13), secreted proteins (p-value< 8E-9), and to a lesser extent CAZymes (p-value< 8E-4) and cytochrome P450 enzymes (p-value< 3E-3), were overrepresented in the pool of AS genes in isolates Race 5 and/or Race 4 that produced multiple protein isoforms (Table 4.S11). For instance, in isolate Race 5 and Race 4, 432 (12.2%) and 424 (11.9%) of such AS genes, encoded secreted proteins, respectively (Table 4.S10). Interestingly, there were 134 effector-encoding genes among the pool AS genes yielding distinct protein isoforms, although on a genome-wide level no enrichment for genes encoding candidate effectors was observed (Table 4.S11). Included among those were the previously described effector genes *Ecp1* (Van den Ackerveken et al. 1993), *Ecp5* (Stergiopoulos et al. 2007a), *Ecp6* (Bolton et al. 2008), and *Ecp12* (Mesarich et al. 2018) (Fig 4.S8).

The vast majority of the AS genes producing multiple protein isoforms (i.e. 2,498 (70.5%) genes in isolate Race 5 and 2,540 (71.5%) AS genes in isolate Race 4) were predicted to yield just two isoforms (Fig 4.5A) and only a relatively a small number of AS genes (i.e. 117 (3.3%) in isolate Race 5 and 134 (3.8%) in isolate Race 4) were predicted to encode five or more distinct protein isoforms (Fig 4.5A). Along the same lines, of the 2,950 AS genes that were common to both isolates, 2,207 (74.8%) of them were predicted to yield a similar number of distinct protein isoforms, indicating a similar contribution to proteome diversity (Fig 4.5B, Table 4.S10). Two notable examples of such AS genes are *CLAFUR5_09979* and *CLAFUR5_09583*, both of which encode putative transcription factors (TFs). Interestingly, *CLAFUR5_09979* which was predicted to yield 13 distinct protein isoforms in both isolates Race 5 and Race 4 (Fig 4.S9), the highest number of distinct protein isoforms produced by an AS gene in the genome of *C. fulvum,* is an ortholog (44.1% aa identity) of the C2H2 finger domain TF CON7 (accession G4N5Q2) required for appressorium formation and pathogenicity of *Magnaporthe oryzae* (Odenbach et al. 2007; Shi et al. 1998). Similarly, *CLAFUR5_09583*, which was predicted to yield 10 distinct protein isoforms in both isolates Race 5 and Race 4 (Fig 4.S9), is an ortholog (46% aa identity) of the ascospore maturation 1 protein (Asm-1; accession XP_960837.1) that regulates sexual and asexual reproduction in *Neurospora crassa* (Aramayo et al. 1996). The contribution of these two genes in virulence of *C. fulvum* is unknown, but since both genes were AS in both isolates during infection and were predicted to yield the same number of protein isoforms, it is possible that some of the produced isoforms are biologically meaningful for infections.

We finally examined whether next to putatively increasing proteome diversity in *C. fulvum*, AS also altered the type of proteins produced during infection. To do so, we examined for gain or loss of conserved domains or of a signal peptide within the pool of protein isoforms that were predicted to be produced by AS genes in isolates Race 5 and/or Race 4. A total of 1,664 AS genes in isolate Race 5 and 1,841 AS genes in isolate Race 4 produced multiple protein isoforms that showed presence/absence variation in their PFAM domains (Table 4.S12). PFAM domains that varied the most among the protein isoforms were the major facilitator superfamily domain (PF07690; 67 AS genes in Race 5 and 81 AS genes in Race 4), the fungal Zn(2)-Cys(6) binuclear cluster domain (PF00172; 78 AS genes in Race 5 and 74 AS genes in Race 4), the short chain dehydrogenase domain (PF00106; 41 AS genes in Race 5 and 48 AS genes in Race 4), and the fungal specific TF domain (PF04082; 33 AS genes in Race 5 and 36 AS genes in Race 4). Along the same lines, of the 1,302 and 1,272 AS genes in isolates Race 5 and Race 4 respectively that were predicted to produce protein isoforms with an N-terminal signal peptide (i.e. 2,358 and 2,161 isoforms, respectively), 355 AS genes in Race 5 and 362 AS genes in Race 4 (intersection size = 270) yielded isoforms with presence/absence variation of signal peptide (Table 4.S12). Among these, there were 94 CAZyme and 76 candidate effector encoding genes which included the previously described *Ecp30, Ecp42, Ecp53-1, Ecp33,* and *Ecp10-3* effectors (Mesarich et al. 2018).

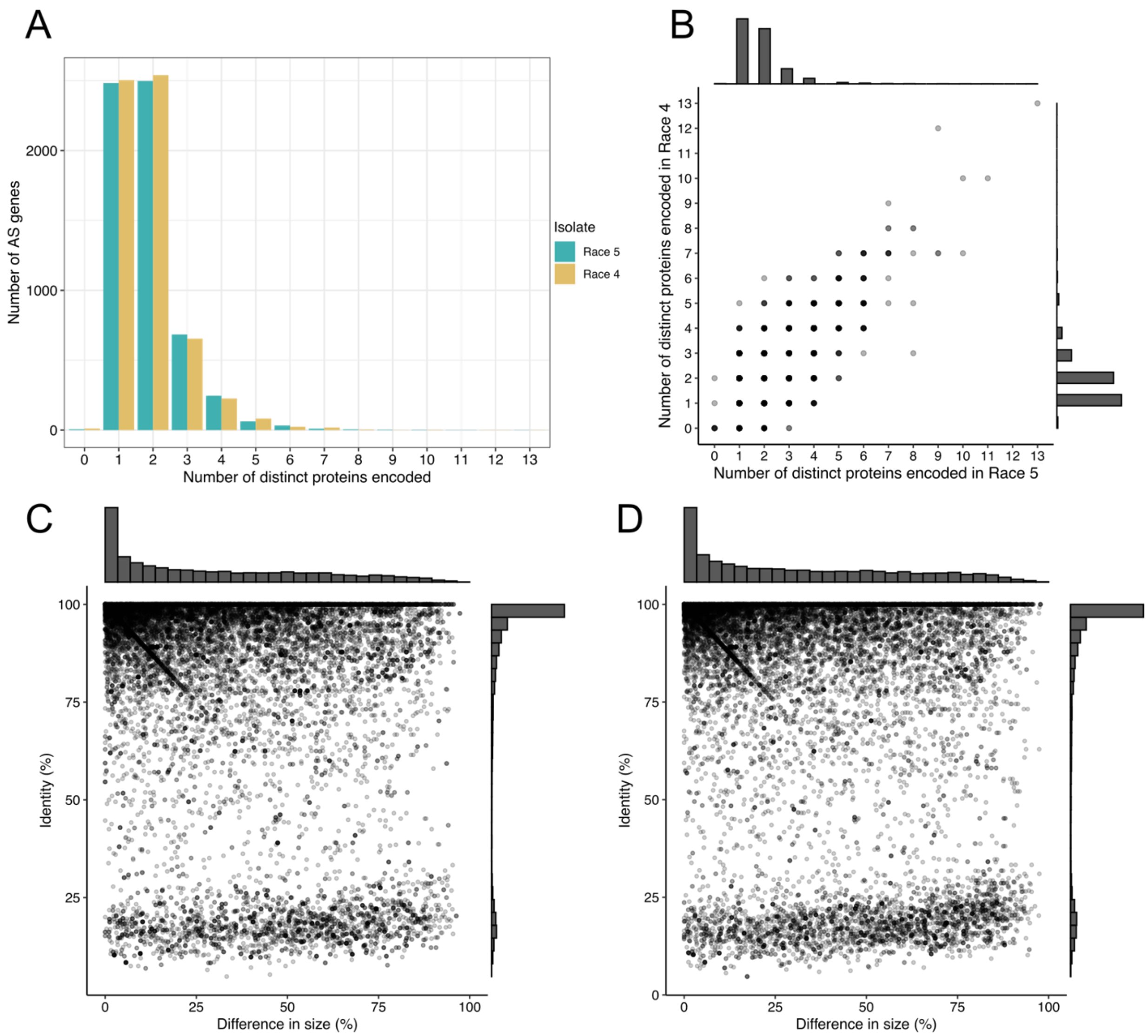


Figure 4.5: Alternative splicing (AS) in pathogen genes during tomato infections increases the proteome diversity of *Cladosporium fulvum* isolates Race 5 and Race 4. (A) Bar chart showing the number of genes in isolates Race 5 and Race 4 that are putatively producing through AS 0, 1, or more distinct protein isoforms (x-axis). (B) Scatter plot showing that a positive correlation exists among pairs of orthologous AS genes in *C. fulvum* isolates Race 5 and Race 4 that produce similar numbers of distinct protein isoforms. Each point represents a gene, and the plot shows a total of 6,493 AS genes in either isolate Race 5 or Race 4. (C and D) Scatter plots showing the diversity of protein isoforms produced by AS genes in isolate Race 5 (C) and Race 4 (D). Each point represents a pairwise alignment between the different protein isoforms that are produced by a single AS gene. Only AS genes predicted to yield two or more distinct protein isoforms are shown in the scatter plots. The Y-axis shows the percent amino acid identity among the protein isoforms, and the X-axis shows the difference in size between the aligned protein isoforms as the percentage of the longest aligned isoform. Alignments were generated based on the local-global strategy, which is based on aligning the longest sequences locally, and the shortest sequences globally.

4.2.7 Differential isoform usage across different infection timepoints is mostly isolate-specific

Differential isoform usage refers to statistically significant changes in the relative abundance of the AS isoforms produced by a gene across different conditions or timepoints. Such changes could be biologically relevant, as they might signify the preferential production of isoforms with varying functional potential at different stages of the infection. Therefore, a differential isoform usage analysis was performed in order to detect significant changes in the usage of the different transcripts produced by AS genes during the course of the infection.

To look for cases of differential usage of isoforms produced by individual AS genes of *C. fulvum* during disease progression on tomato, the expression values of the 26,818 and 26,397 assembled transcripts from isolates Race 5 and Race 4 were estimated at each of the seven timepoints for which RNA was sequenced. Then, differential isoform usage analyses were carried out between all possible pairwise comparisons among the seven timepoints. By doing so, a total of 401 transcripts from 246 genes of isolate Race 5 and 166 transcripts from 103 genes of isolate Race 4 showed significant (p-value< 0.01) changes in abundance during different timepoints of the infection (Fig 4.6A and 4.6B). Moreover, of the AS genes producing transcripts with evidence of differential usage, 111 genes in isolate Race 5 and 42 genes in isolate Race 4 had AS events affecting their open reading frame, and thereby putatively resulted in the production of protein isoforms with altered levels of abundance during disease progression (Table 4.S13). Most of these genes (i.e. 76 and 34 genes in isolates Race 5 and Race 4, respectively) produced just two protein isoforms but 35 genes in isolate Race 5 and seven genes in isolate Race 4 produced more than 3 isoforms. The majority also of the AS genes producing isoforms with differential usage during the infection process encoded hypothetical proteins (i.e. 55 in isolate Race 5 and 26 in isolate race 4) but 17 and 11 genes in isolate Race 5 and Race 4, respectively encoded for secreted proteins of which 2 and 2, respectively were candidate effectors (Fig 4.6C and 4.6D). When considering the intersection between isolate Race 5 and Race 4, then only 17 genes with differential usage of isoforms at the transcript level during disease progression were common to both isolates (Fig 4.S10, Fig 4.S11, Table 4.S13) and only five had AS events in their coding sequences and could thereby yield multiple protein isoforms. These five genes encoded a tRNA (uracil-O(2)-)-methyltransferase (*CLAFUR5_01082*), a lactose permease (*CLAFUR5_11255*), and hypothetical proteins (*CLAFUR5_08245, CLAFUR5_09805, CLAFUR5_20329*), and resulted in the production of two protein isoforms each in both isolates Race 5 and Race 4 (*CLAFUR5_01082, CLAFUR5_08245, CLAFUR5_20329*), or in a different number of protein isoforms in the two isolates (*CLAFUR5_11255, CLAFUR5_09805*). It is currently unknown whether differential isoform usage in these genes has any functional consequences for infections or if it is just transcriptional noise. However, given the low number of common between the two isolates genes with differential isoform usage, relatively to the total number of genes with differential usage of isoforms in each isolate, it can be assumed that the potential impact of differential isoform usage on host infections may be isolate-specific rather than at the species level.

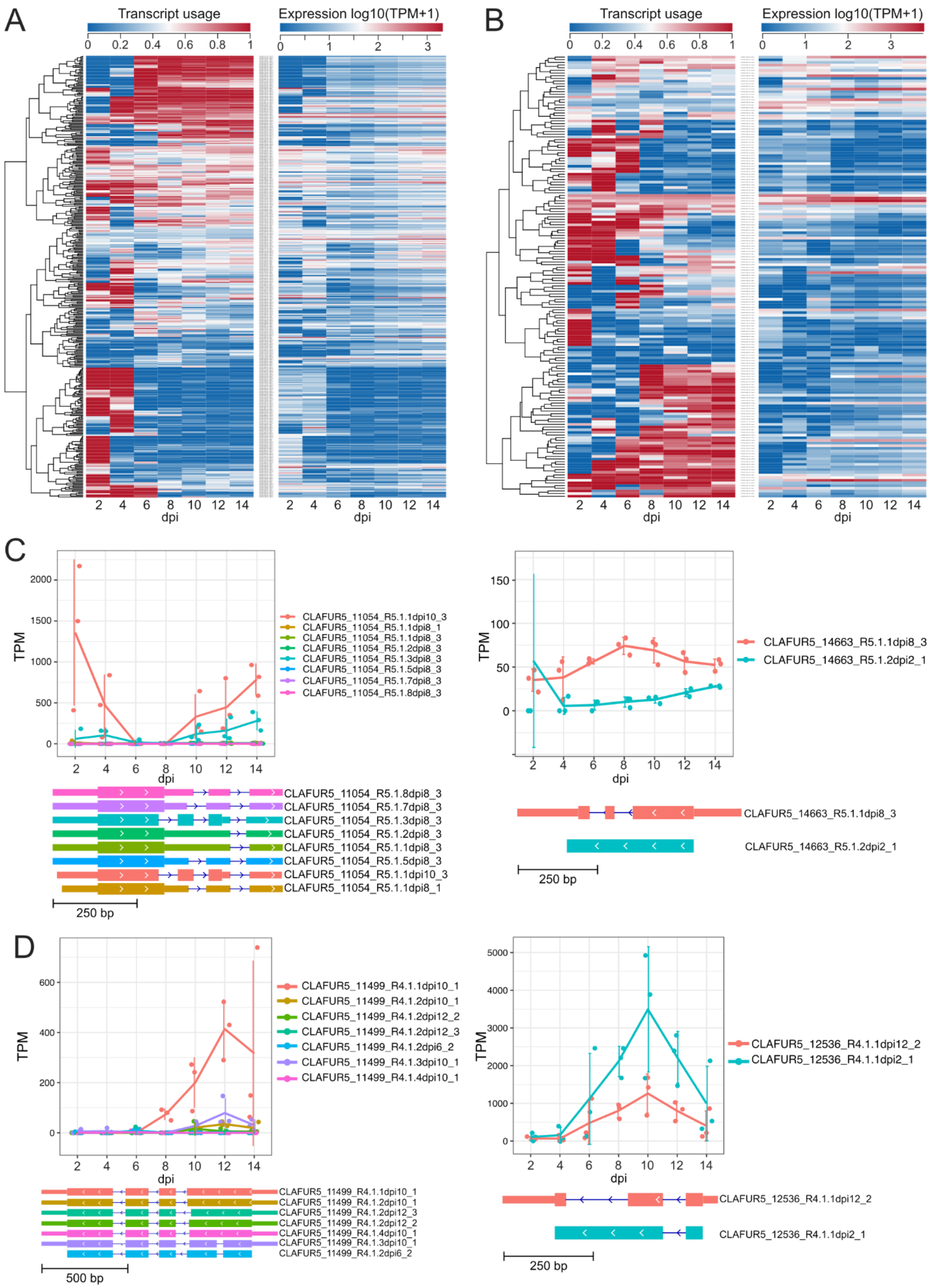

A
Transcript usage
0 0.2 0.4 0.6 0.8 1
Expression log10(TPM+1)
0 1 2 3
2 4 6 8 10 12 14
dpi
B
Transcript usage
0 0.2 0.4 0.6 0.8 1
Expression log10(TPM+1)
0 1 2 3
2 4 6 8 10 12 14
dpi
C
TPM
0 500 1000 1500 2000
CLAFUR5_11054_R5.1.1dpi10_3
CLAFUR5_11054_R5.1.1dpi8_1
CLAFUR5_11054_R5.1.1dpi8_3
CLAFUR5_11054_R5.1.2dpi8_3
CLAFUR5_11054_R5.1.3dpi8_3
CLAFUR5_11054_R5.1.5dpi8_3
CLAFUR5_11054_R5.1.7dpi8_3
CLAFUR5_11054_R5.1.8dpi8_3
2 4 6 8 10 12 14
dpi
CLAFUR5_11054_R5.1.8dpi8_3
CLAFUR5_11054_R5.1.7dpi8_3
CLAFUR5_11054_R5.1.3dpi8_3
CLAFUR5_11054_R5.1.2dpi8_3
CLAFUR5_11054_R5.1.1dpi8_3
CLAFUR5_11054_R5.1.5dpi8_3
CLAFUR5_11054_R5.1.1dpi10_3
CLAFUR5_11054_R5.1.1dpi8_1
250 bp
TPM
0 50 100 150
CLAFUR5_14663_R5.1.1dpi8_3
CLAFUR5_14663_R5.1.2dpi2_1
2 4 6 8 10 12 14
dpi
CLAFUR5_14663_R5.1.1dpi8_3
CLAFUR5_14663_R5.1.2dpi2_1
250 bp
D
TPM
0 200 400 600
CLAFUR5_11499_R4.1.1dpi10_1
CLAFUR5_11499_R4.1.2dpi10_1
CLAFUR5_11499_R4.1.2dpi12_2
CLAFUR5_11499_R4.1.2dpi12_3
CLAFUR5_11499_R4.1.2dpi6_2
CLAFUR5_11499_R4.1.3dpi10_1
CLAFUR5_11499_R4.1.4dpi10_1
2 4 6 8 10 12 14
dpi
CLAFUR5_11499_R4.1.1dpi10_1
CLAFUR5_11499_R4.1.2dpi10_1
CLAFUR5_11499_R4.1.2dpi12_3
CLAFUR5_11499_R4.1.2dpi12_2
CLAFUR5_11499_R4.1.4dpi10_1
CLAFUR5_11499_R4.1.3dpi10_1
CLAFUR5_11499_R4.1.2dpi6_2
500 bp
TPM
0 1000 2000 3000 4000 5000
CLAFUR5_12536_R4.1.1dpi12_2
CLAFUR5_12536_R4.1.1dpi2_1
2 4 6 8 10 12 14
dpi
CLAFUR5_12536_R4.1.1dpi12_2
CLAFUR5_12536_R4.1.1dpi2_1
250 bp

Figure 4.6: Differential isoform usage in AS genes of *Cladosporium fulvum* isolates Race 5 and Race 4 during the time course of tomato infections. (A and B) Transcript usage and relative expression of the 401 transcripts from 246 genes of isolate Race 5 (A) and 166 transcripts from 103 genes of isolate Race 4 (B) with intersect levels of abundance, when originating from the same gene, during differen t timepoints of the infection. In both (A) and (B), the left hand-side heat map shows transcript usage as the fraction of the sum of the expression from all transcripts from the gene considering the average expression values of three replicates. The heat map on the right shows the expression in transcripts per million (TPM) for the transcripts. (C and D) Examples of AS genes encoding candidate effectors from *C. fulvum* isolates Race 5 and Race 4 with differential usage of isoforms during the course of the infection. The line graphs show three AS genes, i.e. CLAFUR5_11054 and CLAFUR5_14663, from isolate Race 5 (C) and three AS genes, i.e. CLAFUR5_11499 and CLAFUR5_12536, from isolate Race 4 (D) that produce transcripts whose relative abundance intersects during the course of the infection. In the line graphs, the points represent the expression values in TPM (transcripts per million) of the individual transcripts at different timepoints of the infection. Standard deviation in the TPM values from three infections (i.e. biological replicates) is shown as vertical lines. The trends of transcript expression across time are shown as thick lines connecting the average TPM values for each individual transcript. The exon/intron structures of the transcripts are shown at the bottom of each line graph, with the predicted coding sequences represented as thicker boxes.

4.3 Discussion

Alternative splicing (AS) is one of the major phenomena that increases proteome diversity in eukaryotes. However, AS in fungal plant pathogens remains largely unexplored, particularly in the context of host-microbe interactions. In this study, we sequenced the transcriptomes of isolates Race 5 and Race 4 of the fungal pathogen *C. fulvum* and followed the landscape of AS events occurring in pathogen genes nearly throughout the infection process and in three independent tomato infections. This was done in order to assess the reproducibility of AS events and, based on their conservation acro ss different fungal isolates, infections, and infection stages, identify the ones that are most likely to be functional relevant for virulence rather than being splicing noise. Our studies revealed that approximately 40% of the genes in each of the two isolates of *C. fulvum* were AS at some point during host infection. However, when analyzing the transcript isoforms that were assembled from pathogen genes at different infections and infections stages, then an extensive heterogeneity was observed, suggesting that most isoforms were likely the result of aberrant splicing and therefore not recurrently present from one isolate or infection to another. The overall picture that emerged from our studies is that although AS in pathogen genes during host infections generates an extensive diversity of transcript isoforms, the majority of these seem to be the result of stochastic noise in the splicing machinery, as extensive presence/absence variability exists in transcript isoforms in different isolates and infections. Similar observations were made in the entomopathogenic fungus *Beauveria bassiana* (Dong et al. 2017). In this study, the authors analyzed two biological replicates of its transcriptome during interaction with insect hosts, and observed an overall poor overlap of 556 AS genes between the 1,290 and 970 AS genes identified in each biological replicate, despite the replicates having a high Pearson's correlation coefficient > 0.95 (Dong et al. 2017). The dynamic nature of AS in cells and the complexity of the splicing machinery means that stochastic fluctuations in the splicing output are to an extent expected. Another factor contributing to the lack of consensus in AS events across different isolates and infections could be that even small changes in the infection conditions and in the host responses towards the pathogen may have a larger effect on the regulation of AS in pathogen genes and its eventual output. Finally, it cannot be excluded that factors that are difficult to control such as inaccuracies in transcriptome sequencing and assembly, stochasticity in splicing and transcript kinetics, cell-to-cell variability in splicing and transcription, etc, contribute to transcriptome plasticity and the discrepancies observed in the splicing outputs. StringTie, which was used in our analyses, is considered one of the best reference-based transcriptome assemblers, despite showing overall low precision levels of between 29% and 59% at the transcript level (Yu et al. 2020; Voshall et al. 2021). Based on this, it is possible that StringTie, as well as other transcriptome assemblers, produce false positives that are poorly consistent among replicates and thus contribute to the heterogeneity seen in the data.

In our analyses, 40.3% and 40.5% of the protein-coding genes in isolates Race 5 and Race 4 were AS during infection of tomato. These numbers are high compared to what was reported in other filamentous fungal pathogens, which typically have less than 30% of their genes undergoing AS (Muzafar et al. 2021; Fang et al. 2020). For example, in *B. bassiana*, 5.4% of all genes undergo AS (Dong et al. 2017). In the rice blast fungus *Magnaporthe oryzae*, between 16% and 18% of the genes go through AS during interaction with rice

(Jeon et al. 2022), and in the cereal pathogen *Fusarium graminearum*, 29% of all genes were reported to experience AS based on PacBio Iso-seq data (Lu et al. 2022). Our results indicate that the frequency of AS in *C. fulvum* is higher compared to other fungi but nonetheless similar to *Trichoderma longibrachiatum*, in which 48.9% of its genes were reported as AS (Xie et al. 2015). Despite such differences in the percentage of AS genes reported in different fungal species, most studies, including this one, are in agreement that intron retention is the most frequent type of AS in fungi (Dong et al. 2017; Gehrmann et al. 2016; Ibrahim et al. 2021; Jin et al. 2017; Xie et al. 2015; Jeon et al. 2022). This is in contrast to mammals, in which exon skipping is typically the prevalent type of AS (Black 2003). Currently, there is no conclusive explanation for this contrasting frequency in AS types between different groups of organisms. One possibility, however, is that introns in fungi are more frequently retained due to their small length (for instance, the average intron length in *C. fulvum* genes is 79 bp (Zaccaron and Stergiopoulos 2024)), which makes it less likely to introduce in-frame premature stop codons (Grützmann et al. 2014; Sieber et al. 2018; Fang et al. 2020). Another plausible explanation for the high prevalence of intron retention events reported herein could be the presence of intermediate RNA splicing products. It is possible that some of the observed intron retention events are snapshots of pre-mRNAs in the process of being spliced, rather than final, mature transcripts (Mayer et al. 2015).

AS in fungi is known to influence their cellular, physiological, and stress responses, including host infections (Fang et al. 2020). A comparative analysis of AS in seven human fungal pathogens showed that genes subjected to AS during host infections were mostly associated with the functionality of the cell membrane, whereas AS under environmental stress conditions mainly affected gen es with diverse regulatory functions. Collectively, these results suggested that AS could have a functional impact for host infections and adaptation to the host environment (Sieber et al. 2018). Various other studies have likewise highlighted that AS differentially affects different gene categories in response to various stresses and stimuli. For instance, in the rice-blast fungus *M. oryzae*, it was shown that genes that were AS during infections were mostly enriched for TFs and phospho-transferases (Jeon et al. 2022). In the insect pathogen *B. brassiana*, AS genes were enriched for different categories related to metabolism, protein synthesis, cellular communication, and cellular differentiation, but also transcription and cellular transport (Dong et al. 2017). In the mushroom-forming fungus *Schizophyllum commune*, there was an enrichment of genes encoding cytochrome P450s among the AS genes, but not TFs (Gehrmann et al. 2016). In the stem-rot fungus *Sclerotinia sclerotiorum,* AS genes during infection of oilseed rape (*Brassica napus*) were enriched in processes such as cellular modified amino acid metabolic process, organonitrogen compound catabolic process, and heterocycle biosynthetic process (Cheng et al. 2022). Taken together, the above indicate that the complement of genes undergoing AS in fungi varies among species and possibly relates to their lifestyle as well (Fang et al. 2020; Sieber et al. 2018; Grützmann et al. 2014). In our studies, we found that in *C. fulvum,* AS frequently affects genes encoding TFs, suggesting that it may have a predominantly regulatory effect, by affecting the interaction of the TFs with downstream *cis*-regulatory elements and thereby reprogramming gene expression in pathogen genes during host infections. The genes *CLAFUR5_09583* and *CLAFUR5_09979* were notable examples of AS genes with a potential impact on gene regulation, as they encode orthologs of the Asm-1 (Aramayo et al. 1996) and CON7 (Odenbach et al. 2007) TFs, respectively and each is yielding large numbers of 10 or more distinct protein isoforms. Recently, CON7 has been shown to be a key transcriptional regulator in *Fusarium graminearum*, affecting genes that play important roles for conidium production, virulence, sexual development, and vegetative growth (Shin Soobin et al. 2024). It is possible that the large number of CON7 isoforms generated in *C. fulvum* through AS alters the regulation of a wide range of genes involved in similar processes. Overall, given the general role of TFs is regulating cellular responses to diverse stimuli and stresses, including host-infection responses, the prevalence of splicing events in TF-encoding genes supports that AS has a role in *C. fulvum* in modulating infections on its tomato host.

Other functional gene category enriched among AS genes in *C. fulvum* were MFS transporters, sugar-like transporters, and cytochrome P450 enzymes. The functional significance of AS in these gene categories is perhaps less clear but given their general involvement in nutrition, metabolism, and cellular detoxification processes it may imply that AS in these genes promotes adaptation to the host environment as well. MFS transporters form the largest superfamily of secondary active transporters that can collectiv ely transport (i.e. import or export) a very a broad spectrum of substrates, thereby participating in diverse physiological processes and stress responses, including nutrient acquisition (e.g. sugar uptake), resistance against oxidative stress and xenobiotic compounds, acid tolerance, secretion of endogenously produced secondary metabolites and toxins, and others (Xu et al. 2014; Hayashi et al. 2002; Alexander et al. 1999; Perlin et al. 2014). Not surprisingly thus, they are often reported as virulence factors in plant and human pathogenic fungi (Costa et al. 2014; Callahan et al. 1999; Santos et al. 2017; Liu et al. 2017). In a similar way, fungal sugar transporters, many of which belong to the MFS superfamily, are involved in the uptake of small plant-derived sugar molecules, thereby playing an important role in nutrition during infections (Carbó and Rodríguez 2023). Moreover, they may have functions in carbon catabolite repression and the utilization by the fungus of the most favorable carbon source in its environment as well as in sugar sensing, thereby regulating through their interaction with TFs the production of hydrolyzing and other metabolizing enzymes needed for nutrition (Mattam et al. 2022; Wu et al. 2020; Adnan et al. 2018; Kim et al. 2013). Finally, cytochrome P450 monooxygenases are a diverse superfamily of proteins known to be involved in cellular metabolism, xenobiotic detoxification, synthesis of toxins (Shin et al. 2018; Chen et al. 2014b), and other metabolic processes. In plant pathogenic fungi, cytochrome P450s are known to be virulence factors through their involvement, for instance, in the detoxification of plant-derived toxins and the scavenging of reactive active oxygen species during infections (Zhang et al. 2022a; George et al. 1998; Shin et al. 2017). Collectively, the above suggest that in *C. fulvum*, AS in MFS transporters, sugar-like transporters, and cytochrome P450 enzymes during tomato infections may offer a means to augment and fine-tune virulence on the host at multiple levels, including metabolic adaptation to the host's carbon and nutrient environment, protection against oxidative stress and plant defense compounds, production of toxins during pathogenesis, and others. However, another possibility is that the frequent AS events observed in genes encoding TFs, MFS transporters, sugar-like transporters, and cytochrome P450 could be due to potential high turnover of transcripts from these gene categories. Transcripts with high turnover rates are often associated with high transcription rates and rapid degradation (Wada and Becskei 2017). This means that even though these transcripts are degraded quickly, they are also synthesized rapidly, resulting in high concentrations of splicing intermediates that will be detected by sequencing. This complicates the interpretation of alternative splicing from RNA-seq data.

The genome of *C. fulvum* is rich in transposable elements (TEs) and exhibits a bipartite architecture that resembles the 'two-speed genome' model of evolution with candidate effector genes enriched in TE-rich regions (Zaccaron et al. 2022). To investigate a possible connection between genome organization and occurrence of AS, we compared the intergenic regions of AS genes with non-AS genes. By doing so, we discovered that AS genes were more frequent in repeat-rich chromosomes, with the exception of the dispensable chromosomes Chr14, most likely because the majority of the genes in this chromosomes are transcriptionally inactive during interaction with tomato (Zaccaron and Stergiopoulos 2024). This observation suggests a connection between TEs and AS genes. Indeed, we showed that AS genes have significantly longer upstream intergenic regions with higher repetitive DNA content, but the same was not true for downstream intergenic regions. Insertion of TEs in intergenic or intronic regions has been previously associated with changes in AS patterns in plants and humans (Zhang et al. 2022b; Clayton et al. 2020; Varagona et al. 1992). Therefore, one possibility is that the insertion of TEs in the proximal promoter regions of genes in *C. fulvum* could trigger the occurrence of AS.

By comparing the expression of pathogen-derived transcripts at seven timepoints during host infections, we revealed several cases of differential isoform usage in AS genes of *C. fulvum*. Most of these genes encoded hypothetical proteins, but a few of those are putatively directly involved in virulence, as they encode for effector proteins. One such gene, for instance, is *Ecp41*, which produces multiple protein isoforms in both isolate Race 5 and Race 4, and there is evidence of differential isoform usage in at least isolate Race 5. Other effector-encoding genes with evidence of differential isoform usage at the transcript or proteome level were identified as well in isolates Race 5 or Race 4, but none that was common to both isolates. On the one hand, such data suggest that infection stage-dependent modulation of AS in effector-encoding genes could possibly prime infections of the host by the selective production of specific functional isoforms of the effectors, as a means of promoting virulence and/or avoiding recognition by the host. On the other hand, the data may further imply that the augmentation and fine-tuning of the infection process may proceed via alternative routes depending on the isolate, infection conditions, or other sources of stochastic variation.

Interestingly, ~41% of AS genes in *C. fulvum* and over 50% of the AS genes in isolates Race 5 and Race 4 that produce transcripts with evidence of differential usage during the infection had splicing events in their 5' or 3' UTRs, thereby producing just a single protein isoform. Such splicing events, although they do not result in aberrant proteins, they increase the functional diversity in the 5' or 3' UTRs, which could lead to significant alterations in gene expression and protein translation. 5’UTRs, for instance, contain cis-regulatory elements that affect mRNA stability, splicing, translation initiation, and efficiency, and whose alteration may significantly impact gene transcription and translation (Wieder et al. 2024; Ryczek et al. 2023). In a similar way, 3’ UTRs contain elements that regulate stability, localization, and translation of mRNA and thus play a significant role in the post-transcriptional control of gene expression (Chan et al. 2022; Mayr 2019; Hong and Jeong 2023). Moreover, alterations in the 3’UTR are also shown to promote the differential cellular localization of protein isoforms, through the formation of 3'UTR-protein complexes (Ciolli Mattioli et al. 2019). Collectively, the high frequency of AS events in the 5' or 3' UTRs of *C. fulvum* genes during infections may suggest an additional level of post-transcriptional control of the infection process that was so far remains largely unexplored.

In this study, we revealed the landscape of AS in two isolates of the tomato pathogen *C. fulvum* during the course of their interaction with the tomato host. We show that AS is prevalent in pathogen genes during infections and that, despite its heterogeneity among isolates and infections, it is likely to have a multifactorial effect on host infections.

4.4 Materials and Methods

4.4.1 Inoculations of tomato plants with *C. fulvum* isolates

Cladosporium fulvum isolates Race 5 (Stergiopoulos et al. 2007b) and Race 4 (Boukema 1981) were grown in half-strength potato dextrose agar (PDA) for 2 weeks at 23°C. Tomato plants (*Solanum lycopersicum* cv. Moneymaker) were performed by conidia spray-inoculations as previously described (De Wit 1977; van Esse et al. 2007) and infected leaves were harvested at 2, 4, 6, 8, 10, 12, and 14 dpi. Collected samples were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction.

4.4.2 RNA extraction and sequencing

Samples were ground to a powder in liquid nitrogen, and total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA, USA). RNA quality was measured using a Qubit fluorometer (Life Technologies, New York, NY, USA) and the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Preparation and sequencing of the polyA-selected RNA-seq libraries were outsourced to the DNA Technologies and Expression Analysis Core Laboratory at the UC Davis Genome Center (https://dnatech.genomecenter.ucdavis.edu/). Libraries were sequenced on an Illumina NovaSeq 6000 instrument (PE150 format).

4.4.3 Comparison of introns and splicing sites

The number of introns supported by RNA-seq reads in *C. fulvum* Race 4 was obtained by analyzing the splicing junction table generated by STAR v2.7.9a, after mapping the RNA-seq reads to the genome of *C. fulvum* Race 4 (Zaccaron and Stergiopoulos 2024). To estimate conservation of number and size of introns, one-to-one pairs of orthologous genes from *C. fulvum* Race 5 and Race 4 were obtained with OrthoFinder v.2.5.4 (Emms and Kelly 2015). Number and size of introns of orthologous genes were then investigated with support of the script *agat_sp_add_introns.pl* from AGAT v1.2 package (Dainat et al. 2023) to add introns to the gene annotation files. To investigate whether the ortholog genes share the same intron start and end coordinates, the gene annotation of isolate Race 4 was mapped to the genome of isolate Race 5 using Liftoff v1.6.3 (Shumate and Salzberg 2020). This resulted in a new annotation with coordinates of genes, exons, and introns from isolate Race 4 in the genome of isolate Race 5. The new annotation generated by this procedure allowed direct comparison of the intron coordinates based on the reference genome of isolate Race 5.

4.4.4 Estimation of percentages of RNA-seq reads from *C. fulvum* and tomato genomes

Prior to read mapping, the RNA-seq reads were processed with the script *bbduk.sh* from BBMap package v38.90 (Bushnell 2014) to remove remaining adapter sequences (parameters: *ktrim=r*, *k=23*, *mink=11*, *hdist=1*, *minlength=40*, *tpe*, and *tbo*). To quantify and remove reads that originated from tomato, the trimmed reads were processed with *seal.sh* from BBMap package v38.90 (parameters: *ambiguous=toss*, and *k=27*), using as reference the genomes of *C. fulvum* Race 5 (Zaccaron et al. 2022) (GenBank accession GCA_020509005.2) and *Solanum lycopersicum* SL4.0 (Hosmani et al. 2019). Briefly, RNA-seq reads that had better match to the *S. lycopersicum* genome instead of the *C. fulvum* genome and reads that were equally well matched to both genomes, were filtered out. RNA-seq reads that had a better match to the *C. fulvum* genome and reads considered unmatched to any genome were used for downstream analysis.

4.4.5 Read mapping and transcriptome assembly

The filtered RNA-seq reads were mapped to the *C. fulvum* Race 5 genome with STAR v2.7.9a (Dobin et al. 2013) in 2-pass mode (parameters: *--twopassMode Basic*, *--alignIntronMin 20*, *--alignIntronMax 2000*, *--outSAMtype BAM SortedByCoordinate*). The mapped reads were assembled into full length transcripts with StringTie v2.1.7 (Pertea et al. 2015), using the gene annotation of *C. fulvum* Race 5 as reference (parameter: *-G*). Because the genome of *C. fulvum* Race 5 has many genes physically close to each other, with median intergenic region size of only 646 bp (Zaccaron et al. 2022), overlapping of untranslated regions (UTRs) between neighboring genes is common, which can result in chimeric transcripts during assembly. To minimize this issue, a gene-by-gene transcriptome assembly strategy was utilized. This strategy consisted of assembling the transcripts for each individual sample by extracting the RNA-seq reads from the respective sample mapped to the annotated gene space using SAMtools v1.9 (Li et al. 2009), and then assembled into full length transcripts using StringTie. By doing so, transcripts for each gene were obtained for each sample (7 timepoints x 3 replicates x 2 isolate = 42). To facilitate downstream analyses, assembled transcripts were assigned IDs that contained the name of the sample from which they originated.

4.4.6 Detection of alternative splicing genes

Using the GTF files of the assembled transcripts and the reference genomes of *C. fulvum* Race 5, transcript sequences were extracted with gffread v0.12.7 (Pertea and Pertea 2020). The sequences were then clustered with cd-hit v4.8.1 (parameters: *-T 8 -M 2048 -c 1 -d 0*) (Li and Godzik 2006), such that identical or fully contained transcripts were organized into the same cluster. Thus, each cluster represents a unique transcript. A table with the clusters were obtained with the script *cluster2txt* that comes with cd-hit. The organization of transcripts into clusters allowed to identify whether the cluster included transcripts assembled using RNA-seq reads from a specific biological replicate from a timepoint. Specifically, the representative transcript *t* of the cluster *c* was considered present in a sample *s* if there was at least one transcript *u* assembled using reads from *s* such that *u* was present in *c*. This strategy allowed the identification of transcripts supported by multiple replicates, and whether they were shared or unique to the isolates Race 5 and Race 4. Because transcripts supported by only one sample are likely random transcriptional events, only transcripts supported by at least two samples, i.e., transcripts that could be replicated, were considered for downstream analyses. After that, one additional filtering step was applied. First, the script *agat_sp_add_introns.pl* from AGAT v1.2 package (Dainat et al. 2023) was used to add intron coordinates to the GTF files containing the assembled transcripts. Then, the intron coordinates were compared among isoforms using a custom bash script. The isoforms containing the same number of introns and the same coordinates were considered duplicates, and only the longest one was kept. Finally, genes were considered undergoing AS, if they encoded at least two distinct transcripts that remained after the filtering steps.

4.4.7 Classification of AS types and gene enrichment

AS events were classified into Skipping Exon (SE), alternative 5'/3' Splice Sites (A5/A3), mutually exclusive exons (MX), intron retained (IR) , and alternative first/last exons (AF/AL) with the command *generateEvents* from SUPPA v2.3 (Trincado et al. 2018) with default settings, except for IR events, which were identified with the "variable boundary" parameter (*-b V*) set to 50 to relax the restrictive default behavior of SUPPA2 to identify IR events. Gene enrichments were performed for conserved PFAM domains, GO terms, and functional gene categories. Enrichment for PFAM domains was conducted with clusterProfiler v.4.6.2 (Yu et al. 2012) with Benjamini-Hochberg adjusted p-value threshold of 0.05 based on PFAM domains identified with InterProScan v5.59-91.0 (Jones et al. 2014). Enrichment for GO terms was performed with topGO v2.52 (Alexa and Rahnenführer 2009) with p-value threshold of 0.01 based on GO terms identified with PANNZER2 (Törönen et al. 2018), using minimum Positive Predictive Value (PPV) of 0.4. Enrichment of functional gene categories was performed with hypergeometric tests using the R function *phyper* based on the lists of gene category reported previously (Zaccaron and Stergiopoulos 2024). All three types of enrichment were conducted within R v4.3.1.

4.4.8 Prediction of ORFs and functional impact of AS

Before predicting ORFs within the transcripts, the transcripts in GFF format obtained from Race 4 using Race 5 genome as reference, were mapped to the genome of Race 4 using liftoff v1.6.3 (Shumate and Salzberg 2020) with default settings. By doing so, a new GFF file with transcripts coordinates in the genome of Race 4 was obtained. The nucleotide sequences of the transcripts from Race 5 and Race 4 were extracted using gffread v0.12.7 (Pertea and Pertea 2020). ORFs were predicted with ORFanage v1.2.0 (Varabyou et al. 2023) using parameters adjusted to finding ORFs of at least 120 bp that best matched the gene annotation (parameters: *--best* and *--minlen 120*). Transcripts with no predicted ORF were further processed with ORFfinder v0.4.3 with parameters adjusted to predict ORFs of at least 120 bp and starting only with ATG (parameters: *-s 0*, *-ml 120*, *-strand plus*). Only the longest ORFs predicted by ORFfinder were retained. For each isolate Race 5 and Race 4, the predicted protein sequences were clustered with cd-hit v4.8.1 (Li and Godzik 2006) such that identical or fully contained protein sequences were present in the same group (parameter *-c 1*). Genes encoding distinct proteins were identified based on the cd-hit results. Specifically, the protein sequences encoded by the isoforms of a gene grouped in distinct cd-hit clusters, then the gene was considered to encode distinct proteins. To investigate to what extent protein sequences encoded by isoforms of the same gene differ in amino acid sequence, the protein sequences encoded by isoforms were aligned in a pairwise manner using the *pairwiseAlignment* function within the R package Biostrings v2.68.1 (Pagès et al. 2022) using the "local-global" alignment strategy, such that the gaps at the end of the alignment do not penalize the alignment score. To investigate gain or loss of conserved motifs among isoforms, conserved PFAM domains within the protein sequences were identified with InterProScan v5.59-91.0 (Jones et al. 2014). Presence/absence variation of PFAM domains among protein isoform was obtained by analyzing the output of InterProScan using a custom R script within R v4.3.1. Proteins with signal peptide were predicted using SignalP v6 (Teufel et al. 2022). AS genes encoding at least one protein with a predicted signal peptide and at least one protein without a signal peptide were considered to exhibit gain or loss of signal peptide.

4.4.9 Transcript quantification and differential transcript usage

To estimate expression levels of the assembled transcripts, first their nucleotide sequences were extracted with BEDtools v2.29.0 (Quinlan and Hall 2010) and used to build an index with Salmon v1.10.0 (Patro et al. 2017) (parameter *--keepDuplicates*). After that, Salmon v1.10.0 was used in mapping-based mode (parameters: *-l IU*, *--gcBias*, *--seqBias*) to calculate the expression levels of the transcripts using the reads after trimming with *bbduk.sh* and selecting those with a *k*-mer matching with the *C. fulvum* genome. By doing so, Salmon generated expression values in Transcripts Per Million (TPM), which were used in SUPPA2 v2.3 (Trincado et al. 2018) to predict differential transcript usage among all possible pairs of timepoints following SUPPA2's specification. More precisely, isoform inclusion levels were quantified with the command *psiPerIsoform* using the TPM values, followed by the command *diffSplice* to calculate differential splicing between conditions with replicates (parameters: *--area 1000, --lower-bound 0.05*, *--combination*, *--tpm-threshold 2*, *--gene-correction*). A filtering step was carried out to keep only events with p-value < 0.01 and differential splicing value (dPSI) > 0.2.

4.5 Data availability

The raw RNA-seq reads were deposited in the NBCI's sequence reads archive (SRA) under accessions SRR29437234-SRR29437254 for isolate Race 5, and SRR29424125-SRR29424145 for isolate Race 4. The assembled transcripts and their expression values were deposited in a public repository, available at https://zenodo.org/records/11176736.

4.6 References

Adnan, M., Zheng, W., Islam, W., Arif, M., Abubakar, Y. S., Wang, Z., and Lu, G. 2018. Carbon Catabolite Repression in Filamentous Fungi. Int. J. Mol. Sci. 19

Alexa, A., and Rahnenführer, J. 2009. Gene set enrichment analysis with topGO. Bioconductor Improv. 27:1–26

Alexander, N. J., McCormick, S. P., and Hohn, T. M. 1999. TRI12, a trichothecene efflux pump from *Fusarium sporotrichioides*: gene isolation and expression in yeast. Mol. Gen. Genet. MGG. 261:977–984

Aramayo, R., Peleg, Y., Addison, R., and Metzenberg, R. 1996. *Asm-1* +, a *Neurospora crassa* Gene Related to Transcriptional Regulators of Fungal Development. Genetics. 144:991–1003

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Çolak, R., Kim, T., Misquitta-Ali, C. M., Wilson, M. D., Kim, P. M., Odom, D. T., Frey, B. J., and Blencowe, B. J. 2012. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science. 338:1587–1593

Black, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. 72:291–336

Bolton, M. D., Van Esse, H. P., Vossen, J. H., De Jonge, R., Stergiopoulos, I., Stulemeijer, I. J., Van Den Berg, G. C., Borrás-Hidalgo, O., Dekker, H. L., De Koster, C. G., and others. 2008. The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. Mol. Microbiol. 69:119–136

Boukema, I. 1981. Races of *Cladosporium fulvum* Cke.(*Fulvia fulva*) and genes for resistance in the tomato (Lycopersicon Mill.). Pages 287--292 in: Genetics and breeding of tomato: proceedings of the meeting of the Eucarpia Tomato Working Group, Avignon-France, May 18-21, 1981, Versailles, France: Institut national de la recherche agronomique, 1981.

Bushnell, B. 2014. *BBMap: a fast, accurate, splice-aware aligner*. Lawrence Berkeley National Lab (LBNL), Berkeley, CA.

Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., Knight, H., Nimmo, H. G., Zhang, R., and Brown, J. W. S. 2018. Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome. Plant Cell. 30:1424–1444

Callahan, T. M., Rose, M. S., Meade, M. J., Ehrenshaft, M., and Upchurch, R. G. 1999. *CFP*, the Putative Cercosporin Transporter of *Cercospora kikuchii*, Is Required for Wild Type Cercosporin Production, Resistance, and Virulence on Soybean. Mol. Plant-Microbe Interactions®. 12:901–910

Carbó, R., and Rodríguez, E. 2023. Relevance of Sugar Transport across the Cell Membrane. Int. J. Mol. Sci. 24

Chan, J. J., Zhang, B., Chew, X. H., Salhi, A., Kwok, Z. H., Lim, C. Y., Desi, N., Subramaniam, N., Siemens, A., Kinanti, T., Ong, S., Sanchez-Mejias, A., Ly, P. T., An, O., Sundar, R., Fan, X., Wang, S., Siew, B. E., Lee, K. C., Chong, C. S., Lieske, B., Cheong, W.-K., Goh, Y., Fam, W. N., Ooi, M. G., Koh, B. T. H., Iyer, S. G., Ling, W. H., Chen, J., Yoong, B.-K., Chanwat, R., Bonney, G. K., Goh, B. K. P., Zhai, W., Fullwood, M. J., Wang, W., Tan, K.-K., Chng, W. J., Dan, Y. Y., Pitt, J. J., Roca, X., Guccione, E., Vardy, L. A., Chen, L., Gao, X., Chow, P. K. H., Yang, H., and Tay, Y. 2022. Pan-cancer pervasive upregulation of 3' UTR splicing drives tumourigenesis. Nat. Cell Biol. 24:928–939

Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A., and Urrutia, A. O. 2014a. Correcting for Differential Transcript Coverage Reveals a Strong Relationship between Alternative Splicing and Organism Complexity. Mol. Biol. Evol. 31:1402–1413

Chen, W., Lee, M.-K., Jefcoate, C., Kim, S.-C., Chen, F., and Yu, J.-H. 2014b. Fungal Cytochrome P450 Monooxygenases: Their Distribution, Structure, Functions, Family Expansion, and Evolutionary Origin. Genome Biol. Evol. 6:1620–1634

Cheng, X., Zhao, C., Gao, L., Zeng, L., Xu, Y., Liu, F., Huang, J., Liu, L., Liu, S., and Zhang, X. 2022. Alternative splicing reprogramming in fungal pathogen *Sclerotinia sclerotiorum* at different infection stages on *Brassica napus*. Front. Plant Sci. 13:1008665

Ciolli Mattioli, C., Rom, A., Franke, V., Imami, K., Arrey, G., Terne, M., Woehler, A., Akalin, A., Ulitsky, I., and Chekulaeva, M. 2019. Alternative 3' UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. Nucleic Acids Res. 47:2560–2573

Clayton, E. A., Rishishwar, L., Huang, T.-C., Gulati, S., Ban, D., McDonald, J. F., and Jordan, I. K. 2020. An atlas of transposable element-derived alternative splicing in cancer. Philos. Trans. R. Soc. B Biol. Sci. 375:20190342

Costa, C., Dias, P. J., Sá-Correia, I., and Teixeira, M. C. 2014. MFS multidrug transporters in pathogenic fungi: do they have real clinical impact? Front. Physiol. 5:92186

Costanzo, S., and Jia, Y. 2009. Alternatively spliced transcripts of *Pi-ta* blast resistance gene in *Oryza sativa*. Plant Sci. 177:468–478

Dainat, J., Hereñú, D., Murray, D. K. D., Davis, E., Crouch, K., LucileSol, Agostinho, N., pascal-git, Zollman, Z., and tayyrov. 2023. NBISweden/AGAT: AGAT-v1.2.0.

De Wit, P. J. G. M. 1977. A light and scanning-electron microscopic study of infection of tomato plants by virulent and avirulent races of *Cladosporium fulvum*. Neth. J. Plant Pathol. 83:109–122

De Wit, P. J., Joosten, M. H., Thomma, B. H., and Stergiopoulos, I. 2009. Gene for gene models and beyond: the *Cladosporium fulvum*-Tomato pathosystem. Pages 135–156 in: Plant relationships, Springer.

De Wit, P. J., Van Der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., Bahkali, A. H., Beenen, H. G., Chettri, P., Cox, M. P., and others. 2012. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. PLoS Genet. 8:e1003088

Dinesh-Kumar, S., and Baker, B. J. 2000. Alternatively spliced *N* resistance gene transcripts: their possible role in tobacco mosaic virus resistance. Proc. Natl. Acad. Sci. 97:1908–1913

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21

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

Dong, W.-X., Ding, J.-L., Gao, Y., Peng, Y.-J., Feng, M.-G., and Ying, S.-H. 2017. Transcriptomic insights into the alternative splicing-mediated adaptation of the entomopathogenic fungus *Beauveria bassiana* to host niches: autophagy-related gene 8 as an example. Environ. Microbiol. 19:4126–4139

Emms, D. M., and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:1–14

van Esse, H. P., Bolton, M. D., Stergiopoulos, I., de Wit, P. J. G. M., and Thomma, B. P. H. J. 2007. The Chitin-Binding *Cladosporium fulvum* Effector Protein Avr4 Is a Virulence Factor. Mol. Plant-Microbe Interactions®. 20:1092–1101

Fang, S., Hou, X., Qiu, K., He, R., Feng, X., and Liang, X. 2020. The occurrence and function of alternative splicing in fungi. Fungal Biol. Rev. 34:178–188

Freitag, J., Ast, J., and Bölker, M. 2012. Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi. Nature. 485:522–525

Gehrmann, T., Pelkmans, J. F., Lugones, L. G., Wösten, H. A. B., Abeel, T., and Reinders, M. J. T. 2016. *Schizophyllum commune* has an extensive and functional alternative splicing repertoire. Sci. Rep. 6:33640

George, H. L., Hirschi, K. D., and VanEtten, H. D. 1998. Biochemical properties of the products of cytochrome P450 genes (*PDA*) encoding pisatin demethylase activity in *Nectria haematococca*. Arch. Microbiol. 170:147–154

Grützmann, K., Szafranski, K., Pohl, M., Voigt, K., Petzold, A., and Schuster, S. 2014. Fungal Alternative Splicing is Associated with Multicellular Complexity and Virulence: A Genome-Wide Multi-Species Study. DNA Res. 21:27–39

Hayashi, K., Schoonbeek, H., and De Waard, M. A. 2002. *Bcmfs1*, a novel major facilitator superfamily transporter from *Botrytis cinerea*, provides tolerance towards the natural toxic compounds camptothecin and cercosporin and towards fungicides. Appl. Environ. Microbiol. 68:4996–5004

Hong, D., and Jeong, S. 2023. 3'UTR Diversity: Expanding Repertoire of RNA Alterations in Human mRNAs. Mol. Cells. 46:48–56

Hooks, K. B., Delneri, D., and Griffiths-Jones, S. 2014. Intron Evolution in Saccharomycetaceae. Genome Biol. Evol. 6:2543–2556

Hoppins, S. C., Go, N. E., Klein, A., Schmitt, S., Neupert, W., Rapaport, D., and Nargang, F. E. 2007. Alternative Splicing Gives Rise to Different Isoforms of the *Neurospora crassa* Tob55 Protein That Vary in Their Ability to Insert β-Barrel Proteins Into the Outer Mitochondrial Membrane. Genetics. 177:137–149

Hosmani, P. S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., van Haarst, J., Cordewener, J., Sanchez-Perez, G., Peters, S., Fei, Z., Giovannoni, J. J., Mueller, L. A., and Saha, S. 2019. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv. :767764

Hossain, M. A., Rodriguez, C. M., and Johnson, T. L. 2011. Key features of the two-intron *Saccharomyces cerevisiae* gene *SUS1* contribute to its alternative splicing. Nucleic Acids Res. 39:8612–8627

Ibrahim, H. M. M., Kusch, S., Didelon, M., and Raffaele, S. 2021. Genome-wide alternative splicing profiling in the fungal plant pathogen *Sclerotinia sclerotiorum* during the colonization of diverse host families. Mol. Plant Pathol. 22:31–47

Jeon, J., Kim, K.-T., Choi, J., Cheong, K., Ko, J., Choi, G., Lee, H., Lee, G.-W., Park, S.-Y., Kim, S., and others. 2022. Alternative splicing diversifies the transcriptome and proteome of the rice blast fungus during host infection. RNA Biol. 19:373–386

Jin, L., Li, G., Yu, D., Huang, W., Cheng, C., Liao, S., Wu, Q., and Zhang, Y. 2017. Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus *Verticillium dahliae*. BMC Genomics. 18:130

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., and others. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236–1240

Kim, J.-H., Roy, A., Jouandot, D., and Cho, K. H. 2013. The glucose signaling network in yeast. Biochim. Biophys. Acta BBA - Gen. Subj. 1830:5204–5210

Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M. J. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat. Rev. Mol. Cell Biol. 14:153–165

Laloum, T., Martín, G., and Duque, P. 2018. Alternative splicing control of abiotic stress responses. Trends Plant Sci. 23:140–150

Leal, J., Squina, F. M., Freitas, J. S., Silva, E. M., Ono, C. J., Martinez-Rossi, N. M., and Rossi, A. 2009. A splice variant of the *Neurospora crassa hex-1* transcript, which encodes the major protein of the Woronin body, is modulated by extracellular phosphate and pH changes. FEBS Lett. 583:180–184

Leviatan, N., Alkan, N., Leshkowitz, D., and Fluhr, R. 2013. Genome-Wide Survey of Cold Stress Regulated Alternative Splicing in *Arabidopsis thaliana* with Tiling Microarray. PLOS ONE. 8:e66511

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

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

Lin, F., Zhang, Y., and Jiang, M.-Y. 2009. Alternative Splicing and Differential Expression of Two Transcripts of Nicotine Adenine Dinucleotide Phosphate Oxidase B Gene from *Zea mays*. J. Integr. Plant Biol. 51:287–298

Liu, L., Yan, Y., Huang, J., Hsiang, T., Wei, Y., Li, Y., Gao, J., and Zheng, L. 2017. A novel MFS transporter gene *ChMfs1* is important for hyphal morphology, conidiation, and pathogenicity in *Colletotrichum higginsianum*. Front. Microbiol. 8:1953

Lopes, M. E. R., Bitencourt, T. A., Sanches, P. R., Martins, M. P., Oliveira, V. M., Rossi, A., and Martinez-Rossi, N. M. 2022. Alternative Splicing in *Trichophyton rubrum* Occurs in Efflux Pump Transcripts in Response to Antifungal Drugs. J. Fungi. 8

Lu, P., Chen, D., Qi, Z., Wang, H., Chen, Y., Wang, Q., Jiang, C., Xu, J.-R., and Liu, H. 2022. Landscape and regulation of alternative splicing and alternative polyadenylation in a plant pathogenic fungus. New Phytol. 235:674–689

Mastrangelo, A. M., Marone, D., Laidò, G., De Leonardis, A. M., and De Vita, P. 2012. Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity. Plant Sci. 185–186:40–49

Mattam, A. J., Chaudhari, Y. B., and Velankar, H. R. 2022. Factors regulating cellulolytic gene expression in filamentous fungi: an overview. Microb. Cell Factories. 21:44

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J. A., and Churchman, L. S. 2015. Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. Cell. 161:541–554

Mayr, C. 2019. What are 3' UTRs doing? Cold Spring Harb. Perspect. Biol. 11:a034728

Merkin, J., Russell, C., Chen, P., and Burge, C. B. 2012. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science. 338:1593–1599

Mesarich, C. H., Barnes, I., Bradley, E. L., de la Rosa, S., de Wit, P. J. G. M., Guo, Y., Griffiths, S. A., Hamelin, R. C., Joosten, M. H. A. J., Lu, M., McCarthy, H. M., Schol, C. R., Stergiopoulos, I., Tarallo, M., Zaccaron, A. Z., and Bradshaw, R. E. 2023. Beyond the genomes of *Fulvia fulva* (syn. *Cladosporium fulvum*) and *Dothistroma septosporum*: New insights into how these fungal pathogens interact with their host plants. Mol. Plant Pathol. 24:474–494

Mesarich, C. H., Ökmen, B., Rovenich, H., Griffiths, S. A., Wang, C., Karimi Jashni, M., Mihajlovski, A., Collemare, J., Hunziker, L., Deng, C. H., and others. 2018. Specific hypersensitive response–associated recognition of new apoplastic effectors from *Cladosporium fulvum* in wild tomato. Mol. Plant. Microbe Interact. 31:145–162

Michael Weaver, L., Swiderski, M. R., Li, Y., and Jones, J. D. G. 2006. The *Arabidopsis thaliana* TIR-NB-LRR R-protein, RPP1A; protein localization and constitutive activation of defence by truncated alleles in tobacco and Arabidopsis. Plant J. 47:829–840

Muzafar, S., Sharma, R. D., Chauhan, N., and Prasad, R. 2021. Intron distribution and emerging role of alternative splicing in fungi. FEMS Microbiol. Lett. 368:fnab135

Nilsen, T. W., and Graveley, B. R. 2010. Expansion of the eukaryotic proteome by alternative splicing. Nature. 463:457–463

Odenbach, D., Breth, B., Thines, E., Weber, R. W. S., Anke, H., and Foster, A. J. 2007. The transcription factor Con7p is a central regulator of infection-related morphogenesis in the rice blast fungus *Magnaporthe grisea*. Mol. Microbiol. 64:293–307

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. 2022. Biostrings: Efficient manipulation of biological strings. R Package Version 2640.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods. 14:417–419

Perlin, M. H., Andrews, J., and San Toh, S. 2014. Chapter Four - Essential Letters in the Fungal Alphabet: ABC and MFS Transporters and Their Roles in Survival and Pathogenicity. Pages 201–253 in: Advances in Genetics, T. Friedmann, J.C. Dunlap, and S.F. Goodwin, eds. Academic Press.

Pertea, G., and Pertea, M. 2020. GFF Utilities: GffRead and GffCompare [version 1; peer review: 3 approved]. F1000Research. 9

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290–295

Preker, P. J., Kim, K. S., and Guthrie, C. 2002. Expression of the essential mRNA export factor Yra1p is autoregulated by a splicing-dependent mechanism. Rna. 8:969–980

Qin, F., Kakimoto, M., Sakuma, Y., Maruyama, K., Osakabe, Y., Tran, L.-S. P., Shinozaki, K., and Yamaguchi-Shinozaki, K. 2007. Regulation and functional analysis of <i>ZmDREB2A<i> in response to drought and heat stresses in *Zea mays* L. Plant J. 50:54–69

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

Ryczek, N., Łyś, A., and Makałowska, I. 2023. The Functional Meaning of 5'UTR in Protein-Coding Genes. Int. J. Mol. Sci. 24

Santos, R., Costa, C., Mil-Homens, D., Romão, D., de Carvalho, C. C. C. R., Pais, P., Mira, N. P., Fialho, A. M., and Teixeira, M. C. 2017. The multidrug resistance transporters CgTpo1_1 and CgTpo1_2 play a role in virulence and biofilm formation in the human pathogen *Candida glabrata*. Cell. Microbiol. 19:e12686

Shaul, O. 2017. How introns enhance gene expression. Splicing. 91:145–155

Shi, Z., Christian, D., and Leung, H. 1998. Interactions Between Spore Morphogenetic Mutations Affect Cell Types, Sporulation, and Pathogenesis in *Magnaporthe grisea*. Mol. Plant-Microbe Interactions®. 11:199–207

Shin, J., Kim, J.-E., Lee, Y.-W., and Son, H. 2018. Fungal Cytochrome P450s and the P450 Complement (CYPome) of *Fusarium graminearum*. Toxins. 10

Shin, J. Y., Bui, D.-C., Lee, Y., Nam, H., Jung, S., Fang, M., Kim, J.-C., Lee, T., Kim, H., Choi, G. J., Son, H., and Lee, Y.-W. 2017. Functional characterization of cytochrome P450 monooxygenases in the cereal head blight fungus *Fusarium graminearum*. Environ. Microbiol. 19:2053–2067

Shin Soobin, Park Jiyeun, Yang Lin, Kim Hun, Choi Gyung Ja, Lee Yin-Won, Kim Jung-Eun, and Son Hokyoung. 2024. Con7 is a key transcription regulator for conidiogenesis in the plant pathogenic fungus *Fusarium graminearum*. mSphere. 0:e00818-23

Shumate, A., and Salzberg, S. L. 2020. Liftoff: accurate mapping of gene annotations. Bioinformatics.

Sieber, P., Voigt, K., Kämmer, P., Brunke, S., Schuster, S., and Linde, J. 2018. Comparative Study on Alternative Splicing in Human Fungal Pathogens Suggests Its Involvement During Host Invasion. Front. Microbiol. 9

Singh, P., and Ahi, E. P. 2022. The importance of alternative splicing in adaptive evolution. Mol. Ecol. 31:1928–1938

Staal, J., and Dixelius, C. 2008. RLM3, a potential adaptor between specific TIR-NB-LRR receptors and DZC proteins. Commun. Integr. Biol. 1:59–61

Stergiopoulos, I., De Kock, M. J., Lindhout, P., and De Wit, P. J. 2007a. Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. Mol. Plant. Microbe Interact. 20:1271–1283

Stergiopoulos, I., Groenewald, M., Staats, M., Lindhout, P., Crous, P. W., and De Wit, P. J. 2007b. Mating-type genes and the genetic structure of a world-wide collection of the tomato pathogen *Cladosporium fulvum*. Fungal Genet. Biol. 44:415–429

Strijbis, K., van den Burg, J., Visser, W. F., van den Berg, M., and Distel, B. 2012. Alternative splicing directs dual localization of *Candida albicans* 6-phosphogluconate dehydrogenase to cytosol and peroxisomes. FEMS Yeast Res. 12:61–68

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 40:1023–1025

Thomma, B. P., Van Esse, H. P., Crous, P. W., and de Wit, P. J. 2005. *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. Mol. Plant Pathol. 6:379–393

Törönen, P., Medlar, A., and Holm, L. 2018. PANNZER2: a rapid functional annotation web server. Nucleic Acids Res. 46:W84–W88

Trevisan, G. L., Oliveira, E. H. D., Peres, N. T. A., Cruz, A. H. S., Martinez-Rossi, N. M., and Rossi, A. 2011. Transcription of *Aspergillus nidulans pacC* is modulated by alternative RNA splicing of *palB*. FEBS Lett. 585:3442–3445

Trincado, J. L., Entizne, J. C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D. J., and Eyras, E. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol. 19:1–11

Van den Ackerveken, G., Van Kan, J. A., Joosten, M., Muisers, J. M., Verbakel, H. M., and De Wit, P. 1993. Characterization of two putative pathogenicity genes of the fungal tomato pathogen *Cladosporium fulvum*. Mol Plant-Microbe Interact. 6:210–215

Varabyou, A., Erdogdu, B., Salzberg, S. L., and Pertea, M. 2023. Investigating open reading frames in known and novel transcripts using ORFanage. Nat. Comput. Sci. 3:700–708

Varagona, M. J., Purugganan, M., and Wessler, S. R. 1992. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. Plant Cell. 4:811–820

Voshall, A., Behera, S., Li, X., Yu, X.-H., Kapil, K., Deogun, J. S., Shanklin, J., Cahoon, E. B., and Moriyama, E. N. 2021. A consensus-based ensemble approach to improve transcriptome assembly. BMC Bioinformatics. 22:513

Wada, T., and Becskei, A. 2017. Impact of Methods on the Measurement of mRNA Turnover. Int. J. Mol. Sci. 18

Wang, S., Liang, H., Li, G., and Zhang, S.-H. 2021. Alternative splicing of *MoPTEN* is important for growth and pathogenesis in *Magnaporthe oryzae*. Front. Microbiol. 12:715773

Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M., and Wang, X.-Z. 2015. Mechanism of alternative splicing and its regulation. Biomed. Rep. 3:152–158

Wieder, N., D'Souza, E. N., Martin-Geary, A. C., Lassen, F. H., Talbot-Martin, J., Fernandes, M., Chothani, S. P., Rackham, O. J. L., Schafer, S., Aspden, J. L., MacArthur, D. G., Davies, R. W., and Whiffin, N. 2024. Differences in 5'untranslated regions highlight the importance of translational regulation of dosage sensitive genes. Genome Biol. 25:111

de Wit, P. J. 2016. *Cladosporium fulvum* effectors: weapons in the arms race with tomato. Annu. Rev. Phytopathol. 54:1–23

Wu, V. W., Thieme, N., Huberman, L. B., Dietschmann, A., Kowbel, D. J., Lee, J., Calhoun, S., Singan, V. R., Lipzen, A., Xiong, Y., Monti, R., Blow, M. J., O'Malley, R. C., Grigoriev, I. V., Benz, J. P., and Glass, N. L. 2020. The regulatory and transcriptional landscape associated with carbon utilization in a filamentous fungus. Proc. Natl. Acad. Sci. 117:6003–6013

Xie, B.-B., Li, D., Shi, W.-L., Qin, Q.-L., Wang, X.-W., Rong, J.-C., Sun, C.-Y., Huang, F., Zhang, X.-Y., Dong, X.-W., Chen, X.-L., Zhou, B.-C., Zhang, Y.-Z., and Song, X.-Y. 2015. Deep RNA sequencing reveals a high frequency of alternative splicing events in the fungus *Trichoderma longibrachiatum*. BMC Genomics. 16:54

Xing, Y., and Lee, C. 2006. Alternative splicing and RNA selection pressure — evolutionary consequences for eukaryotic genomes. Nat. Rev. Genet. 7:499–509

Xu, X., Chen, J., Xu, H., and Li, D. 2014. Role of a major facilitator superfamily transporter in adaptation capacity of *Penicillium funiculosum* under extreme acidic stress. Fungal Genet. Biol. 69:75–83

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J. Integr. Biol. 16:284–287

Yu, T., Mu, Z., Fang, Z., Liu, X., Gao, X., and Liu, J. 2020. TransBorrow: genome-guided transcriptome assembly by borrowing assemblies from different assemblers. Genome Res. 30:1181–1190

Zaccaron, A. Z., Chen, L.-H., Samaras, A., and Stergiopoulos, I. 2022. A chromosome-scale genome assembly of the tomato pathogen *Cladosporium fulvum* reveals a compartmentalized genome architecture and the presence of a dispensable chromosome. Microb. Genomics. 8:000819

Zaccaron, A. Z., and Stergiopoulos, I. 2024. Analysis of five near-complete genome assemblies of the tomato pathogen *Cladosporium fulvum* uncovers additional accessory chromosomes and structural variations induced by transposable elements effecting the loss of avirulence genes. BMC Biol. 22:25

Zhang, J., Jin, X., Wang, Y., Zhang, B., and Liu, T. 2022a. A Cytochrome P450 Monooxygenase in Nondefoliating Strain of *Verticillium dahliae* Manipulates Virulence via Scavenging Reactive Oxygen Species. Phytopathology®. 112:1723–1729

Zhang, L., Qian, J., Han, Y., Jia, Y., Kuang, H., and Chen, J. 2022b. Alternative splicing triggered by the insertion of a CACTA transposon attenuates *LsGLK* and leads to the development of pale-green leaves in lettuce. Plant J. 109:182–195

Zhang, M.-Y., and Miyake, T. 2009. Development and Media Regulate Alternative Splicing of a Methyltransferase Pre-mRNA in *Monascus pilosus*. J. Agric. Food Chem. 57:4162–4167

4.7 Supplementary materials

4.7.1 Supplementary figures

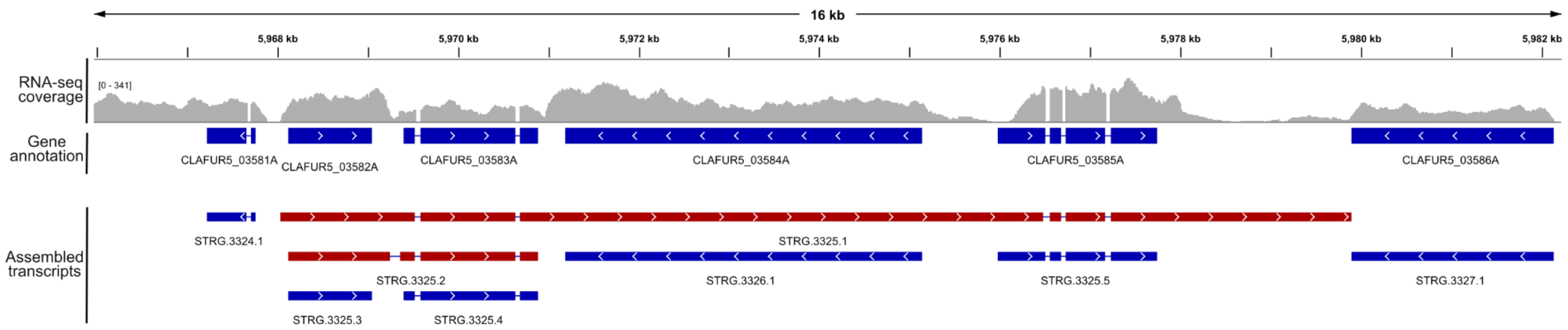


Figure 4.S1: Preliminary transcriptome assembly generated chimeric transcripts spanning genes physically close in the genome. The figure shows a region of 16 kb in chromosome 1 of *Cladosporium fulvum* Race 5 containing six predicted genes. RNA-seq coverage of reads from the first of the three infections that were performed (i.e. biological replicate 1) at 14 dpi during interaction with tomato is shown above the predicted genes. Reference-based transcriptome assembly based on the reads shown resulted in chimeric transcripts, shown in red, that span multiple genes.

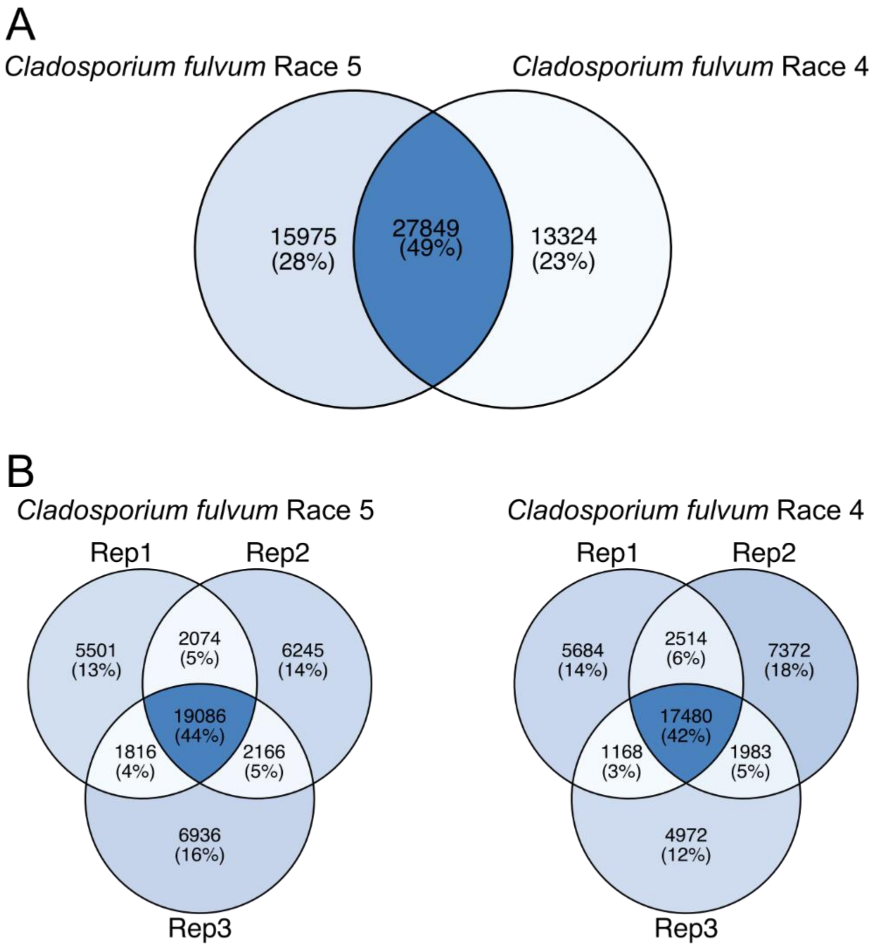


Figure 4.S2: A high heterogeneity in transcripts produced by *Cladosporium fulvum* isolates Race 5 and Race 4 during tomato infections is seen between the two isolates and the three different infections that were performed with each isolate. (A) The total number of transcripts that are shared between isolates Race 5 and Race 4. The Venn diagram shows all uniquely assembled transcripts, after combining all transcripts assembled across three independent tomato infections (i.e. biological replicates) and the seven timepoints sampled in each infection. (B) The total number of assembled transcripts shared by all biological replicates (Rep1, Rep2, and Rep3) for isolates Race 5 and Race 4. The Venn diagrams show all unique transcripts assembled for each infection and isolate, after combining the assembled transcripts from all seven sampled timepoints during the infection. Darker colors of intersections indicate higher numbers.

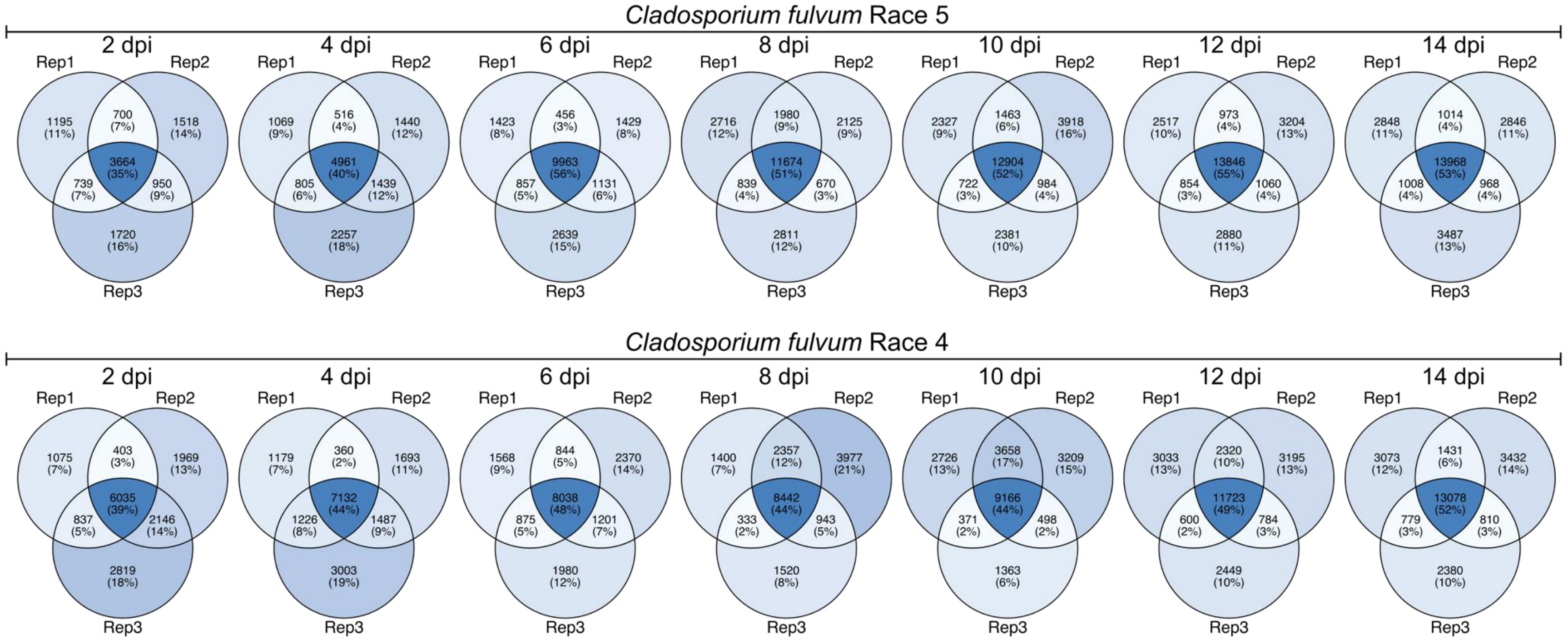


Figure 4.S3: An overall low number of transcripts are constitutively present in all three tomato infections performed either with *Cladosporium fulvum* isolates Race 5 or Race 4 and in every of the seven infection timepoints that were sampled. Venn diagrams showing the number of transcripts supported by one, two, or all three biological replicates (Rep1, Rep2, and Rep3) at each sampled timepoint (2, 4, 6, 8, 10, 12, and 14 dpi) for isolates Race 5 and Race 4. A biological replicate represents a different infection experiment with isolates Race 5 and Race 4. Darker colors of intersections indicate higher numbers. To identify transcripts supported by each replicate, all assembled transcripts were organized into clusters, such that each cluster contained identical or fully contained transcripts, and thus each cluster represented a unique transcript. A transcript is considered supported by all three replicates if the corresponding cluster contained transcripts assembled from all three replicates.

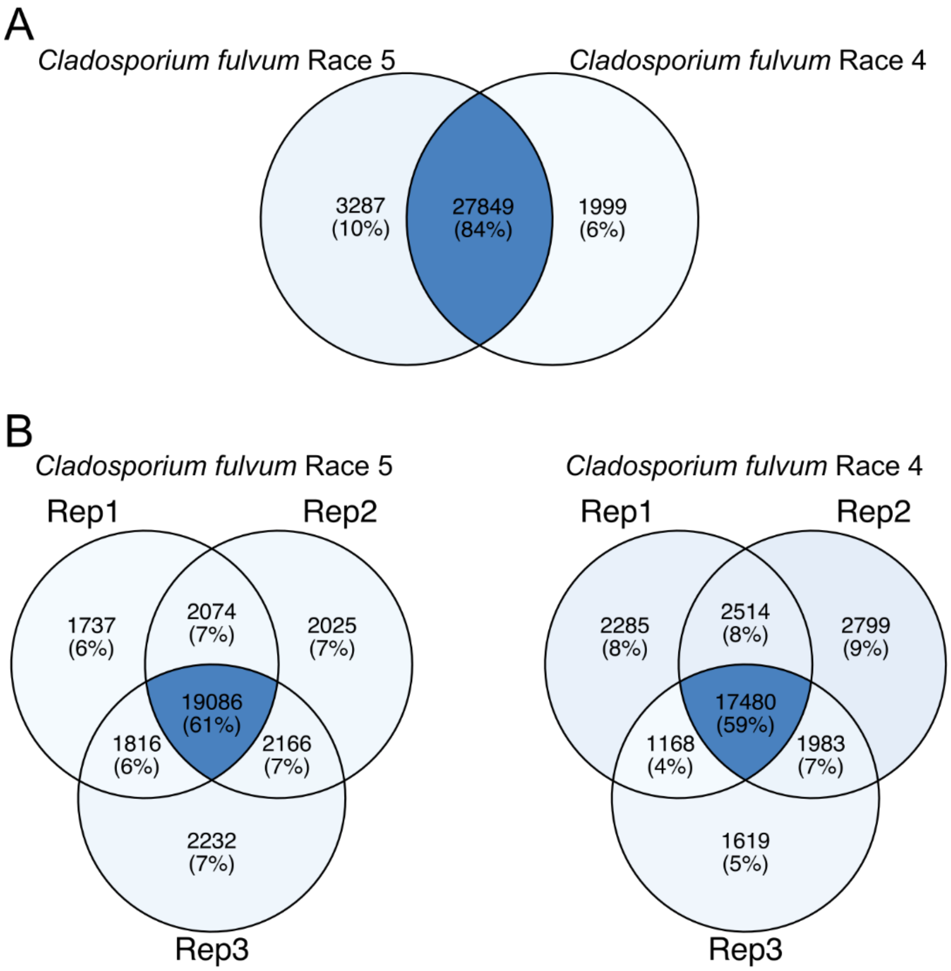


Figure 4.S4: The number of transcripts shared by *Cladosporium fulvum* isolates Race 5 and Race 4 increases when transcripts that could not be reproduced between infections were filtered out. (A) The total number of transcripts shared between isolates Race 5 and Race 4. The Venn diagram shows all uniquely assembled transcripts after combining all transcripts assembled across the sampled timepoints and infections (i.e. biological replicates), and subsequently removing transcripts that were present in only one sample, i.e., present in only one replicate, in one timepoint, for one isolate. (B) The total number of assembled transcripts shared among biological replicates (Rep1, Rep2, and Rep3) for isolates Race 5 and Race 4. The Venn diagrams show all unique transcripts assembled for each replicate after combining all timepoints and removing transcripts that were present in only one sample. Darker colors of intersections indicate higher numbers.

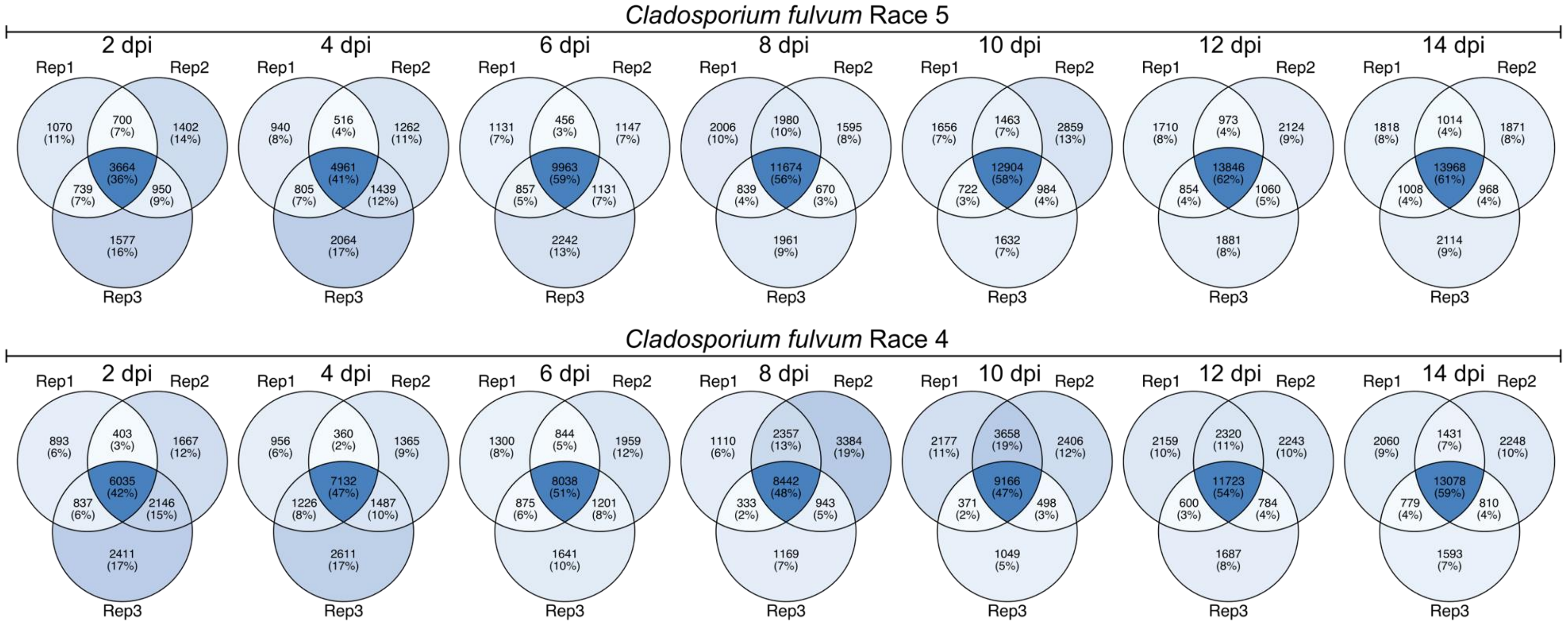


Figure 4.S5: The number of transcripts common in all three tomato infections performed either with *Cladosporium fulvum* isolates Race 5 or Race 4 and in each of the seven sampled infection timepoints, after filtering out transcripts that were present in only one sample. Venn diagrams showing the number of transcripts supported by one, two, or all three biological replicates (Rep1, Rep2, and Rep3) at each timepoint (2, 4, 6, 8, 10, 12, and 14 dpi) for isolates Race 5 and Race 4 after removing transcripts that were present in only one sample, i.e., present in only one replicate, in one timepoint, for one isolate. Darker colors of intersections indicate higher numbers.

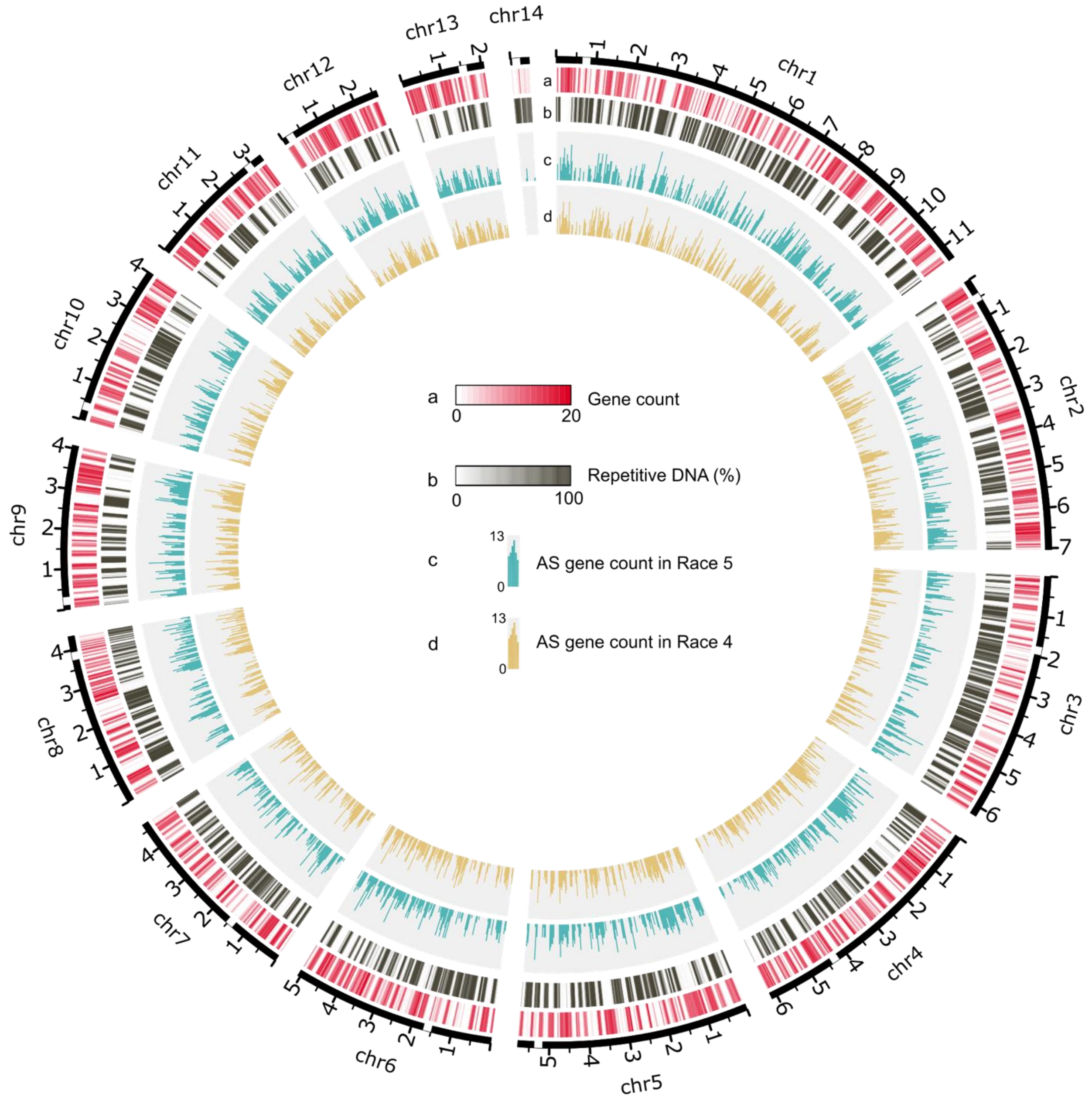


Figure 4.S6: The distribution of genes predicted to undergo alternative splicing (AS) in the genomes of *Cladosporium fulvum* isolates Race 5 and Race 4 during tomato infections. Circos plot showing the chromosomes of the reference genome of isolate Race 5. The two outermost tracks show the gene and repetitive DNA content, respectively. The predicted locations of centromeres are indicated with white rectangles on the outermost axis. The histograms in the two innermost tracks represent the number of AS genes (0 to 13) along the chromosomes for isolates Race 5 and Race 4. Numbers in all tracks were calculated using a sliding window of 30 kb.

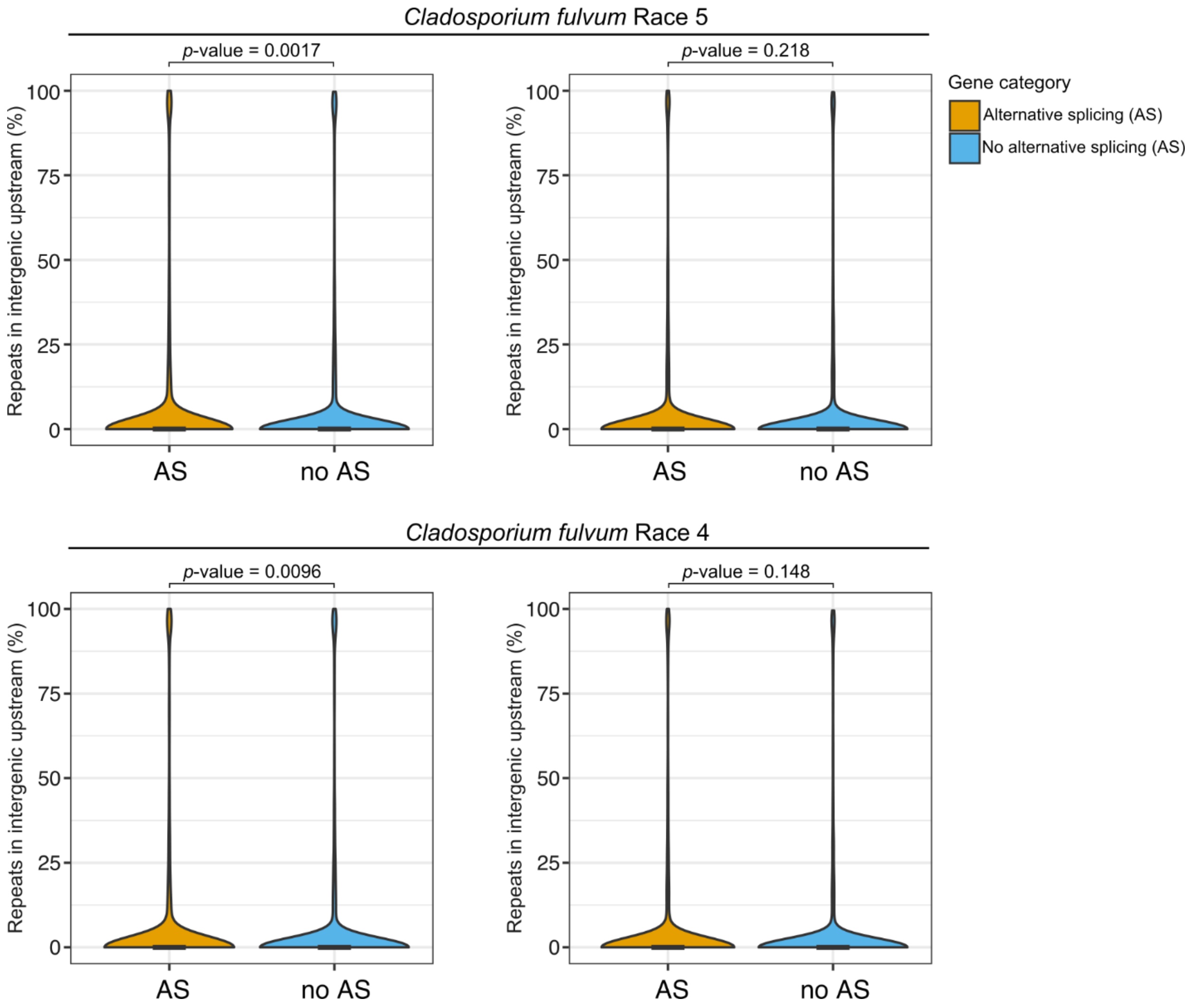


Figure 4.S7: The upstream regions of genes predicted to undergo alternative splicing (AS) in the genomes of *Cladosporium fulvum* isolates Race 5 and Race 4 during tomato infections, have higher amounts of repetitive DNA than genes with no evidence of AS. The violin plots show the distribution of the amount of repetitive DNA (i.e., predicted transposable elements) present in the up - and downstream intergenic regions of genes with and without evidence of AS in the genomes of isolates Race 5 and Race 4. The figure shows that the repetitive DNA content of upstream intergenic regions of AS genes in Race 5 (average= 6.26%) is significantly larger compared to the genes with no evidence of AS (average= 5.5%). Similar significance level was observed for upstream intergenic regions of genes from Race 4 with evidence of AS (average= 6.13%) and with no evidence of AS (average= 5.59%). In contrast, no significant differences of amount of repetitive DNA in the downstream intergenic regions of AS genes (Race 5 average= 5.19%; Race 4 average = 5.23%) compared to genes with no evidence of AS (Race 5 average = 4.96%; Race 4 average = 4.94%). The p-values were obtained with the Wilcoxon rank sum test.

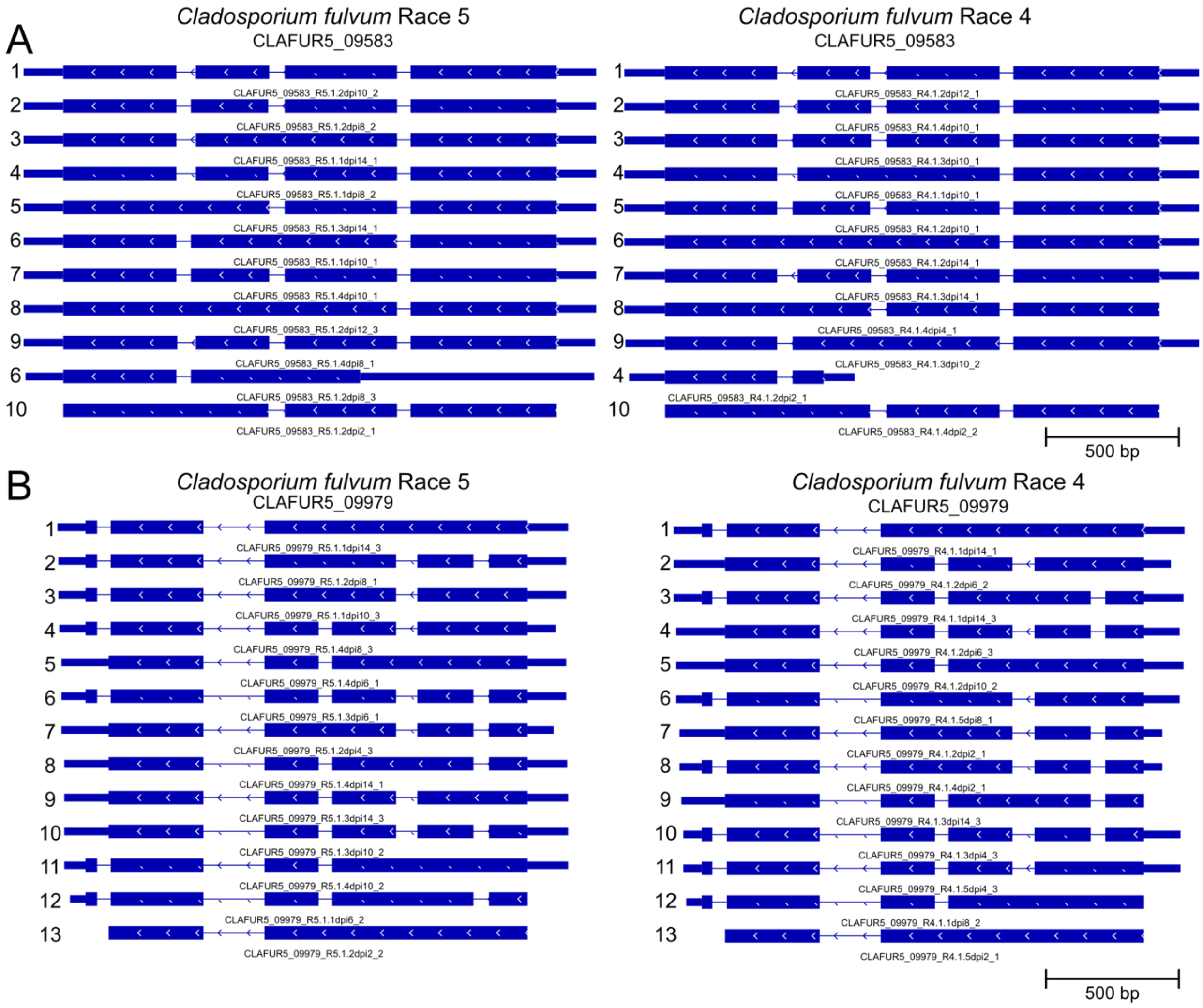

A
Cladosporium fulvum Race 5
CLAFUR5_09583
1 2 3 4 5 6 7 8 9 6 10
CLAFUR5_09583_R5.1.2dpi10_2
CLAFUR5_09583_R5.1.2dpi8_2
CLAFUR5_09583_R5.1.1dpi14_1
CLAFUR5_09583_R5.1.1dpi8_2
CLAFUR5_09583_R5.1.3dpi14_1
CLAFUR5_09583_R5.1.1dpi10_1
CLAFUR5_09583_R5.1.4dpi10_1
CLAFUR5_09583_R5.1.2dpi12_3
CLAFUR5_09583_R5.1.4dpi8_1
CLAFUR5_09583_R5.1.2dpi8_3
CLAFUR5_09583_R5.1.2dpi2_1
Cladosporium fulvum Race 4
CLAFUR5_09583
1 2 3 4 5 6 7 8 9 4 10
CLAFUR5_09583_R4.1.2dpi12_1
CLAFUR5_09583_R4.1.4dpi10_1
CLAFUR5_09583_R4.1.3dpi10_1
CLAFUR5_09583_R4.1.1dpi10_1
CLAFUR5_09583_R4.1.2dpi10_1
CLAFUR5_09583_R4.1.2dpi14_1
CLAFUR5_09583_R4.1.3dpi14_1
CLAFUR5_09583_R4.1.4dpi4_1
CLAFUR5_09583_R4.1.3dpi10_2
CLAFUR5_09583_R4.1.2dpi2_1
CLAFUR5_09583_R4.1.4dpi2_2
500 bp
B
Cladosporium fulvum Race 5
CLAFUR5_09979
1 2 3 4 5 6 7 8 9 10 11 12 13
CLAFUR5_09979_R5.1.1dpi14_3
CLAFUR5_09979_R5.1.2dpi8_1
CLAFUR5_09979_R5.1.1dpi10_3
CLAFUR5_09979_R5.1.4dpi8_3
CLAFUR5_09979_R5.1.4dpi6_1
CLAFUR5_09979_R5.1.3dpi6_1
CLAFUR5_09979_R5.1.2dpi4_3
CLAFUR5_09979_R5.1.4dpi14_1
CLAFUR5_09979_R5.1.3dpi14_3
CLAFUR5_09979_R5.1.3dpi10_2
CLAFUR5_09979_R5.1.4dpi10_2
CLAFUR5_09979_R5.1.1dpi6_2
CLAFUR5_09979_R5.1.2dpi2_2
Cladosporium fulvum Race 4
CLAFUR5_09979
1 2 3 4 5 6 7 8 9 10 11 12 13
CLAFUR5_09979_R4.1.1dpi14_1
CLAFUR5_09979_R4.1.2dpi6_2
CLAFUR5_09979_R4.1.1dpi14_3
CLAFUR5_09979_R4.1.2dpi6_3
CLAFUR5_09979_R4.1.2dpi10_2
CLAFUR5_09979_R4.1.5dpi8_1
CLAFUR5_09979_R4.1.2dpi2_1
CLAFUR5_09979_R4.1.4dpi2_1
CLAFUR5_09979_R4.1.3dpi14_3
CLAFUR5_09979_R4.1.3dpi4_3
CLAFUR5_09979_R4.1.5dpi4_3
CLAFUR5_09979_R4.1.1dpi8_2
CLAFUR5_09979_R4.1.5dpi2_1
500 bp

Figure 4.S8: Two genes encoding putative transcription factors in *Cladosporium fulvum* isolates Race 5 and Race 4, and their predicted protein isoforms produced via alternative splicing (AS) events. (A) AS in gene *CLAFUR5_09583* leads to 10 different protein isoforms in both isolates Race 5 and Race 4. (B) AS in gene *CLAFUR5_09979* leads to 13 different protein isoforms in both isolates Race 5 and Race 4. The number on the left-hand side counts the number of distinct protein isoforms produced via AS by each gene in each isolate. Image is exported from IGV. Thick rectangles represent coding sequences, whereas thinner rectangles represent untranslated regions, and lines represent introns.

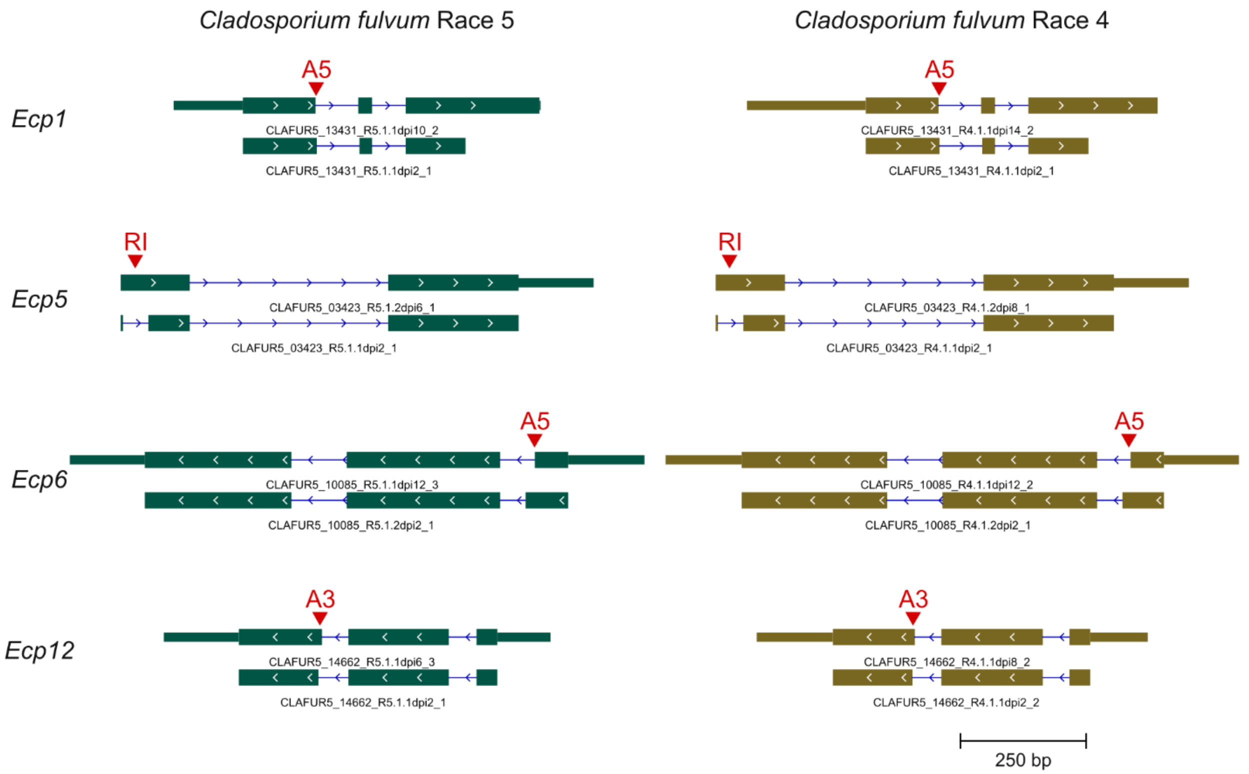

Cladosporium fulvum Race 5
Cladosporium fulvum Race 4
Ecp1
A5
CLAFUR5_13431_R5.1.1dpi10_2
CLAFUR5_13431_R5.1.1dpi2_1
A5
CLAFUR5_13431_R4.1.1dpi14_2
CLAFUR5_13431_R4.1.1dpi2_1
Ecp5
RI
CLAFUR5_03423_R5.1.2dpi6_1
CLAFUR5_03423_R5.1.1dpi2_1
RI
CLAFUR5_03423_R4.1.2dpi8_1
CLAFUR5_03423_R4.1.1dpi2_1
Ecp6
A5
CLAFUR5_10085_R5.1.1dpi12_3
CLAFUR5_10085_R5.1.2dpi2_1
A5
CLAFUR5_10085_R4.1.1dpi12_2
CLAFUR5_10085_R4.1.2dpi2_1
Ecp12
A3
CLAFUR5_14662_R5.1.1dpi6_3
CLAFUR5_14662_R5.1.1dpi2_1
A3
CLAFUR5_14662_R4.1.1dpi8_2
CLAFUR5_14662_R4.1.1dpi2_2
250 bp

Figure 4.S9: Alternative splicing (AS) events in genes encoding effectors in *Cladosporium fulvum* isolates Race 5 and Race 4. The figure shows AS events identified in the *Ecp1*, *Ecp5*, *Ecp6*, and *Ecp12* effector-encoding genes. In all cases, AS altered the coding sequence of the genes, and AS events were conserved between the two isolates. In *Ecp1*, an alternative 5' splice site (A5) event results in frameshift. In *Ecp5*, an intro retention (IR) event results in the modification of the 17 amino acids after the splice site. In *Ecp6*, an alternative 5' splice site (A5) event results in the modification of 6 amino acids after the splice site. In *Ecp12*, an alternative 3' splice site (A3) event results in the modification of 2 amino acids after the splice site.

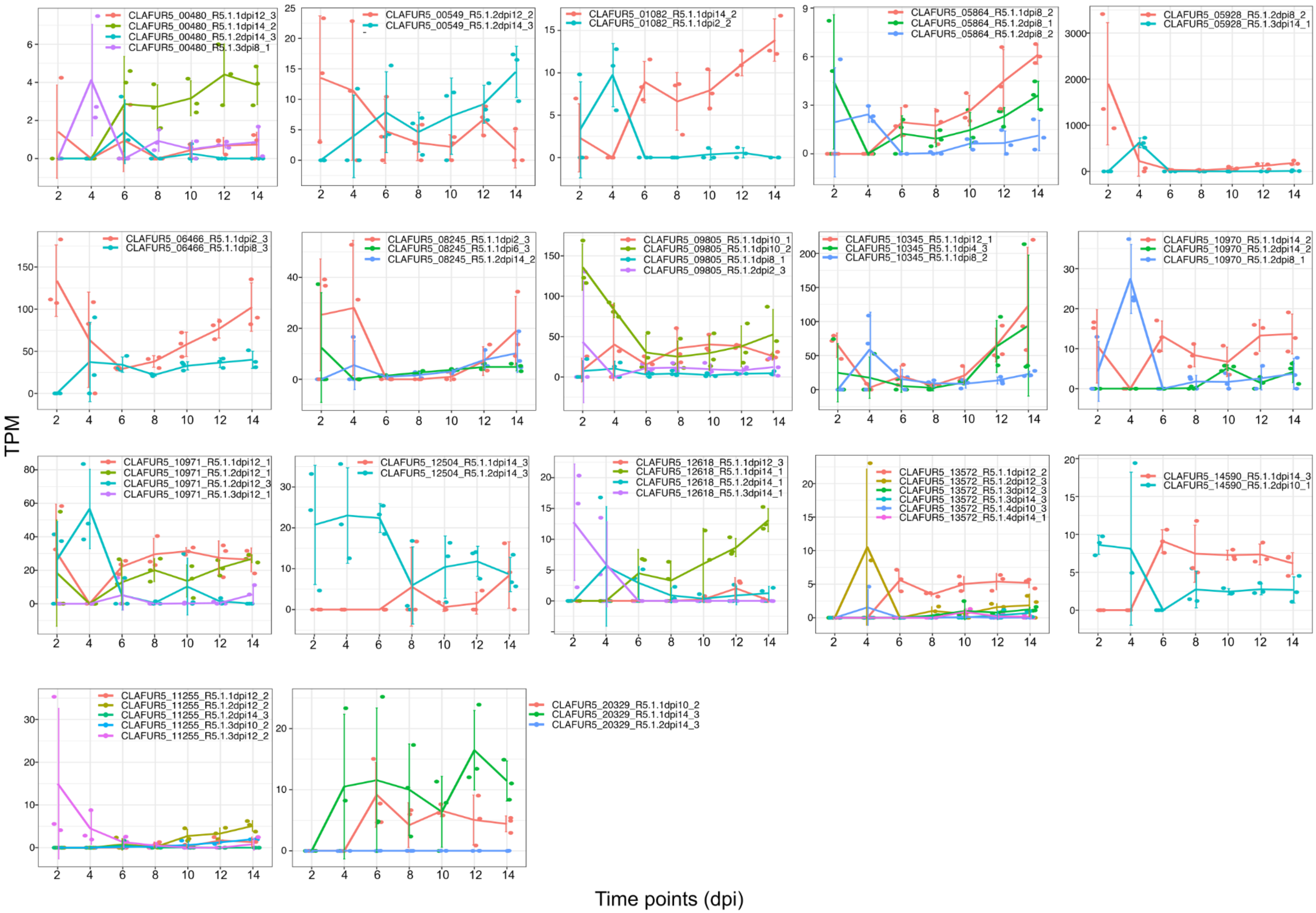

CLAFUR5_00480_R5.1.1dpi12_3
CLAFUR5_00480_R5.1.1dpi14_2
CLAFUR5_00480_R5.1.2dpi14_3
CLAFUR5_00480_R5.1.3dpi8_1
CLAFUR5_00549_R5.1.2dpi12_2
CLAFUR5_00549_R5.1.2dpi14_3
CLAFUR5_01082_R5.1.1dpi14_2
CLAFUR5_01082_R5.1.1dpi2_2
CLAFUR5_05864_R5.1.1dpi8_2
CLAFUR5_05864_R5.1.2dpi8_1
CLAFUR5_05864_R5.1.2dpi8_2
CLAFUR5_05928_R5.1.2dpi8_2
CLAFUR5_05928_R5.1.3dpi14_1
CLAFUR5_06466_R5.1.1dpi2_3
CLAFUR5_06466_R5.1.1dpi8_3
CLAFUR5_08245_R5.1.1dpi2_3
CLAFUR5_08245_R5.1.1dpi6_3
CLAFUR5_08245_R5.1.2dpi14_2
CLAFUR5_09805_R5.1.1dpi10_1
CLAFUR5_09805_R5.1.1dpi10_2
CLAFUR5_09805_R5.1.1dpi8_1
CLAFUR5_09805_R5.1.2dpi2_3
CLAFUR5_10345_R5.1.1dpi12_1
CLAFUR5_10345_R5.1.1dpi4_3
CLAFUR5_10345_R5.1.1dpi8_2
CLAFUR5_10970_R5.1.1dpi14_2
CLAFUR5_10970_R5.1.2dpi14_2
CLAFUR5_10970_R5.1.2dpi8_1
TPM
CLAFUR5_10971_R5.1.1dpi12_1
CLAFUR5_10971_R5.1.2dpi12_1
CLAFUR5_10971_R5.1.2dpi12_3
CLAFUR5_10971_R5.1.3dpi12_1
CLAFUR5_12504_R5.1.1dpi14_3
CLAFUR5_12504_R5.1.2dpi14_3
CLAFUR5_12618_R5.1.1dpi12_3
CLAFUR5_12618_R5.1.1dpi14_1
CLAFUR5_12618_R5.1.2dpi14_1
CLAFUR5_12618_R5.1.3dpi14_1
CLAFUR5_13572_R5.1.1dpi12_2
CLAFUR5_13572_R5.1.2dpi12_3
CLAFUR5_13572_R5.1.3dpi12_3
CLAFUR5_13572_R5.1.3dpi14_3
CLAFUR5_13572_R5.1.4dpi10_3
CLAFUR5_13572_R5.1.4dpi14_1
CLAFUR5_14590_R5.1.1dpi14_3
CLAFUR5_14590_R5.1.2dpi10_1
CLAFUR5_11255_R5.1.1dpi12_2
CLAFUR5_11255_R5.1.2dpi12_2
CLAFUR5_11255_R5.1.2dpi14_3
CLAFUR5_11255_R5.1.3dpi10_2
CLAFUR5_11255_R5.1.3dpi12_2
CLAFUR5_20329_R5.1.1dpi10_2
CLAFUR5_20329_R5.1.1dpi14_3
CLAFUR5_20329_R5.1.2dpi14_3
2 4 6 8 10 12 14
Time points (dpi)

Figure 4.S10: Genes from *Cladosporium fulvum* isolate Race 5 with significant evidence of differential isoform usage at the transcript level during disease progression, which are common to both *C. fulvum* isolates Race 5 and Race 4. The line graphs show 17 AS genes from isolate Race 5 that produce transcripts whose relative abundance significantly changes during the course of the infection. In the line graphs, the points represent the expression values in TPM (transcripts per million) of the individual transcripts across different timepoints of the infection. Standard deviation in the TPM values from three infections (i.e. biological replicates) is shown as vertical lines. The trends of transcript expression across time are shown as thick lines connecting the average TPM values for each individual transcript.

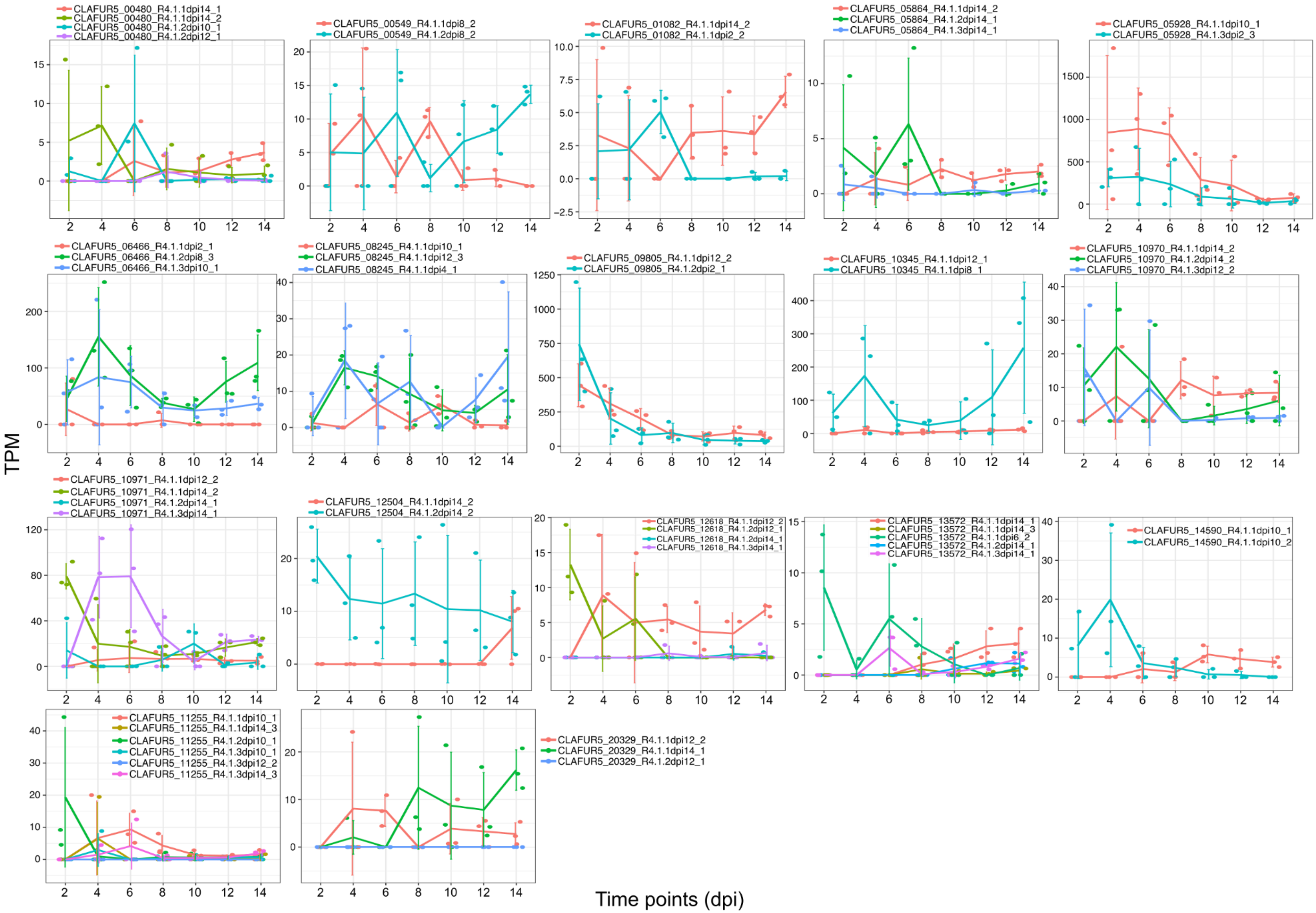

TPM
Time points (dpi)
CLAFUR5_00480_R4.1.1dpi14_1
CLAFUR5_00480_R4.1.1dpi14_2
CLAFUR5_00480_R4.1.2dpi10_1
CLAFUR5_00480_R4.1.2dpi12_1
CLAFUR5_00549_R4.1.1dpi8_2
CLAFUR5_00549_R4.1.2dpi8_2
CLAFUR5_01082_R4.1.1dpi14_2
CLAFUR5_01082_R4.1.1dpi2_2
CLAFUR5_05864_R4.1.1dpi14_2
CLAFUR5_05864_R4.1.2dpi14_1
CLAFUR5_05864_R4.1.3dpi14_1
CLAFUR5_05928_R4.1.1dpi10_1
CLAFUR5_05928_R4.1.3dpi2_3
CLAFUR5_06466_R4.1.1dpi2_1
CLAFUR5_06466_R4.1.2dpi8_3
CLAFUR5_06466_R4.1.3dpi10_1
CLAFUR5_08245_R4.1.1dpi10_1
CLAFUR5_08245_R4.1.1dpi12_3
CLAFUR5_08245_R4.1.1dpi4_1
CLAFUR5_09805_R4.1.1dpi12_2
CLAFUR5_09805_R4.1.2dpi2_1
CLAFUR5_10345_R4.1.1dpi12_1
CLAFUR5_10345_R4.1.1dpi8_1
CLAFUR5_10970_R4.1.1dpi14_2
CLAFUR5_10970_R4.1.2dpi14_2
CLAFUR5_10970_R4.1.3dpi12_2
CLAFUR5_10971_R4.1.1dpi12_2
CLAFUR5_10971_R4.1.1dpi14_2
CLAFUR5_10971_R4.1.2dpi14_1
CLAFUR5_10971_R4.1.3dpi14_1
CLAFUR5_12504_R4.1.1dpi14_2
CLAFUR5_12504_R4.1.2dpi14_2
CLAFUR5_12618_R4.1.1dpi12_2
CLAFUR5_12618_R4.1.2dpi12_1
CLAFUR5_12618_R4.1.2dpi14_1
CLAFUR5_12618_R4.1.3dpi14_1
CLAFUR5_13572_R4.1.1dpi14_1
CLAFUR5_13572_R4.1.1dpi14_3
CLAFUR5_13572_R4.1.1dpi6_2
CLAFUR5_13572_R4.1.2dpi14_1
CLAFUR5_13572_R4.1.3dpi14_1
CLAFUR5_14590_R4.1.1dpi10_1
CLAFUR5_14590_R4.1.1dpi10_2
CLAFUR5_11255_R4.1.1dpi10_1
CLAFUR5_11255_R4.1.1dpi14_3
CLAFUR5_11255_R4.1.2dpi10_1
CLAFUR5_11255_R4.1.3dpi10_1
CLAFUR5_11255_R4.1.3dpi12_2
CLAFUR5_11255_R4.1.3dpi14_3
CLAFUR5_20329_R4.1.1dpi12_2
CLAFUR5_20329_R4.1.1dpi14_1
CLAFUR5_20329_R4.1.2dpi12_1
2 4 6 8 10 12 14

Figure 4.S11: Genes from *Cladosporium fulvum* isolate Race 4 with significant evidence of differential isoform usage at the transcript level during disease progression, which are common to both *C. fulvum* isolates Race 5 and Race 4. The line graphs show 17 AS genes from isolate Race 4 that produce transcripts whose relative abundance significantly changes during the course of the infection. In the line graphs, the points represent the expression values in TPM (transcripts per million) of the individual transcripts across different timepoints of the infection. Standard deviation in the TPM values from three infections (i.e. biological replicates) is shown as vertical lines. The trends of transcript expression across time are shown.

4.7.2 Supplementary tables

Table 4.S1: Number (No.) of raw paired-end reads obtained for *Cladosporium fulvum* isolates Race 5 and Race 4, during interaction with tomato (Solanum lycopersicum) cv. Moneymaker. Tomato infections with the two isolates were done three times, with each infection representing a biological replicate of the experiment (Rep.1-3), and sampling was performed at seven times points (i.e. at 2, 4, 6, 8, 10, 12, and 14 days post inoculations). The total number of reads and number of reads remaining after quality control (QC) are shown. This table is available at https://zenodo.org/records/11211529.

Table 4.S2: Number and percentage of paired-end reads obtained for *Cladosporium fulvum* isolates Race 5 and Race 4 that had a k-mer matching to the genomes of *C. fulvum* or tomato (Solanum lycopersicum) cv. Moneymaker. Numbers and percentages were obtained with the alignment-free method implemented in the script seal.sh from the BBMap package that assigned reads after quality control to either the genome of *C. fulvum* or the genome of tomato. This table is available at https://zenodo.org/records/11211529.

Table 4.S3: The number (No.) of protein-coding genes in the genomes of *Cladosporium fulvum* isolates Race 5 and Race 4 with the specified number of introns.

No. of introns	No. of genes in isolate Race 5	No. of genes in isolate Race 4
0 (intronless)	5,637	5,617
1	5,046	5,015
2	2,527	2,505
3	994	983
4	414	408
5	187	185
6	93	92
7	42	41
8	27	23
9	7	7
10	8	8
11	3	3
12	4	4
14	1	1
15	2	2
20	1	1
Total number of genes	14,993	14,895
Total number of genes with introns	9,356	9,278

Table 4.S4: Number (No.) of transcript isoforms assembled for *Cladosporium fulvum* isolates Race 5 and Race 4, at each of the seven time points during interaction with tomato. Numbers for each of the three independent tomato infections that were done (i.e. biological replicates; Rep.1-3), are shown.

	***Cladosporium fulvum* isolate Race 5**						***Cladosporium fulvum* isolate Race 4**					
Time point	**Rep. 1**		**Rep. 2**		**Rep. 3**		**Rep. 1**		**Rep. 2**		**Rep. 3**	
	No. of transcripts	**% transcripts**	**No. of transcripts**	**% transcripts**	**No. of transcripts**	**% transcripts**	**No. of transcripts**	**% transcripts**	**No. of transcripts**	**% transcripts**	**No. of transcripts**	**% transcripts**
2 dpi	6415	6.32	6953	6.60	7194	6.80	8613	8.85	10955	10.26	12236	12.84
4 dpi	7506	7.40	8545	8.11	9694	9.16	10174	10.45	11027	10.33	13294	13.95
6 dpi	13100	12.91	13392	12.70	15105	14.28	11694	12.02	12949	12.13	12499	13.11
8 dpi	17882	17.62	17092	16.21	16550	15.64	12919	13.27	16316	15.29	11628	12.20
10 dpi	18075	17.81	20034	19.00	17628	16.66	16493	16.95	17179	16.10	11772	12.35
12 dpi	18889	18.61	19858	18.84	19371	18.31	18353	18.86	18783	17.60	16170	16.97
14 dpi	19624	19.34	19553	18.55	20245	19.14	19080	19.60	19525	18.29	17711	18.58
Total	101491	100	105427	100	105787	100	97326	100	106734	100	95310	100

Table 4.S5: Number of unique transcript isoforms assembled from alternatively spliced (AS) genes in *Cladosporium fulvum* isolates Race 5 or Race 4. Genes that produce more than one transcript isoforms are conciderted as AS. Note that some genes may be AS in one isolate but not in the other. Also, some genes may not be expressed and thus not produce any isoform in one isolate. The differences in the n umber of transcript isoforms generated by each gene in each isolate are shown as well. This table is available at https://zenodo.org/records/11211529.

Table 4.S6: The subset of alternatively spliced (AS) genes that are common between *Cladosporium fulvum* isolates Race 5 and Race 4. The number (No.) of unique transcript isoforms generated by each AS gene in isolates Race 5 and Race 4, and the differences in the number of transcript isoforms produced by each AS gene in each isolate are shown. This table is available at https://zenodo.org/records/11211529.

Table 4.S7: Number (No.) and percentages of alternative splicing (AS) events in *Cladosporium fulvum* isolates Race 5 and Race 4. The table shows the number and percentages of AS events in isolates Race 5 and Race 4 classified in one of the major types of AS, i.e., 5'/3' splice sites (A5/A3), alternative first/last exons (AF/AL), mutually exclusive exons (MX), intron retention (IR), and skipping exon (SE). The numbers of genes containing the AS events are also shown.

Event type	Isolate Race 5		Isolate Race 4		Isolate Race 5		Isolate Race 4	
	No. of events	% events	No. of events	% events	No. of genes*	% genes**	No. of genes*	% genes**
A3	1429	13.4	1261	13.4	1111	22.2	1012	21.5
A5	1256	11.8	1129	12	1014	20.3	898	19.1
AF	176	1.6	163	1.7	114	2.3	103	2.2
AL	64	0.6	74	0.8	36	0.7	35	0.7
MX	2	0	5	0.1	2	0	3	0.1
IR	7593	71.2	6667	70.7	4148	82.9	3912	83.1
SE	151	1.4	128	1.4	113	2.3	97	2.1
Total	10671	-	9427	-	5004	-	4710	-

* Number of genes containing the specific type of AS. Each gene may contain multiple types of AS. The total indicates the total number of unique genes containing any of the seven types of AS.

** Percentages based on the total number of unique genes containing any of the seven types of AS.

Table 4.S8: Enrichment analysis of alternatively spliced (AS) genes in *Cladosporium fulvum* isolates Race 5 and Race 4. The table shows enrichment results for (A) gene ontology (GO) terms, (B) conserved PFAM domains, and (C) functional gene categories. This table is available at https://zenodo.org/records/11211529.

Table 4.S9: Number and percentages of genes from *Cladosporium fulvum* isolates Race 5 and Race 4 harboring different types of alternative splicing (AS) events. The table shows the number and percentages of genes from different chromosomes harboring major types of AS events, i.e., 5'/3' splice sites (A5/A3), alternative first/last exons (AF/AL), mutually exclusive exons (MX), intron retention (IR), and skipping exon (SE). The statistics of the chromosomes of were obtained from Zaccaron et al. (2022) and Zaccaron and Stergiopoulos (2024). This table is available at https://zenodo.org/records/11211529.

Table 4.S10: Alternatively spliced (AS) genes in *Cladosporium fulvum* isolates Race 5 and Race 4, producing multiple distinct protein isoforms. The table shows the number (No.) of unique protein isoforms putatively produced by each AS gene, the functional annotations of the AS genes, as well as whether the AS genes encode carbohydrate-active enzymes (CAZymes), secreted proteins, or candidate effectors. This table is available at https://zenodo.org/records/11211529.

Table 4.S11: Enrichment analysis of alternatively spliced (AS) genes in *Cladosporium fulvum* isolates Race 5 and Race 4 producing multiple distinct protein isoforms. Enrichment results based on (A) gene ontology (GO) terms, (B) conserved PFAM domains, and (C) functional gene categories, are shown. This table is available at https://zenodo.org/records/11211529.

Table 4.S12: Alternatively spliced (AS) genes in *Cladosporium fulvum* isolates Race 5 and Race 4, producing multiple distinct protein isoforms with presence/absence variation in PFAM domains and signal peptides (SP). The table shows the gene IDs of the AS genes, the conserved PFAM domains showing presence/absence variation among the produced protein isoforms from a single AS gene, and whether these isoforms vary in SPs. Blank cells indicate no presence/absence variation in PFAM domains or SPs. The functional annotations of the AS genes are also provided, as well as whether the genes encode carbohydrate-active enzymes (CAZymes) or candidate effectors. This table is available at https://zenodo.org/records/11211529.

Table 4.S13. Alternatively spliced (AS) genes in *Cladosporium fulvum* isolates Race 5 and Race 4, producing transcript isoforms with (Yes) or without (No) evidence of switching during the seven time points of the infection process that were analyzed. The gene IDs are shown in the first column. When AS is in gene coding regions, thereby resulting in the production of multiple distinct protein isoforms, then the number of these isoforms are indicated as well. Functional annotations of the AS genes are provided too, as well as whether the genes encode candidate effectors, cytochrome P450 enzymes, major facilitator superfamily (MFS) transporters, carbohydrate-active enzymes (CAZymes), or secreted proteins. This table is available at https://zenodo.org/records/11211529.

Chapter 5

A chromosome-scale genome assembly of the grape powdery mildew pathogen *Erysiphe necator* reveals its genomic architecture and previously unknown features of its biology

Alex Z. Zaccaron
Tara Neill
Jacob Corcoran
Walter F. Mahaffee
Ioannis Stergiopoulos

Leading author statement

I contributed to the conceptualization of this chapter, performed all analyses, generated all figures, contributed to the investigation and interpretation, wrote the original draft, and contributed to the review and editing of the final draft.

This chapter was published in: *mBio*, June 2023, volume 14, number 4 (https://doi.org/10.1128/mbio.00645-23).

Abstract

Erysiphe necator is an obligate fungal pathogen that causes grape powdery mildew, globally the most important disease on grapevines. Previous attempts to obtain a quality genome assembly for this pathogen were hindered by its high repetitive DNA content. Here, we combined chromatin conformation capture (Hi-C) with long-read PacBio sequencing to obtain a chromosome-scale assembly and a high-quality annotation for *E. necator* isolate EnFRAME01. The resulting 81.1 Mb genome assembly is 98% complete and consists of 34 scaffolds, 11 of which represent complete chromosomes. All chromosomes contain large centromeric-like regions and lack synteny to the 11 chromosomes of the cereal PM pathogen *Blumeria graminis*. Further analysis of their composition showed that repeats and transposable elements (TEs) occupy 62.7% of their content. TEs were almost evenly interspersed outside centromeric and telomeric regions and massively overlapped with regions of annotated genes, suggesting that they could have a significant functional impact. Abundant gene duplicates were observed as well, particularly in genes encoding candidate secreted effector proteins. Moreover, younger in age gene duplicates exhibited more relaxed selection pressure and were more likely to be located physically close in the genome than older duplicates. A total of 122 genes with copy number variations among six isolates of *E. necator* were also identified and were enriched in genes that were duplicated in EnFRAME01, indicating they may reflect an adaptive variation. Taken together, our study illuminates higher-order genomic architectural features of *E. necator* and provides a valuable resource for studying genomic structural variations in this pathogen.

5.1 Introduction

Erysiphe necator (Ascomycetes; Leotiomycetes, Erysiphaceae) is an obligate biotrophic fungal pathogen that causes grapevine powdery mildew (GPM), one of the most common and economically important fungal diseases in vineyards around the globe (Gadoury et al. 2012). The pathogen can significantly reduce grape yield and quality, and most cultivated varieties of grapevine (*Vitis vinifera*) are susceptible to it (Gadoury et al. 2012; Gaforio et al. 2011; Qiu et al. 2015). As a consequence, GPM is commonly managed by fungicides, which dramatically increase the overall production costs and the risk of resistance development (Fuller et al. 2014; Kunova et al. 2021).

The obligate nature of powdery mildews (PMs) prohibits the functional analysis of their genes by means of standard genetic manipulations. Instead, comparative and population genomics have been used as alternatives for studying the molecular mechanisms underlying obligate biotrophy, pathogenicity, and other aspects of the biology of these pathogens (Frantzeskakis et al. 2018; Jones et al. 2014a; Liang et al. 2018; Müller et al. 2019; Spanu et al. 2010; Wu et al. 2018; Zaccaron and Stergiopoulos 2021). To date, the genomes of at least 16 species or formae speciales of PMs have been obtained, including three monocotyledonous-infecting and 13 dicotyledonous-infecting species. However, the highly repetitive nature of these genomes has posed major challenges to the construction of high-quality genome assemblies based on short sequencing reads alone and has further hindered in-depth comparative genome analyses. As a result, chromosome-scale genome assemblies have, so far, only been obtained from just two monocot-infecting species of PM, namely the wheat pathogen *Blumeria graminis* f.sp. *tritici* and the triticale pathogen *Blumeria graminis* f.sp. *triticale* (Müller et al. 2021, 2019), but none from dicot-infecting PMs.

Despite challenges in obtaining high-quality genome assemblies and annotations for PM fungi, analysis of their genomic content has shown that they possess some of the largest genomes among filamentous ascomycetes, with sizes typically ranging from 120 Mb to 180 Mb (Bindschedler et al. 2016; Frantzeskakis et al. 2018; Müller et al. 2019; Hacquard 2014). The increase is due to the extensive proliferation of transposable elements (TEs) in their genomes, which typically comprise up to 85% of their genomic content (Frantzeskakis et al. 2018; Jones et al. 2014a; Müller et al. 2019; Wu et al. 2018). However, contrary to their enlarged genomes, PMs have a reduced number of circa (ca.) 7,000 genes (Wu et al. 2018), which is considerably smaller compared to the ca. 11,000 genes typically present in non-obligate fungal plant pathogens (Aylward et al. 2017). The reduction is due to marked losses in genes encoding key enzymes in primary and secondary metabolism as well as in hydrolytic enzymes that cause damage to host cells during infection, a hallmark of their obligate biotrophic lifestyle (Spanu et al. 2010; Spanu 2012). PMs also lack a repeat-induced point mutation (RIP) defense mechanism against the deleterious effects caused by TE replication in their genomes (Irelan and Selker 1996; Selker 1990). As a consequence, their genomes experience higher rates of TE and gene duplication and retention, as these are more prone to pseudogenization in species with an active RIP mechanism. An examination of their genome architecture has further shown that PM genomes are generally deprived of large-scale compartmentalization, AT-rich isochores, and accessory chromosomes, which constitute signatures of 'plastic' or 'two-speed' genomes (Müller et al. 2019; Frantzeskakis et al. 2018). Instead, they adhere mostly to the 'one-speed' model of genome evolution, in which gene duplication is an important mechanism of evolution and adaptation (Frantzeskakis et al. 2019).

Previous efforts to sequence the genome of *E. necator* were constrained by its highly repetitive nature and the limitations of the short-read sequencing technologies used at the time. Consequently, the current reference genome of *E. necator* isolate C-strain is estimated to be 36.5-48.6% complete and is assembled into 5,935 scaffolds, which forbids a rigorous analysis of its architecture (Jones et al. 2014a). In this study, we present a chromosome-scale genome assembly and gene annotation for *E. necator* isolate EnFRAME01. The new reference genome of *E. necator* presented herein is the first chromosome-scale assembly obtained for a dicot-infecting PM species and elucidates major aspects of their biology.

5.2 Results

5.2.1 The genome of *E. necator* consists of 11 chromosomes with large centromeric-like regions

The genome of *E. necator* isolate EnFRAME01 was assembled using a combination of PacBio reads and Hi-C data into 34 scaffolds, totaling 81.1 Mb in size. This is a considerable improvement over the previous reference genome of *E. necator* isolate C-strain that was fragmented into 5,935 scaffolds (Table 5.S1 and Supplementary Results). Of the 34 assembled scaffolds, 11 embodied distinct chromosomes (Chr1-to-Chr11) (Fig 5.1, Fig 5.S1 and Table 5.1), 22 were unplaced scaffolds, and one scaffold represented the complete mitochondrial genome (Fig 5.S2). The size of the 11 chromosomes ranged from 11.3 Mb (Chr1) to 3.3 Mb (Chr11) and all were putatively assembled telomere-to-telomere, containing only five collapsed regions (Fig 5.S3). All chromosomes had 22 to 31 copies of the canonical telomeric repeat 5'-TTAGGG-3' at their ends and predicted centromeric regions with high inter-chromosomal Hi-C contact frequency (Fig 5.S1A), as previously observed in other fungi (Seidl et al. 2020; Winter et al. 2018). However, in contrast to other ascomycetes (Smith et al. 2011; Schotanus et al. 2015; Seidl et al. 2020; Yadav et al. 2019; King et al. 2015), the predicted centromeres of *E. necator* were large segments that accounted for 15.8% of the genome (Table 5.1 and Table 5.S2). Centromeric regions of similar sizes have been reported before for the wheat PM *B. graminis* f.sp. *tritici* (Müller et al. 2019) but whole-genome alignment showed that the predicted centromeric regions of *E. necator* are poorly conserved in *B. graminis* f.sp. *tritici*. Moreover, although both PM species have 11 chromosomes, they exhibited an overall low synteny as no one-to-one chromosome match was observed between them (Fig 5.S4). These results indicate poor conservation of centromeric regions and extensive inter-chromosomal rearrangements between *E. necator* and *B. graminis* f.sp. *tritici*.

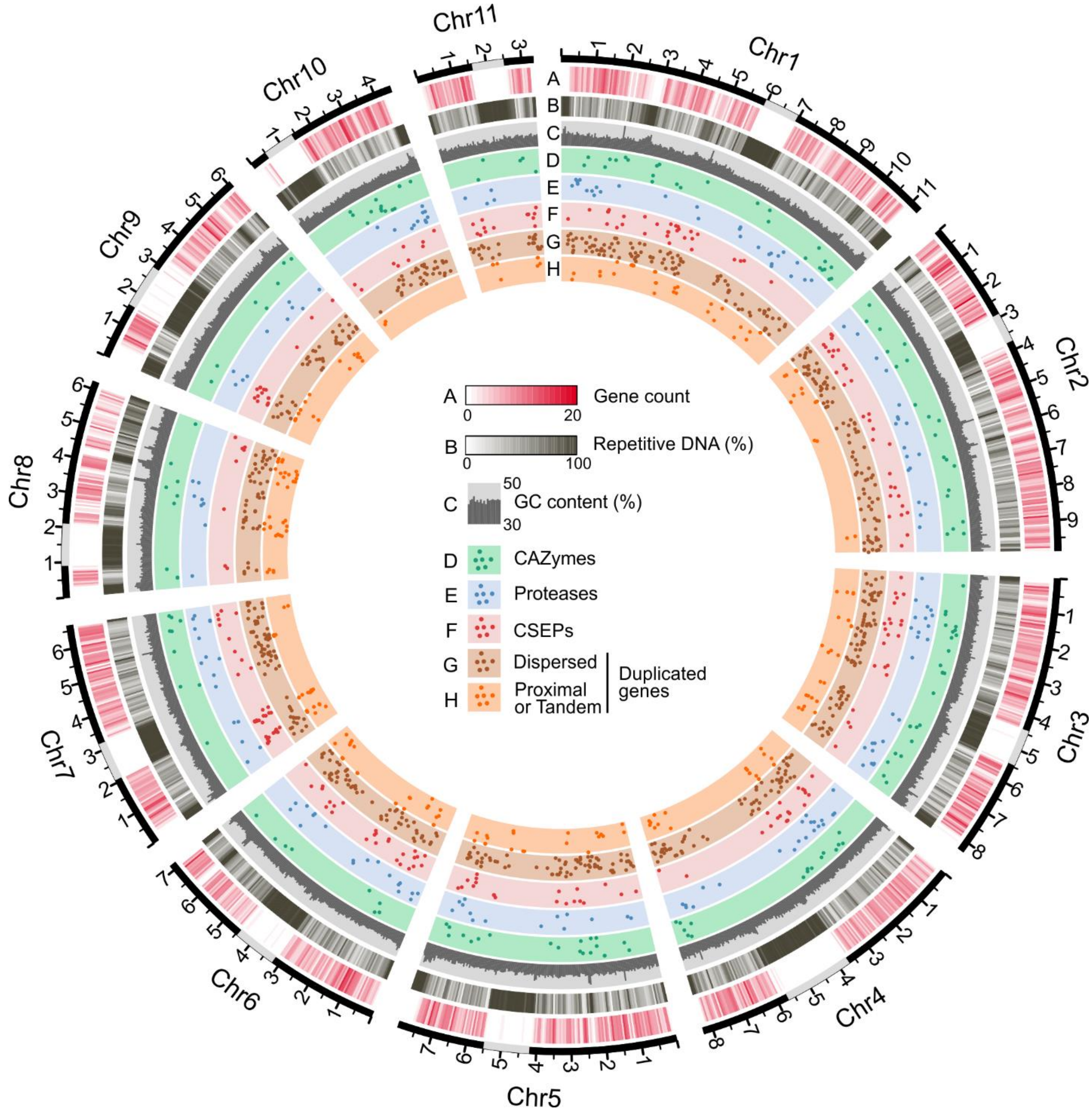


Figure 5.7: Schematic representation of the 11 chromosomes of *Erysiphe necator* isolate EnFRAME01. The Circos plot shows the assembled chromosomes as solid black lines with major tick marks representing Mb. Predicted location of centromeric regions is indicated with grey rectangles. The outermost-to-innermost tracks represent (A) density of protein-coding genes, (B) repetitive DNA content, (C) GC content from 30% to 50%, (D) location of genes encoding carbohydrate-active enzymes (CAZymes), (E) location of genes encoding proteases, (F) location of genes encoding candidate secreted effector proteins (CSEPs), (G) location of dispersed gene duplicates (i.e. gene copies located in different chromosomes or separated by more than 10 genes), and (H) location of proximal or tandem gene duplicates (i.e. gene copies located less than 10 genes apart or next to each other). Gene locations are represented by bullet points on the perpendicular axis. Gene count, repetitive DNA, and GC content were determined using a sliding window of 50 kb. The figure shows that the chromosomes of EnFRAME01 contain long centromeric regions, which are abundant in repeats and nearly devoid of protein-coding genes.

Table 5.10: Size and content of the 11 chromosomes of *Erysiphe necator* isolate EnFRAME01.

Chromosome	Size (Mb)	GC (%)	Centromere size (Mb)	Predicted genes	Genes per Mb	Median intergenic size (bp)	Repeats (%)
Chr1	11.30	39.8	1.00	1000	88	4,333	61.7
Chr2	9.87	39.6	0.85	1060	107	3,391	55.6
Chr3	8.28	39.5	1.05	840	101	3,502	58.4
Chr4	8.27	39.6	2.00	684	83	4,000	64.4
Chr5	7.98	39.8	1.30	758	95	3,286	61.8
Chr6	7.15	39.8	1.25	608	85	4,402	64.9
Chr7	6.66	39.4	1.12	641	96	3,639	60.5
Chr8	6.21	40.2	1.20	393	63	5,218	70.0
Chr9	6.07	39.9	1.25	474	78	3,249	69.4
Chr10	4.49	39.5	0.90	387	86	3,536	63.6
Chr11	3.34	39.4	0.90	261	78	3,666	68.3

5.2.2 A reduced gene complement underlies the obligate biotrophic lifestyle of *E. necator*

A total of 7,146 protein-coding genes were predicted in the genome of EnFRAME01, with a BUSCO completeness of 98.2%. This gene number is comparable to the 6,046-to-8,470 genes reported in other PM fungi (Frantzeskakis et al. 2018; Müller et al. 2021) and a notable improvement over the gene annotation of *E. necator* C-strain, for which 6,484 genes were predicted with an estimated completeness of 90.1% (Table 5.S3). Functional gene annotations showed that the genome of EnFRAME01 contained 174 proteases (Table 5.S4), eight key enzymes for secondary metabolism (Table 5.S5), 11 cytochrome P450s (Fig 5.S5 and Table 5.S6), 1,238 putative transporters (Fig 5.S6 and Table 5.S7), 160 carbohydrate-active enzymes (CAZymes) (Table 5.S8), and 527 secreted proteins (SPs) (Table 5.S9) of which 234 were candidate secreted effector proteins (CSEPs) (Table 5.S10) (Supplementary Results). The number of genes in these functional categories is low compared to other plant pathogenic Ascomycete fungi (Liang et al. 2018; Cissé et al. 2014) but similar to PMs (Liang et al. 2018; Hacquard 2014). Consistent also with an obligate biotrophic lifestyle, 181 core genes were identified that are typically present in *Saccharomyces cerevisiae* and non-obligate biotrophic fungi but were missing in EnFRAME01 (Table 5.S11). Included in these genes were 95 of the 99 so-called “missing ascomycete pathogen core genes” (MACGs), generally reported as absent in PMs (Spanu et al. 2010) (Table 5.S12). Based on KEGG orthology (KO) identifiers (Kanehisa et al.

2022), the 181 genes are predicted to partake in 47 conserved pathways (Table 5.S13), of which 23 were significantly enriched in genes missing in EnFRAME01 (Supplementary Results). The two pathways most affected by gene losses were thiamine and sulfur metabolism, in accordance to other obligate biotrophic fungi (Spanu et al. 2010; Spanu 2012; Cissé et al. 2014). Absence of a sterol O-acyltransferase (EC:2.3.1.26) gene, and of the *ERG5* (C-22 sterol desaturase; EC:1.14.19.41) and *ERG4* (EC:1.3.1.71) genes whose products catalyze the last two steps of ergosterol biosynthesis in yeast (Hu et al. 2017), was also observed. Collectively, these results indicate that the obligate lifestyle of *E. necator* is driven by losses in genes involved in several biochemical pathways, in accordance to what has been observed in other PMs and obligate biotrophs (Spanu et al. 2010; Spanu 2012).

5.2.3 *Erysiphe necator* harbors a reduced arsenal of CSEPs

The small number of 234 CSEP-encoding genes identified in the genome of EnFRAME01 is in line with reports from dicot-infecting PMs but in contrast to monocot-infecting PMs such as different *B. graminis* formae speciales (Frantzeskakis et al. 2018; Müller et al. 2019). Of the 234 CSEPs, 49 (20.9%) were species-specific and 185 (79.1%) had homologs in PMs (*n*=183) and/or non-PM fungi (*n*=86) (Fig 5.S7). Moreover, 86 (36.7%) contained the Y/F/WxC sequence motif that is typically found in CSEPs of *B. graminis* and other PMs (Fig 5.S8 and Table 5.S10). PM fungi are also known to harbor many ribonuclease-like effectors that belong to a large family of catalytically inactive RNAses, known as RALPHs (RNase-Like Proteins associated with haustoria) (Spanu 2017; Pennington et al. 2019; Pedersen et al. 2012). A genome-wide search in EnFRAME01 identified 38 genes encoding RALPH-like proteins, 24 of which could also be classified as CSEPs (Table 5.S14). A phylogenetic analysis grouped the 38 RALPH-like proteins into two major clades, whose members differed in average protein size and the location in the genome of their encoding genes (Fig 5.S9). An Egh16-like virulence factor domain (PF11327) was also commonly found in the *E. necator* CSEPs. CSEPs with an Egh16 domain are members of a multigene family in fungi (Xue et al.

2002; Grell et al. 2003) and often play a role during the early stages of host infection. A total of 11 genes encoding Egh16-like proteins that could be further clustered into two clades (Fig 5.S10) were identified in EnFRAME01, but only four of these were classified as CSEPs (Table 5.S15). Finally, an analysis of the localization of the 234 CSEP-encoding genes of *E. necator* on the 11 chromosomes of the fungus showed there was no enrichment of genes encoding CSEPs in sub-telomeric regions, as has been observed in other fungi (Zaccaron et al. 2022; Gan et al. 2020) (Fig 5.1).

5.2.4 TE bursts have drastically shaped the genome of *E. necator*

The genome of EnFRAME01 is highly repetitive, with repeats accounting for 62.7% (50.8 Mb) of its DNA content. Class I retrotransposons, such as long terminal repeat (LTR)-retrotransposons (26.5%, 21.5 Mb) and non-LTR retrotransposons (16.8%, 13.6 Mb), were more abundant than Class II DNA transposons (6.3%, 5.1 Mb) and unclassified interspersed repeats (13.1%, 10.6 Mb) (Table 5.S16). This is consistent with most fungi (Castanera et al. 2016; Amselem et al. 2015) but in contrast to cereal PMs, whose genomes are mainly dominated by non-LTR retrotransposons (Frantzeskakis et al. 2018; Müller et al. 2019). TEs were fairly evenly dispersed outside centromeric and subtelomeric regions, which generally contained smaller amounts of non-LTR elements and exhibited an overall lower TE divergence. A similar pattern was also observed in genomic islands rich in rolling-circle (RC) elements (Fig 5.2A). Collectively, these observations indicate that younger TEs accumulated preferentially in centromeric and subtelomeric regions and that RC elements are younger than other TEs (Fig 5.2A). Interestingly, an examination of the nucleotide divergence among TE copies revealed a bimodal distribution with two peaks of contrasting TE composition. This suggests the presence of two TE burst events in the evolutionary history of *E. necator* that involved different TE classes (Fig 5.2B and Supplementary Results). A similar pattern was also observed in the genomes of different *B. graminis* formae speciales, although the bimodal peaks were less pronounced and lacked RC elements (Fig 5.2B and Table 5.S16). In addition, highly divergent TEs were in all genomes enriched in non-

LTR rather than LTR elements, whereas the opposite was observed for TEs with low divergence. By using the *E. necator* repeat library to mask the genomes of the cereal PMs, and *vice-versa*, nearly all low-divergence TEs were left unmasked (Fig 5.S11). These observations suggest that *E. necator* and *B. graminis* underwent a similar burst of non-LTR TEs, possibly prior to their divergence, followed by clade-specific proliferation of LTR-retrotransposons and, in the case of *E. necator*, of RC elements as well.

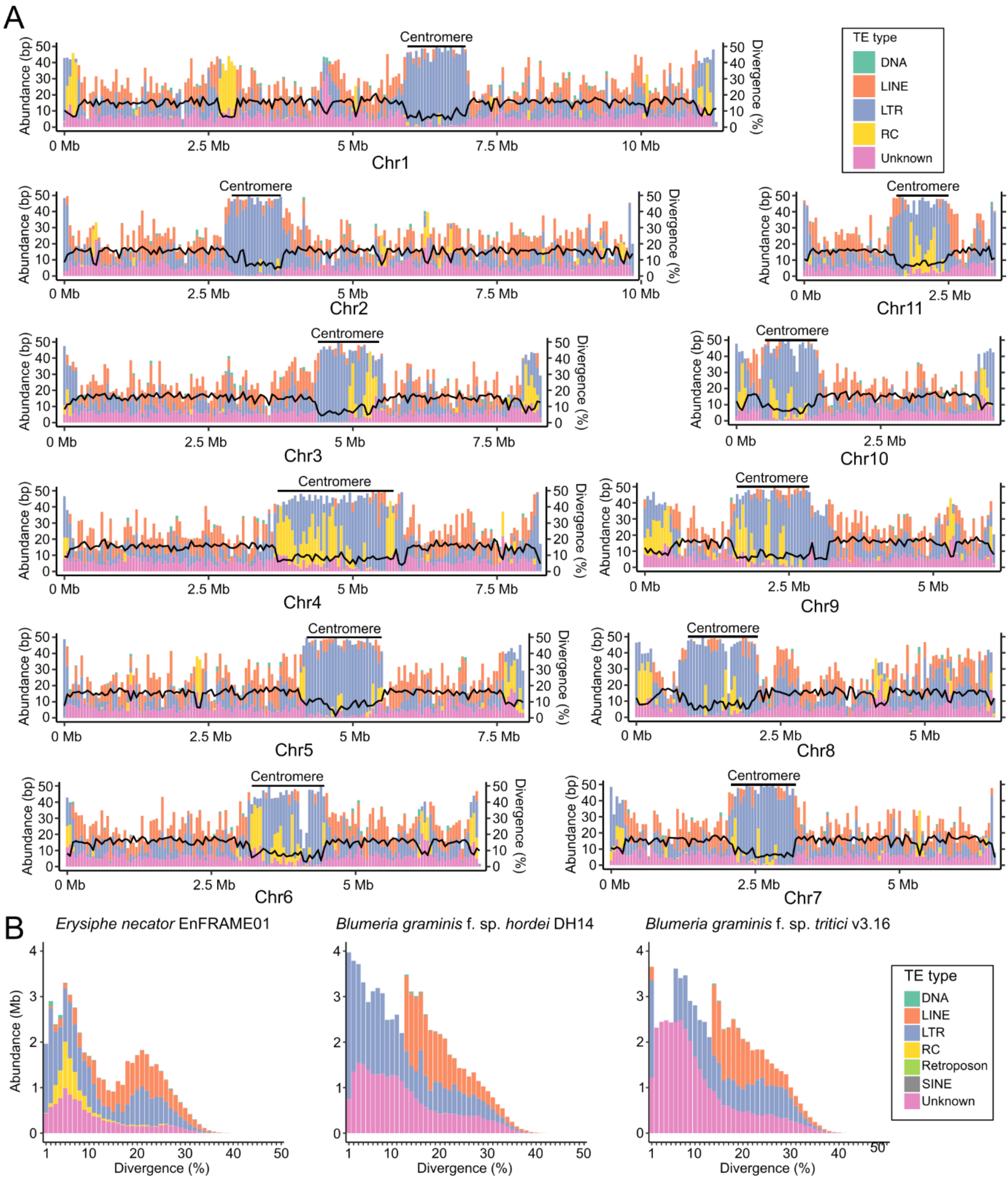


Figure 5.8: The transposable element (TE) composition of *Erysiphe necator* differs from that of the cereal powdery mildew (PM) pathogens *Blumeria graminis* f.sp. *hordei* and *B. graminis* f.sp. *tritici*. (A) Distribution of TEs in the 11 chromosomes of *E. necator* isolate EnFRAME01. The figure shows the abundance of the different TE classes, represented as stacked bar plots along the chromosomes. Overall divergence of TE families is indicated by solid black lines along the chromosomes. Predicted centromeric regions are indicated as well. The figure shows high abundance of repeats near chromosome ends and at centromeres. Predicted centromeric regions are enriched mainly in long terminal repeat (LTR)

retrotransposons with overall low sequence divergence compared to the rest of the genome. Rolling circle (RC) elements are also abundantly found in centromeres and have an overall low sequence divergence. TE abundance and divergence were calculated using a sliding window of 50 kb. (B) Repetitive DNA landscape represented as bar plots showing the number of bases covered by predicted TEs from different (sub)classes. The predicted divergence of the TEs is shown on the *x*-axis. The figure shows a bimodal repeat divergence landscape with peaks for *E. necator* at approximately 5% and 21% divergence. The two peaks differ in their composition, with the peak at 5% divergence being dominated by LTRs, RCs, and unknown elements, and the peak at 21% divergence being dominated by LTR and long interspersed nuclear elements (LINE). The landscape of TE divergence of the cereal PM pathogens also follows a bimodal distribution, but it is less pronounced as compared to *E. necator* and the peaks are void of RCs.

5.2.5 The genome of *E. necator* exhibits small-scale compartmentalization

An examination of the distribution of repeats and of protein coding genes on the 11 chromosomes of EnFRAME01 revealed large differences in gene density among the chromosomes, and an inverse correlation between density of protein coding genes and repetitive DNA content (Fig 5.1, Fig 5.S12 and Supplementary Results). An assessment of whether certain gene categories were associated with specific TE superfamilies showed no major differences in TE content within the flanking regions of genes encoding CAZymes, proteases, CSEPs, and non-CSEP secreted proteins (Fig 5.S13). However, the intergenic regions of CSEP genes were significantly longer, richer in repetitive DNA, and had a different TE composition as compared to other functional gene categories (Fig 5.3A, Fig 5.3B and Table 5.S17). Collectively, these observations indicate the absence of large-scale compartmentalization in gene-dense and gene-sparse regions in the genome of *E. necator* (Fig 5.3C and Fig 5.3D), consistent with the 'one-speed' genome hypothesis suggested for PM species (Frantzeskakis et al. 2018, 2019a). Instead, small-scale compartmentalization of CSEP-encoding genes was seen, which were preferentially located in somewhat gene-sparse and repeat-rich regions, as commonly reported in other fungal pathogens (Dong et al. 2015; Raffaele and Kamoun 2012).

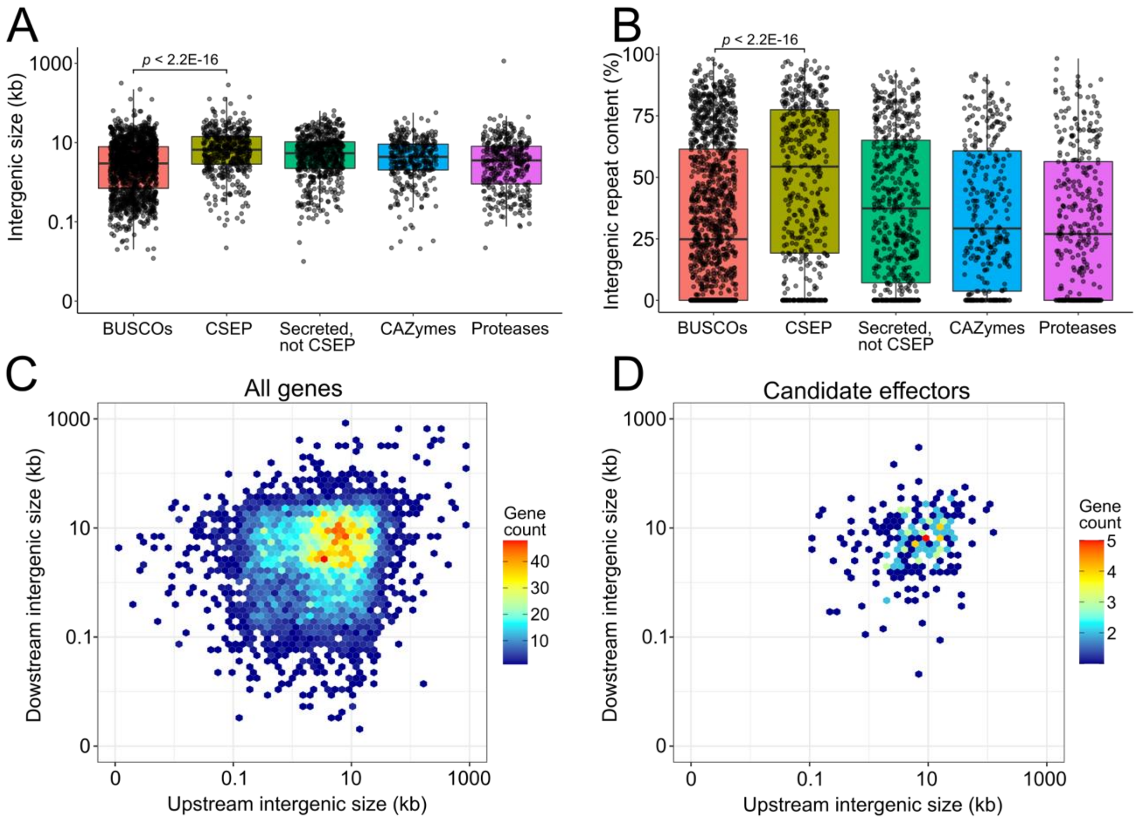


Figure 5.9: The genome of *Erysiphe necator* isolate EnFRAME01 exhibits small-scale compartmentalization of genes encoding candidate secreted effector proteins (CSEPs) in repeat-rich genomic regions. (A and B) Boxplots showing the size distribution and repetitive DNA content of upstream (panel A) and downstream (panel B) intergenic regions flaking BUSCO genes, genes encoding CSEPs, genes encoding secreted proteins not classified as CSEPs, genes encoding carbohydrate-active enzymes (CAZymes), and genes encoding proteases. The figure shows that intergenic regions of CSEP-encoding genes typically have higher repetitive DNA content compared to genes in the other categories. The *p*-values shown in panels A and B were obtained with the Wilcoxon rank sum test. (C and D) Heatmaps of the number of protein coding genes (panel C) and CSEPs (panel D) with certain sizes of upstream (*y*-axis) and downstream (*x*-axis) intergenic regions. The figure shows that the genome of EnFRAME01 does not exhibit large-scale compartmentalization of CSEP-encoding genes in gene-sparse regions.

5.2.6 Gene duplication asymmetrically affects different functional gene categories in *E. necator* and their genomic organization

A self-BLASTp search revealed a total of 941 genes (13.1%) duplicated in the genome of EnFRAME01, with CSEP-encoding genes experiencing significantly (*p*-value=1.8E-35) higher rates of gene duplications, as compared to genes in other functional categories (Fig 5.4A, Table 5.S18 and Supplementary Results). A conserved domain enrichment analysis further identified 30 domains that were significantly enriched among duplicated genes (adjusted *p*-value< 0.01), with the two most significantly enriched being the microbial ribonuclease (cl00212) and the Egh16-like virulence factor (PF11327) domains that are associated with CSEPs as well (Fig 5.4B and Table 5.S19). When considering the arrangement of the 941 gene duplicates in the genome of *E. necator*, the majority were dispersed gene duplicates (DGDs; *n*=712; 75.6%), as opposed to being proximal (PGDs; *n*=139; 14.8%) or tandem gene duplications (TGDs; *n*=90; 9.5%) (Fig 5.4A). However, genes encoding CSEPs significantly deviated from this pattern as they exhibited almost equal frequencies of dispersed (*n*=38; 35.5%), proximal (*n*=34; 31.8%), and tandem (*n*=35; 32.7%) duplications (Fig 5.4A). Indeed, several of the multi-copy CSEP-encoding genes, including the RALPH-like and Egh16-like CSEPs, were found to be tandemly arranged in clusters (Fig 5.S14), suggesting that CSEPs families expand by frequent local duplications in *E. necator*. A prominent example of this trend was the discovery of a 350 kb region on Chr1 that harbored 20 copies of a CSEP-encoding gene (i.e. *HI914_00480*), which were tandemly arranged in the same orientation on the same DNA strand (Fig 5.S15), and with 15 consecutive copies encoding identical proteins. Our analyses also suggested that local gene duplicates (i.e. PGDs and TGDs) are more conserved and thus more likely to contribute to genetic redundancy than DGDs, which due to their higher divergence, are likely to contribute more to functional diversification (Fig 5.4C). Similarly, when examining the rate of synonymous (K_S) and nonsynonymous substitutions (K_A), local gene duplicates exhibited overall lower K_S and higher K_A/K_S values as compared to DGDs (Fig 5.4D and Fig 5.S16). This indicated that PGDs and TGDs were likely more recent duplicates and were under more relaxed selection pressure as compared to DGDs. Collectively the above results indicate that gene duplication is a driver of genome evolution in *E. necator* that has differentially affected different gene categories, thereby leading to differences in their mode of evolution and organization of their paralogs in the genome.

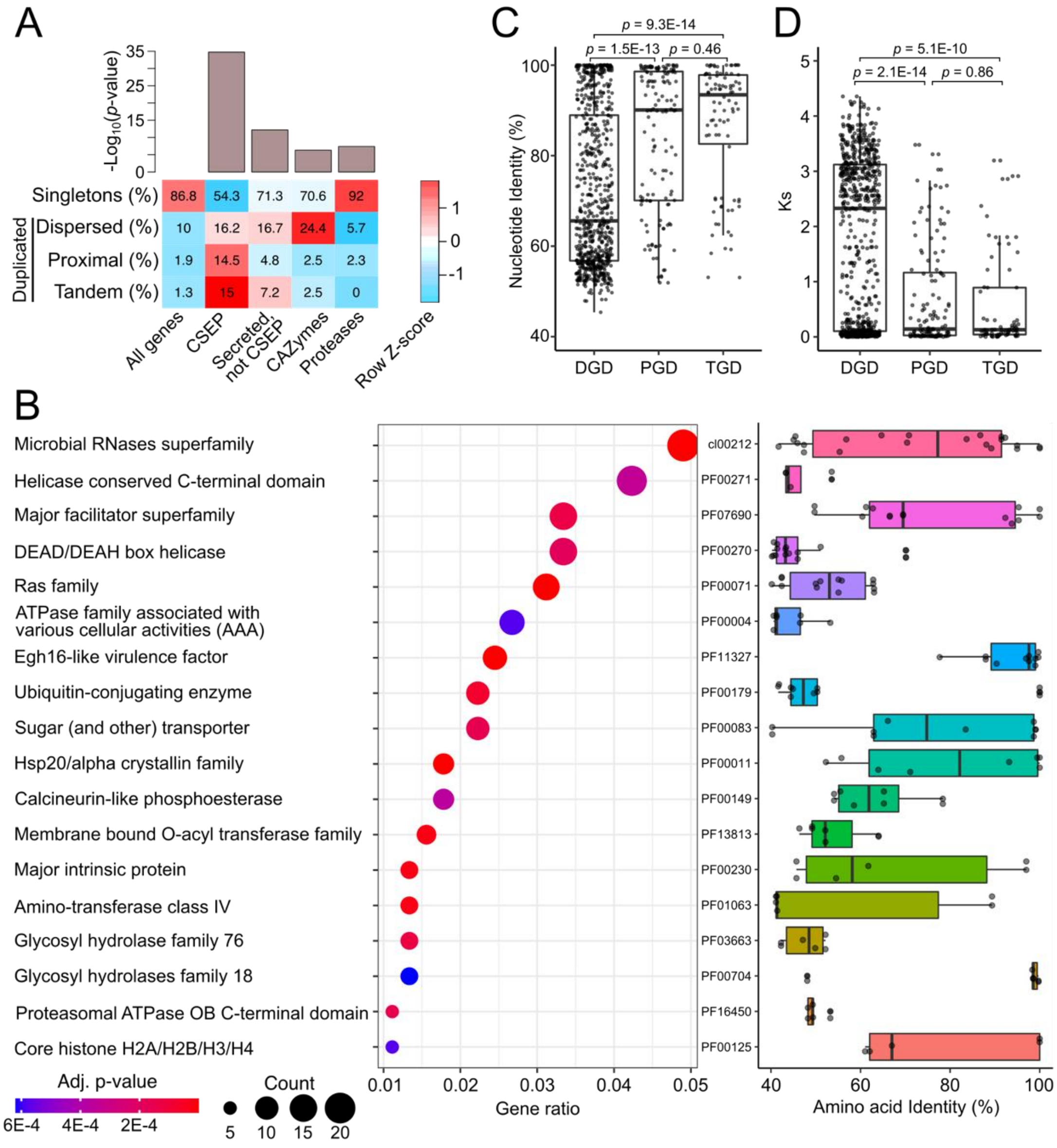


Figure 5.10: Landscape of gene duplications in the genome of *Erysiphe necator* isolate EnFRAME01. (A) Heatmap showing the percentage of genes in different functional categories that are singletons, dispersed duplications, proximal duplications, and tandem duplications. The bar chart shows *p*-values for enrichment of duplicated genes based on hypergeometric tests. The figure shows that ~13% of the genes are duplicated, and that the percentages of duplicated genes encoding candidate secreted effector proteins (CESPs) are significantly higher than genes encoding secreted proteins not classified as CSEPs, carbohydrate-active enzymes (CAZymes), and proteases. (B) The dot plot shows conserved domains significantly enriched within duplicated genes. The size of the dots corresponds to the number of duplicated genes containing the respective domain. The *x*-axis shows the proportion of the duplicated genes containing the respective domain that contributes to all duplicated genes containing a conserved domain. Dots are color-coded based on enrichment *p*-values adjusted using the Benjamini-Hochberg method. Distributions of pairwise identity values of duplicated copies are shown on the right hand -side based on top BLASTp hit.

The *p*-values shown in (A) and (B) were obtained with the Wilcoxon rank sum test. (C) Boxplots showing the distribution of pairwise nucleotide identity values of dispersed gene duplicates (DGD), proximal gene duplicates (PGD), and tandem gene duplicates (TGD). Each point represents a duplicated gene with the percent identity of its top BLASTn hit shown in the *y*-axis. The figure shows that copies of DGDs share significantly less nucleotide identity (median=65.6%), than copies of PGDs (median=90.1%) and TGDs (median=93.4%). (D) Boxplots showing the distribution of K_A/K_S values for DGDs, PGDs, and TGDs. Each point represents a duplicated gene with the K_A/K_S value of its top BLASTp hit shown in the *y*-axis. The figure shows that copies of PGDs and TGDs share higher conservation of K_A/K_S values than copies of DGDs.

5.2.7 Duplicated genes in EnFRAME01 frequently vary in copy number among *E. necator* isolates

The whole-genome sequencing (WGS) data of five *E. necator* isolates (Jones et al. 2014a) were used to identify genomic regions in EnFRAME01 with copy number variation (CNV) (i.e. deleted or duplicated). A total of 1,760 distinct CNV regions were identified, of which 1,589 (90.3% with an average size of 2.9 kb) were deletions and only 171 (9.7% with an average size of 5.6 kb) were duplications (Fig 5.S17A and Table 5.S20). CNV regions were dispersed throughout the 11 chromosomes of EnFRAME01 (Fig 5.S18), and were more frequently located near chromosome ends rather than gene-rich and repeat-rich regions (Fig 5.S17B). However, despite the lack of enrichment of CNV regions in repeat-rich regions, 80.5% (*n*=1,279) of the deleted and 70.1% (*n*=120) of the duplicated regions overlapped with predicted TEs (Fig 5.S17C). Moreover, 122 of the CNV regions overlapped with protein-coding genes and could therefore be considered as CNV genes (Supplementary Results). Of these, 53 genes were duplicated (average of 0.3 duplicated gene per all duplicated regions) and 69 genes were deleted (average of 1.5E-5 deleted gene per all deleted regions) among the *E. necator* isolates (Table 5.S21), indicating that genes with CNV were most likely to be affected by duplications rather than deletions. Most CNV regions also typically affected single genes rather than groups of genes, and no significant over- or under-representation of CSEPs, CAZymes, and proteases was observed among CNV genes. Instead, genes with CNV were significantly (*p*-value = 1.7E-27) enriched with the 941 genes predicted to be duplicated in EnFRAME01 (Table 5.S22). A notable example is the CSEP-encoding gene *HI914_00480*, which is present in 20 copies in isolate EnFRAME01 and eight-to-twelve copies in other strains (Table 5.S23). This indicated that rates of gain, retention, and loss of duplicated genes were asymmetric among different isolates of *E. necator*.

5.2.8 *Erysiphe necator* exhibits extensive CNV of a novel and PM-specific carboxylesterase

An inspection of the 122 genes with CNVs among the isolates of *E. necator* showed that gene *HI914_00624,* encoding a predicted secreted carboxylesterase (CE), exhibited the most dynamic changes in copy numbers, ranging from one in isolate EnFRAME01 to 31 in isolate Lodi (Supplementary Results). In all isolates the duplication affected the same 9.5 kb fragment that contained only the *HI914_00624* gene and was flanked by short direct repeats (Fig 5.5A). A blast search within the NCBI nr database indicated that homologs of HI914_00624 are abundantly present both within PM and non-PM fungal species (Fig 5.5B). However, a phylogenetic tree constructed using the top 400 best blastp hits, representing at least 195 distinct fungal species, showed that HI914_00624 belonged to a distinct clade that included 22 CEs, all from PM species (Fig 5.5C). Moreover, a multiple sequence alignment showed that the catalytic triad Ser-Asp/Glu-His, that is indispensable to the function of CEs (Sood et al. 2018; Oakeshott et al. 2005), is poorly conserved in these 22 PM-specific CEs (Fig 5.S19 and Table 5.S24). This suggests that HI914_00624 is a member of new clade of potentially non-catalytically active CEs or CEs with a modified enzymatic activity (Alam et al. 2002).

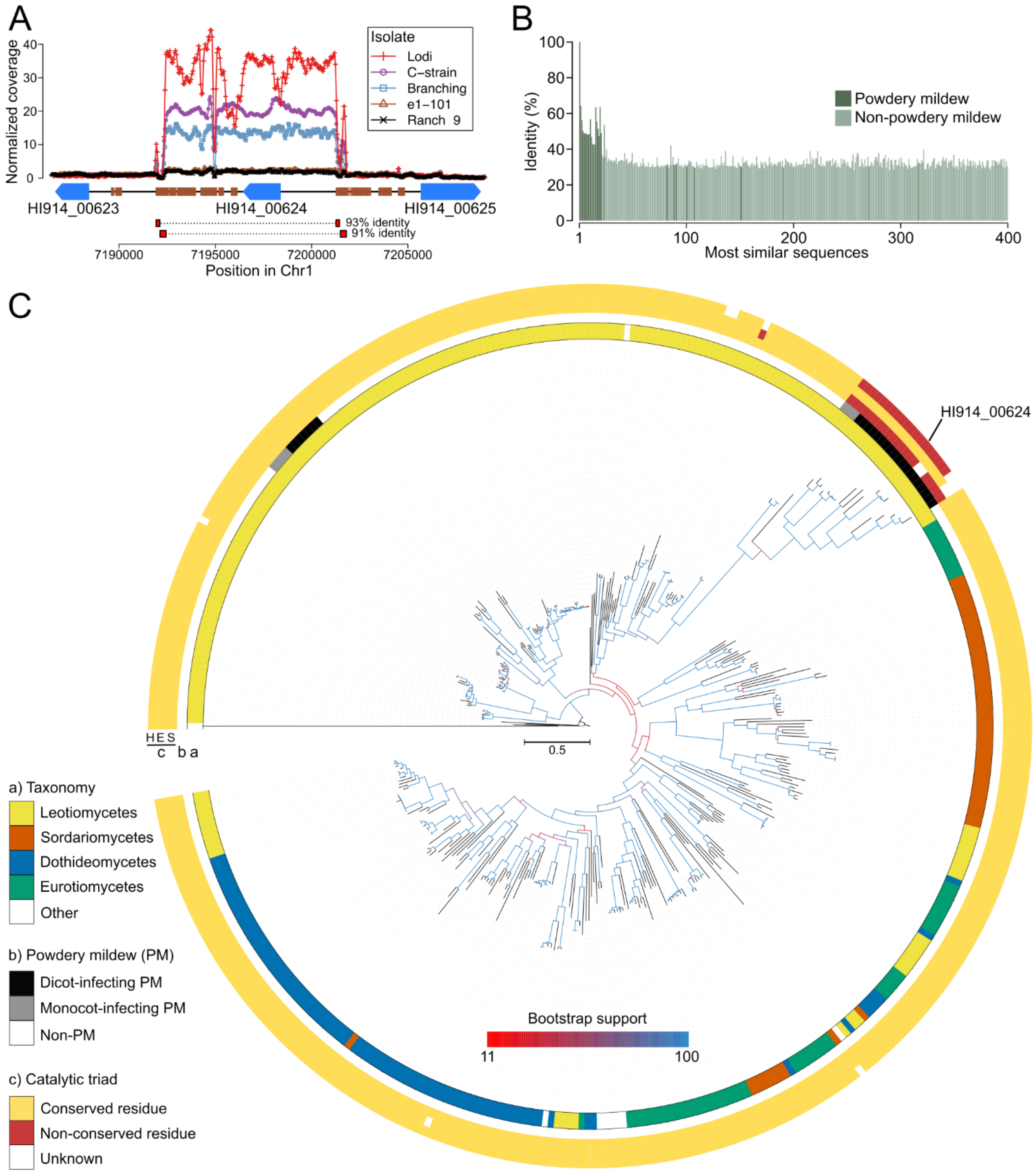


Figure 5.11: *Erysiphe necator* shows extensive copy number variation (CNV) of a putative secreted carboxylesterase (CE) that is poorly conserved in non-powdery mildew (PM) fungi. (A) Region of chromosome 1 (Chr1) in the genome of *E. necator* isolate EnFRAME01 containing the gene *HI914_00624* encoding a putative secreted CE. Genes are represented as blue arrows and repetitive DNA as small brown rectangles. Lines above the genes indicate estimated copy numbers of the region in five different isolates. The figure shows that isolates Lodi, C-strain, and Branching have more than ten predicted copies of

HI914_00624. The duplicated segment is flanked by short direct repeats of more than 90% identity. The figure also shows that genes flanking *HI914_00624* are not duplicated in the isolates analyzed. (B) Percent identity values of most similar sequences to the HI914_00624 protein sequence based on BLASTp searches. The figure shows that nearly all sequences from non-PM species have less than 40% amino acid identity. (C) Maximum likelihood phylogenetic tree of HI914_00624 and its most similar protein sequences from GenBank (2022-08-13) and EnFRAME01. The protein sequence of the acetylcholinesterase DmAChE from *Drosophila melanogaster* (Harel et al. 2000) was included as an outgroup. Tree branches are color-coded based on their support of 1000 bootstrap replicates. The tree was rooted at DmAChE. Track (a) shows the distribution of taxonomy classes of the sequences. Track (b) indicates sequences from monocot-infecting and dicot-infecting PMs. Track (c) shows the conservation of the Ser, Asp/Glu, and His residues that comprise the catalytic triad conserved in CEs. The figure shows that homologs of HI914_00624 are conserved in other PMs, but poorly conserved in non-PM fungal species. The figure also shows that homologs of HI914_00624 in PMs lack the Ser and His residues of the catalytic triad, which are largely conserved in predicted CEs from other fungal species.

5.3 Discussion

In this study we obtained a chromosome-scale genome for the grape PM *E. necator*, the first one for a dicot PM. The 81.1 Mb genome of *E. necator* is organized into 11 chromosomes, which are broadly characterized by the presence of large centromeric-like regions rich in repetitive DNA, a high content of retrotransposons and unclassified repeats that are mostly evenly dispersed outside their centromeric and telomeric regions, and the lack of compartmentalization in repeat-rich/gene-sparse regions, in agreement with the "one-speed genome" model of evolution. Moreover, *E. necator* had a reduced complement of genes encoding lytic enzymes (e.g. CAZymes and proteases) and those involved in carbohydrate metabolism, amino acid and purine metabolism, thiamine biosynthesis, and assimilation of inorganic nitrogen and sulfur. Loss of genes affecting these pathways is a characteristic feature of obligate biotrophs (Baxter et al. 2010; Spanu 2012; Cissé et al. 2014; Liang et al. 2018; Duplessis et al. 2011) and likely reflects a lack of selective pressure to retain these genes when the end products of the impaired pathways are available through leaky metabolic processes in the host (RoyChowdhury et al. 2022; Morris et al. 2012; Hauser 2014; Spanu 2012). This model of reductive genome evolution, in which organisms lose genes needed to synthesize metabolites that can be obtained directly through the host environment is frequently observed in nature, including obligate biotrophic fungi (RoyChowdhury et al. 2022; Cissé et al. 2014; Morris et al. 2012; Morris 2015).

The search for genes that are missing in *E. necator* but are commonly present in other ascomycete fungi also revealed that the fungus lacks *ERG5* and *ERG4*, whose products catalyze the two last steps, respectively, of ergosterol biosynthesis in fungi (Hu et al. 2017). This corroborates previous reports that ergosterol is essentially absent in this fungus as well as in PMs in general (Loeffler et al. 1992; Debieu et al. 1995). The deletion of *ERG5* and/or *ERG4* is typically not lethal to fungi but leads to an altered ergosterol biosynthesis that may affect their physiology, increase their sensitivity to multiple chemicals, and generally decrease their fitness under stress conditions (Kodedova and Sychrova 2015; Aguilar et al. 2010; Hu et al. 2018; Liu et al. 2017; Bhattacharya et al. 2018). For instance, both genes are required for conidiation of *Aspergillus fumigatus* (Long et al. 2017; Long and Zhong 2022), *ERG4* is crucial for vegetative differentiation and virulence in *Fusarium graminearum* (Liu et al. 2013), and deletion of *ERG4* increased the production of extracellular pigments in *Monascus purpureus* (Liu et al. 2019). Likewise, the deletion of *ERG5* increased the susceptibility of *Candida albicans, Neurospora crassa,* and *Fusarium verticillioides* to azole antifungals (Sun et al. 2013; Martel et al. 2010). Loss of *ERG4* and *ERG5* is unlikely to have a fitness effect on *E. necator* or impact major features of its physiology, and it is thus intriguing to speculate possible biological explanations for the loss of ergosterol biosynthesis. As ergosterol is an inducer of innate immunity in plants (Klemptner et al. 2014), its absence may help *E. necator* avoid sterol-induced immunity in grapes. This might be possible as the activation in *Vitis vinifera* of the type I lipid transfer protein VvLTP1 by ergosterol treatment (Laquitaine et al. 2006; Gomès et al. 2003) leads to the induction of the stilbene synthase gene *Vst1* (Laquitaine et al. 2006). *Vst1*, in turn, regulates the biosynthesis of the phytoalexin resveratrol that enhances resistance against *Botrytis cinerea* (Coutos-Thévenot et al. 2001) and *E. necator* (Schnee et al. 2008).

The overall genomic characteristics of *E. necator* conform to those reported for other PM pathogens, including different formae speciales of the cereal PM pathogen *B. graminis* (Liang et al. 2018; Spanu et al. 2010; Bindschedler et al. 2016). However, the 11 chromosomes of *E. necator* are not syntenic to the 11 chromosomes of *B. graminis*, indicating rapid diversification of their genomic architecture following speciation. Moreover, the genome of *E. necator* has a different TE complement compared to *B. graminis*, as it contains mostly LTR retrotransposons rather than the non-LTR retrotransposons. TEs are a major force of evolution and adaptation to stressful environments, as their bursts and mobilization provoke chromosomal reorganization and phylogenetic divergence (Muszewska et al. 2019; Belyayev 2014). Our analysis showed that both species have experienced not one, as previously reported (Frantzeskakis et al. 2018), but at least two bursts of TEs in their evolutionary history. The first burst possibly preceded their divergence and involved mostly non-LTR retrotransposons, and the second burst likely took place after their speciation and involved LTR-retrotransposons, but also RCs (i.e. Helitrons) in *E. necator*. Such differences in TE bursts in *E. necator* and *B. graminis* are likely to have restructured their genomes, accelerated their speciation, and influenced their adaptation on different hosts by, among others, affecting virulence-associated genes such as CSEPs. Indeed, despite the lack of large-scale compartmentalization in *E. necator*, its CSEP-encoding genes exhibited significantly higher duplication rates as compared to other functional gene categories and were embedded in larger intergenic regions that were richer in TEs. The increase in duplication rates could have been prompted by the presence of TEs, as the repetitive nature of transposons provides a substrate for non-allelic homologous recombination that would typically generate tandemly arranged gene copies in their flanking regions (Kuzmin et al. 2021; Hastings et al. 2009; Cerbin and Jiang 2018). TEs have also been hypothesized to mediate the duplication and proliferation of CSEPs in *B. graminis* (Pedersen et al. 2012; Müller et al. 2019), as CSEP-encoding genes in this species are frequently duplicated and present in tandem in physical proximity to similar repetitive DNA (Pedersen et al. 2012; Müller et al. 2019). Thus, next to promoting chromosomal reorganization, TEs seem to have had a major role in shaping the evolution of *E. necator* and *B. graminis* as plant pathogens by providing a favorable environment for CSEP duplication.

The inflation of the *E. necator* genome by TEs was further accompanied by high rates of gene duplication, which likely contributed further to its genomic plasticity and genetic diversity. Gene duplication is a major force of evolution as it provides material for functional, regulatory, and transcriptional divergence through the generation of new genes and their subsequent neo-, sub-, or hypo-functionalization (Magadum et al. 2013; Birchler and Yang 2022). A variety of mechanisms can trigger gene duplications, with different mechanisms creating suites of duplicated genes in different configurations within a genome, which in turn contribute differentially to functional innovation and redundancy (Qiao et al. 2018; Wang et al. 2011). Our analysis indicated that in *E. necator,* genes from different functional categories exhibited different rates and modes of duplications, and that different modes of gene duplication were under different strengths of selection pressure. These features were again more prominent with CSEP-encoding genes, whose gene duplicates were more likely to be in close (i.e. tandem or proximal) physical location in the genome of EnFRAME01 than the copies of other gene classes and had on average higher K_A/K_S values, indicating that duplicated CSEP genes are potentially subject to higher rates of evolution. This is consistent with the role of effectors on host adaptation and overcoming of the host immune system, and indicates an ongoing arms-race between *E. necator* and its grapevine host (Müller et al. 2019).

Next to gene duplications, CNVs within a species population can significantly affect its fitness (Katju and Bergthorsson 2013). It has been shown, for example, that an increase in *CYP51* (*ERG11*) copy numbers, the gene encoding a key enzyme for ergosterol biosynthesis, creates a gene dosage effect that reduces the sensitivity of *E. necator* to demethylase inhibitor fungicides (Jones et al. 2014a). In *B. graminis* f.sp. *hordei* (Menardo et al. 2017; Pedersen et al. 2012) and *B. graminis* f.sp. *tritici* (Müller et al. 2019), high levels of CNV in genes encoding CSEPs are thought to be major drivers of virulence and rapid adaptation to host genotypes. Our CNV analysis revealed that the CE-encoding gene *HI914_00624* exhibited the most dynamic changes in copy numbers, suggesting that it is a target of natural selection. Moreover, we found that *HI914_00624* is a member of a novel family of CE-encoding genes with multiple duplications in PM species. CEs are a large superfamily of structurally diverse, multifunctional enzymes that hydrolyze carboxylesters in natural and synthetic molecules (Oakeshott et al. 2005), including pharmaceutical drugs, pesticides, environmental pollutants, and toxins. Due to their catalytic flexibility, they may have crucial roles in detoxifying cells from harmful compounds and metabolites, but also in physiological processes such as lipid metabolism and energy homeostasis (Ross et al. 2010). We speculate that the putative CE encoded by *HI914_00624*, and its homologs in PM species, represent a new family of non-catalytic CEs, as they were poorly conserved in non-PM fungi and lack the conserved Ser-Asp/Glu-His amino-acid triad required for their proper function (Oakeshott et al. 2005). Catalytic competence is thought to be the ancestral state of CEs, but several non-catalytic clades that have acquired new functions are present in higher Eukaryotes (Oakeshott et al. 2005, 1999). Among fungi, *vdtD* from the opportunistic human pathogen *Paecilomyces variotii* encodes a putative non-catalytic CE that is part of a gene cluster mediating the biosynthesis of the antibacterial viriditoxin (Urquhart et al. 2019). It has been suggested that instead of acting as a hydrolase, vdtD could bind to the compound to protect the methyl ester from being hydrolyzed by endogenous hydrolases (Hu et al. 2019). These examples highlight the capability of CEs to evolve new functions, and it is therefore possible that the putative CE encoded by *HI914_00624* and its homologs in PMs have evolved new functions compared to ancestral CEs.

5.4 Materials and Methods

A detailed version of materials and methods is provided in Supplementary Results at the end of this chapter.

5.4.1 Fungal isolate, nucleic acid extraction and sequencing

Erysiphe necator isolate EnFRAME01 was isolated from greenhouse-grown grapes in Corvallis Oregon, USA, in 2018. EnFRAME01 was propagated by dusting from detached leaves (Miles et al. 2021) on *Vitis vinifera* L., cv. 'Chardonnay' seedlings grown hydroponically in half-strength Hoagland's solution (Hoagland and Arnon 1950). High-molecular weight DNA was obtained from conidia as in (Feehan et al. 2017) with modifications. PacBio library construction and sequencing was outsourced to the DNA Technologies and Expression Analysis Core Laboratory at the UC Davis Genome Center. The constructed library was sequenced using two SMRT Cells 1M v2 on a Sequel Chemistry v2 platform (Pacific Biosciences, Menlo Park, CA). Extracted DNA was also used to generate an Illumina WGS library and a Hi-C library using the Proximo Hi-C Kit (microbial) (Phase Genomics), according to the manufacturer's instructions. Both Illumina libraries were sequenced on a NovaSeq 6000 instrument (PE150 format). To assist gene prediction, total RNA was extracted from conidia of six *E. necator* isolates (EnFRAME01, BPPQ1B.3, BPPQ1B.5, DDOME-1, DDOME-2 and HO2) held at the USDA – ARS Horticultural Crops Disease and Pest Management Research Unit in Corvallis, Oregon, using Trizol reagent (ThermoFisher) according to the manufacturer's instructions. Sample integrity analysis, cDNA library preparation, and sequencing on the Illumina NovaSeq6000 platform using the paired-end (PE150) format were carried out at Novogene, Inc (Sacramento, California).

5.4.2 Genome assembly and annotation of repetitive DNA

PacBio reads were assembled with Canu v1.8 (Koren et al. 2017) and then used to polish the contigs with pbmm2 and Arrow from the GenomicConsensus package v2.3.3. The assembly was further polished with Pilon v1.23 (Walker et al. 2014) after mapping the Illumina reads with BWA-MEM v0.7.17 (Li and Durbin 2009). Hi-C reads were mapped with BWA-MEM v0.7.17 and chromatin interaction frequencies were estimated with the 3D-DNA package (Dudchenko et al. 2017). They were then visualized with Juicebox v1.11.08 (Durand et al. 2016), which allowed the grouping of contigs into putative chromosomes. Repetitive regions were identified with RepeatModeler v2.0.2a (Flynn et al. 2020) and masked with RepeatMasker v4.1.2-p1. The repeat divergence landscape was estimated with the script parseRM.pl v5.8.2.

5.4.3 Gene prediction

RNA-seq reads were mapped to the genome assembly with HISAT2 v2.2.0 (Kim et al. 2015) and transcripts were reconstructed with Stringtie v2.1.1 (Pertea et al. 2015) and Trinity v2.9.1 (Grabherr et al. 2011). Genes were predicted with Maker v2.31.10 (Cantarel et al. 2008) by integrating i) the trained *ab initio* predictors GeneMark-ES v4.57 (Lukashin and Borodovsky 1998), SNAP v2013-11-29 (Korf 2004), and Augustus v3.2.3 (Stanke et al. 2006), ii) gene models generated with GeMoMa (Keilwagen et al. 2019), iii) assembled transcripts, and iv) protein sequences from close relative species.

5.4.4 Homology-based functional annotations

Conserved PFAM domains were identified with InterProScan v5.32-71.0 (Jones et al. 2014b) or the NCBI CDD database (Marchler-Bauer et al. 2017). Carbohydrate-active enzymes (CAZymes) were predicted with dbCAN2 (Zhang et al. 2018, 2). Proteases and transporters were classified based on the top BLASTp hit (E-value< 1E-10) against the MEROPS database v12.1 (Rawlings et al. 2014) and the TCDB database version of 2020-07-12 (Saier Jr et al. 2014), respectively. Secreted proteins were predicted with SignalP v5.0 (Armenteros et al. 2019). Membrane-bound proteins were predicted with PredGPI (Pierleoni et al. 2008) and TMHMM v2.0 (Krogh et al. 2001). Secreted proteins were classified into CSEPs based on three lines of evidence (Fig 5.S20): i) EffectorP v2.0 (Sperschneider et al. 2016); ii) proteins shorter than 250 aa with at least 2% cysteines; iii) proteins with no homologs in Leotiomycetes, except Erysiphales, based on a BLASTp search (E-value< 1E-3) against 93 Leotiomycetes genomes.

5.4.5 Identification of core genes missing in EnFRAME01

Protein sequences from *E. necator*, *B. graminis* f.sp. *hordei*, *B. cinerea, Zymoseptoria tritici*, *Aspergillus niger*, *N. crassa*, and *S. cerevisiae* were organized into orthogroups with OrthoFinder v2.5.4 (Emms and Kelly 2015). Orthogroups containing proteins from all non-PM species, but not from *E. necator* were considered core genes missing in *E. necator*.

5.4.6 Classification and enrichment of duplicated genes

Duplicated genes were identified based on an *all-vs-all* BLASTp (e-value < 1E-5) search, with minimum identity of 40% and minimum coverage of 50%. The script *duplicate_gene_classifier* from MCScanX (Wang et al. 2012) was used to classify gene duplications into dispersed, proximal, or tandem. Enrichment of gene categories within duplicated genes was performed with hypergeometric tests using the *phyper* function within R v4.1.2. Pairwise K_A/K_S ratios were estimated with K_A/K_S_calculator v3 (Zhang 2022). Conserved domain enrichment was performed with the *enricher* function from the R package clusterProfiler v4.2.2 (Yu et al. 2012) within R v4.1.2 with adjusted *p*-value< 0.01.

5.4.7 Identification of CNVs

Whole-genome sequencing reads of five *E. necator* isolates (Jones et al. 2014a) were mapped to the genome with BWA-MEM v0.7.17 (Li and Durbin 2009). PCR duplicates were marked with samblaster v0.1.24 (Faust and Hall 2014) and removed with SAMtools v1.9 (Li et al. 2009). CNV regions were identified with CNVnator (Abyzov et al. 2011). Genes with at least 80% overlapping with CNV regions were considered CNV genes.

5.4.8 Comparative analysis of carboxylesterases

The predicted carboxylesterase HI914_00624 was queried with BLASTp against the NCBI nr database (2022-08-13) and proteins from EnFRAME01 and the 400 most similar sequences (e-value< 1E-50) were obtained. The acetylcholinesterase DmAChE from *Drosophila melanogaster* (1QO9) (Harel et al. 2000) was included as an outgroup and also used as reference to identify conserved residues. The 401 amino acid sequences were aligned with MAFFT v7.490 (Katoh et al. 2002) and sites composed of more than 50% gaps were removed with trimAl v1.4 (Capella-Gutiérrez et al. 2009). The phylogenetic tree was inferred with IQ-TREE v1.6.12 (Nguyen et al. 2015) using the built-in ModelFinder (Kalyaanamoorthy et al. 2017) and 1000 rapid bootstrap replicates (Hoang et al. 2018). The tree was visualized and edited with iTOL (Letunic and Bork 2021). Quantitative PCR (qPCR) and quantitative reverse transcription PCR (RT-qPCR) were used to determine the copy number and gene expression of the *HI914_00624* gene, respectively, in six isolates of

E. necator. qPCR reactions were run in triplicate on an Applied Biosystems QuantStudio5 qPCR machine using PerfeCTa qPCR ToughMix Low ROX (Quantabio) and the primers and probes listed in Table 5.S26.

5.5 Data availability

Raw sequencing reads generated in this study were deposited at NCBI SRA under accessions SRR18712274 through SRR18712279 (BioProject PRJNA627990). The annotated genome of *E. necator* EnFRAME01 was deposited at NCBI under accession JABETL000000000.1 (E101OR).

Acknowledgments

This research was funded by a USDA-NIFA Specialty Crop Research Initiative Award (No. 2018-03375) titled "FRAME: Fungicide Resistance Assessment, Mitigation and Extension Network for Wine, Table, and Raisin Grapes", United States Department of Agriculture, Agriculture Research Service CRIS 5358-22000-041-00D. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement of approval by the United States Department of Agriculture (USDA) or the Agricultural Research Service (ARS) of any product or service to the exclusion of others that may be suitable. The authors declare no conflict of interest.

Author contributions

Conceptualization: AZZ, IS; Data curation: AZZ; Formal Analysis: AZZ; Funding acquisition: IS, WFM; Investigation: AZZ; Methodology: AZZ; Project administration: IS; Resources: TN, JC; Software: AZZ; Supervision: IS, WFM; Validation: AZZ; Visualization: AZZ; Writing – original draft: AZZ, IS; Writing – review & editing: AZZ, TN, JC, WFM, IS.

5.6 References

Abyzov, A., Urban, A. E., Snyder, M., and Gerstein, M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21:974–984

Aguilar, P. S., Heiman, M. G., Walther, T. C., Engel, A., Schwudke, D., Gushwa, N., Kurzchalia, T., and Walter, P. 2010. Structure of sterol aliphatic chains affects yeast cell shape and cell fusion during mating. Proc. Natl. Acad. Sci. 107:4170–4175

Alam, M., Vance, D. E., and Lehner, R. 2002. Structure−function analysis of human triacylglycerol hydrolase by site-directed mutagenesis: identification of the catalytic triad and a glycosylation site. Biochemistry. 41:6679–6687

Amselem, J., Lebrun, M.-H., and Quesneville, H. 2015. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. BMC Genomics. 16:141

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420–423

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D., and Wingfield, M. J. 2017. A plant pathology perspective of fungal genome sequencing. IMA Fungus. 8:1–15

Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., and others. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. science. 330:1549–1551

Belyayev, A. 2014. Bursts of transposable elements as an evolutionary driving force. J. Evol. Biol. 27:2573–2584

Bhattacharya, S., Esquivel, B. D., and White, T. C. 2018. Overexpression or deletion of ergosterol biosynthesis genes alters doubling time, response to stress agents, and drug susceptibility in Saccharomyces cerevisiae. MBio. 9

Bindschedler, L. V., Panstruga, R., and Spanu, P. D. 2016. Mildew-omics: how global analyses aid the understanding of life and evolution of powdery mildews. Front. Plant Sci. 7:123

Birchler, J. A., and Yang, H. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. Plant Cell. 34:2466–2474

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., and Yandell, M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188–196

Capella-Gutiérrez, S., Silla-Martínez, J. M., and 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., and Ramírez, L. 2016. Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles C. Feschotte, ed. PLOS Genet. 12:e1006108

Cerbin, S., and Jiang, N. 2018. Duplication of host genes by transposable elements. Curr. Opin. Genet. Dev. 49:63–69

Cissé, O. H., Pagni, M., and Hauser, P. M. 2014. Comparative genomics suggests that the human pathogenic fungus *Pneumocystis jirovecii* acquired obligate biotrophy through gene loss. Genome Biol. Evol. 6:1938–1948

Coutos-Thévenot, P., Poinssot, B., Bonomelli, A., Yean, H., Breda, C., Buffard, D., Esnault, R., Hain, R., and Boulay, M. 2001. In vitro tolerance to *Botrytis cinerea* of grapevine 41B rootstock in transgenic plants expressing the stilbene synthase *Vst1* gene under the control of a pathogen-inducible PR 10 promoter. J. Exp. Bot. 52:901–910

Debieu, D., Corio-Costet, M.-F., Steva, H., Malosse, C., and Leroux, P. 1995. Sterol composition of the vine powdery mildew fungus, *Uncinula necator*: comparison of triadimenol-sensitive and resistant strains. Phytochemistry. 39:293–300

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

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., and others. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 356:92–95

Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., and others. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. 108:9166–9171

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., and Aiden, E. L. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3:99–101

Emms, D. M., and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:1–14

Faust, G. G., and Hall, I. M. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 30:2503–2505

Feehan, J. M., Scheibel, K. E., Bourras, S., Underwood, W., Keller, B., and Somerville, S. C. 2017. Purification of high molecular weight genomic DNA from powdery mildew for long-read sequencing. J. Vis. Exp. :e55463

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., and Smit, A. F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. 117:9451–9457

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., and 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

Frantzeskakis, L., Kusch, S., and Panstruga, R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. Mol. Plant Pathol. 20:3–7

Fuller, K. B., Alston, J. M., and Sambucci, O. S. 2014. The value of powdery mildew resistance in grapes: evidence from California. Wine Econ. Policy. 3:90–107

Gadoury, D. M., Cadle-Davidson, L., Wilcox, W. F., Dry, I. B., Seem, R. C., and Milgroom, M. G. 2012. Grapevine powdery mildew (*Erysiphe necator*): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. Mol. Plant Pathol. 13:1–16

Gaforio, L., Garcia-Munoz, S., Cabello, F., and Munoz-Organero, G. 2011. Evaluation of susceptibility to powdery mildew (*Erysiphe necator*) in *Vitis vinifera* varieties. Vitis. 50:123–126

Gan, P., Hiroyama, R., Tsushima, A., Masuda, S., Shibata, A., Ueno, A., Kumakura, N., Narusaka, M., Hoat, T. X., Narusaka, Y., and others. 2020. Subtelomeric regions and a repeat-rich chromosome harbor multicopy effector gene clusters with variable conservation in multiple plant pathogenic *Colletotrichum* species. bioRxiv. :2020.04.28.061093

Gomès, E., Sagot, E., Gaillard, C., Laquitaine, L., Poinssot, B., Sanejouand, Y.-H., Delrot, S., and Coutos-Thévenot, P. 2003. Nonspecific lipid-transfer protein genes expression in grape (*Vitis* sp.) cells in response to fungal elicitor treatments. Mol. Plant. Microbe Interact. 16:456–464

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., and others. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29:644

Grell, M. N., Mouritzen, P., and Giese, H. 2003. A *Blumeria graminis* gene family encoding proteins with a C-terminal variable region with homologues in pathogenic fungi. Gene. 311:181–192

Hacquard, S. 2014. The genomics of powdery mildew fungi: past achievements, present status and future prospects. Adv. Bot. Res. 70:109–142

Harel, M., Kryger, G., Rosenberry, T. L., Mallender, W. D., Lewis, T., Fletcher, R. J., Guss, J. M., Silman, I., and Sussman, J. L. 2000. Three-dimensional structures of *Drosophila melanogaster* acetylcholinesterase and of its complexes with two potent inhibitors. Protein Sci. 9:1063–1072

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. 2009. Mechanisms of change in gene copy number. Nat. Rev. Genet. 10:551–564

Hauser, P. M. 2014. Genomic insights into the fungal pathogens of the genus *Pneumocystis*: obligate biotrophs of humans and other mammals. PLoS Pathog. 10:e1004425

Hoagland, D. R., and Arnon, D. I. 1950. *The water-culture method for growing plants without soil*. 2nd ed. Circular 347. California agricultural experiment station, University of California, Berkeley, CA.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522

Hu, C., Zhou, M., Wang, W., Sun, X., Yarden, O., and Li, S. 2018. Abnormal ergosterol biosynthesis activates transcriptional responses to antifungal azoles. Front. Microbiol. 9:9

Hu, J., Li, H., and Chooi, Y.-H. 2019. Fungal dirigent protein controls the stereoselectivity of multicopper oxidase-catalyzed phenol coupling in viriditoxin biosynthesis. J. Am. Chem. Soc. 141:8068–8072

Hu, Z., He, B., Ma, L., Sun, Y., Niu, Y., and Zeng, B. 2017. Recent advances in ergosterol biosynthesis and regulation mechanisms in *Saccharomyces cerevisiae*. Indian J. Microbiol. 57:270–277

Irelan, J. T., and Selker, E. U. 1996. Gene silencing in filamentous fungi: RIP, MIP and quelling. J. Genet. 75:313–324

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K. C. H., McGuire, B., Gubler, W. D., Walker, M. A., and Cantu, D. 2014a. Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. BMC Genomics. 15:1081

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., and others. 2014b. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236–1240

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermiin, L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589

Kanehisa, M., Sato, Y., and Kawashima, M. 2022. KEGG mapping tools for uncovering hidden features in biological data. Protein Sci. 31:47–53

Katju, V., and Bergthorsson, U. 2013. Copy-number changes in evolution: rates, fitness effects and adaptive significance. Front. Genet. 4:273

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

Keilwagen, J., Hartung, F., and Grau, J. 2019. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol. Biol. Clifton NJ. 1962:161–177

Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 12:357–360

King, R., Urban, M., Hammond-Kosack, M. C. U., Hassani-Pak, K., and Hammond-Kosack, K. E. 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. BMC Genomics. 16:544

Klemptner, R. L., Sherwood, J. S., Tugizimana, F., Dubery, I. A., and Piater, L. A. 2014. Ergosterol, an orphan fungal microbe-associated molecular pattern (MAMP). Mol. Plant Pathol. 15:747–761

Kodedova, M., and Sychrova, H. 2015. Changes in the sterol composition of the plasma membrane affect membrane potential, salt tolerance and the activity of multidrug resistance pumps in *Saccharomyces cerevisiae*. PLoS One. 10:e0139306

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736

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

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

Kunova, A., Pizzatti, C., Saracchi, M., Pasquali, M., and Cortesi, P. 2021. Grapevine powdery mildew: fungicides for its management and advances in molecular detection of markers associated with resistance. Microorganisms. 9:1541

Kuzmin, E., Taylor, J. S., and Boone, C. 2021. Retention of duplicated genes in evolution. Trends Genet. 38:59–72

Laquitaine, L., Gomès, E., François, J., Marchive, C., Pascal, S., Hamdi, S., Atanassova, R., Delrot, S., and Coutos-Thévenot, P. 2006. Molecular basis of ergosterol-induced protection of grape against *Botrytis cinerea*: induction of type I LTP promoter activity, WRKY, and stilbene synthase gene expression. Mol. Plant. Microbe Interact. 19:1103–1112

Letunic, I., and Bork, P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49:W293–W296

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25:1754–1760

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

Liang, P., Liu, S., Xu, F., Jiang, S., Yan, J., He, Q., Liu, W., Lin, C., Zheng, F., Wang, X., and others. 2018. Powdery mildews are characterized by contracted carbohydrate metabolism and diverse effectors to adapt to obligate biotrophic lifestyle. Front. Microbiol. 9:3160

Liu, G., Chen, Y., Færgeman, N. J., and Nielsen, J. 2017. Elimination of the last reactions in ergosterol biosynthesis alters the resistance of *Saccharomyces cerevisiae* to multiple stresses. FEMS Yeast Res. 17

Liu, J., Chai, X., Guo, T., Wu, J., Yang, P., Luo, Y., Zhao, H., Zhao, W., Nkechi, O., Dong, J., and others. 2019. Disruption of the ergosterol biosynthetic pathway results in increased membrane permeability, causing overproduction and secretion of extracellular *Monascus* pigments in submerged fermentation. J. Agric. Food Chem. 67:13673–13683

Liu, X., Jiang, J., Yin, Y., and Ma, Z. 2013. Involvement of *FgERG4* in ergosterol biosynthesis, vegetative differentiation and virulence in *Fusarium graminearum*. Mol. Plant Pathol. 14:71–83

Loeffler, R. T., Butters, J. A., and Hollomon, D. W. 1992. The sterol composition of powdery mildews. Phytochemistry. 31:1561–1563

Long, N., Xu, X., Zeng, Q., Sang, H., and Lu, L. 2017. *Erg4A* and *Erg4B* are required for conidiation and azole resistance via regulation of ergosterol biosynthesis in *Aspergillus fumigatus*. Appl. Environ. Microbiol. 83:e02924-16

Long, N., and Zhong, G. 2022. The C-22 sterol desaturase Erg5 is responsible for ergosterol biosynthesis and conidiation in *Aspergillus fumigatus*. J. Microbiol. :1–7

Lukashin, A. V., and Borodovsky, M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26:1107–1115

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., and Ravikesavan, R. 2013. Gene duplication as a major force in evolution. J. Genet. 92:155–161

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., and Bryant, S. H. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45:D200–D203

Martel, C. M., Parker, J. E., Bader, O., Weig, M., Gross, U., Warrilow, A. G., Kelly, D. E., and Kelly, S. L. 2010. A clinical isolate of <i>Candida albicans<i> with mutations in *ERG11* (encoding sterol 14α-demethylase) and *ERG5* (encoding C22 desaturase) is cross resistant to azoles and amphotericin B. Antimicrob. Agents Chemother. 54:3578–3583

Menardo, F., Praz, C. R., Wicker, T., and Keller, B. 2017. Rapid turnover of effectors in grass powdery mildew (*Blumeria graminis*). BMC Evol. Biol. 17:1–14

Miles, T. D., Neill, T. M., Colle, M., Warneke, B., Robinson, G., Stergiopoulos, I., and Mahaffee, W. F. 2021. Allele-specific detection methods for QoI fungicide-resistant *Erysiphe necator* in vineyards. Plant Dis. 105:175–182

Morris, J. J. 2015. Black Queen evolution: the role of leakiness in structuring microbial communities. Trends Genet. 31:475–482

Morris, J. J., Lenski, R. E., and Zinser, E. R. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio. 3:e00036-12

Müller, M. C., Kunz, L., Graf, J., Schudel, S., and Keller, B. 2021. Host adaptation through hybridization: genome analysis of triticale powdery mildew reveals unique combination of lineage-specific effectors. Mol. Plant. Microbe Interact. 34:1350–1357

Müller, M. C., Praz, C. R., Sotiropoulos, A. G., Menardo, F., Kunz, L., Schudel, S., Oberhänsli, S., Poretti, M., Wehrli, A., Bourras, S., Keller, B., and Wicker, T. 2019. A chromosome-scale genome assembly reveals a highly dynamic effector repertoire of wheat powdery mildew. New Phytol. 221:2176–2189

Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., and Ginalski, K. 2019. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci. Rep. 9:4307

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274

Oakeshott, J., Claudianos, C., Campbell, P., Newcomb, R., and Russell, R. 2005. Biochemical genetics and genomics of insect esterases. Pages 309–381 in: Biochemical genetics and genomics of insect esterases, Elsevier, Oxford, Amsterdam.

Oakeshott, J., Claudianos, C., Russell, R., and Robin, G. 1999. Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. Bioessays. 21:1031–1042

Pedersen, C., van Themaat, E. V. L., McGuffin, L. J., Abbott, J. C., Burgis, T. A., Barton, G., Bindschedler, L. V., Lu, X., Maekawa, T., Weßling, R., and others. 2012. Structure and evolution of barley powdery mildew effector candidates. BMC Genomics. 13:694

Pennington, H. G., Jones, R., Kwon, S., Bonciani, G., Thieron, H., Chandler, T., Luong, P., Morgan, S. N., Przydacz, M., Bozkurt, T., and others. 2019. The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA. PLoS Pathog. 15:e1007620

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290–295

Pierleoni, A., Martelli, P. L., and Casadio, R. 2008. PredGPI: a GPI-anchor predictor. BMC Bioinformatics. 9:392

Qiao, X., Yin, H., Li, L., Wang, R., Wu, J., Wu, J., and Zhang, S. 2018. Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*). Front. Plant Sci. 9:161

Qiu, W., Feechan, A., and Dry, I. 2015. Current understanding of grapevine defense mechanisms against the biotrophic fungus (*Erysiphe necator*), the causal agent of powdery mildew disease. Hortic. Res. 2

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

Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42:D503–D509

Ross, M. K., Streit, T. M., Herring, K. L., and Xie, S. 2010. Carboxylesterases: dual roles in lipid and pesticide metabolism. J. Pestic. Sci. 35:257–264

RoyChowdhury, M., Sternhagen, J., Xin, Y., Lou, B., Li, X., and Li, C. 2022. Evolution of pathogenicity in obligate fungal pathogens and allied genera. PeerJ. 10:e13794

Saier Jr, M. H., Reddy, V. S., Tamang, D. G., and Västermark, Å. 2014. The transporter classification database. Nucleic Acids Res. 42:D251–D258

Schnee, S., Viret, O., and Gindro, K. 2008. Role of stilbenes in the resistance of grapevine to powdery mildew. Physiol. Mol. Plant Pathol. 72:128–133

Schotanus, K., Soyer, J. L., Connolly, L. R., Grandaubert, J., Happel, P., Smith, K. M., Freitag, M., and Stukenbrock, E. H. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin. 8:41

Seidl, M. F., Kramer, H. M., Cook, D. E., Fiorin, G. L., Berg, G. C. M. van den, Faino, L., and Thomma, B. P. H. J. 2020. Repetitive elements contribute to the diversity and evolution of centromeres in the fungal genus *Verticillium*. mBio. 11:e01714-20

Selker, E. U. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu. Rev. Genet. 24:579–613

Smith, K. M., Phatale, P. A., Sullivan, C. M., Pomraning, K. R., and Freitag, M. 2011. Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. Mol. Cell. Biol. 31:2528–2542

Sood, S., Sharma, A., Sharma, N., and Kanwar, S. S. 2018. Carboxylesterases: sources, characterization and broader applications. Insights Enzyme Res. 01:2

Spanu, P. D. 2017. Cereal immunity against powdery mildews targets RNase-Like Proteins associated with Haustoria (RALPH) effectors evolved from a common ancestral gene. New Phytol. 213:969–971

Spanu, P. D. 2012. The genomics of obligate (and nonobligate) biotrophs. Annu. Rev. Phytopathol. 50:91–109

Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., van Themaat, E. V. L., Brown, J. K., Butcher, S. A., Gurr, S. J., and others. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 330:1543–1546

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743–761

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439

Sun, X., Wang, W., Wang, K., Yu, X., Liu, J., Zhou, F., Xie, B., and Li, S. 2013. Sterol C-22 desaturase ERG5 mediates the sensitivity to antifungal azoles in *Neurospora crassa* and *Fusarium verticillioides*. Front. Microbiol. 4:127

Urquhart, A. S., Hu, J., Chooi, Y.-H., and Idnurm, A. 2019. The fungal gene cluster for biosynthesis of the antibacterial agent viriditoxin. Fungal Biol. Biotechnol. 6:9

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J. C., and Paterson, A. H. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40:e49–e49

Wang, Y., Wang, X., Tang, H., Tan, X., Ficklin, S. P., Feltus, F. A., and Paterson, A. H. 2011. Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms S.R. Proulx, ed. PLoS ONE. 6:e28150

Winter, D. J., Ganley, A. R., Young, C. A., Liachko, I., Schardl, C. L., Dupont, P.-Y., Berry, D., Ram, A., Scott, B., and Cox, M. P. 2018. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. PLoS Genet. 14:e1007467

Wu, Y., Ma, X., Pan, Z., Kale, S. D., Song, Y., King, H., Zhang, Q., Presley, C., Deng, X., Wei, C.-I., and Xiao, S. 2018. Comparative genome analyses reveal sequence features reflecting distinct modes of host-adaptation between dicot and monocot powdery mildew. BMC Genomics. 19:705

Xue, C., Park, G., Choi, W., Zheng, L., Dean, R. A., and Xu, J.-R. 2002. Two novel fungal virulence genes specifically expressed in appressoria of the rice blast fungus. Plant Cell. 14:2107–2119

Yadav, V., Yang, F., Reza, Md. H., Liu, S., Valent, B., Sanyal, K., and Naqvi, N. I. 2019. Cellular dynamics and genomic identity of centromeres in cereal blast fungus A. Idnurm, ed. mBio. 10:e01581-19

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J. Integr. Biol. 16:284–287

Zaccaron, A. Z., Chen, L.-H., Samaras, A., and Stergiopoulos, I. 2022. A chromosome-scale genome assembly of the tomato pathogen *Cladosporium fulvum* reveals a compartmentalized genome architecture and the presence of a dispensable chromosome. Microb. Genomics. 8:000819

Zaccaron, A. Z., and Stergiopoulos, I. 2021. Characterization of the mitochondrial genomes of three powdery mildew pathogens reveals remarkable variation in size and nucleotide composition. Microb. Genomics. 7

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46:W95–W101

Zhang, Z. 2022. KaKs_calculator 3.0: calculating selective pressure on coding and non-coding sequences. Genomics Proteomics Bioinformatics.

5.7 Supplementary materials

5.7.1 Supplementary results

S1.1 Obtaining a chromosome-level genome assembly for EnFRAME01. The genome of *E. necator* isolate EnFRAME01 was sequenced with a PacBio Sequel I platform using two SMRT cells, which produced a total of 991,705 reads with an average length of 16 kb and an estimated coverage of 127x. The PacBio reads were assembled with Canu (Koren et al. 2017) into a preliminary genome assembly of 83.0 Mb containing 71 contigs with an L50 of 7. Two rounds of polishing using 40 M paired-end Illumina reads (75x coverage) performed 1,170 and 35 corrections in the assembly, respectively, indicating that further polishing would not have provided any major further improvements. A total of 135 M Illumina paired-end reads generated based on the chromatin conformation capture technique Hi-C (Lieberman-Aiden et al. 2009) were used to merge the polished contigs into predicted chromosomes using Proximo (Phase Genomics), followed by manual adjustments using Juicebox v1.11.08 (Durand et al. 2016). The resulting and final genome assembly of EnFRAME01 contained 81.1 Mb organized into 34 scaffolds with an L50 of 5 and an N50 of 7.98 Mb (Fig 5.1 and Table 5.S1). The size of the genome assembly obtained for EnFRAME01 is significantly smaller than that of C-strain, which has estimated size of 126±18 Mb (Jones et al. 2014a). However, k-mer counting of the Illumina reads predicted a genome size of 82.7 Mb, which is consistent with the genome assembly obtained in this study.

Of the 34 scaffolds, 11 represented full chromosomes, 22 remained unplaced, and one corresponded to the mitochondrial genome of the fungus, which was 99.7% similar to the mitochondrial genome of C-strain (Zaccaron et al. 2021) (Fig 5.S2 and Table 5.1). To evaluate the integrity of the EnFRAME01 genome assembly, a BUSCO analysis in genome mode (Simão et al. 2015) was next conducted, which indicated 98.2% completeness with only 0.3% duplication and 0.8% fragmentation (Table 5.S1). Moreover, Illumina and PacBio reads were mapped to the assembly to identify collapsed regions. A total of 99.29% of the trimmed Illumina reads mapped to the assembly, and 97.36% mapped in proper pair. From the PacBio reads, 93.19% mapped to the assembly. Coverage analysis revealed only five putative collapsed regions, which were located in repeat-rich regions (>88% repetitive) of five different chromosomes and totaling 450 kb in size (Fig 5.S3). Two of the collapsed regions co-localized with regions containing clusters of 5S and 18S-5.8S-28S rDNA, and the three others were located at or physically close to the predicted centromeres.

Next to the absence of major assembly gaps, all chromosomes were also putatively assembled telomere-to-telomere and had 22 to 31 copies of the canonical telomeric repeat 5´-TTAGGG-3´ at both their ends. Centromeric regions were also predicted in all chromosomes and based on their positioning, chromosomes could be considered as metacentric, except for Chr8 which was submetacentric and Chr10 which was subtelocentric (Levan et al. 1964) (Fig 5.S1). Moreover, centromeric regions corresponded to large segments of the chromosomes that accounted for 12.8 Mb (15.8%) of the genome and ranged in size from 0.85 Mb in Chr2 to 2.0 Mb in Chr4 (Table 5.1). Similar-sized centromeres were reported for the wheat powdery mildew (PM) *B. graminis* f.sp. *tritici* (Müller et al. 2019) but they are poorly conserved between the two species and both species have overall low synteny (Fig 5.S4). Instead, long syntenic regions were present between Chr1, Chr2, and Chr5 of *E. necator,* and chr-05, chr-08, and chr-01 of *B. graminis* f.sp. *tritici*, respectively (Fig 5.S4). Collectively, these results indicate that centromeric regions are poorly conserved between *E. necator* and *B. graminis* f.sp. *tritici* and exhibit extensive inter-chromosomal rearrangements. Similar results were obtained by aligning the genome of *E. necator* with those of the cereal PM pathogens *B. graminis* f.sp. *triticale* (Müller et al. 2021) and *B. graminis* f.sp. *hordei* (Frantzeskakis et al. 2018).

S1.2 Genes encoding proteases in the genome of EnFRAME01. The genome of EnFRAME01 contained a total of 174 genes encoding proteases, which is at the lower end compared to other fungal genomes (Iqbal et al. 2018). The proteases present in the genome of EnFRAME01 could be further classified into serine (*n*=51), cysteine (*n*=50), metallo (*n*=47), threonine (*n*=17), aspartic (*n*=7), and inhibitor proteases (*n*=2) (Table 5.S4). The protease encoded by the gene *HI914_05529* was also classified as a CAZyme, which was a serine protease predicted to belong to the CE1 acetylxylan esterase CAZyme family. A total of 27 proteases were predicted to be secreted. Most of the secreted proteases (*n*=19) were serine proteases, followed by aspartic (*n*=5), metallo (*n*=2), and cysteine (*n*=1) proteases (Table 5.S4).

S1.3 Secondary metabolism in EnFRAME01. The genome of EnFRAME01 contained only eight genes encoding key enzymes required for the biosynthesis of secondary metabolites, including one type I polyketide synthase (PKS), one type III PKS, three terpene synthases (TSs), one non-ribosomal peptide synthetase (NRPS), and two NRPS-like enzymes (Table 5.S5). The single Type I PKS in *E. necator* (*HI914_06546*) had the typical domain architecture of non-reducing PKSs, and was composed of a starter unit ACP transacylase domain at its N-terminus, followed by a beta-ketoacyl synthase domain, an acyl transferase domain, a polyketide synthase dehydratase domain, and a phosphor-pantetheine attachment site at its C-terminus. Of the three TS-encoding genes in EnFRAME01, one (i.e. *HI914_04683*) encoded a squalene synthase that is homologous the to the *S. cerevisiae* ERG9 (47.1% aa product identity; e-value= 2.88E-139) that catalyzes two molecules of farnesyl pyrophosphate to form squalene, one of the main steps in the sterol biosynthesis pathway (Chambon et al. 1990). The other two TSs encoded by *HI914_04296* and *HI914_01229*, respectively were homologous to Al-2 (P37295; 48.8% aa product identify; e-value= 0) and Al-3 (P24322; 69.5% aa product identity; e-value= 3.46e-165) encoded by the *albino-2* and *albino-3* genes from *Neurospora crassa*, respectively. Both *albino-2* and *albino-3* genes encode key enzymes for carotenoid biosynthesis, and have been named after the albino phenotype of their mutant alleles in *N. crassa*, caused by absence of carotenoid pigments (Schmidhauser et al. 1994; Sandmann et al. 1993). The NRPS-encoding gene *HI914_04002* present in EnFRAME01 was homologous to the *Fusarium graminearum NPS2* gene (I1RN14; 34.3% aa product identity; e-value= 0) that encodes an enzyme involved in the biosynthesis of intracellular siderophore ferricrocin (Tobiasen et al. 2007). Finally, of the two NRPS-like genes present in the genome of EnFRAME01, one (*HI914_00236*) was homologous to the *S. cerevisiae Lys2* (P07702; 48.2% aa product identity; e-value= 0) that encodes an L-2-aminoadipate reductase involved in the biosynthesis of lysine from L-alpha-aminoadipate (Ehmann et al. 1999). The product of the second gene (*HI914_04388*) had low homology (aa identity< 30% product aa identity) with other characterized proteins, but the putative pathways associated with them is unknown.

S1.4 Genes encoding cytochrome P450s in the genome of EnFRAME01. A total of 11 genes encoding cytochrome P450s were identified in the genome of EnFRAME01 (Table 5.S6 and Fig 5.S5), based on a search for the signature domain of cytochrome P450 (PF00067) among its proteins. This is again on the lower end as compared to other fungal genomes (Durairaj et al. 2016). Among the genes encoding cytochrome P450s in EnFRAME01 was *CYP51* (*HI914_01650*), which encodes for the sterol 14-alpha-demethylase enzyme in the sterol biosynthesis pathway that is the target of sterol demethylation inhibitor (DMI) fungicides. Increase in copy number of *CYP51* in *E. necator* has been associated with increase in tolerance to DMI fungicides (Jones et al. 2014a). However, *CYP51* is present as a single-copy gene in the genome of EnFRAME01. The rest of the genes encoding cytochrome P450s present in EnFRAME01 were members of the *CYP52* (*n*=1), *CYP53* (*n*=1), *CYP544* (*n*=1), and *CYP617* (*n*=1) families, and six could not be classified because they lacked homology (aa product identity< 45%) to the curated fungal cytochrome P450s (https://drnelson.uthsc.edu/P450seqs.dbs.html). Notably, the *E. necator CYP52* (*HI914_06503*) is homologous to the *g430* from *Fusarium* sp. strain RK97-94 (A0A6S6AA17; 57.1% aa product identity; e-value= 0) whose product is involved in the biosynthesis of 1233A (Kato et al. 2020), an inhibitor of HMG-CoA synthase in the mevalonate pathway (Omura et al. 1987) with antibacterial and antifungal activities (Tomoda et al. 1988) (Kato et al. 2020). Also, the *E. necator CYP53* (*HI914_00178*) is homologous to the *bphA* (P17549; 56.3% aa product identity; e-value= 0) from *Aspergillus niger*, whose product is a benzoate 4-monooxygenase that converts benzoic acids, natural phenolic defense compounds produced by plants (Widhalm and Dudareva 2015), into the less toxic *p*-hydroxybenzoic acid (Lubbers et al. 2019).

S1.5 Genes encoding transporters in the genome of EnFRAME01. In order to identify the arsenal of transporters present in EnFRAME01, the predicted proteins of EnFRAME01 were queried with BLASTp against the Transporter Classification Database (TCDB) (e-value< 1E-10). This revealed a total of 1,238 genes (16.5% of all genes) encoding putative transporters from 258 distinct families in the genome of EnFRAME01 (Table 5.S7). Two families stood out for their high frequency, i.e. the eukaryotic nuclear pore complex (E-NPC) family (*n*=125) and the major facilitator superfamily (MFS) (*n*=64). NPCs are large proteinaceous structures that mediate bidirectional nucleocytoplasmic transport (Lin and Hoelz 2019). MFS transporters, on the other hand, are involved in nutrient uptake, multidrug resistance, and metabolite extrusion (Yan 2015), which makes them of particular importance for fungal pathogens. Most MFS transporters in EnFRAME01 are from three subfamilies, the sugar porter family (*n*=19), the Drug:H+ Antiporter 1 (DHA1) family (*n*=16), and the Drug:H+ Antiporter 2 (DHA2) family (*n*=9). The DHA1 and DHA2 families of MFS transporters are typically drug-efflux pumps associated with the extrusion of toxic compounds out of the cell. Another category of drug-efflux pumps commonly associated with antimicrobial resistance is the ATP-binding cassette (ABC) transporters. The genome of EnFRAME01 contained a total of 21 genes encoding ABC transporters from seven subfamilies, i.e. ABCB (*n*=5), ABCF (*n*=5), ABCC (*n*=4), ABCG (*n*=3), ABCD (*n*=2), ABCE (*n*=1), and ABCI (*n*=1) (Fig 5.S6). All 21 ABC transporters contained at least one nucleotide binding domain (NBD), which is characteristic of this family. The seven ABC transporters from subfamilies ABCF, ABCE, and ABCI lacked transmembrane domains (TMDs), and are likely involved in functions not related to transmembrane transport. All the remaining 14 ABC transporters from families ABCB, ABCC, ABCD, and ABCG, contained transmembrane domains. Of these, eight were half-length transporters, of which six had the domain topology TMD-NBD and two had the reverse topology NBD-TMD. The other six ABC transporters were full-length transporters containing two TMD domains and two NBD domains each. Of these, five had the domain topology $(TMD-NBD)_2$ and one had the reverse topology $(NBD-TMD)_2$.

S1.6 Genes encoding carbohydrate-active enzymes (CAZymes) in the genome of EnFRAME01. The genome of EnFRAME01 contained a total of 160 genes predicted to encode CAZymes, representing 171

CAZyme modules from five major classes, i.e., glycoside hydrolases (GH), auxiliary activity (AA), glycosyl transferases (GT), carbohydrate esterases (CE), and carbohydrate-binding modules (CBM), whereas no CAZyme from the sixth major class, polysaccharide lyases (PL), was present in EnFRAME01 (Table 5.S8). The most abundant CAZyme module was GT2 (*n*=11), followed by GH16 (*n*=10), GH18 (*n*=9), and GH76 (*n*=7). There were 11 CAZymes containing a domain architecture composed of more than one CAZyme module. Most of the CAZymes (*n*=7) contained a chitin-binding module (CBM18) combined either with a GH18 chitinase domain (*n*=3), an AA5 copper radical oxidase domain (*n*=2), a GH16 β-1,3-glucanase domain (*n*=1), or a CE4 acetylxylan esterase domain (*n*=1). The other four multi-module CAZymes included a putative glycoamylase with a starch-binding domain (GH15+CMB20), a putative 1,4-alpha-glucan-branching enzyme with a glycogen-binding domain (GH13+CBM48), a putative β-1,3-glucanosyltransglycosylase with a β-1,3-glucan-binding domain (GH72+CBM43), and a gene similar to N-acetylgalactosamine deacetylase with an α-1,4-N-acetylgalactosamine-binding domain (CE18+CBM87). A total of 60 CAZymes were predicted to be secreted, which included seven from the AA class, seven from the CE class, 43 from the GH class, and three from the GT class. Among the 60 secreted CAZymes, 12 were predicted to act on plant cell walls, including hemicellulose (*n*=9), cellulose (*n*=2), and pectin ((*n*=1) (Hage and Rosso 2021) (Table 5.S8).

S1.7 Genes encoding candidate secreted effector proteins (CSEPs) in the genome of EnFRAME01.

Typical criteria used to classify proteins as CSEPs in PMs include the presence of a signal peptide (SP), no transmembrane domains (TMDs), and no homology to proteins outside the Erysiphales (Spanu et al. 2010). However, these criteria might underestimate the true number of CSEPs in PMs. For example, the barley PM CSEP BEC1019 (AHZ59730.1) is a metalloprotease (Zhang et al. 2019) with homologs in several non-PM pathogens, including *Botrytis spp*. and *Fusarium spp*. To overcome such drawbacks in the classification of secreted proteins (SPs) as CSEPs in *E. necator*, we used three lines of evidence, i.e. i) the lack of homology to proteins outside Erysiphales, ii) their classification as candidate effectors based on the EffectorP software (Sperschneider et al. 2016), and iii) their domain and amino acid composition (i.e. having a SP and being shorter than 250 aa, being cysteine-rich, having no TMDs and no GPI-anchor). Among the 527 SPs predicted in EnFRAME01 (Table 5.S9), 152 could be classified as CSEPs based on lack of homology to proteins from non-Eurotiomycetes, 151 based on EffectorP classification, and 120 based on their domain composition. The union of these three sets that satisfied all of our parameters for classifying SPs as CSEPs, yielded a total of 234 CSEPs in *E. necator* (Fig 5.S20 and Table 5.S10). As expected, this number is significantly larger than the 150 CSEPs reported for the *E. necator* isolate C-strain (Jones et al. 2014a), but still considerably lower than the 844 CSEPs reported for the cereal PM *B. graminis* f.sp. *tritici* (Müller et al. 2019). Also, all 234 CSEP-encoding genes could be mapped to the genome of the five *E. necator* isolates sequenced before (Jones et al. 2014a), although five of them were pseudogenised in one or more *E. necator* isolates by the introduction of a premature stop codon in their sequence (Table 5.S10).

Homology searches in the NCBI nr database showed that of the 234 CSEPs predicted in EnFRAME01, 49 were species-specific and 183 had homologs in other fungi, including PM fungi (*n*=183) and/or non-PM fungi (*n*=86) (Fig 5.S7). The majority of CSEPs shared with other PMs had homologs only in dicot-infecting PMs (*n*=77) or both in dicot and monocot-infecting PMs (*n*=105), as compared to having homologs only in monocot-infecting PMs (*n*=1) (Table 5.S10). Twenty-four of the CSEPs also had a homolog in the Pathogen Host Interaction (PHI) database (e-value< 1E-5), of which 17 had a match to BEC1019 from *B. graminis*, GAS1 from *Magnaporthe oryzae*, Mras7 from *Metarhizium robertsii*, bim1 from *Cryptococcus neoformans*, cypB from *Beauveria bassiana*, and PBC1 from *Pyrenopeziza brassicae*, for which mutants exhibited reduced pathogenicity (Table 5.S10).

Further sequence analysis of the EnFRAME01 CSEPs showed that of the 234 CSEPs, 38 had a conserved domain in their sequence, including a microbial-type ribonuclease (cl00212) domain (Table 5.S10). Ribonuclease-like effectors belong to a large family of catalytically inactive RNAses in cereal PMs that are generally referred to as RNase-Like proteins associated with haustoria (RALPHs) (Spanu 2017; Pennington et al. 2019; Pedersen et al. 2012). A genome-wide search for RALPHs in EnFRAME01 identified a total of 38 genes encoding RALPH-like proteins. Of the 38 RALPH-like proteins, 24 fulfilled the criteria used to classify secreted proteins as CSEPs and could thus be further considered as RALPH-like CSEPs (Table 5.S14). A phylogenetic analysis showed that the 38 RALPH-like proteins could be organized into two clades (Fig 5.S9). One of the clades consisted of 22 long RALPH-like proteins of 405-636 aa in size and their encoding genes mostly clustered towards the end of Chr8 (*n*=13) and Chr11 (*n*=4). The other clade contained 16 short RALPH-like proteins of 172-219 aa in size and their encoding genes mostly clustered at an 85 kb region of Chr7 (*n*=10) (Fig 5.S9). Other domains most commonly found in the *E. necator* CSEPs were an Egh16-like virulence factor domain (PF11327) (*n*=4) and a P24 protein family domain (PF01105) (*n*=4). Phylogenetic analysis of the 11 Egh16-like proteins showed that they too formed two major clades, with one clade containing six Egh16-like proteins that were highly similar to the gEgh16 of *B. graminis* f.sp. *hordei* (Justesen et al. 1996) and whose encoding genes mostly clustered in Chr4 (*n*=5). The other clade contained five Egh16-like proteins that were highly similar to the Egh16H1 of *B. graminis* f.sp. *hordei* (Grell et al. 2003) and their encoding genes clustered in Chr2 (Fig 5.S10 and Table 5.S15).

It is often the case that genes encoding fungal effector proteins are located in dynamic parts of the chromosomes such as subtelomeric regions (Zaccaron et al. 2022; Gan et al. 2020). Analysis of the localization of the 234 CSEP-encoding genes on the 11 chromosomes of *E. necator* showed that although some genes encoding RALPH-like CSEPs (*n*=6) were located within 500 kb of the end of chromosomes Chr8, Chr10, or Chr11 (Fig 5.S9), there was no enrichment of genes encoding CSEPs in sub-telomeric regions (i.e. no CSEP-encoding gene within 50 kb of chromosome ends), as has been observed in other fungi (Zaccaron et al. 2022; Gan et al. 2020), and all were totally absent from centromeric regions (Fig 5.1).

S1.8 "Missing ascomycete pathogen core genes" (MACGs) in the genome of EnFRAME01. To understand the molecular basis of obligate biotrophy in *E. necator*, the 181 distinct core genes that were not present in the genome of EnFRAME01 but conserved in *Saccharomyces cerevisiae* and other non-obligate fungi were further analyzed (Table 5.S11). These 181 core missing genes included 95 of the 99 MACGs that were previously reported as being absent in PMs, but present in other ascomycetes (Table 5.S12). Of the remaining four MACGs, two are likely pseudogenes in *S. cerevisiae* S288c and two, namely *HEM4* (*HI914_06207*) encoding a uroporphyrinogen-III synthase that is involved in methionine metabolism and (siro-) heme biosynthesis and *ADH4* (*HI914_02737*) encoding an alcohol dehydrogenase 4 that is involved in alcohol metabolism, were present in EnFRAME01. Of these two MACGs, *HEM4* was also present in the pea PM *Erysiphe pisi* (Sharma et al. 2019), indicating that its presence in EnFRAME01 is not unique.

In accordance with previous reports that PMs lack biosynthesis of glycerol from glycolytic intermediates and biosynthesis of thiamine (Spanu 2012; Spanu et al. 2010), we observed the absence of a gene encoding glycerol 3-phosphatase (EC:3.1.3.21) that is required for glycerol biosynthesis in *S. cerevisiae* (Påhlman et al. 2001), and the absence of six genes involved in thiamine biosynthesis, namely hydroxymethylpyrimidine/phosphomethylpyrimidine kinase *THI20* (EC:2.7.1.49), 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase *THI11*, repressible alkaline phosphatase *PHO8* (EC:3.1.3.1), thiamine biosynthetic bifunctional enzyme *THI6* (EC:2.5.1.3), thiamine thiazole synthase *THI4* (EC:2.4.2.60), and low molecular weight phosphotyrosine protein phosphatase *LTP1* (EC:3.1.3.48) (Table 5.S13). In terms of nitrogen metabolism, EnFRAME01 lacked transporters from the nitrate/nitrite porter (NNP) family (TCDB family ID 2.A.1.8) as well as a nitrite reductase (EC:1.7.1.4), a nitrate reductase (EC:1.7.1.3), and a NADP-specific glutamate dehydrogenase (EC:1.4.1.4) (Table 5.S13). A reduced arsenal of genes involved in amino acid biosynthesis was observed too, as EnFRAME01 lacked the *CAR1* (EC:3.5.3.1)*, GDH1* (EC:1.4.1.4)*,* and *GDH3* (EC:1.4.1.4) genes involved in arginine biosynthesis as well as the *ARO8* gene that encodes a general aromatic amino acid transaminase (EC:2.6.1.57 2.6.1.39 2.6.1.27 2.6.1.5) involved in phenylalanine, tyrosine, tryptophan and lysine biosynthesis (Table 5.S13). Significant losses were also observed in genes involved in metabolism of several amino acids, including tryptophan, arginine, proline, tyrosine, alanine, aspartate, glutamate, phenylalanine, glycine, serine, threonine, cysteine, and methionine (Table 5.S13). Nucleotide metabolism was also impacted by gene losses, including of genes encoding guanine deaminase (EC:3.5.4.3), cytosine deaminase (EC:3.5.4.1), cytidine deaminase (EC:3.5.4.5), and uridine nucleosidase (EC:3.2.2.3). A considerable reduction in genes involved in sulfur metabolism was also observed. Specifically, the *MET3*, *MET10*, *MET14*, and *MET16* genes that are part of the sulfate assimilation pathway and collectively synthesize sulfite from sulfate were absent in *E. necator*. Lastly, a reduction in genes for sterol biosynthesis was observed as well. Specifically, *E. necator* lacked genes encoding sterol O-acyltransferase (EC:2.3.1.26), and the genes *ERG4* (EC:1.3.1.71) and *ERG5* (EC:1.14.19.41) that are required for ergosterol biosynthesis in yeast and other fungi (Hu et al. 2017) (Table 5.S13).

S1.9 Analysis of repetitive DNA. A *de novo* annotation of repetitive DNA revealed that 62.7% (50.8 Mb) of the EnFRAME01 genome was composed of transposable elements (TEs), which is in agreement with the previous estimate of 62.9±3% for C-strain (Jones et al. 2014a). The abundant presence of TEs in the genome of *E. necator* suggests that genome defense mechanisms against TE activity are inactive in this species. Indeed, a tBLASTn search (e-value < 1E-3) for homologs of the *RID1* (XP_011392925.1), *Dim-2* (XP_959891.1), *MASC1* (AAC49849.1), and *MASC2* (AAC03766.1) genes that mediate RIP and MIP (Methylation Induced Premeiotically) in fungi (Gladyshev 2017) showed that these are absent from the genome of *E. necator,* indicating that RIP and MIP have likely no effect on TE activity or diversification of their nucleotide sequences. Even so, an examination of the nucleotide divergence of TEs in EnFRAME01 exhibited bimodal distribution with two peaks, at approximately 5% and 21% of divergence, respectively (Fig 5.2B). The first peak at ~5% divergence was dominated by LTRs (1.2 Mb, 35.5%), RC (1.0 Mb, 30.7%), and unknown elements (0.8 Mb, 30.0%); whereas the second peak at ~21% divergence was dominated by LTRs (0.8 Mb, 46.2%) and the non-LTR long interspersed nuclear elements (LINEs; 0.8 Mb, 43.7%) (Fig 5.2B). This bimodal distribution contrasts the pattern reported previously in *B. graminis* f.sp. *hordei*, for which a single recent burst of TE proliferation was proposed (Frantzeskakis et al. 2018). However, our analysis of the genomic data of different *B. graminis* formae speciales (Müller et al. 2021, 2019) also revealed bimodal distributions of TE divergence in these genomes.

S1.10 Distribution of genes and repeats on the chromosomes of EnFRAME01. An examination of the distribution of repeats and protein coding genes on the 11 chromosomes of EnFRAME01 showed large differences in gene density among the chromosomes (Fig 5.1). Specifically, gene density was highest in Chr2 (107 genes per Mb) and lowest in Chr8 (63 genes per Mb), with an overall average of 88 genes per Mb for the entire genome. Likewise, the density of repetitive DNA also differed among the chromosomes and was highest in Chr8 (70.0%) and lowest in Chr2 (55.6%). Although no major differences in TE content within the flanking regions of genes encoding CAZymes, proteases, CSEPs, and non-CSEP secreted proteins was observed, intergenic regions of CSEP genes were significantly longer and richer in repetitive DNA (Fig 5.3A to 3B and Table 5.S17). For instance, while intergenic regions of CSEP-coding genes averaged 11.4 kb in size and 48.6% in repeat content, intergenic regions of BUSCO genes averaged 6.1 kb in size and 31.7% in repeat content. Moreover, intergenic regions downstream of CSEP-encoding genes had the highest content of LTR retrotransposons and RC DNA transposons as compared to genes from the above four functional categories (Fig 5.S13). This indicates that TEs were essentially evenly distributed in gene-flanking regions, but small differences in TE content could be observed in the vicinity of CSEP-encoding genes. Overall, these results indicate that the genome of *E. necator* exhibits local instead of large-scale compartmentalization.

S1.11 Gene duplications. The absence of RIP in *E. necator* suggests that, similar to TEs, duplicated genes could also be retained at a higher rate within the genome of the fungus, as compared to other ascomycete fungi. A large number of 941 gene duplicates were identified in the genome of *E. necator* by self-BLASTp search, while an enrichment analysis further showed that different functional gene categories exhibiting different rates and modes of gene duplication (Fig 5.4A and Table 5.S18). This was most evident in CSEPs, which experienced higher rates of gene duplications and, in contrast to other functional gene categories, almost equal levels of dispersed, proximal, and tandem gene duplications (Fig 5.4A).

An examination of the nucleotide divergence among copies of genes with different duplication modes showed that dispersed gene duplicates (DGDs) had significantly higher levels of nucleotide divergence (median=65.6%), as compared to proximal gene duplicates (PGDs; median=90.1%) (*p*-value=1.55E-13) and tandem gene duplicates (TGDs; median=93.4%) (*p*-value=9.26E-14), which did not statistically differ from each other (*p*-value=0.46) (Fig 5.4C). This indicates that gene copies physically close in the genome of *E. necator* are more conserved and thus more likely to contribute to genetic redundancy than dispersed copies that are likely to contribute more to functional diversification. Alternatively, by considering nucleotide sequence similarity as a proxy for the age of gene duplication events (Wang et al. 2011; Blanc and Wolfe 2004), these results suggest that copies of recently duplicated genes of EnFRAME01 were more likely to be physically close in the genome than older duplications. To assess the age of the duplication events, we examined the rate of synonymous substitutions per synonymous site (K_S) among the gene duplicates in the different categories of gene duplication modes. We also examined the rates of nonsynonymous (K_A) to synonymous nucleotide substitutions (K_A/K_S) in order to evaluate differences in selection pressure among them (Lynch and Conery 2003). The distribution of K_S and K_A/K_S values differed among genes with different duplication modes, with TGDs and PGDs experiencing significantly lower median K_S values and higher median K_A/K_S values than DGDs (Fig 5.4D and Fig 5.S16A), suggesting that they are more recent duplicates and under more relaxed selection pressure. However, DGDs showed a distinctive bimodal distribution of K_S values with one peak at K_S≈3-3.5 and a second peak at K_S≈0.05, suggesting that a substantial number of DGDs were younger in age and were generated by recent gene duplication events (Fig 5.4C and Fig 5.S16A). A similar observation was made when plotting the frequency of the K_S estimates obtained for each gene duplication pair, in which case the K_S distribution did not follow the expected exponential decrease in frequency with the age of the duplicates (Lynch and Conery 2003) but instead, a secondary peak was observed at the tail of the distribution at K_S≈3 (Fig 5.S16B). This suggested that at least two bursts of large gene duplication events took place in the evolutionary history of *E. necator* (Blanc and Wolfe 2004; Tiley et al. 2018; Vanneste et al. 2013). Finally, when examining nucleotide divergence among gene copies in each functional gene category, the CSEP-encoding genes again differed markedly from the other gene functional categories as they exhibited significantly lower median K_S values and higher median K_A/K_S values (Fig 5.S16C), suggesting that CSEP gene duplicates were mostly younger in age and evolved faster as compared to other gene categories. Taken together, these results suggest that gene duplication contributes to genome plasticity in *E. necator* but differentially affects different functional gene categories, including their rate of duplication, mode of evolution, and organization of their paralogs in the genome.

S1.12 Genes and regions with copy number variations (CNVs) among isolates of *E. necator*. In order to identify regions with CNVs, whole-genome sequencing reads of *E. necator* isolates Branching, C-strain, e1-101, Lodi, and Ranch9 (Jones et al. 2014a) were mapped to the genome assembly of EnFRAME01, and CNV regions were subsequently identified based on differences in coverage depth. A total of 1,760 CNV regions were identified, covering 5.5 Mb (6.9%) of the EnFRAME01 genome (Fig 5.S17 and Table 5.S20). Regions with CNV were fairly dispersed throughout the 11 chromosomes of EnFRAME01 but mostly overlapped with predicted TEs. Interestingly, while 7.5% of the total deleted regions overlapped with RC elements, only 1.2% of the duplicated regions were associated with RCs (Fig 5.S17C), indicating a differential loss and acquisition by duplication of RC elements among the strains.

When considering genes only with CNVs, then these were most likely to be affected by duplications rather than deletions. Indeed, a total of 122 CNV genes were identified among the five isolates analyzed (Table 5.S21), which is slightly less than the 135 CNV genes reported previously among the same five isolates (Jones et al. 2014a). However, among the 122 CNV genes identified, 66 were absent in the previous reference genome annotation of *E. necator* isolate C-strain. Moreover, from the 135 CNV genes previously reported using the genome of *E. necator* C-strain as reference, 75 were absent in the genome annotation of EnFRAME01, with the majority of them (*n*=60) mapping to highly repetitive regions (>90%). This suggests that the previous CNV analysis (Jones et al. 2014a) missed several CNV genes and reported many transposon-like CNV genes. Of the 122 CNV genes, only 37 had a neighboring CNV gene, indicating that CNV regions typically affected single genes instead of groups of genes. From the 122 CNV genes identified, 14 formed two large clusters of consecutive CNV genes. The largest of these clusters contained eight consecutive genes located within a 22.2 kb region of Chr2, all of which had two copies in isolate Ranch9. Among them, *HI914_01688* is predicted to encode a diacylglycerol O-acyltransferase (DGAT) that catalyzes the last step in triacylglycerol (TGA) synthesis from diacylglycerol and fatty acyl-CoA as substrates, *HI914_01692* encodes a putative zinc finger C2H2 transcription factor homologous to crzA that plays a role in proper chitin and glycan incorporation into the cell wall, and *HI914_01695* is predicted to encode an alternative oxidase (AOX) that participates in the electron transfer chain (ETC) by providing an alternative route for electrons. The other cluster of CNV genes contained six consecutive genes located within a 11.4 kb region of Chr10 and had five copies in isolate C-strain, four copies in isolates Ranch9 and e1-101, and three copies in isolate Lodi (Table 5.S21). Among these six genes, *HI914_07035* showed homology to the yeast *RRP8* gene (P38961) that encodes an rDNA methyltransferase responsible for the N1-methylation of adenine 645 of 25S rRNA, *HI914_07036* showed homology to the yeast *RPB5* (P20434) gene that encodes the subunit RPB5 of the DNA-directed RNA polymerases I, II, and III, *HI914_07038* encoded a putative mitochondrial GTPase, and *HI914_07039* encoded a secreted protein containing an ankyrin repeat domain (PF12796). The rest 48 CNV genes formed 19 smaller clusters containing two to four genes, whereas 60 CNV genes were not clustered, i.e. there were no genes with CNV within the next 10 genes up- and downstream. This indicates that duplicated segments in the genome of EnFRAME01 are typically short and affect only one or a few genes at a time, instead of long segmental duplications that could affect several genes at once.

S1.13 A putative PM-specific carboxylesterase varying in copy number. An inspection of the 122 genes with CNVs among the isolates of *E. necator* showed that gene *HI914_00624* was the one exhibiting the most dynamic changes in copy numbers. Predicted copy numbers of this gene ranged from one in isolate EnFRAME01, two in isolates Ranch9 and e1-101, 13 in isolate Branching, 20 in isolate C-strain, and 31 in isolate Lodi. Functional annotations of *HI914_00624* using InterProScan showed that it encodes a predicted secreted carboxylesterase (CE). CEs (E.C. 3.1.1.1) form a large multimember family of ubiquitous enzymes that play crucial roles in endo- and xenobiotic metabolism (Putterill et al. 2003; Wang et al. 2018; Bornscheuer 2002). Not surprisingly, a search for homologues of HI914_00624 in the NCBI nr database showed that these were abundantly present both within PM and non-PM fungi (Fig 5.5B). However, the phylogenetic tree constructed using the top 400 best blastp hits in the NCBI nr database (2022-06-20), representing at least 195 distinct fungal species, showed that HI914_00624 belonged to a distinct phylogenetic clade that included 22 CEs from only PM species (Fig 5.5C). The monocot-infecting PMs *B. graminis* f.sp. *hordei*, *B. graminis* f.sp. *tritici, and B. graminis* f.sp. *triticale* each had only a single homolog of HI914_00624 from this clade. In contrast, the dicot-infecting PMs *Erysiphe neolycopersici, Erysiphe pulchra, Golovinomyces cichoracearum,* and *Golovinomyces magnicellulatus,* had multiple homologs of HI914_00624 (Table 5.S23). This indicates that duplication of *HI914_00624* occurs more frequently in dicot-infecting than monocot-infecting PMs. To further explore the characteristics of this novel PM-specific clade of CEs, we searched for the Ser-Asp/Glu-His catalytic triad that is indispensable to the function of CEs as well as for the consensus Gly-x-Ser-x-Gly pentapeptide motif within which the Ser heading the catalytic triad residue is generally fixed in CEs (Sood et al. 2018; Oakeshott et al. 2005). A multiple sequence alignment of the 22 CEs included in the HI914_00624 clade showed that these two motifs were poorly conserved in HI914_00624 and its homologs (Fig 5.S19 and Table 5.S23). Collectively, these results suggest that HI914_00624 is a member of new clade of potentially non-catalytically active CEs (Alam et al. 2002) that experience dynamic CNV in dicot-infecting PMs.

5.7.2 Supplementary Methods

S2.1 Fungal isolate, nucleic acid extraction and sequencing. The sequenced *E. necator* strain EnFRAME01 is sensitive to (Fungicide Resistance Action Committee) group 3, 7, 11, 13, 50, and U6 fungicides. To obtain High-molecular weight (HMW) DNA, EnFRAME01 was propagated on *Vitis vinifera* L., cv. 'Chardonnay' grown in a Percival biocontainment growth chamber (Percival Scientific, Perry, IA) at 22°C with a 16 h photoperiod. Briefly, conidia were harvested from infected grape plants 7 to 10 days after inoculation and again every 5 to 7 days for 30 days. A vacuum (Shop-Vac model 2030100, Williamsport, PA) reduced from 18mm to 8mm ID hose with a Whatman #1 filter paper (Whatman International Ltd., Maidstone, England) placed in-line at the reducer (resulting 11t m/sec flowrate) was used to collect conidia from leaves. The collected conidia were immediately suspended in 0.05% Tween 20 (Sigma-Aldrich, Saint Louis, MO) in nuclease-free water, and passed through a 70 µm cell strainer (Fisher Scientific, Pittsburgh, PA) to remove debris. The conidial suspension was then centrifuged for 5 min at 4696 x g at 4°C, the supernatant was poured off, and the pellet flash-frozen in liquid nitrogen and stored at -80°C. For DNA extraction, the ball milling step was omitted and no more than 100 mg of the conidia were combined with 700 µL of pre-warmed 65°C lysis buffer and 300 µL 65°C 5% (w/v) Sarcosyl as in (Feehan et al. 2017), followed by a 40 min incubation at 65°C. The chloroform:isoamyl alcohol (24:1) DNA extraction step was performed as in (Feehan et al. 2017) but the isopropanol precipitation step was incubated for 10 min at room temperature and the resulting pellet was dissolved in 300 µL pH 8.0 TE buffer by incubating for 20 min at 37°C. Removal of RNA was accomplished by the addition of 1.5 µL RNase A (100 mg/ml, Qiagen, Germantown, MD) to each 300 µL and incubated for 2 hr at 37°C. Phenol:chloroform:isoamyl alcohol (25:24:1) extractions were performed in 5PRIME Phase Lock Heavy Gel tubes (Quantabio, Beverly, MA) and the tubes were centrifuged for 5 min at 12,000 x g at room temperature following the Phase Lock protocol. The DNA was precipitated as in (Feehan et al. 2017) with the addition of three extra -20°C 70% ethanol washes of the DNA pellets followed by air drying at room temperature in a laminar flow hood. The DNA pellets were resuspended in 50 µL 10 mM pH8.0 Tris-HCL and allowed to dissolve for 30 min at room temperature and stored at 4°C. All DNA transfer steps were done using wide-bore low-retention nuclease-free filter pipet tips. The quality and quantity of DNA was initially assessed using a NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and Qubit 2.0 fluorometer using the dsDNA High Sensitivity kit (Invitrogen/Thermo Fisher Scientific, Waltham, MA) and 20 µL (~100 ng) of DNA was run on a 0.7% agarose gel at 90V for 1 hr.

Total RNA from conidia of the *E. necator* isolates EnFRAME01, BPPQ1B.3, BPPQ1B.5, DDOME-1, DDOME-2 and HO2 was performed. Briefly, for each isolate, grape leaves with actively growing colonies were held above a 6" x 6" sterile glass slide and gently flicked to dislodge and spread the conidia into a semi-confluent monolayer on the slide. Conidia were incubated in a sealed chamber with a water reservoir for 8 hours, after which the conidia were removed with a razor blade and immediately transferred to Eppendorf tubes and snap frozen in liquid nitrogen. Assessment of the samples quality was carried out on Qubit and Agilent 2100 analyzers, and 250-300 bp cDNA libraries were constructed and sequenced on Illumina NovaSeq6000 platform, after which the raw sequencing data were filtered to remove low-quality reads and adapter sequences. A minimum of 157,294,292 clean reads were produced from each RNA sample, all with Q30 values of 95% or higher.

S2.2 Genome assembly. Canu v1.8 (Koren et al. 2017) was used to assemble PacBio reads with parameters genomeSize=90m corOutCoverage=60 minReadLength=5000 minOverlapLength=3000 corMinCoverage=5 corMhapSensitivity=normal correctedErrorRate=0.03. To predict chromosomes based on chromatin conformation capture, Hi-C reads were mapped to the assembly with BWA-MEM v0.7.17 (Li and Durbin 2009) with parameters -5 -S , which allow the mapping of each read end individually. The mapped Hi-C reads were processed with samblaster v0.1.24 (Faust and Hall 2014) to mark PCR duplicates, and with SAMtools v1.9 (Li et al. 2009) to remove reads with one of the ends unmapped and to remove supplementary alignments (SAM flag = 2216). The resulting mapped Hi-C reads were used to generate links with the *matlock* script (https://github.com/phasegenomics/matlock). Links were then sorted by the mapping position of the left end, and then by the mapping position of the right end. The assembled contigs were used to generate an assembly file with the scripts makeAgpFromFasta.py and agp2assembly.py (https://github.com/phasegenomics/juicebox_scripts). Finally, the assembly file and the sorted links were used by the script run-assembly-visualizer.sh from the 3D-DNA package (Dudchenko et al. 2017) to produce a hic file, which was visualized and edited with Juicebox v1.11.08 (Durand et al. 2016). Based on Hi-C interaction frequency visualized with Juicebox, contigs were grouped into putative chromosomes. To evaluate the assembly integrity, the genome assembly completeness was estimated with BUSCO v5.2.1 in genome mode using as reference the fungi_odb10 database (*n*=758 BUSCO genes). Moreover, the WGS Illumina and PacBio reads were mapped to the genome assembly with BWA-MEM v0.7.17-r1188 (Li and Durbin 2009) and minimap2 v2.17-r941 (Li 2018). Read coverage across the chromosomes was calculated with mosdepth v0.3.2 (Pedersen and Quinlan 2018) using a sliding window of 50 kb. Windows with more than 2x the median coverage of the entire genome were considered collapsed regions. Circular representation of the chromosomes was generated with circos v0.69-8 (Krzywinski et al. 2009). Because the resulting genome assembly (81.1 Mb) was considerably smaller than the previous estimate of 126±18 Mb (Jones et al. 2014a), the genome size was estimated based on the WGS Illumina reads. First, reads matching the mitochondrial genome were removed with the script bbduk.sh script from BBmap v38.90 using k-mer value of 29 (k=29). Then, genome size was estimated using the script kmercountexact.sh from BBmap v38.90 with parameters k=29 and peaks=out_peaks.txt.

S2.3 Gene predictions. RNA-seq reads were mapped to the genome assembly with HISAT2 v2.2.0 (Kim et al. 2015) with parameters --max-intronlen 3000 --dta. Transcripts were reconstructed with Trinity v2.9.1 (Grabherr et al. 2011) in genome-guided mode with the parameter --jaccard_clip, and with Stringtie v2.1.1 (Pertea et al. 2015). Assembled transcripts were processed with the script dedupe.sh from BBMap v38-72 to remove identical or fully contained transcripts. Transcripts were mapped to the genome assembly with

GMAP v2019-09-12 (Wu and Watanabe 2005) with parameters --min-intronlength=20 --intronlength=3000 --format=gff3_gene to obtain location of introns to support training of the gene predictor GeneMark-ES as described below.

Initially, gene models were generated to train *ab initio* gene predictors. Gene models were generated with GeMoMa v1.6.3 (Keilwagen et al. 2019) using the mapped RNA-seq reads in BAM format and protein sequences from *Oidium neolycopersici* (Wu et al. 2018). Briefly, GeMoMa maps the protein sequences to a reference genome and uses RNA-seq reads to call gene models with accurate exons and introns. Gene models were filtered with GeMoMa with parameters tpc=1 tie=1, and putative duplicated genes were removed by identifying highly similar proteins with cd-hit v4.7 (Li and Godzik 2006) at the identity threshold of 80%, resulting in 4,279 gene models. These gene models and location of introns identified with GMAP were used to train GeneMark-ES v.4.57 (Lukashin and Borodovsky 1998) with parameters --fungus –training --soft_mask auto. Gene models were converted to ZFF format with the script gff3_to_zff.pl that is incorporated in the SNAP v2013-11-29 software (Korf 2004). The gene models in ZFF format were filtered with the script fathom that is incorporated in SNAP. The scripts forge and hmm-assembler.pl were then used to train SNAP. The same gene models used to train GeneMark were used to train the Augustus v3.2.3 software (Stanke et al. 2006). To reduce computational time, 1,626 models were randomly selected from the total 4,279. An initial training was performed with the script etraining that is incorporated in Augustus. The 1,626 models were randomly split into 1,326 models for parameter optimization and 300 models for benchmarking. After parameter optimization, sensitivity and specificity were equal to 0.838 and 0.823 at the exon level, and 0.700 and 0.667 at the gene level, respectively.

As an additional line of gene evidence, more gene models were generated with GeMoMa, using the mapped RNA-seq reads and protein sequences from *E. necator* C-strain (Jones et al. 2014a). Finally, Maker v2.31.10 (Cantarel et al. 2008) was used to generate gene models for EnFRAME01. To do so, Maker used i) the trained *ab initio* predictors GeneMark-ES, SNAP, and Augustus, ii) the gene models generated with GeMoMa, iii) the assembled transcripts, iv) protein sequences from *Blumeria graminis* f.sp. *hordei* isolate DH14 (GCA_000151065.2), *O. neolycopersici* isolate UMSG2 (GCA_003610855.1), and *Botrytis cinerea* isolate B05.10 (GCF_000143535.2), and v) the location of repetitive regions in GFF format identified with RepeatMasker. Predicted genes entirely contained within repetitive regions were further verified and removed if the encoded proteins are hypothetical or contained conserved domains typically associated with transposable elements (Min and Choi 2019). Completeness of gene prediction was estimated with BUSCO v5.2.1 in proteome mode using as reference the fungi_odb10 database (n = 758 BUSCO genes).

S2.4 Characterization of repeats and of transposable elements (TEs). *De novo* TE libraries were generated for the genomes of *Erysiphe necator* isolate EnFRAME01, *Blumeria graminis* f.sp. *hordei* isolate DH14 (Frantzeskakis et al. 2018), *Blumeria graminis* f.sp. *tritici* isolate v3.16 (Müller et al. 2019), and *B. graminis* f.sp. *triticale* isolate THUN12 (Müller et al. 2021) using RepeatModeler v2.0.2a (Flynn et al. 2020). RepeatModeler was configured to use RECON v1.08 (Bao and Eddy 2002) and RepeatScout v1.0.6 (Price et al. 2005), and was executed with the parameter -LTRStruct to run the LTR structural discovery pipeline using LTR_retriever v2.9.0 (Ou and Jiang 2018) and LTRharvest (Ellinghaus et al. 2008) that is part of the GenomeTools suite v1.6.2. To classify predicted TE families, RepeatModeler used rmblast v2.11.0+ as the search engine against known TE families in the Dfam database release 3.5 (October 2021) (Storer et al. 2021). The custom *de novo* TE libraries were then used to mask the genomes with RepeatMasker 4.1.2-p1 with parameters -xsmall -gff -s -a and using rmblast v2.11.0+ as the search engine. The output alignment files generated by RepeatMasker were parsed with the script parseRM.pl v5.8.2 (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs) with parameters --land 50,1 --parse --fa --nrem to estimate the abundance (bp) of each TE superfamily, to generate TE divergence landscape, and to estimated divergence of each TE family.

To estimate differences of TE divergence across the chromosomes of EnFRAME01, the GFF file produced by RepeatMasker was parsed with the script genomecov from BEDtools v2.29 (Quinlan and Hall 2010) to find genomic regions masked more than one time, i.e., overlapping TE regions. These regions were subsequently removed from the TE annotations with the subtract script from BEDtools. Non -overlapping windows of 50 kb along the 11 chromosomes of EnFRAME01 were generated with the script makewindows from BEDtools. TE families overlapping each window and their abundances were extracted with the script intersect from BEDtools. To estimate overall divergence of TEs within each window, the divergence of TE families overlapping each window were obtained, as previously estimated with the script parseRM.pl. For each TE family *f* overlapping window *w*, the normalized divergence *d(w)* of TEs in the window *w* was estimated as $\sum_{i=0}^{n} \frac{d(f_i)\cdot abun(f_{iw})}{\sum_{k=1}^{n} abun(f_{kw})}$, where *n* is the total number of unique TE families that overlap the window *w*, $d(f_i)$ is the divergence of the TE family f_i across the entire genome, and $abund(f_{iw})$ is the total number of bases masked in *w* by the TE family f_i. In other words, the divergence of the TE families are multiplied by their respective abundances in the window, then divided by the total number of masked bases in the window.

S2.5 Homology-based functional annotation of predicted genes. The predicted protein sequences were queried with BLASTp against the UniProt/SwissProt database (e-value< 1E-10) and conserved PFAM domains were identified with InterProScan v5.32-71.0 (Jones et al. 2014b). The outputs of BLASTp and InterProScan were processed with ANNIE (http://genomeannotation.github.io/annie/) to generate a 3-column table with functional descriptions, which was then used by GAG (https://genomeannotation.github.io/GAG/) to assign functional descriptions and conserved domains to the predicted genes, and to generate a feature table file (.tbl). The feature table file was parsed with the script tbl2asn (https://ftp.ncbi.nih.gov/toolbox/ncbi_tools/converters/by_program/tbl2asn/) to validate gene models, to generate a sequin file (.sqn) ready for submission to NCBI, and to generate a GenBank format (.gb) file.

The generated GenBank file was uploaded to antiSMASH fungal version v5.1.2 webserver to predict genes encoding key enzymes for secondary metabolism. The predicted secondary metabolite genes were queried with BLASTp (e-value< 1E-10) against the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database v2 to search for homologous genes.

Genes encoding carbohydrate-active enzymes (CAZymes) were predicted with the dbCAN2 web server (Zhang et al. 2018, 2) using the HMM database v9 and three tools: HMMER (e-value< 1E-15, coverage> 0.35), DIAMOND (e-value< 1E-102), and Hotpep (Frequency> 2.6, Hits> 6). Following the dbCAN's manual, CAZymes predicted by only one tool were discarded to reduce the number of false positives. Proteases were identified by querying all the proteins against the MEROPS database v12.1 (Rawlings et al. 2014) with BLASTp (e-value< 1E-10), and classified based on the most homologous sequence (i.e. top BLASTp hit).

Genes encoding transporters were identified by querying the proteins against the Transporter Classification Database (TCDB) version of 2020-07-12 (Saier Jr et al. 2014) with BLASTp (e-value< 1E-10), and classified based on the most homologous sequence. The ATP-binding cassette (ABC) transporters and the major facilitator superfamily (MFS) were further classified. ABC transporters (TCDB ID: 3.A.1) were classified into ABCB, ABCC, ABCD, ABCE, ABCF, and ABCI families based on their respective top BLASTp hits against the TCDB or based on their domain architecture (Víglaš and Olejníková 2021). A maximum likelihood phylogenetic tree of the ABC transporters was inferred by aligning the protein sequences with MAFFT v7.475 (parameters: --maxiterate 1000 --localpair). All positions in the alignment containing gaps were removed with trimAl v1.4.1. A tree was then constructed with IQTREE v2.1.2 using ModelFinder to determine the best substitution model (best-fit model LG+I+G4) and to perform 1000 rapid bootstrap replicates (parameters: -m MFP -B 1000). For better visualization, the tree was rooted at the ABCI-type transporter HI914_00568. Domains were identified with a batch search (e-value< 1E-3) against the PFAM domain data base (http://pfam.xfam.org). Transmembrane domains were predicted with TMHMM v2.0

(http://www.cbs.dtu.dk/services/TMHMM/). The MFS transporters were further classified into subfamilies based on their top BLASTp hit against the TCDB database.

Genes encoding cytochrome P450s were identified by querying the predicted proteins with the script hmmsearch from HMMER v3.3.2 with maximum e-value of 0.001 (parameters: -E 0.001) using the HMM model for cytochrome P450 (PF00067) obtained from the PFAM website (http://pfam.xfam.org/family/PF00067/hmm). Cytochrome P450s were classified based on BLASTp searches (e-value< 1E-10) against the Dr. Nelson's database of curated fungal cytochrome P450s (https://drnelson.uthsc.edu/P450seqs.dbs.html). Cytochrome P450s with a BLASTp hit with more than 45% identity and at least 70% coverage were classified according to its top BLASTp hit. Protein sequences of the cytochrome P450 were aligned with MAFFT v7.475 (parameters: --maxiterate 1000 --localpair), and all positions in the alignment containing gaps were removed with trimAl v1.4.1. The trimmed alignment was then used to infer a tree with IQTREE v2.1.2 using ModelFinder to determine the best substitution model (best-fit model LG+R2) and to perform 1000 rapid bootstrap replicates (parameters: -m MFP -B 1000).

S2.6 Prediction and analysis of genes encoding candidate secreted effector proteins (CSEPs). Secreted proteins were predicted with SignalP v5.0 (Armenteros et al. 2019) with parameter -org euk. Secreted proteins were considered candidate effectors based on three lines of evidence: i) secreted proteins classified as effectors with EffectorP v2.0 (Sperschneider et al. 2016); ii) secreted proteins shorter than 250 aa, with cysteine content of at least 2%, no glycosylphosphatidylinositol (GPI) anchor as determined with PredGPI (FP rate > 0.005) (Pierleoni et al. 2008), and no transmembrane domain in the mature protein as determined with TMHMM v2.0 (Krogh et al. 2001); iii) proteins with no homologs in Leotiomycetes, except Erysiphales, based on a BLASTp search (e-value< 1E-3) against 93 annotated genomes of Leotiomycetes available at NCBI on 2021-03-18 (Fig 5.S10). All secreted proteins that satisfied any of these three lines of evidence were classified as candidate secreted effector proteins (CSEPs). CSEPs were subsequently queried with BLASTp (e-value< 1E-5) against the Pathogen Host Interaction (PHI) database v4.13 (Winnenburg 2006) to identify putative homologs.

The Y/F/WxC motif (Godfrey et al. 2010) was identified among the CSEPs by parsing a single-line fasta file containing mature CSEPs (i.e. signal peptide trimmed) using the Unix command sed 's/[W|F|Y][A-Z]C/|/g', then searching for the presence of the pipe ("|") character.

To examine the conservation of the CSEPs in other *E. necator* isolates, the CSEP-encoding genes were mapped to the genome of isolates branching, C-strain, e1-101, lodi, and Ranch9 (Jones et al. 2014a) using Liftoff v1.6.3 (Shumate and Salzberg 2020). CSEPs that failed to map with Liftoff were queried with exonerate v2.4.0 (Slater and Birney 2005) with parameters -m protein2genome --bestn 1 --showtargetgff yes and mapped to the *E. necator* genomes.

A total of 22 secreted proteins contained a microbial RNase domain (cl00212) and were further classified as RNase-like proteins associated with haustoria (RALPHs) (Spanu 2017). Because e-values of identified microbial RNase domains in these 22 predicted RALPHs were high (i.e. 1E-4 – 1E-09), we reasoned that RALPH-like proteins without a predicted microbial RNase domain above the e-value threshold (i.e. 1E-3) could be present among the secreted proteins. To search for such RALPH-like proteins, the 22 predicted RALPH proteins with a conserved microbial RNase domain were queried with BLASTp against all predicted proteins of EnFRAME01. Matched proteins with e-value< 1E-5 and at least 40% aa identity were also considered RALPH-like proteins. By doing so, the total number of RALPH-like proteins increased from 22 to 38. These 38 RALPH-like proteins were aligned with MAFFT v7.455 (Katoh et al. 2002) with parameters --maxiterate 1000 --localpair. The resulting 817 aa in size alignment was processed with trimAl (Capella-Gutiérrez et al. 2009) to remove sites containing 50% or more gaps, resulting in a trimmed alignment of 465 aa. The trimmed alignment was used to infer a phylogenetic tree with IQ-TREE v1.6.11 (Nguyen et al. 2015) using 1000 ultrafast bootstrap replicates (Hoang et al. 2018) and the substitution model FLU+F+G4 selected with the built-in ModelFinder (Kalyaanamoorthy et al. 2017).

A total of 11 Egh16-like virulence factor proteins were identified based on the presence of the Egh16-like virulence factor conserved domain (PF11327). Differently from RALPH-like proteins, all proteins containing an Egh16-like virulence factor domain also contained a predicted signal peptide, and the e-values of the Egh16-like virulence factor domain were lower (between 3.3E-48 and 4.4E-67). Following the same approach described for RALPH-like proteins, a multiple sequence alignment of 499 aa was obtained for the 11 Egh16-like virulence factor proteins. The alignment was trimmed, resulting in an alignment of 257 aa, that was used to infer a phylogenetic tree using the WAG+F+R2 substitution model.

S2.7 Identification of MACGs in *E. necator*. Protein sequences from *E. necator* isolate EnFRAME01, *Blumeria graminis* f.sp. *hordei* isolate DH14 (https://fungi.ensembl.org/Blumeria_graminis/Info/Index), *Botrytis cinerea* isolate B05.10 (GCA_000143535.4), *Zymoseptoria tritici* isolate IPO323 (GCA_000219625.1), *Aspergillus niger* isolate CBS 513.88 (GCA_000002855.2), *Neurospora crassa* isolate OR74A (GCA_000182925.2), and *Saccharomyces cerevisiae* isolate S288C (GCF_000146045.2) were organized into hierarchical orthogroups (HOGs) with OrthoFinder v2.5.4 (Emms and Kelly 2015) with parameters -t 8 -M msa -A mafft -T fasttree. HOGs containing protein sequences from non-obligate fungi analyzed, i.e. *B. cinerea*, *Z. tritici*, *A. niger*, *N. crassa*, and *S. cerevisiae*, but not from *E. necator* were considered as candidate MACGs in *E. necator*. These genes were further filtered based on enzyme commission (EC) numbers. First, all predicted proteins from EnFRAME01 were queried with BlastKOALA webserver (Kanehisa et al. 2016) to obtain EC numbers. Then, corresponding EC numbers of the candidate MACGs in *E. necator* were obtained for the *S. cerevisiae* proteins by searching their accession numbers in the UniProt database (www.uniprot.org). If a gene from *E. necator* had an assigned EC number identical to an EC number of a MACG in *E. necator*, then this gene was no longer considered missing. Finally, a KEGG pathway enrichment analysis of the *E. necator* MACGs was performed with the YeastEnrichr webserver (https://maayanlab.cloud/YeastEnrichr/) (Chen et al. 2013), by using the corresponding *S. cerevisiae* gene names.

S2.8 Compartmentalization of the genome of *E. necator*. First, all intergenic regions were identified with the script complement from BEDtools v2.29.0, which obtained regions (i.e. intervals) of the genome not covered by genes. Then, up- and downstream intergenic regions were identified using the script closest from the BEDtools v2.29.0 software. The script was executed once to identify the downstream intergenic regions (parameters -t first -D a -iu), and then once more to identify the upstream intergenic regions (parameters -t first -D a -id). A heat map of the size of intergenic regions was obtained with the script geom_hex (parameter bins=50) from the R package ggplot2 v3.3.5 within R v4.1.2. Up- and downstream intergenic regions were grouped based on gene categories (i.e. BUSCOs, candidate effectors, secreted not candidate effectors, CAZymes, and proteases). Their sizes were compared, as well as their repetitive DNA content. To estimate their repetitive DNA content, the script coverage from BEDtools was used to calculate the percentage of the intergenic regions covered by repeats, using the repeat coordinates (GFF format) generated by RepeatMasker. Statistical significance was obtained with the Wilcoxon rank sum test implemented in the function wilcox.test within R v4.1.2.

S2.9 Classification, analysis, and enrichment of duplicated genes. An all-vs-all BLASTp search with parameters -max_hsps 5 -max_target_seqs 5 evalue 1E-5 was performed to identify duplicated genes. To reduce the number of false positives, only BLASTp hits with identity values greater than 40% and at least 50% query and subject coverage were retained. The filtered output of BLASTp was used by the script duplicate_gene_classifier from MCScanX (Wang et al. 2012) to classify gene duplications into dispersed, proximal, or tandem. Briefly, dispersed duplications corresponded to gene copies located in different chromosomes or with more than ten genes in between copies. Proximal duplications corresponded to gene copies with one to ten genes in between copies. Tandem duplications corresponded to gene copies with no genes in between copies. Conservation of dispersed, proximal, and tandem duplicated genes at the nucleotide (i.e. coding sequence) and amino acid levels were estimated by aligning each duplicate gene with its corresponding top BLASTp hit (not to itself). To do so, pairwise alignments were obtained with the pairwiseAlignment function (parameter: type = "global-local") from the Biostrings package v2.64.0 (Pagès et al. 2022) within R v4.2.1. The "global-local" alignment strategy was used in which only the shortest of the two sequences is aligned end-to-end. Percent identity values of the alignments were obtained with the function pid (parameter: type = "PID1") from the Biostrings package. The "PID1" was used within pid in order to take into consideration internal gaps, but not gaps at both ends of the alignment. The top BLASTp hits were also used to estimate selection pressure of duplicated genes based on the K_A/K_S ratio (number of substitutions per nonsynonymous sites (K_A) per number of substitutions per synonymous sites (K_S). To do so, the query and its top BLASTp hit were aligned with the Needleman -Wunsch global alignment algorithm (parameters -aformat fasta -gapopen 10 -gapextend 0.5) from EMBOSS v6.6.0. Amino acid alignments were converted to codon alignments with PAL2NAL v14. Codon alignments in FASTA format were then converted to AXT format and given as input to K_A/K_S_calculator v3 to estimate K_A/K_S ratios using the MA method. Conserved domain enrichment within duplicated genes was performed with the function enricher from the R package clusterProfiler v4.2.2 (Yu et al. 2012) within R v4.1.2 with parameters pvalueCutoff=0.01 pAdjustMethod="BH" qvalueCutoff=0.01 minGSSize=1. A dot plot of the significantly enriched domains was generated with the function dotplot from the enrichplot v 1.14.1 software. Statistical significance for different duplication frequencies of different gene categories was obtained with hypergeometric tests implemented in the phyper function within R v4.1.2. Statistical significance for different conservation levels and K_A/K_S ratios for dispersed, proximal, and tandem duplicated genes was obtained the Wilcoxon rank sum test implemented in the function wilcox.test within R v4.1.2.

S2.10 Organization of duplicated genes into orthogroups. To identify potential gene families expanded by gene duplications, all predicted protein sequences of *E. necator* isolate EnFRAME01, *E. necator* isolate C-strain (GCA_000798715.1), *Erysiphe pulchra* isolate Cflorida (GCA_002918395.1), *Golovinomyces cichoracearum* isolate UMSG1 (GCA_003611235.1), *Oidium neolycopersici* (GCA_003610855.1), *Blumeria graminis* f.sp. *hordei* isolate DH14 (https://fungi.ensembl.org/Blumeria_graminis/Info/Index), *B.*

graminis f.sp. *tritici* isolate 96224 (GCA_900519115.1), and *Botrytis cinerea* isolate B05.10 (GCA_000143535.4) were organized into hierarchical orthogroups (HOGs) with OrthoFinder v2.5.4 (Emms and Kelly 2015) with parameters -t 12 -M msa -A mafft -T fasttree. The number of duplicated genes of EnFRAME01 within HOGs were counted. HOGs containing at least three duplicated genes of *E. necator* EnFRAME01 were considered candidate gene families expanded by frequent gene duplications, and were further analyzed. Presence of secreted proteins, CSEPs, and conserved domains of proteins within these HOGs were reported.

S2.11 Identification of CNV genes. Whole-genome sequencing reads of *E. necator* isolates Branching (SRR1448453), C-strain (SRR1448450), e1-101 (SRR1448468), Lodi (SRR1448470), and Ranch9 (SRR1448454) (Jones et al. 2014a) were obtained from the Sequence Read Archive (SRA). Reads were mapped to the genome of EnFRAME01 with BWA-MEM v0.7.17 (Li and Durbin 2009). PCR duplicates were marked with samblaster v0.1.24 (Faust and Hall 2014) and further removed with SAMtools v1.9 (Li et al. 2009). Based on the resulting mapped reads, CNV regions were identified with CNVnator (Abyzov et al. 2011) using bin sizes of 150 bp for strain e1-101, 200 bp for strains branching and C-strain, and 300 bp for strains Lodi and Ranch9. Predicted CNV regions with a reported t-test e-value> 0.001 or with more than 50% reads mapped with mapping quality=0 were removed. Furthermore, only the CNV regions with normalized read depth< 0.2 (putative deletion) or normalized read depth> 1.8 (putative duplication) were retained. Genes that had at least 80% of their sequences overlapping CNV regions, as determined with the script overlap from BEDtools v2.29.0, were considered CNV genes (deleted or duplicated).

S2.12 Comparative analysis of carboxylesterases (CEs). The predicted carboxylesterase encoded by the gene *HI914_00624* was queried with BLASTp against the NCBI nr database (2022-08-13) and proteins from EnFRAME01. The 400 most similar sequences (e-value< 1E-50) were obtained to construct a phylogenetic tree. The acetylcholinesterase DmAChE from *Drosophila melanogaster* (1QO9), for which the 3D structure has been elucidated (Harel et al. 2000), was included as outgroup. The 401 amino acid sequences were then aligned with MAFFT v7.487 with parameters --maxiterate 1000 --localpair --thread 12, resulting in an alignment of 3442 aa. Sites containing 50% or more gaps were removed with trimAl, resulting in an alignment of 564 aa. The trimmed alignment was used to infer a phylogenetic tree with IQ-TREE v1.6.12 (Nguyen et al. 2015) using 1000 rapid bootstrap replicates (Hoang et al. 2018) and the substitution model WAG+F+R10 selected by the built-in ModelFinder (Kalyaanamoorthy et al. 2017). The tree was visualized and edited with iTOL (Letunic and Bork 2021). Identification of conserved residues was performed by analyzing the residues from other sequences aligned to the conserved residues previously described in DmAChE (Harel et al. 2000; Oakeshott et al. 2005). Specifically, the catalytic triad Ser238, Glu367, and His480, the oxyanion hole residues Gly150, Gly151, and Ala239, the nucleophilic elbow formed by the pentapeptide Gly236, Glu237, Ser238, Ala239, and Gly240 (i.e., the GxSxG motif), and the six cysteine residues Cys66, Cys93, Cys292, Cys307, Cys442, and Cys560 that form three disulfide bridges in DmAChE (Harel et al. 2000; Oakeshott et al. 2005). A sequence logo was obtained with WebLogo (Crooks et al. 2004) based on the trimmed alignment.

S2.13 Assessment of *HI914_00624* gene copy number and of its expression. Quantitative PCR (qPCR) and quantitative reverse transcription PCR (RT-qPCR) were used to determine the copy number and gene expression of the *HI914_00624* gene, respectively, using the single-copy EnEF1 elongation factor 1 gene as reference in *E. necator* isolates EnFRAME01, C-strain, HO1, MEN8B, CAT1, and BL1-2. Conidia of the isolates were collected from the grapevine leaves 14 days after inoculation using the method described above. Genomic DNA was extracted from conidia using a rapid Chelex method (Brewer and Milgroom 2010) and total RNA was isolated using the methods described in S2.1. cDNA was generated using SuperScript IV Reverse Transcriptase (Invitrogen), using the provided random hexamers and following the manufacturer's protocol. All qPCR reactions were run in triplicate on an Applied Biosystems QuantStudio5 qPCR machine using PerfeCTa qPCR ToughMix Low ROX (Quantabio) and the primers and probes listed in Table 5.S26. The following cycling conditions were used: 95°C for 2 min followed by 45 cycles of 95°C for 15 sec and 62C for

45 sec. Gene copy number and expression were calculated using the ΔΔCT relative quantification method where RQ=2-ΔΔCT. For these calculations, the EnEF1 single-copy gene, whose expression was shown to be constant across inoculation time points (Jones et al. 2014a), was treated as the control gene and EnFRAME01 was the reference isolate. *HI914_00624* PCR products were confirmed by Sanger sequencing.

5.7.3 Supplementary figures

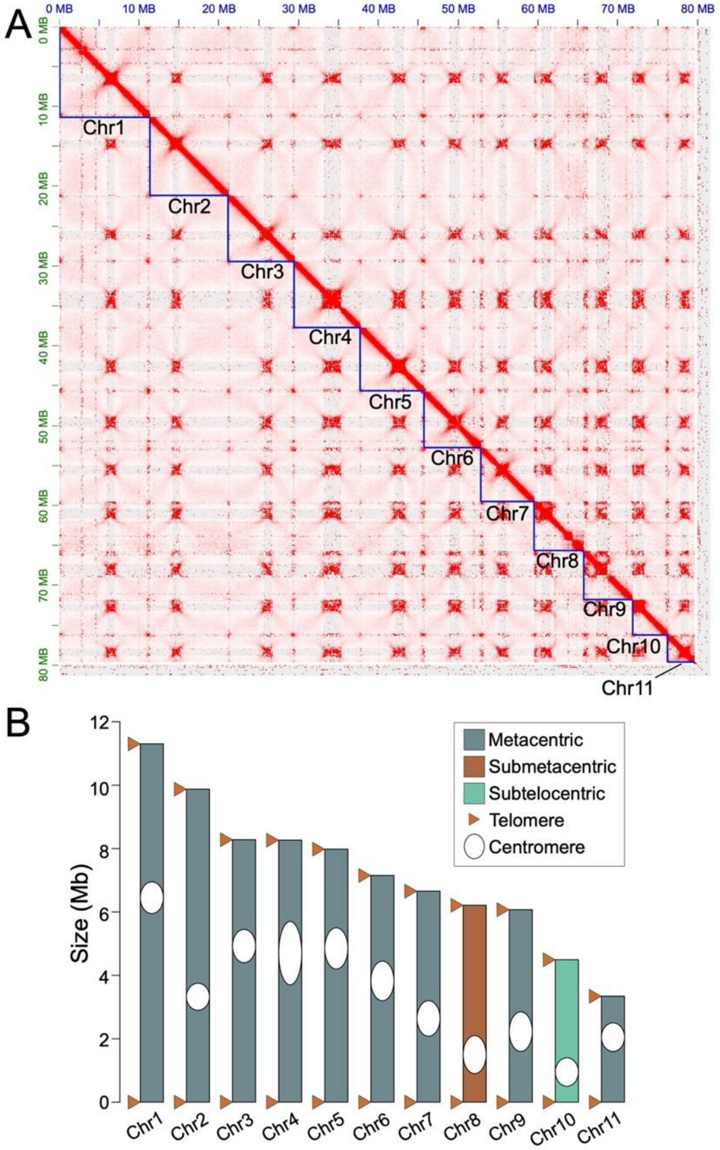


Figure 5.S1: Hi-C contact heatmap and diagram of the predicted chromosomes of *Erysiphe necator* isolate EnFRAME01. (A) The heat map represents the Hi-C contact matrix across the genome of EnFRAME01. High frequency interaction is shown in red scale. The figure shows the 11 predicted chromosomes as regions with high interaction frequency. Predicted centromeres of different chromosomes interact with each other. (B) Diagram of the 11 predicted chromosomes with predicted centromeric regions shown as white ellipses. Based on the ratio (r) of the estimated size of the long and short arms, the chromosomes were classified as metacentric (1.0<r<1.7), submetacentric (1.7<r<3.0), or subtelocentric (3.0<r<7.0), according to (Levan et al. 1964).

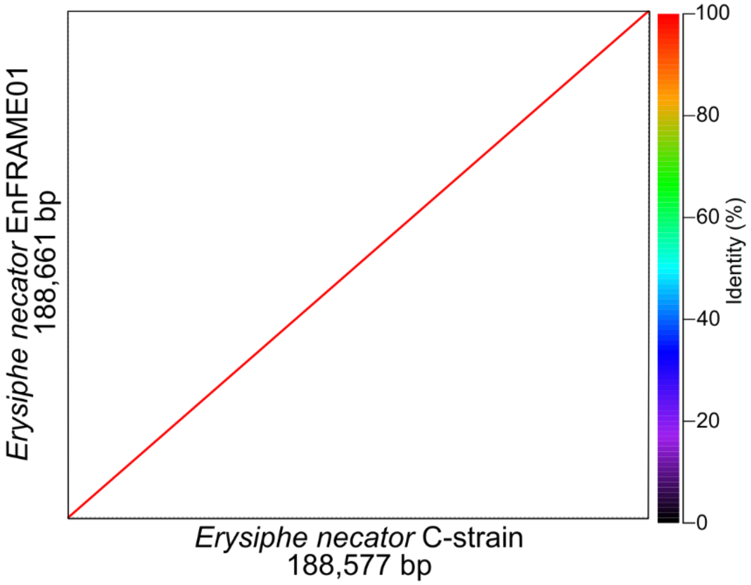


Figure 5.S2: The mitochondrial genomes of *Erysiphe necator* isolates EnFRAME01 and C-strain are highly conserved. The figure shows a dot plot of the alignment between the two mitochondrial genomes. A single aligned block with nucleotide identity of 99.7% covers the entire sequences of both mitochondrial genomes. The mitochondrial genome of *E. necator* isolate C-strain was obtained from NCBI (NC_056146.1) and aligned to that of isolate EnFRAME01 using NUCmer from the MUMmer software package v4.

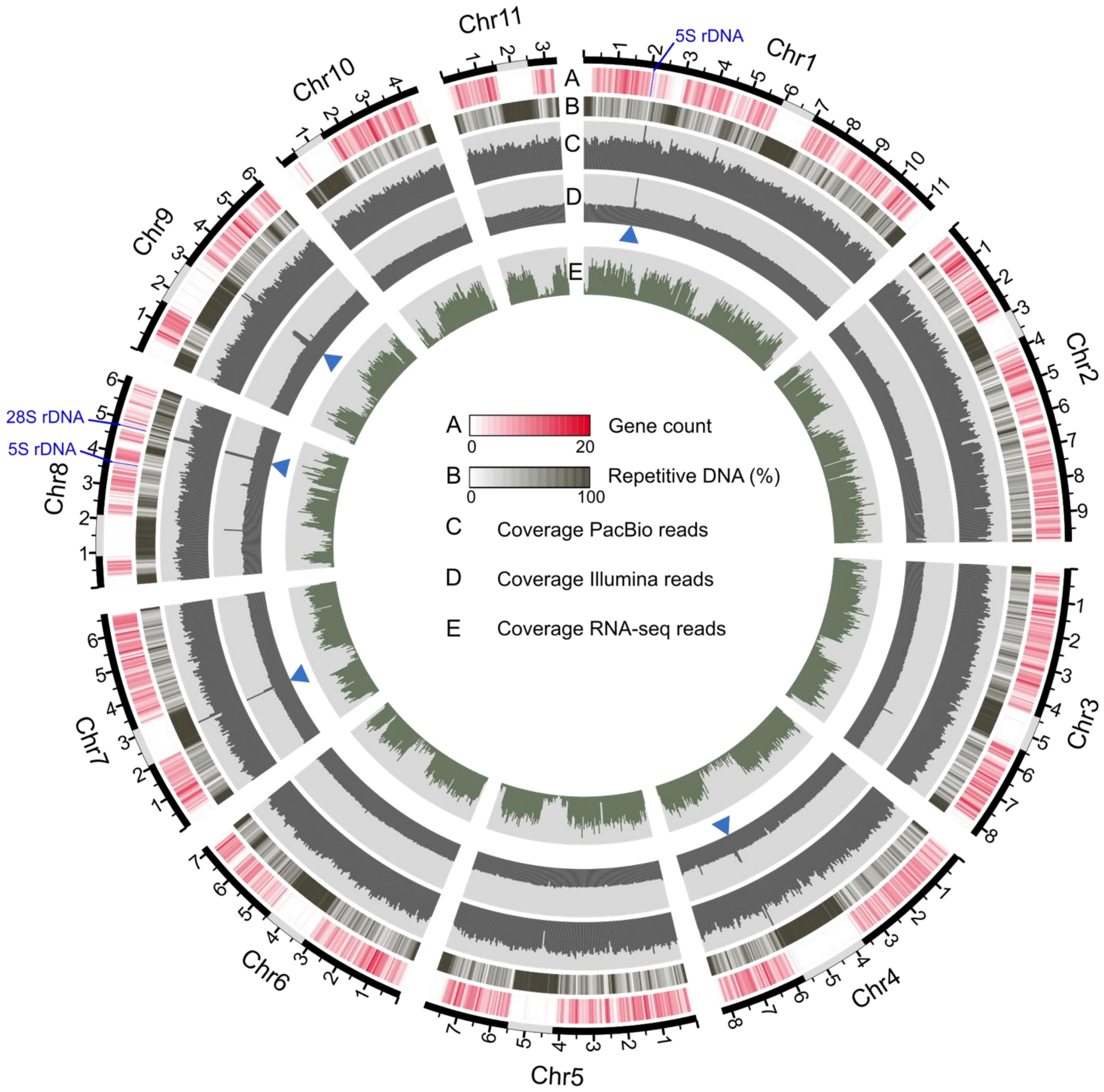


Figure 5.S3: **Coverage of sequenced DNA-seq and RNA-seq reads across the 11 chromosomes of *Erysiphe necator* isolate EnFRAME01.** (A) Heatmap showing the number of predicted protein coding genes. (B) Heatmap showing the percentage of repetitive DNA. (C) Coverage (0x to 200x) of the sequenced PacBio reads used to produce the assembly. (D) Coverage (0x to 200x) of whole-genome sequencing Illumina reads used to polish the assembly. (E) Coverage in log10 scale (0x to 10000x) of RNA-seq reads used to predict the protein coding genes. The five blue triangles indicate regions where PacBio and Illumina coverage exceeds twice the median for the entire genome, and indicates predicted collapsed regions in the assembly. Values in all tracks were determined using a sliding window of 50 kb.

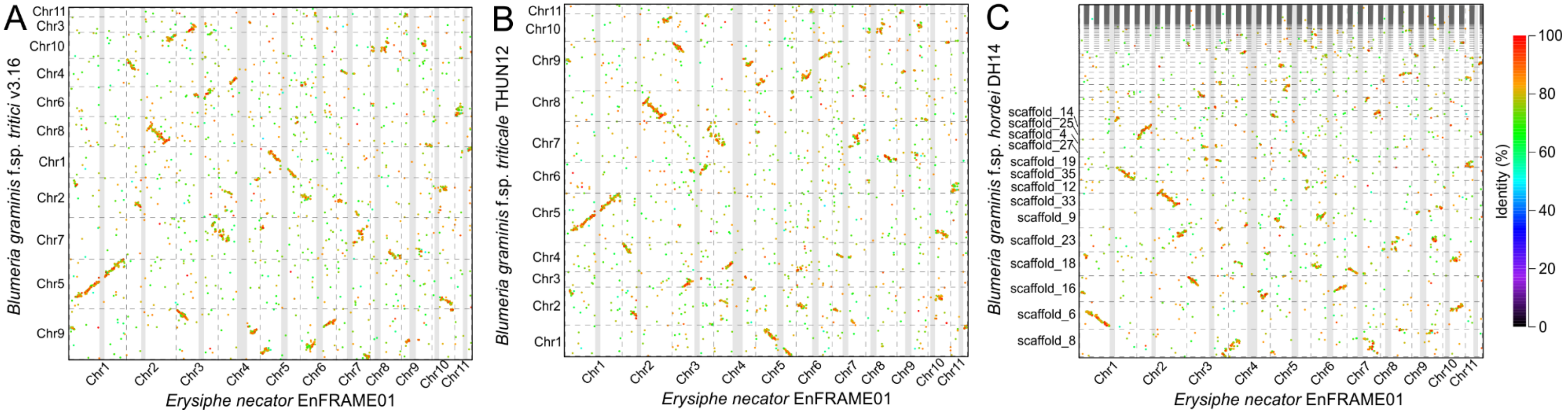


Figure 5.S4: The 11 chromosomes of *Erysiphe necator* are not syntenic to the chromosomes or scaffolds of the cereal powdery mildew pathogens *Blumeria graminis* f.sp. *tritici, B. graminis* f.sp. *triticale*, and *B. graminis* f.sp. *hordei*. Dot plot showing whole-genome alignments at the protein level of the 11 chromosomes of *E. necator* isolate EnFRAME01 and the 11 chromosomes of (A) *B. graminis* f.sp. *tritici* isolate v3.16, (B) the 11 chromosomes of *B. graminis* f.sp. *triticale* isolate THUN12, and (C) the 318 scaffolds of *B. graminis* f.sp. *hordei* isolate DH14. Protein alignments were obtained with PROmer. Predicted centromeres in the chromosomes of *E. necator* are highlighted with grey bars. The figure shows that the genome alignments are largely composed of non-syntenic regions and the predicted centromeric regions of *E. necator* are essentially unaligned. For *B. graminis* f.sp. *hordei* DH14, only the names of the 14 longest scaffolds are shown.

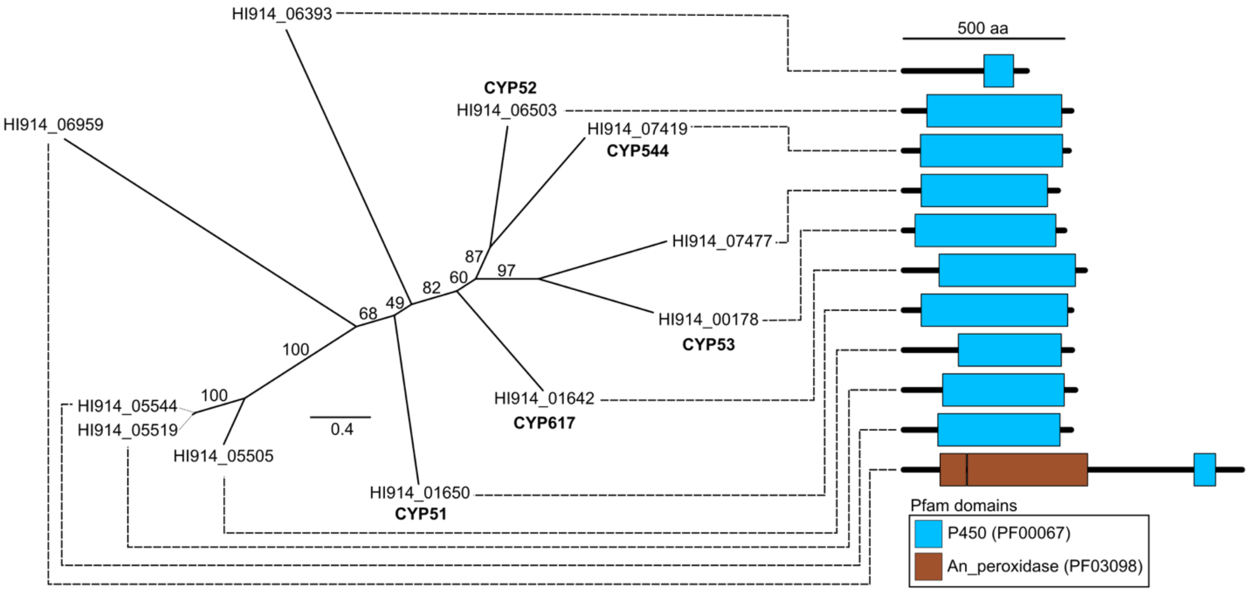


Figure 5.S5: Genes encoding cytochrome P450s in the genome of *Erysiphe necator* isolate EnFRAME01. The figure shows an unrooted maximum likelihood (ML) phylogenetic tree of the cytochrome P450s. Classified P450s have their respective family name in bold next to the encoding gene IDs. Protein size and conserved PFAM domains within them are shown in scale on the right-hand side. All proteins contain a single P450 domain (PF00067). Gene *HI914_06959* encodes a cytochrome P450 with two additional domains corresponding to animal heme-dependent peroxidase (PF03098). The cytochrome P450 genes *HI914_05544, HI914_005519,* and *HI914_05505* are predicted paralogs because they share aa identity of more than 40%.

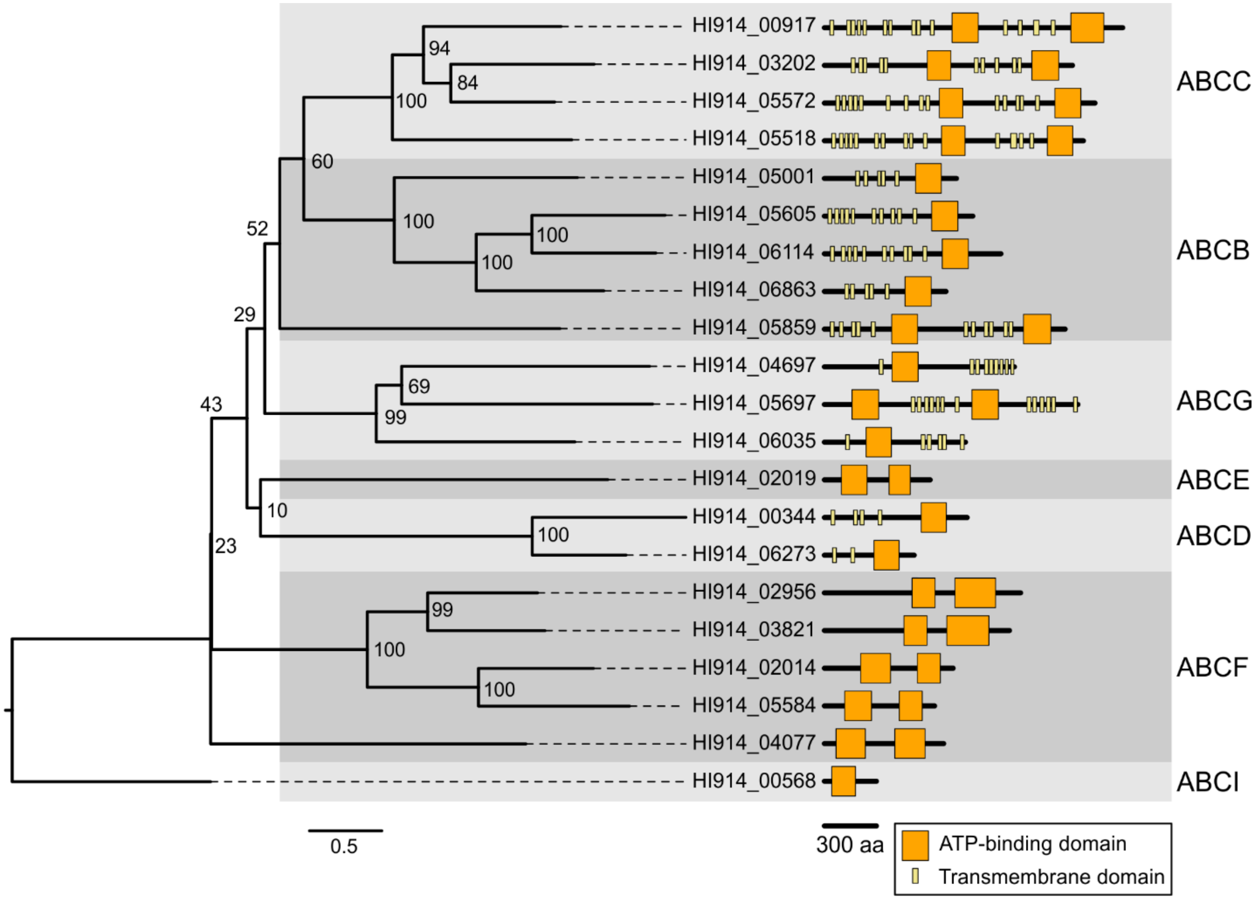


Figure 5.S6: Genes encoding ATP-binding cassette (ABC) transporters in the genome of *Erysiphe necator* isolate EnFRAME01. The figure shows a maximum likelihood phylogenetic tree of the ABC transporters. The domain architectures of the ABC transporters are shown in scale on the right-hand side and include the conserved ATP-binding domain (PF00005) and transmembrane domains. Classification of the ABC transporters into classes ABCC, ABCB, ABCG, ABCE, ABCD, ABCF, and ABCI is shown on the far right-hand side.

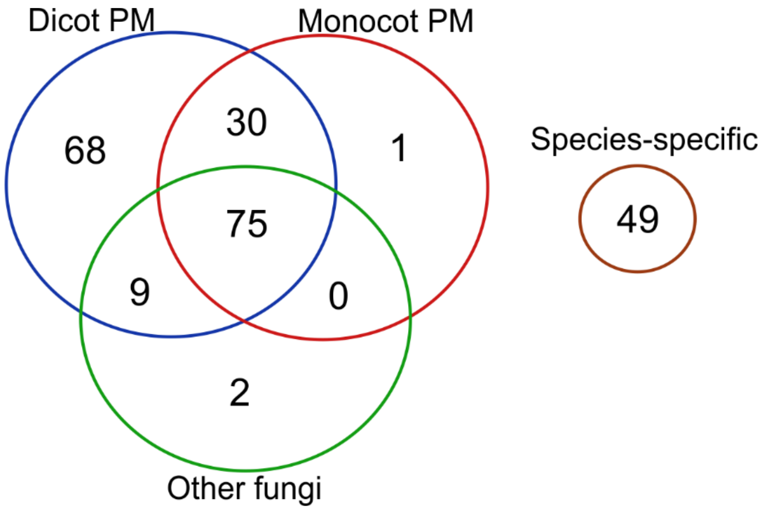


Figure 5.S7: Venn diagram of candidate secreted effector proteins (CSEPs) from *Erysiphe necator* isolate EnFRAME01 that are species-specific or shared with other fungi. CSEPs of EnFRAME01 were considered as shared as if they had BLASTp hits (e-value< 1E-5) to predicted proteins from the dicot-infecting powdery mildews (PM) *Erysiphe pulchra* isolate Cflorida, *Erysiphe neolycopersici* isolate UMSG2, *Golovinomyces cichoracearum* isolate UMSG3, *Golovinomyces cichoracearum* isolate UCSC1, *Golovinomyces cichoracearum* isolate UMSG1*, Golovinomyces magnicellulatus* isolate FPH2017-1, and *Podosphaera aphanis* isolate DRCT72020, the monocot-infecting PMs *Blumeria graminis* f.sp. *hordei* isolate RACE1, *Blumeria graminis* f.sp. *hordei* isolate DH14, *Blumeria graminis* f.sp. *tritici* isolate v3.16, *Blumeria graminis* f.sp. *tritici* isolate 96224, and *Blumeria graminis* f.sp. *triticale* isolate THUN12, or other fungi in the NCBI nr database.

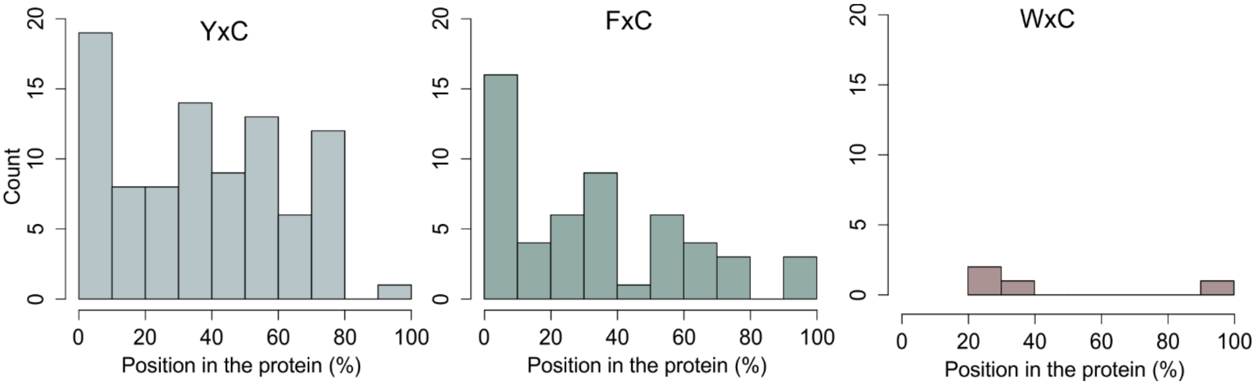


Figure 5.S8: Histograms of Y/F/W-x-C motifs in candidate secreted effector proteins (CSEPs) of *Erysiphe necator* isolate EnFRAME01. The histograms show the number of Y/F/W-x-C motifs located in different regions of mature CSEPs (i.e. after signal peptide cleavage). In the histograms, 0% represents the N-terminus, and 100% represents the C-terminus. Preference of Y/F/W-x-C motifs located near the N-terminus can be observed.

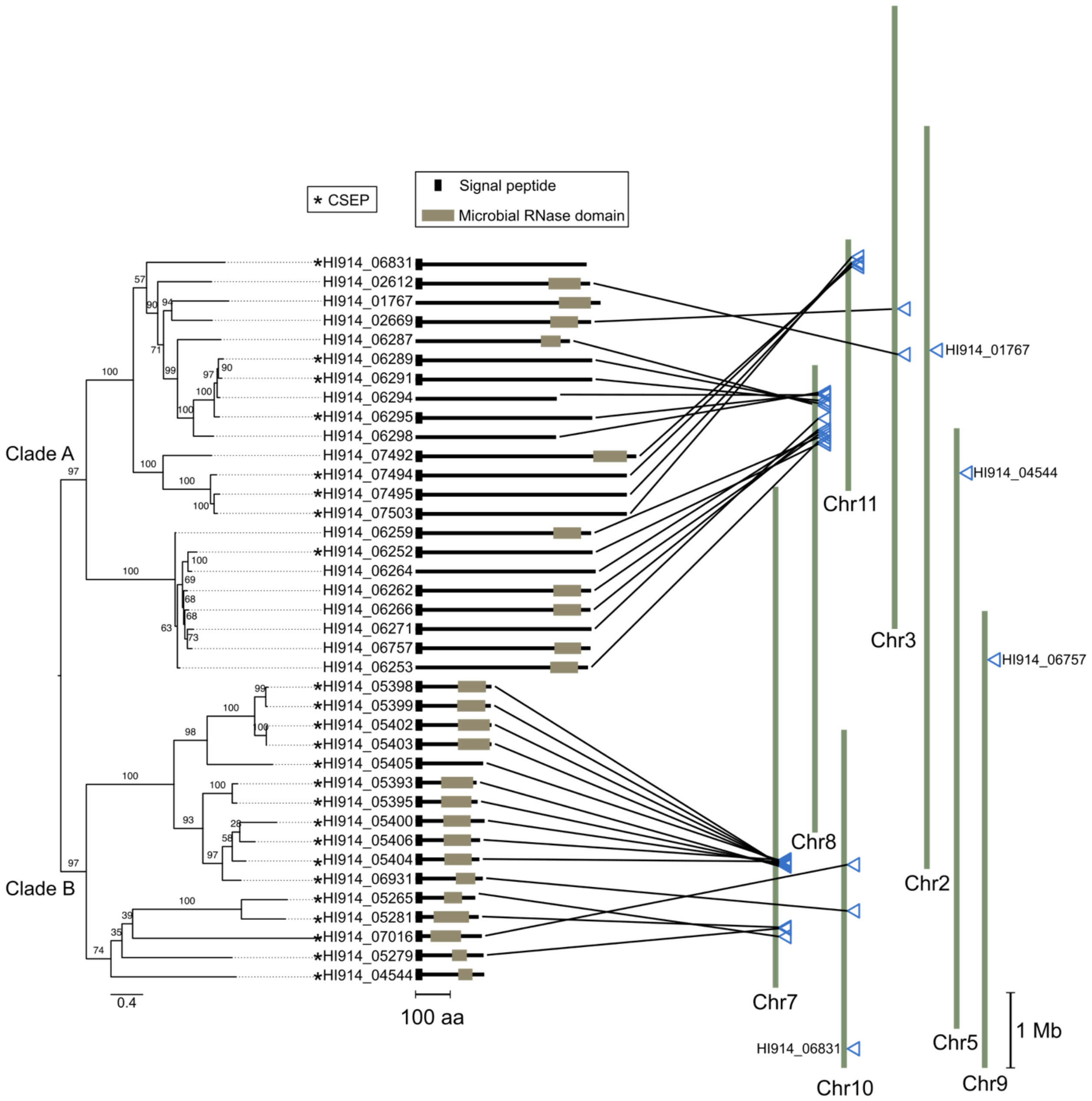


Figure 5.S9: *Erysiphe necator* isolate EnFRAME01 has two major clades of predicted genes encoding RNase-like proteins associated with haustoria (RALPHs). Maximum likelihood (ML) phylogenetic tree of the 38 RALPH-like proteins identified in EnFRAME01. Ultrafast bootstrap support is shown on branches. Protein size, presence of a signal peptide, and location of microbial ribonuclease (RNase) domains (cl00212) are shown next to the encoding gene labels. RALPH-like proteins further classified as candidate secreted effector proteins (CSEPs) are indicated with an asterisk ('*'). The locations on the chromosomes of EnFRAME01 of the genes encoding RALPH-like proteins are shown with hollow triangles connected with lines. The figure shows that genes encoding RALPH-like proteins are organized into two major clades (clade A and B) based on their similarity, size of encoded proteins, and their location on the chromosomes.

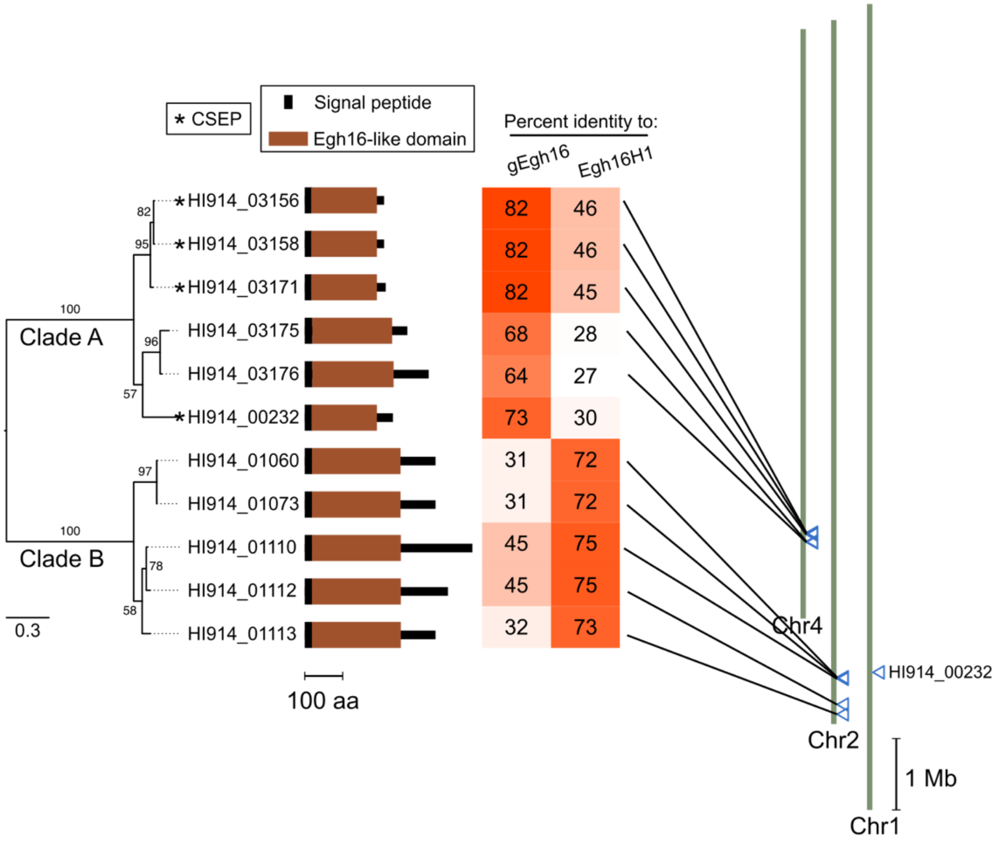


Figure 5.S10: *Erysiphe necator* isolate EnFRAME01 has two major groups of predicted genes encoding Egh16-like proteins. Maximum likelihood phylogenetic tree of the Egh16-like proteins present in EnFRAME01. Ultrafast bootstrap support is shown on the branches. Egh16-like proteins further classified as candidate secreted effector proteins (CSEPs) are indicated with an asterisk ('*'). Protein size, presence of a signal peptide, and location of Egh16-like virulence factor domains (PF11327) are shown next to the encoding gene labels. The heatmap shows the percent amino acid identity values obtained with BLASTp by querying the candidate effectors gEgh16 (JC4750) and Egh16H1 (CCU76428.1) from *Blumeria graminis* f.sp. *hordei*. The locations on the chromosomes of EnFRAME01 of the genes encoding the Egh16-like proteins is shown with hollow triangles connected with lines. The figure shows that genes encoding Egh16-like proteins are organized into two major clades (clade A and B) based on their similarity and location in the chromosomes.

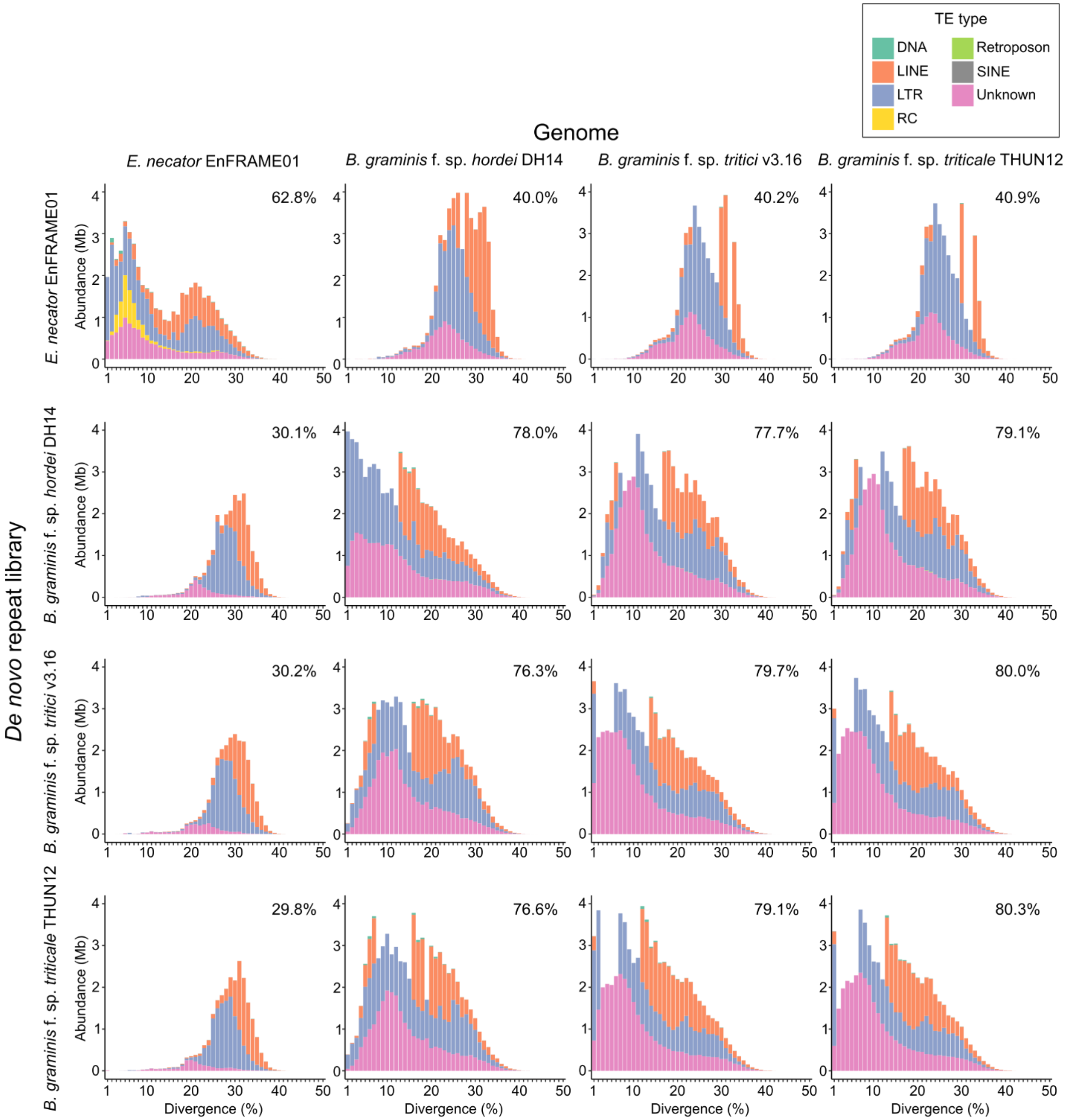


Figure 5.S11: The genome of *Erysiphe necator* has a different transposable element (TE) content compared to the genomes of cereal powdery mildews (PMs). The bar plots represent the repetitive DNA landscape of *E. necator* isolate EnFRAME01, *Blumeria graminis* f.sp. *hordei* isolate DH14, *Blumeria graminis* f.sp. *tritici* isolate v3.16, and *Blumeria graminis* f.sp. *triticale* isolate THUN12. The repetitive DNA landscape for each pathogen was obtained by masking its genome, using its respective custom *de novo* repetitive DNA library. Percentages shown at the top right corner of each plot indicate the amount of masked bases in the genome. The figure shows that, when using TE libraries of the cereal PMs, almost all TEs with low sequence divergence are not identified in the genome of *E. necator*. Similarly, when using the TE library of *E. necator*, almost all TEs with low sequence divergence are not identified in the genomes of the cereal PMs.

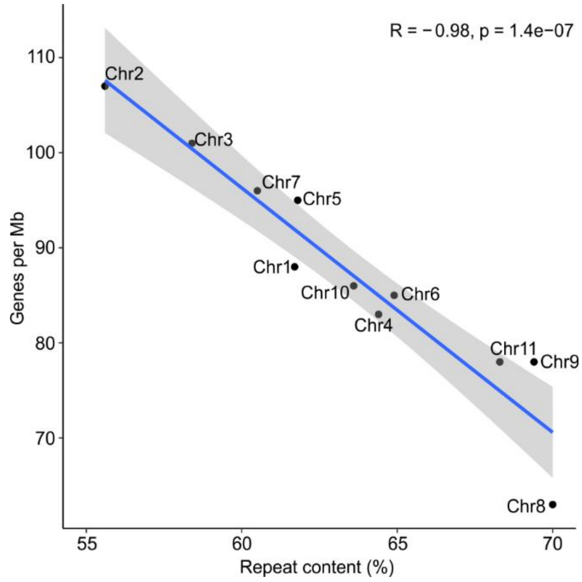


Figure 5.S12: The chromosomes of *Erysiphe necator* isolate EnFRAME01 exhibit negative correlation between gene density and repetitive DNA content. Scatter plot showing the number of genes per Mb and repetitive DNA content of the chromosomes. The regression line is shown in blue and was determined with the geom_smooth function from the R package ggplot2, utilizing the "lm" method. Dark areas represent confidence intervals (95%). Correlation coefficient and *p*-value of the regression line are shown at the top right corner.

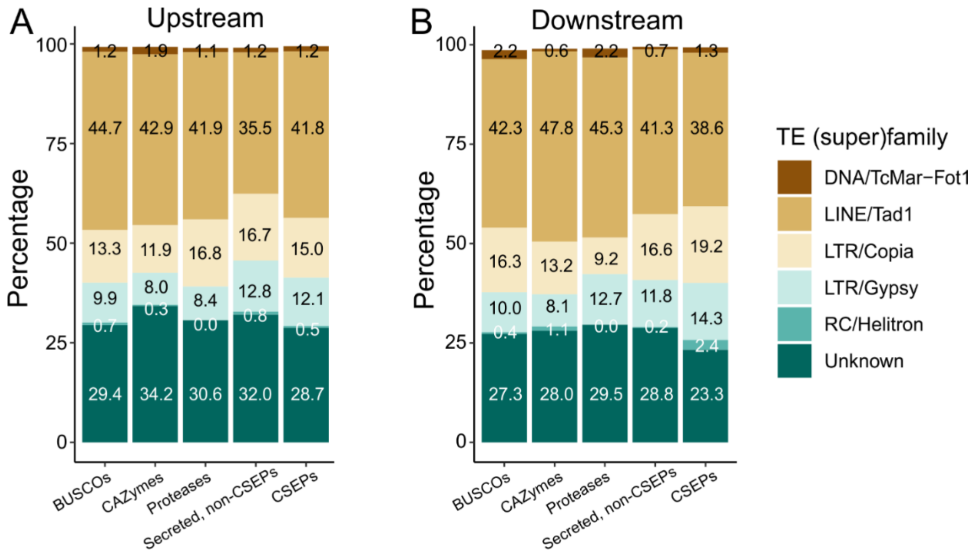


Figure 5.S13: The genome of *Erysiphe necator* isolate EnFRAME01 exhibits no major differences in predicted transposable element (TE) content in intergenic regions of selected functional gene categories. The bar plots show the percentage of the different TE (super)families (A) upstream and (B) downstream of intergenic regions of predicted BUSCO genes, carbohydrate-active enzymes (CAZymes), proteases, secreted proteins not classified as candidate secreted effector proteins (CSEP), and CSEPs.

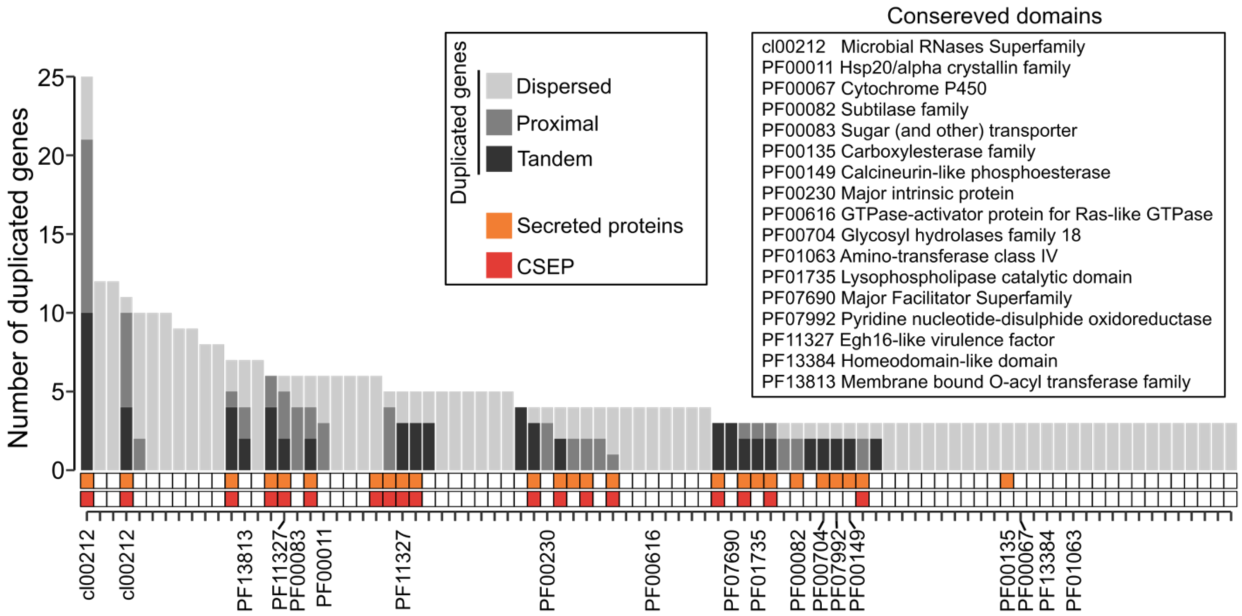
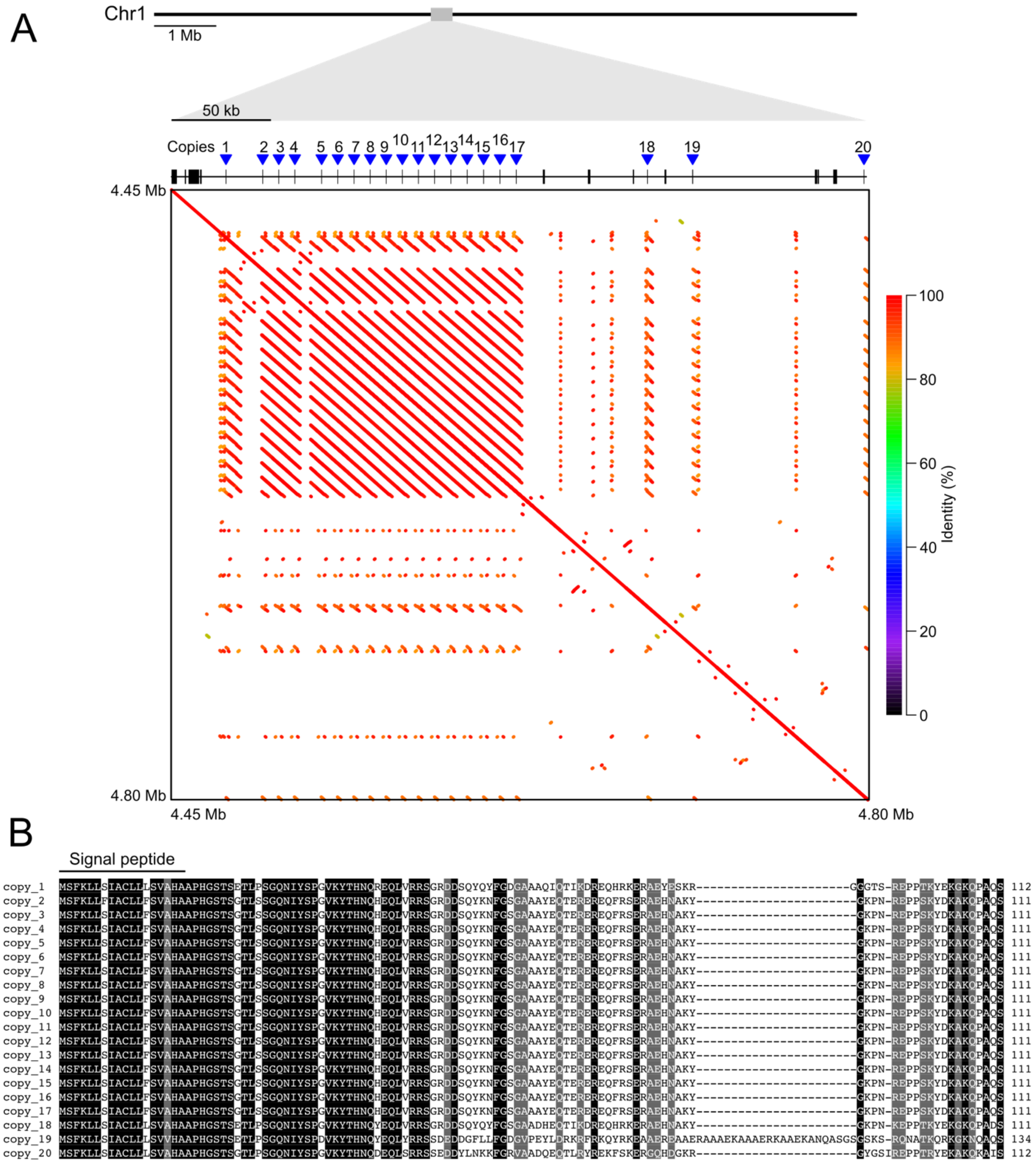


Figure 5.S14: Genes encoding secreted proteins and candidate secreted effector proteins (CSEPs) are expanded by frequent local duplications in *Erysiphe necator* isolate EnFRAME01. The bar plot shows the number of duplicated genes organized in 88 hierarchical orthogroups (HOGs). Duplicated gene copies are classified into dispersed, proximal, or tandem. HOGs that contain at least one gene encoding a secreted protein or a CSEP are indicated with rectangles below the bars. HOGs are assigned to a conserved domain if this domain is present in at least one of the proteins in the respective HOG. The figure shows that duplicated genes encoding secreted proteins and CSEPs are preferentially organized into HOGs that include genes locally duplicated.



```
copy_1  MSFKLLSIACLLLSVAHAAPHGSTSETLPSGQNIYSPGVKYTHNQREQLVRRSGRDDSQYQYFGDGAAAQIQTIKDREQHRKERAEYESKR----------------------GGGTS-REPPTKYEKGKQPAQS 112
copy_2  MSFKLLFIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_3  MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_4  MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_5  MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_6  MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_7  MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_8  MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_9  MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_10 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_11 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_12 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_13 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_14 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_15 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_16 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_17 MSFKLLSIACLLFSVAHAAPHGSTSGTLSSGQNIYSPGVKYTHNQHEQLVRRSGRDDSQYKNFGSGAAAYEQTEREREQFRSERAEHNAKY-----------------------GKPN-REPPSKYDKAKQPAQS 111
copy_18 MSFKLLSIACLLLSVAHAAPHGSTSETLPSGQNIYSPGVKYTHNQYEQLVRRSGRDDSQYQYFGSGAADHEQTIKDREQHRKERAEHNAKY-----------------------GKPN-KEPPTKYEKGKQPADS 111
copy_19 MSFKLLSIACLLLSVVHAAPHGSTSETLPSGQNIYSPDVKYTHNQYEQLYRRSSDEDDGFLLFGDGVPEYLDRKRFRKQYRKEAAEREAAERAAAEKAAAERKAAEKANQASGSGSKS-RQNATKQRKGKNQAQS 134
copy_20 MSFKLLSIACLLLSVAHAAPHGSTSGTLPSGQNIYSPGVKYTHNQDEQLSRRSSEDDYLNKKFGRVAADQEQTLRYREKFSKERGQHDGKR-----------------------GYGSIREPPTRYEKAKQKAIS 112
```

Figure 5.S15: The chromosome 1 (Chr1) of *Erysiphe necator* isolate EnFRAME01 harbors a genomic region that contains 20 copies of a gene encoding a candidate secreted effector protein (CSEP). (A) The figure shows a 350 kb zoomed-in region in chromosome 1 containing 20 copies of a CSEP. Genes within the locus are represented as black rectangles and the CSEP copies indicated with triangles. The dot plot shows the locus aligned to itself, with repetitive regions shown as dots not on the main diagonal. (B) Multiple sequence alignment of the 20 CSEP copies. All 20 copies have the same DNA strand orientation and all were confirmed to be absent of introns based on RNA-seq data. Copies 3 to 17 encode identical proteins.

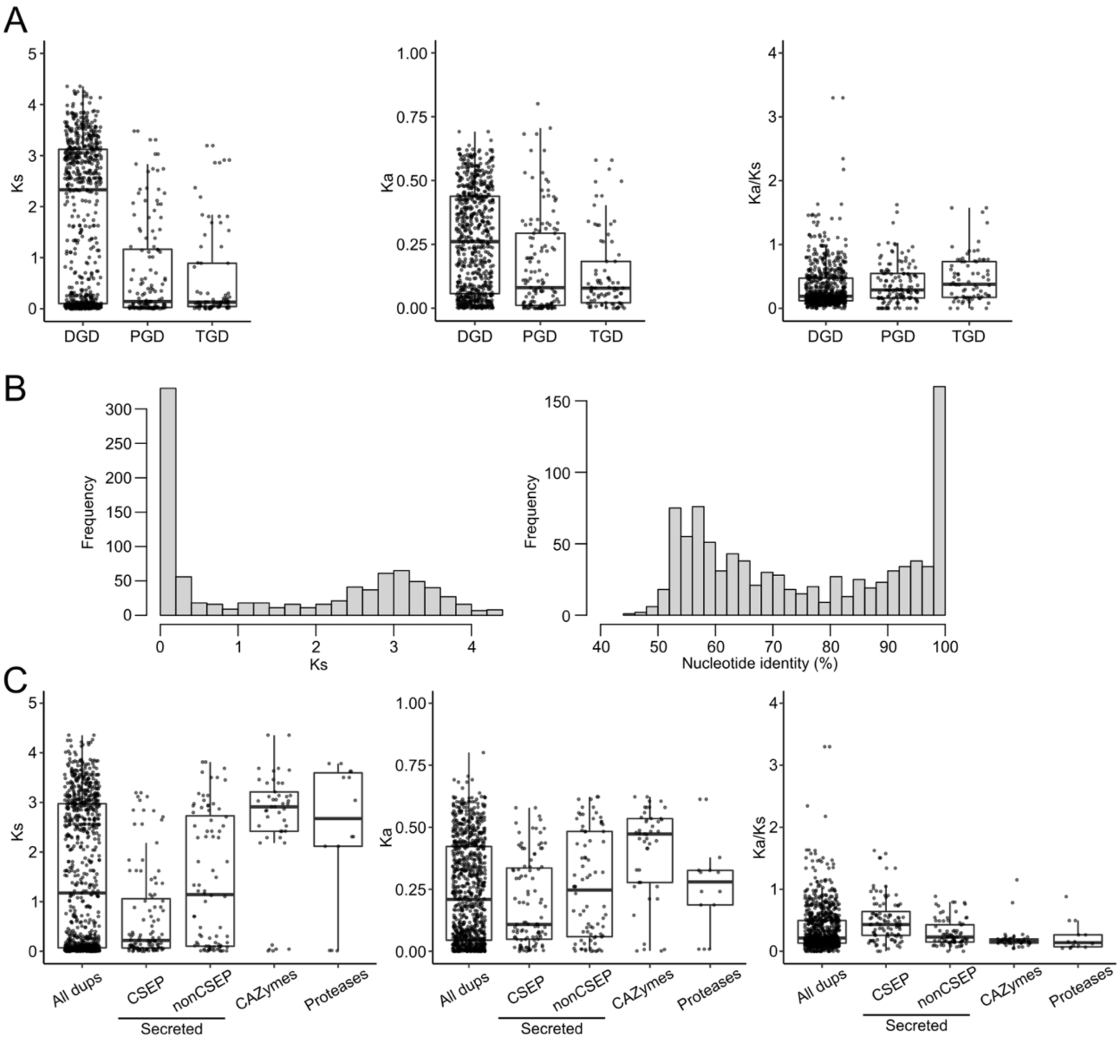


Figure 5.S16: Local gene duplicates in the genome of *Erysiphe necator* isolate EnFRAME01 are likely younger in age, and exhibit more relaxed selection pressure than dispersed gene duplicates. (A) Box-plots showing the distribution values of synonymous substitutions per synonymous site (K_S), non-synonymous substitutions per non-synonymous site (K_A), and of the K_A/K_S ratio, calculated for pairs of gene duplicates in the genome of EnFRAME01. Proximal gene duplicates (PGDs) and tandem gene duplicates (TGDs) have lower median K_S values as compared to the dispersed gene duplicates (DGDs), which further experience a bimodal distribution with peaks at K_S≈3-3.5 and K_S≈0.05 in the distribution of K_S values. Moreover, PGDs and TGDs exhibit higher median K_A/K_S values that are closer to 1 as compared to DGDs, indicating more relaxed selection pressure. (B) The distributions of K_S (left panel) and nucleotide identity values (right panel) of duplicated genes in the genome of EnFRAME01 do not exhibit an L-shaped pattern as secondary peaks are observed at the tail of the two distributions. K_S and amino-acid identity values were calculated based on pairwise comparisons of each duplicated gene with its most homologous paralog. (C) Duplicated genes encoding candidate secreted effector proteins (CSEPs) are likely of younger age, and under more relaxed selection pressure as compared to other functional gene categories in the genome of EnFRAME01. The box-plots show the distribution of K_S, K_A, and K_A/K_S values calculated for pairs of gene duplicates organized into different gene functional categories.

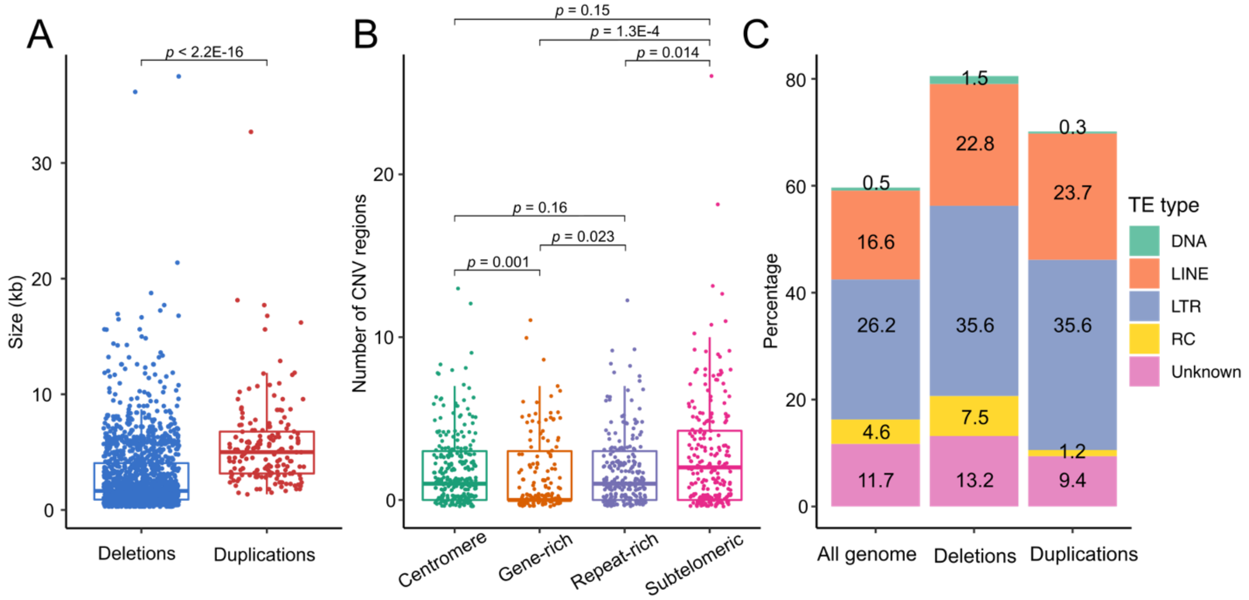


Figure 5.S17: Copy number variation (CNV) regions in the genome of *Erysiphe necator* isolate EnFRAME01 differ in size, repeat content, and are less frequently observed in gene-rich regions. (A) Size distribution of CNV regions. Each point corresponds to a CNV region. (B) Box plots showing the overall number of CNV regions within centromeric regions, gene-rich regions, repeat-rich regions, and subtelomeric regions. Each point corresponds to a 50 kb window containing the number of CNVs shown in the Y-axis. There is a total of 257 windows from centromeric regions, 144 windows from gene-rich regions (i.e. 50 kb windows containing at least 10 predicted genes), 230 windows from repeat-rich regions (i.e. 50 kb windows containing at least 90% repeats), and 220 windows from subtelomeric regions (i.e. within 500 kb from chromosome ends). (C) Bar plots showing the percentage of transposable element (TE) content within CNV regions compared to the entire genome. *P*-values are of Wilcoxon rank sum tests.

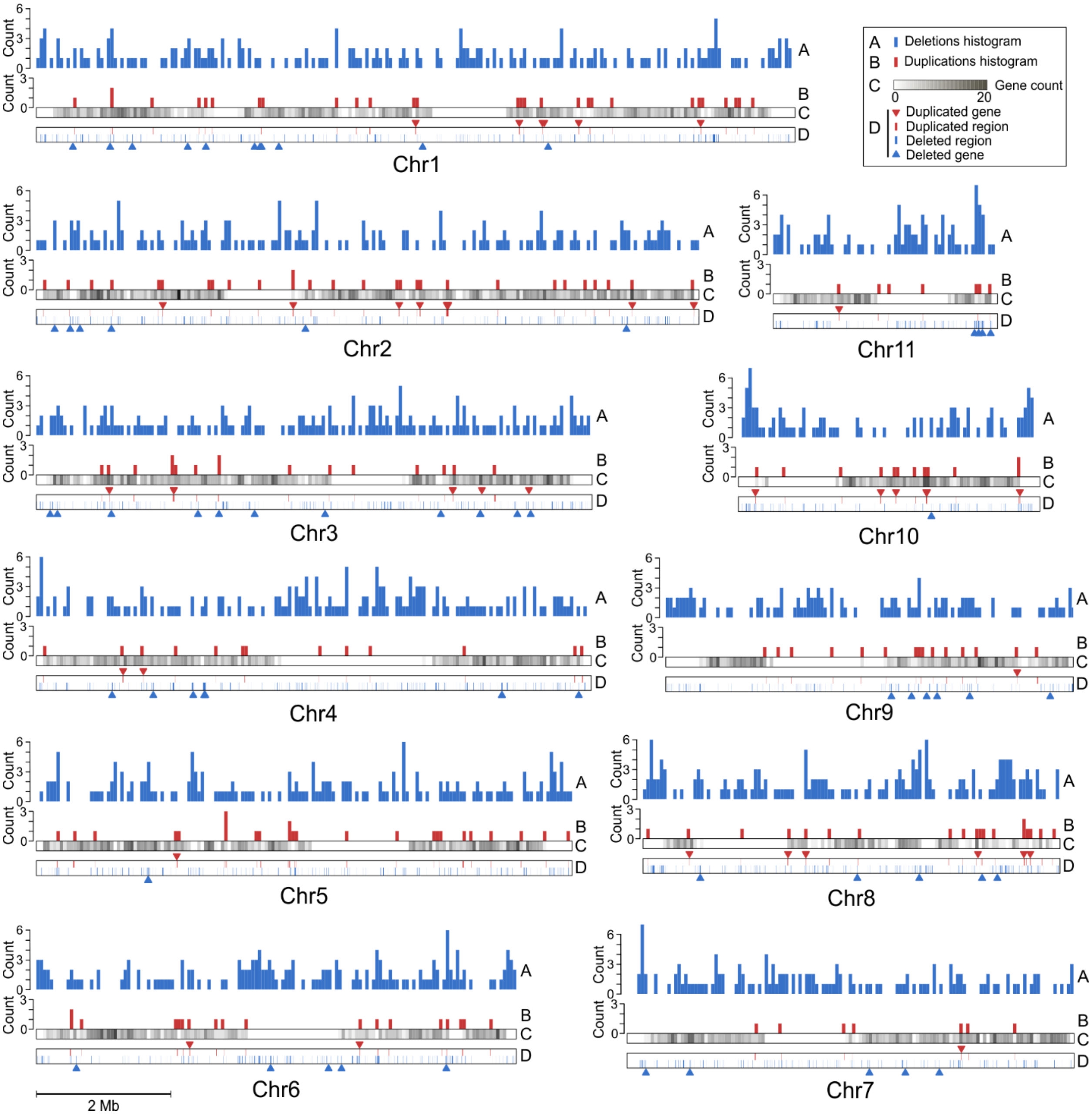


Figure 5.S18: Distribution of copy number variation (CNV) regions in the chromosomes of *Erysiphe necator* isolate EnFRAME01. Tracks (A) and (B) show histograms of the number of deletions and duplications, respectively. Track (C) shows the number of predicted genes in grey scale. Track (D) shows the location of duplications and deletions with rectangles in the upper and lower halves of the track, respectively. Location of protein encoding genes affected by duplications and deletions are indicted with upside down and upright triangles, respectively. Values shown in the tracks (A), (B), and (C) were obtained using a sliding window of 50 kb. CNV regions shown were obtained by merging overlapping CNV regions identified by analyzing the *E. necator* isolates C-strain, e1-101, Lodi, Branching, and Ranch9.

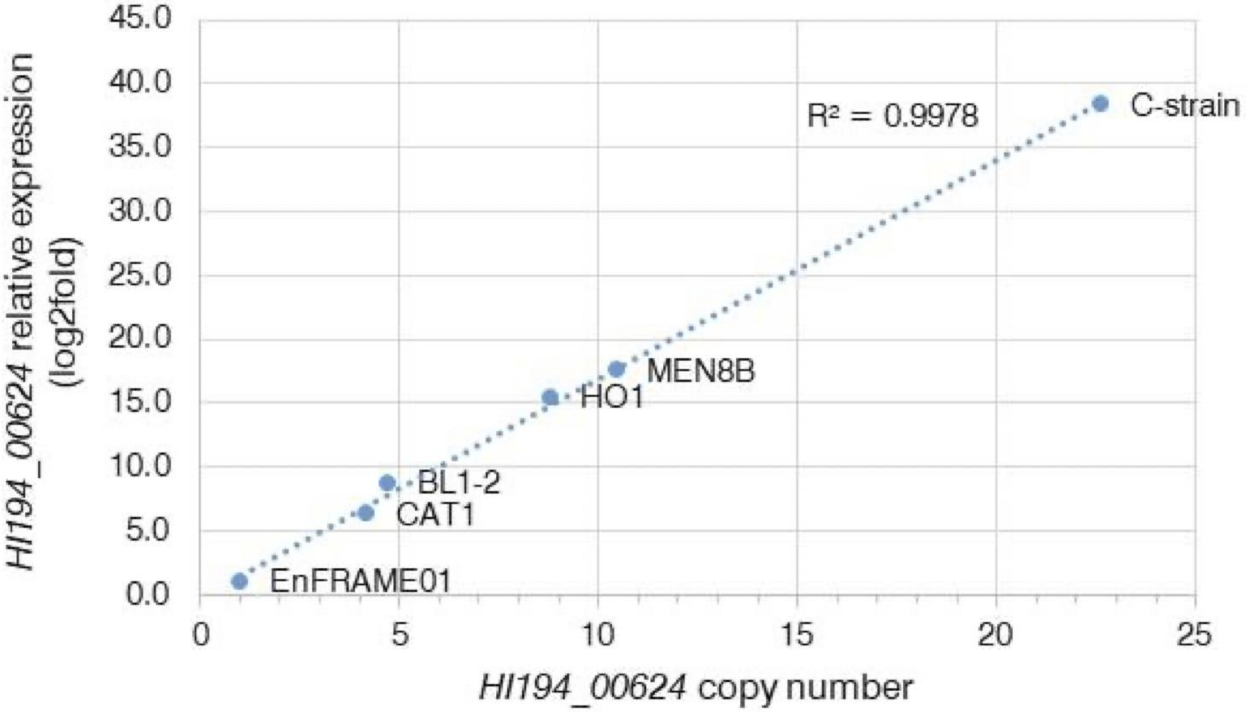


Figure 5.S19: The relative expression of the HI914_00624 gene encoding a secreted carboxylesterase (CE) is strongly correlated to its copy numbers. The scatter plot shows the linear relationship between HI914_00624 copy number (X-axis) as determined by qPCR in six old isolates of *E. necator* (i.e. EnFRAME01, CAT1, BL1-2, HO1, MEN8B, and C-strain) and HI914_00624 expression (Y-axis) as determined by RT-qPCR in conidia of the isolates harvested from detached *Vitis vinifera* L., cv. 'Chardonnay' leaves at 2-weeks post inoculation. Both qPCR assays used the single-copy EnEF1 (*E. necator* elongation factor 1) gene as reference and EnFRAME01 as the reverence isolate. Relative gene expression calculated using the ΔΔCT method where RQ=$2^{-\Delta\Delta CT}$.

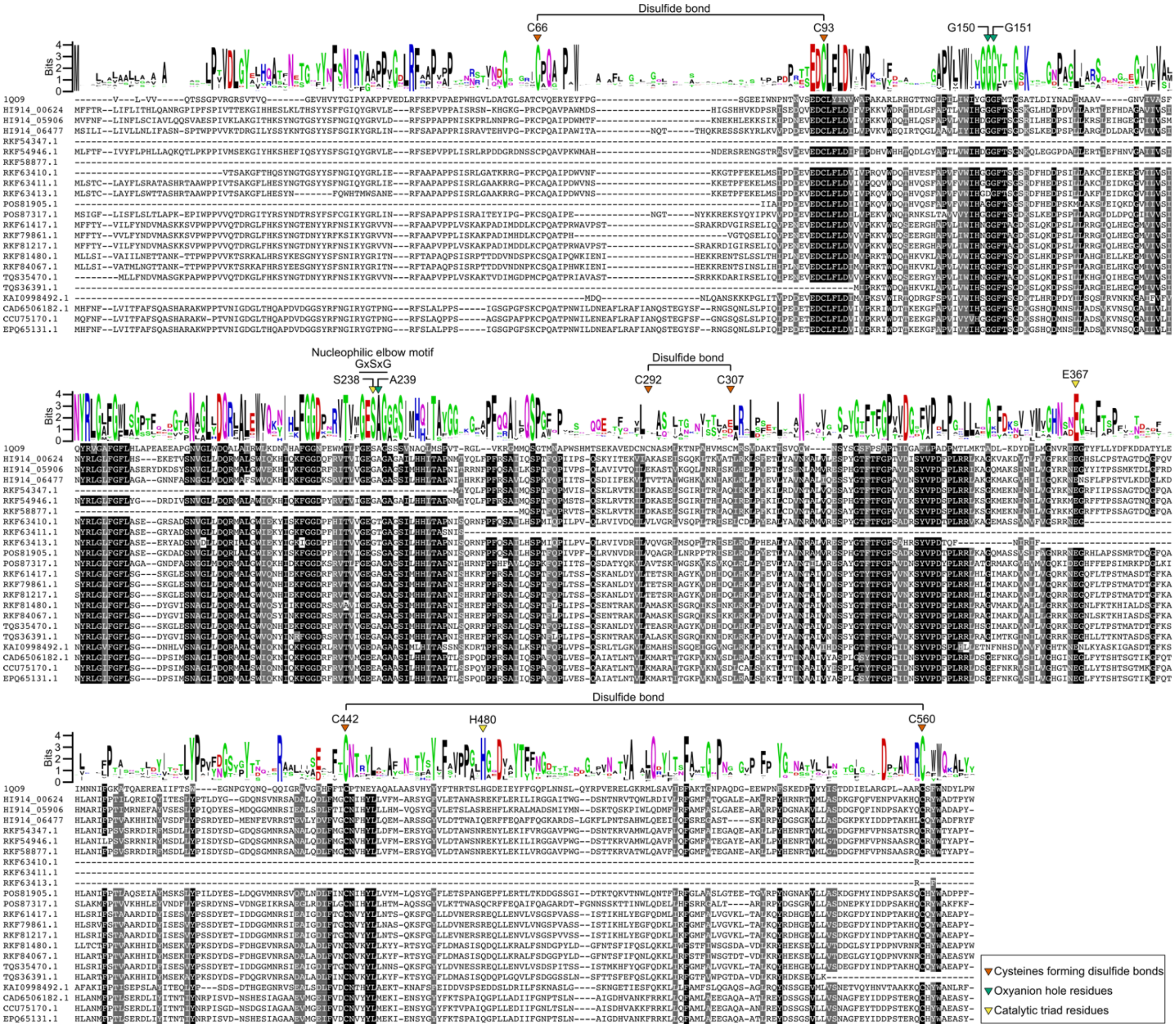


Figure 5.S20: The secreted carboxylesterase (CE) encoded by the *HI914_00624* gene in *Erysiphe necator* isolate EnFRAME01 as well as its homologs in powdery mildew genomes (PM) are likely catalytically inactive. The figure shows an amino-acid alignment of HI914_00624, its homologs in other PM genomes, and the acetylcholinesterase DmAChE from *Drosophila melanogaster* (1QO9), for which the 3D structure has been elucidated. The sequence logo (max = $\log_2 20$ = 4.3 bits) was inferred from the trimmed alignment of HI914_00624 and 400 most similar sequences found in GenBank, together with DmAChE as an outgroup (Fig 5.5C). Sites with at least 80% identity or similarity among the 23 sequences shown are highlighted. Conserved residues in DmAChE are indicated as reported by (Oakeshott et al. 2005). The figure shows that HI914_00624 and its homologs in PM genomes do not have the conserved Ser (S238) and His (H480) residues of the catalytic triad Ser-Asp/Glu-His required for the proper function of CEs. The PM species of each sequence is shown in Table 5.S23.

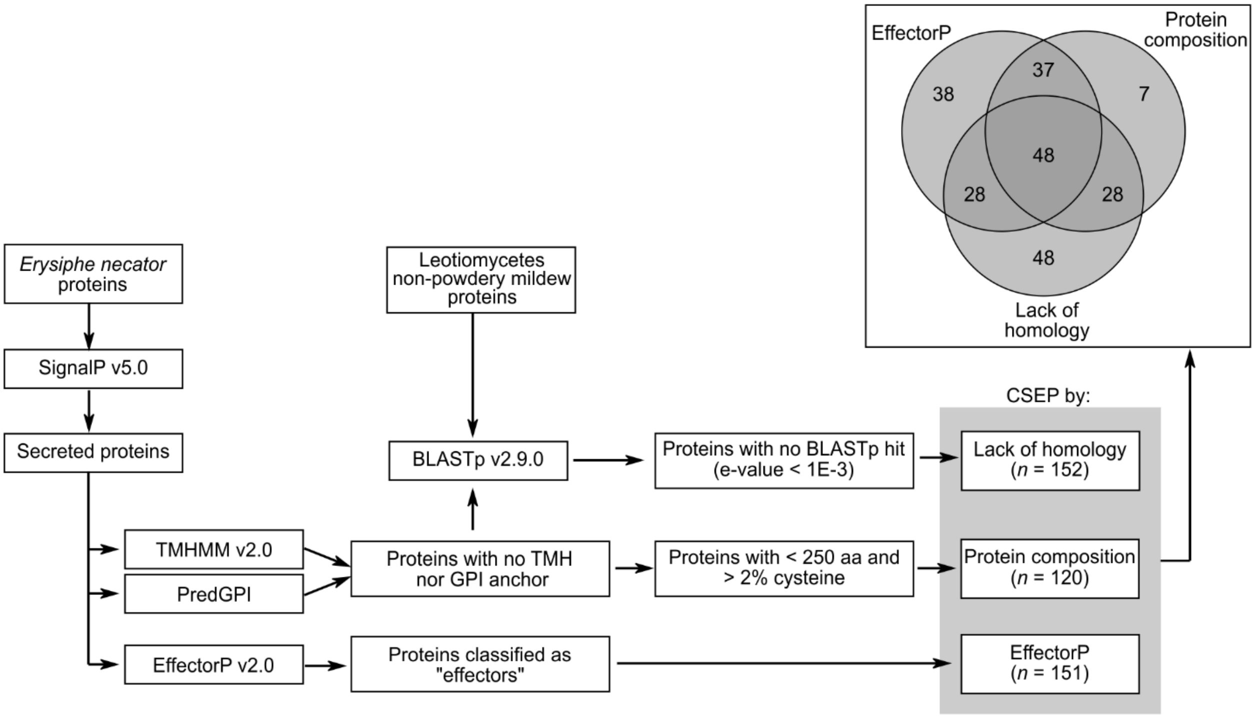

EffectorP
Protein composition
38
37
7
48
28
28
48
Lack of homology
Erysiphe necator proteins
SignalP v5.0
Secreted proteins
Leotiomycetes non-powdery mildew proteins
BLASTp v2.9.0
Proteins with no BLASTp hit (e-value < 1E-3)
CSEP by:
Lack of homology (*n* = 152)
TMHMM v2.0
PredGPI
Proteins with no TMH nor GPI anchor
Proteins with < 250 aa and > 2% cysteine
Protein composition (*n* = 120)
EffectorP v2.0
Proteins classified as "effectors"
EffectorP (*n* = 151)

Figure 5.S21: Workflow used to identify candidate secreted effector proteins (CSEPs) in *Erysiphe necator* isolate EnFRAME01. CSEPs were identified based on three lines of evidence: i) secreted proteins with no transmembrane domain and no glycosylphosphatidylinositol (GPI) anchor with no homologs in non-powdery mildew Leotiomycetes, ii) secreted proteins with no transmembrane domain and no GPI anchor with at most 250 aa and at least 2% cysteine residues, and iii) secreted proteins classified as effector by EffectorP. The number of CSEPs identified by each of these three lines of evidence are indicated on the right-hand side.

5.7.4 Supplementary tables

Table 5.S1: Genome assembly statistics of *Erysiphe necator* isolate EnFRAME01 compared to *E. necator* isolate C-strain, and the cereal powdery mildew pathogens *Blumeria graminis* f.sp. *hordei* isolate DH14 v4, *B. graminis* f.sp. *tritici* isolate v3.16, and *B. graminis* f.sp. *triticale* isolate THUN12. BUSCO completeness was estimated using the fungi orthologous data set odb10 (n=758 BUSCOs) at the genome level.

	E. necator **EnFRAME01**	***E. necator*** **C-strain**	***B. graminis*** **f.sp. *hordei*** **DH14**	***B. graminis*** **f.sp. *tritici*** **v3.16**	***B. graminis*** **f.sp. *triticale*** **THUN12**
Assembly size (bp)	81,105,680	52,505,057	124,489,486	140,802,721	141,403,166
Number of scaffolds	34	5935	318	12	36
Size of longest scaffold (bp)	11,303,845	188,576	9,852,665	19,721,624	19,725,520
GC%	39.8	38.9	43.7	43.7	43.7
N50	7,982,397	21,433	4,574,654	15,587,074	15,157,869
L50	5	710	8	4	4
N90	4,493,462	3975	752,644	10,259,397	10,894,025
L90	10	2775	32	9	9
Number of gaps	7	2659	120	696	18
Gaps (%)	< 0.1	0.5	0.5	0.1	< 0.1
BUSCO completeness:					
Complete	98.2	98.4	99.2	98.7	99.0
Duplicated	0.3	0.1	0.8	0.9	0.8
Fragmented	0.8	0.8	0.4	0.5	0.4
Missing	1.0	0.8	0.4	0.8	0.6

Table 5.S2: Comparison of the predicted centromere sizes of *Erysiphe necator* isolate EnFRAME01 and other species of Ascomycete. Centromere size range shows the size of the smallest and longest centromere for the species shown on the left. Total centromere size shows the total number of nucleotides contained within centromeric regions. Centromeres of *E. necator* are much longer and represent a much higher percentage of the genome compared to other ascomycetes.

Species	Centromere size range (Kb)	Total centromere size (Kb)	Genome size (bp)	Percentage of genome	Reference
Fusarium graminearum PH-1	56 - 65	239	36,563,796	0.7%	King et al. (2015)
Magnaporthe oryzae Guy11	57 - 109	604	37,356,090	1.6%	Yadav et al. (2019)
Verticillium dahliae JR2	94 - 187	1300	36,150,287	3.5%	Seidl et al. (2020)
Zymoseptoria tritici IPO323	6 -14	213	39,686,251	0.5%	Schotanus et al. (2015)
Neurospora crassa	174 - 287	1700	41,102,378	4.2%	Smith et al. (2011)
Erysiphe necator EnFRAME01	850 - 2,000	12800	81,105,680	15.8%	This study

Table 5.S3: Gene annotation statistics of *Erysiphe necator* isolate EnFRAME01 compared to *E. necator* isolate C-strain, *Blumeria graminis* f.sp. *hordei* isolate DH14 v4, B. *graminis* f.sp. *tritici* isolate v3.16, and *B. graminis* f.sp. *triticale* isolate THUN12. BUSCO completeness was estimated using the fungi orthologous data set odb10 (n=758 BUSCOs) at the protein level.

	E. necator **EnFRAME01**	***E. necator*** **C-strain**	***B. graminis*** **f.sp. *hordei* DH14**	***B. graminis*** **f.sp. *tritici* v3.16**	***B. graminis*** **f.sp. *triticale* THUN12**
Number of predicted genes	7146	6484	7118	8348	7993
Average gene size (bp)	1469	1557	2262	1209	1417
Average protein size (aa)	443	472	476	373	428
Average exon size (bp)	478	527	713	516	500
Average intron size (bp)	76	81	73	74	83
Gene density (genes per Mb)	88	123	57	59	56
Number of multi-exon genes	5947	4859	6016	4942	5745
Number of single-exon genes	1199	1625	1102	3406	2248
BUSCO completeness:					
Complete	98.2	90.1	98.7	87.8	98.5
Duplicated	0.4	0.1	0.9	0.1	1.1
Fragmented	0.9	1.2	0.8	3.3	0.9
Missing	0.9	8.7	0.5	8.9	0.6

Table 5.S4: Genes predicted to encode proteases in the genome of *Erysiphe necator* isolate EnFRAME01. The table shows the location of the genes and their classification as a member of the protease classes aspartic (A), cysteine (C), inhibitor (I), metallo (M), serine (S), and threonine (T). This table is available at https://zenodo.org/records/11211529.

Table 5.S5: Genes predicted to encode key enzymes for secondary metabolism in the genome of *Erysiphe necator* isolate EnFRAME01. The table shows the location of the genes and their classification as nonribosomal peptide synthetase (NRPS), NRPS-like, type 1 or type 3 polyketide synthase (PKS), and terpene synthase.

Gene ID	Chromosome	Start	End	Key enzyme
HI914_00236	chr01	1961932	1965546	NRPS-like
HI914_04002	chr05	1718697	1733362	NRPS
HI914_04388	chr05	6407511	6412718	NRPS-like
HI914_06546	chr09	3705386	3711101	Type 1 PKS
HI914_07364	chr11	1058789	1060293	Type 3 PKS
HI914_01229	chr02	1404428	1405617	Terpene synthase
HI914_04296	chr05	5606163	5608122	Terpene synthase
HI914_04683	chr06	1034043	1035624	Terpene synthase

Table 5.S6: Genes predicted to encode cytochrome P450 enzymes in the genome of *Erysiphe necator* isolate EnFRAME01.

Gene	Chromosome	Start	End	Classification
HI914_00178	Chr1	1549481	1551051	CYP53
HI914_01642	Chr2	5637684	5639433	CYP617
HI914_01650	Chr2	5714116	5715797	CYP51
HI914_05505	Chr7	3759151	3760884	unclassified
HI914_05519	Chr7	3882090	3883833	unclassified
HI914_05544	Chr7	4066390	4068133	unclassified
HI914_06393	Chr9	1026701	1027858	unclassified
HI914_06503	Chr9	3309265	3310915	CYP52
HI914_06959	Chr10	2297818	2301211	unclassified
HI914_07419	Chr11	1388030	1389870	CYP544
HI914_07477	Chr11	2907494	2909112	unclassified

Table 5.S7: Genes predicted to encode transporters in the genome of *Erysiphe necator* isolate EnFRAME01. The table shows the location of the genes and their classification according to the transporter classification data base (TCDB). Transporters of the ATP-binding Cassette (ABC) Superfamily (ID 3.A.1) and Major Facilitator Superfamily (MFS) (ID 2.A.1) were further classified into subfamilies. This table is available at https://zenodo.org/records/11211529.

Table 5.S8: Genes predicted to encode carbohydrate-active enzymes (CAZymes) in the genome of *Erysiphe necator* isolate EnFRAME01. A summary shows the number of predicted CAZyme modules from the classes glycoside hydrolase (GH), auxiliary activity (AA), glycosyl transferase (GT), carbohydrate esterase (CE), and carbohydrate-binding module (CBM). CAZymes predicted to act on plant cell wall (i.e. plant cell wall degrading enzymes (PCWDE)) and the corresponding substrates are indicated. The table below shows the genomic location and classification of each predicted CAZyme. This table is available at https://zenodo.org/records/11211529.

Table 5.S9: Genes predicted to encode secreted proteins in the genome of *Erysiphe necator* isolate EnFRAME01. The table shows the location in the chromosomes of all genes encoding secreted proteins. Secreted proteins classified as carbohydrate-active enzymes (CAZymes) or proteases are also shown with their respective classification. This table is available at https://zenodo.org/records/11211529.

Table 5.S10: Genes encoding candidate secreted effector proteins (CSEPs) in the genome of *Erysiphe necator* isolate EnFRAME01. The table shows the location of the genes in the chromosomes, their size, number of cysteines, presence/absence of a Y/F/WxC motif, number of BLASTp hits (e-value < 1E-3) against other Leotiomycetes non-powdery mildews, conserved domains, which ones were mapped to other *E. necator* isolates, and the most similar proteins (BLASTp e-value < 1E-5) in other powdery mildews and in the Pathogen-Host Interaction (PHI) database v4.13. The CSEPs were mapped to the genomes of other *E. necator* isolates using Liftoff or exonerate. This table is available at https://zenodo.org/records/11211529.

Table 5.S11: Predicted core genes missing in the genome of *Erysiphe necator* isolate EnFRAME01. The table shows 195 hierarchical orthogroups (HOGs) containing protein sequences conserved in the five non-obligate fungi analyzed, but missing in the genome of *E. necator*. The number of proteins contained in the HOG are shown. The table also shows descriptions of the *Saccharomyces cerevisiae* genes (i.e. accession numbers, gene names, product, KEGG Ontology IDs, enzyme commission number, pathways associated with the respective genes, and general notes). HOGs containing Missing Ascomycete Core Genes (MACGs) previously described in Spanu et al. (2010) are indicated. This table is available at https://zenodo.org/records/11211529.

Table 5.S12: Presence/absence of Missing Ascomycete Core Genes (MACGs) in *Erysiphe necator* isolate EnFRAME01. The table shows the 99 MACGs missing in *Blumeria graminis* as reported in Spanu et al. (2010), and their presence or absence in *E. necator*. In cases of presence, the gene IDs of *E. necator* are shown. Missing genes predicted with Orthofinder were grouped into hierarchical orthogroups (HOGs) with no representative from *E. necator*. Further BLASTp searched (e-value< 1e-5) were carried out for confirmation. Homologs of MACGs likely pseudogenized in *Saccharomyces cerevisiae* strain S288c were not reported. This table is available at https://zenodo.org/records/11211529.

Table 5.S13: KEGG pathway enrichment analysis of core genes missing in the genome of *Erysiphe necator* isolate EnFRAME01. This table is available at https://zenodo.org/records/11211529.

Table 5.S14: Genes predicted to encode proteins similar to RNase-like proteins expressed in haustoria (RALPHs) in the genome of *Erysiphe necator* isolate EnFRAME01. Genes were classified into two major clades based on a phylogenetic analysis (Fig 5.S9). Genes encoding proteins with a predicted signal peptide or a conserved microbial ribonuclease (RNase) domain (cl00212) are indicated. The right-most columns show the identity percentage of the top BLASTp hit (e-value< 1E-5) against the genomes of other powdery mildews. This table is available at https://zenodo.org/records/11211529.

Table 5.S15: Genes predicted to encode Egh16-like candidate secreted effector proteins (CSEPs) in the genome of *Erysiphe necator* isolate EnFRAME01. Genes were classified into two major clades based on a phylogenetic analysis (Fig 5.S10). Genes encoding proteins with a predicted signal peptide or a conserved Egh16-like virulence factor domain (PF11327) are indicated. The right-most columns show the identity percentage of the top BLASTp hit against the genomes of other powdery mildews. This table is available at https://zenodo.org/records/11211529.

Table 5.S16: Estimated abundances of transposable elements (TEs) in the genomes of *Erysiphe necator* isolate EnFRAME01, *Blumeria graminis* f.sp. *hordei* isolate DH14 v4, *B. graminis* f.sp. *tritici* isolate v3.16, and *B. graminis* f.sp. *triticale* isolate THUN12. The table shows the percentages of the genomes and number of bases masked by TE classes or superfamilies. Bases masked twice (i.e. overlapping repeats) were not considered in order to estimate the percentage of repeats of each class or family of transposable elements more accurately. Custom de novo repeat libraries were generated with RepeatModeler v2 for each genome and then used to mask the genome with RepeatMasker. The output of RepeatMasker was parsed with the script parseRM.pl (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs), which counted the number of masked per TE (super)family. DNA transposons and retrotransposons are further classified into Rolling-Circles (RCs), Long Interspersed Nuclear Elements (LINEs), Long Terminal Repeats (LTRs), Short Interspersed Elements (SINEs), and Retroposons. The respective TE superfamily codes proposed by Wicker et al. (2007) are indicated, if available. This table is available at https://zenodo.org/records/11211529.

Table 5.S17: The genome of *Erysiphe necator* isolate EnFRAME01 exhibits small-scale compartmentalization of candidate effectors in genes-sparse, repeat-rich regions. The table shows the mean and median sizes of intergenic regions, as well as the mean and median percentages of repetitive DNA content in intergenic regions of genes encoding candidate secreted effector proteins (CSEPs), secreted proteins not classified as CSEPs, carbohydrate-active enzymes (CAZymes), proteases, and Benchmarking Universal Single-Copy Orthologs (BUSCO) genes conserved among fungi. The table shows that genes encoding CEP have longer intergenic regions richer in repetitive DNA compared to other gene categories. P-values of Wilcoxon rank sum tests comparing the distributions of intergenic sizes and percentages of repetitive DNA in intergenic regions between CSEP-coding genes and other gene categories are shown.

	Intergenic size (bp)			**Intergenic repeat percentage**		
Gene category	**Mean**	**Median**	**Wilcoxon rank sum test *p*-value**	**Mean**	**Median**	**Wilcoxon rank sum test *p*-value**
CSEP	11755	6611	-	48.6	54.3	-
Secreted, not CSEP	8204	5384	0.001358	37.3	37.8	4.15E-08
CAZymes	6648	4366	1.09E-06	34.9	29.2	2.69E-09
Proteases	9373	3546	2.01E-12	31.1	26.9	5.96E-15
BUSCO	6306	2976	< 2.2e-16	31.7	24.8	< 2.2e-16

Table 5.S18: Genes encoding candidate secreted effector proteins (CSEPs) are significantly enriched with duplicated genes in *Erysiphe necator* isolate EnFRAME01. The table shows p-values of over- and under-representation of gene categories based on hypergeometric tests.

Category	Number of duplicated genes from category	Number of duplicated genes not from category	Total number of genes from category	Total number of genes not from category	Under-representation *p*-value	Over-representation *p*-value
CSEP	107	834	234	6912	1	1.80E-35
Secreted, not CSEP	84	857	293	6853	1	6.35E-13
CAZymes	47	894	160	6986	1	4.05E-08
Proteases	14	927	174	6972	0.02293625	0.9879784

Table 5.S19: Conserved domains enriched within duplicated genes in the genome of *Erysiphe necator* isolate EnFRAME01. The genes containing the respective domains are shown in the last column. This table is available at https://zenodo.org/records/11211529.

Table 5.S20: Summary statistics of copy number variation (CNV) regions in the genome of *Erysiphe necator* isolate EnFRAME01. The table shows the number, total size, average size, and density of deletions and duplications in the 11 chromosomes of EnFRAME01. This table is available at https://zenodo.org/records/11211529.

Table 5.S21: Genes varying in copy number among *Erysiphe necator* isolates. Location of the genes in the genome of *E. necator* isolate EnFRAME01 is shown. Genes predicted to be duplicated or single-copy in the genome of EnFRAME01 are indicated. Consecutive CNV genes in the genome of EnFRAME01 were assigned to clusters (clusters 1 to 12). The predicted copy number of the genes are indicated for isolates C-strain, Branching, Ranch9, e1-101, and Lodi. Genes predicted to encode secreted proteins, candidate secreted effector proteins (CSEP), carbohydrate-active enzymes (CAZymes), and proteases are indicated. The last three columns indicate the conserved domains, KEGG orthology (KO) identifier, and the functional description of the genes. This table is available at https://zenodo.org/records/11211529.

Table 5.S22: Genes varying in copy number among *Erysiphe necator* strains are enriched with genes considered duplicated in *E. necator* isolate EnFRAME01. The table shows p-values of over- and under-representation of gene categories based on hypergeometric tests.

Category	Number of CNV genes from category	Number of CNV genes not from category	Total number of genes from category	Total number of genes not from category	Under-representation *p*-value	Over-representation *p*-value
Duplicated	66	56	941	6205	1	1.76E-27
Duplicated dispersed	41	81	712	6434	1	5.88E-13
Duplicated proximal	17	105	139	7007	1	1.38E-10
Duplicated tandem	8	114	90	7056	0.9999798	1.38E-04
CSEP	6	116	234	6912	0.8951491	0.2100196
Secreted, not CSEP	9	113	527	6619	0.5879494	0.5501289
CAZymes	4	118	160	6986	0.8620421	0.291632
Proteases	3	119	174	6972	0.654019	0.5747693

Table 5.S23: Estimated number of copies of the gene *HI914_00480* in the genome of five isolates of *Erysiphe necator*. The table shows the average read depth at the locus of each copy of *HI914_00480* in the genome of *E. necator* EnFRAME01 (GCA_024703715.1), using whole-genome sequencing Illumina reads from isolates C-strain (SRR1448450), Branching (SRR1448453), Ranch 9 (SRR1448454), e1-101 (SRR1448468), and Lodi (SRR1448470). The copy number of *HI914_00480* in the five *E. necator* isolate was estimated as the sum of the average depth of all 20 copies divided by the coverage of the whole genome ± standard deviation. The coverage for the whole genome was estimated as the average read depth of 741 BUSCO genes (fungi_odb10; https://busco-data.ezlab.org/v4/data/lineages/).

Chromosome	Start	End	Copy #	C-strain	Branching	Ranch 9	e1-101	Lodi
Chr1	4477276	4477612	copy_1	0	4.45	0	42.52	0
Chr1	4495773	4496106	copy_2	0	0	0	0.35	0.3
Chr1	4503875	4504208	copy_3	0	0.72	0.34	0.23	0
Chr1	4511986	4512319	copy_4	0	0.44	0.62	0.81	0.43
Chr1	4525449	4525782	copy_5	0	0.41	0	0.33	0.18
Chr1	4533615	4533948	copy_6	0	0	0.96	0	1.17
Chr1	4541805	4542138	copy_7	0	0.99	0.89	0.89	0
Chr1	4549911	4550244	copy_8	0	0	0	0	0.44
Chr1	4558006	4558339	copy_9	0	0.78	0.46	0.86	0.22
Chr1	4566105	4566438	copy_10	0	0	0.4	0	0.4
Chr1	4574227	4574560	copy_11	0	0.44	0	0	0
Chr1	4582393	4582726	copy_12	0	0	0.79	0	0
Chr1	4590579	4590912	copy_13	0	0.4	0	0	0
Chr1	4598804	4599137	copy_14	0	0	0	0	0.4
Chr1	4607050	4607383	copy_15	0	0	0.39	0	0
Chr1	4615315	4615648	copy_16	0	0.39	0	1.03	0.77
Chr1	4623522	4623855	copy_17	0	0	0	0	0
Chr1	4689606	4689939	copy_18	333.8	173.25	181.64	243.23	231.38
Chr1	4712338	4712740	copy_19	42.84	24.06	19.41	36.55	41.44
Chr1	4798663	4798999	copy_20	45.9	28.49	16.04	28.49	24.99
Sum				422.54	234.82	221.94	355.29	302.12
Whole genome coverage				46.18	28.41	22.09	37.92	34.03
Whole genome coverage standard deviation				7.65	2.42	3.48	4.75	4.3
Estimated copy number of HI914_00480				8 to 11	8 to 9	9 to 12	8 to 11	8 to 10

Table 5.S24: Quantitative PCR (qPCR) analysis of *HI914_00624* gene copy number and gene expression.

	En624-3 FAMHI914-00624 gene			**EnEF1 ABY Single Copy Gene**			**ΔCT**	**ΔΔCT**	**$2^{-\Delta\Delta CT}$**
Isolate	**Ct Mean**	**Ct SD**	**Ct Threshold**	**Ct Mean**	**Ct SD**	**Ct Threshold**	**(*HI914-00624-EnEF1*)**	**(isolate - EnFRAME01)**	**gene copy #/ expression**
RT-qPCR (cDNA template)									
EnFRAME01	34.9	0.3	0.3	24.7	0.1	0.3	10.2	0.0	1
C-Strain	26.4	0.2	0.3	21.5	0.0	0.3	4.9	-5.3	38
HO1	31.6	0.5	0.3	25.4	0.0	0.3	6.2	-4.0	15
MEN8B	27.7	0.3	0.3	21.7	0.2	0.3	6.1	-4.1	18
CAT1	29.5	0.3	0.3	21.9	0.0	0.3	7.5	-2.6	6
BL1-2	32.8	0.3	0.3	25.7	0.1	0.3	7.1	-3.1	9
qPCR (gDNA template)									
EnFRAME01	28.1	0.1	0.3	28.5	0.1	0.3	-0.4	0.0	1
C-Strain	21.7	0.1	0.3	26.6	0.0	0.3	-4.9	-4.5	23
HO1	28.9	0.1	0.3	32.5	0.1	0.3	-3.6	-3.1	9
MEN8B	27.8	0.2	0.3	31.6	0.2	0.3	-3.8	-3.4	10
CAT1	30.2	0.1	0.3	32.6	0.2	0.3	-2.5	-2.1	4
BL1-2	27.6	0.1	0.3	30.3	0.1	0.3	-2.7	-2.2	5

Table 5.S25: Homologs in powdery mildew genomes of the predicted carboxylesterase gene *HI914_00624* varying in copy number in *Erysiphe necator*. The table shows the percent BLASTp identity of the homologs to HI914_00624, and the respective amino acids at the sites expected to have the conserved catalytic triad Ser-Glu-His of carboxylesterases.

Gene	Species	Isolate	Genome accession	% Identity to HI914_00624	Catalytic triad		
					Ser	Glu	His
HI914_00624	*Erysiphe necator*	EnFRAME01	GCA_024703715.1	100	Gly	Glu	Arg
HI914_05906	*Erysiphe necator*	EnFRAME01	GCA_024703715.1	56.2	Gly	Glu	Arg
HI914_06477	*Erysiphe necator*	EnFRAME01	GCA_024703715.1	48.4	Gly	Glu	Gln
EPQ65131.1	*Blumeria graminis* f.sp. *graminis*	96224	GCA_000418435.1	42.6	Glu	Glu	Gln
CCU75170.1	*Blumeria graminis* f.sp. *hordei*	DH14	GCA_000151065.3	42.8	Glu	Glu	Gln
CAD6506182.1	*Blumeria graminis* f.sp. *triticale*	THUN-12	GCA_905067625.1	42.6	Glu	Glu	Gln
RKF54347.1	*Erysiphe neolycopersici*	UMSG2	GCA_003610855.1	63.4	-	Glu	Arg
RKF54946.1	*Erysiphe neolycopersici*	UMSG2	GCA_003610855.1	64.3	Gly	Glu	Arg
RKF58877.1	*Erysiphe neolycopersici*	UMSG2	GCA_003610855.1	63.6	-	Glu	Arg
RKF63410.1	*Erysiphe neolycopersici*	UMSG2	GCA_003610855.1	55.4	Gly	Glu	-
RKF63411.1	*Erysiphe neolycopersici*	UMSG2	GCA_003610855.1	53.7	Gly	-	-
RKF63413.1	*Erysiphe neolycopersici*	UMSG2	GCA_003610855.1	49.1	Gly	-	-
POS81905.1	*Erysiphe pulchra*	Cflorida	GCA_002918395.1	56.8	Gly	Glu	Gly
POS87317.1	*Erysiphe pulchra*	Cflorida	GCA_002918395.1	48.3	Gly	Glu	Gln
RKF61417.1	*Golovinomyces cichoracearum*	UCSC1	GCA_003611215.1	48.7	Gly	Glu	Arg
RKF79861.1	*Golovinomyces cichoracearum*	UMSG1	GCA_003611235.1	48.7	Gly	Glu	Arg
RKF81217.1	*Golovinomyces cichoracearum*	UMSG3	GCA_003611195.1	48.9	Gly	Glu	Arg
RKF81480.1	*Golovinomyces cichoracearum*	UMSG3	GCA_003611195.1	46.8	Gly	Glu	Gln
RKF84067.1	*Golovinomyces cichoracearum*	UMSG3	GCA_003611195.1	48.2	Gly	Glu	Gln
TQS35470.1	*Golovinomyces magnicellulatus*	FPH2017-1	GCA_006912115.1	50.2	Gly	Glu	Arg
TQS36391.1	*Golovinomyces magnicellulatus*	FPH2017-1	GCA_006912115.1	51.5	Gly	Glu	Gln
KAI0998492.1	*Podosphaera aphanis*	DRCT72020	GCA_022627015.1	47.4	Asp	Glu	Glu

Table 5.S26: Quantitative PCR (qPCR) primer and probe sequences, targets, parameters. En-g1817 primer sequences from Jones et al. (2014).

Oligo Name	Sequence (5'-3')	*Erysiphe necator* gene target	Final Concentration in qPCR
EN624-3 Forward	TTCCTTTCCGGTCTGCAATAA	*HI914_00624* multi-copy CE gene	500nM
EN624-3 Reverse	ATTTCGCTGGCCTTAGCTATAA	*HI914_00624* multi-copy CE gene	500nM
EN624-3 FAM Probe	[6FAM] TCAACCAATTATCCCGTCTCAATCGAAGT [QSY]	*HI914_00624* multi-copy CE gene	200nM
EnEF1 Forward	TGGAAAGTCTATTGAGGCAACTCC	*EnEF1/En-g1817* elongation factor single-copy gene	500nM
EnEF1 Reverse	CAACACACATAGGTTTAGATGGAATCA	*EnEF1/En-g1817* elongation factor single-copy gene	500nM
EnEF1 ABY Probe	[ABY] TTAACAATTGCTGCGTCACCAGACTTAA [QSY]	*EnEF1/En-g1817* elongation factor single-copy gene	200nM

5.8 Supplementary References

Abyzov, A., Urban, A. E., Snyder, M., and Gerstein, M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21:974–984

Alam, M., Vance, D. E., and Lehner, R. 2002. Structure−function analysis of human triacylglycerol hydrolase by site-directed mutagenesis: identification of the catalytic triad and a glycosylation site. Biochemistry. 41:6679–6687

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420–423

Bao, Z., and Eddy, S. R. 2002. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. Genome Res. 12:1269–1276

Blanc, G., and Wolfe, K. H. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell. 16:1667–1678

Bornscheuer, U. T. 2002. Microbial carboxyl esterases: classification, properties and application in biocatalysis. FEMS Microbiol. Rev. 26:73–81

Brewer, M. T., and Milgroom, M. G. 2010. Phylogeography and population structure of the grape powdery mildew fungus, *Erysiphe necator*, from diverse Vitis species. BMC Evol. Biol. 10:1–13

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., and Yandell, M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188–196

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

Chambon, C., Ladeveze, V., Oulmouden, A., Servouse, M., and Karst, F. 1990. Isolation and properties of yeast mutants affected in farnesyl diphosphate synthetase. Curr. Genet. 18:41–46

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., and Ma'ayan, A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 14:1–14

Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. 2004. WebLogo: A Sequence Logo Generator. Genome Res. 14:1188–1190

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., and others. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 356:92–95

Durairaj, P., Hur, J.-S., and Yun, H. 2016. Versatile biocatalysis of fungal cytochrome P450 monooxygenases. Microb. Cell Factories. 15:1–16

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., and Aiden, E. L. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3:99–101

Ehmann, D. E., Gehring, A. M., and Walsh, C. T. 1999. Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of α-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. Biochemistry. 38:6171–6177

Ellinghaus, D., Kurtz, S., and Willhoeft, U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 9:1–14

Emms, D. M., and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:1–14

Faust, G. G., and Hall, I. M. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 30:2503–2505

Feehan, J. M., Scheibel, K. E., Bourras, S., Underwood, W., Keller, B., and Somerville, S. C. 2017. Purification of high molecular weight genomic DNA from powdery mildew for long-read sequencing. J. Vis. Exp. :e55463

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., and Smit, A. F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. 117:9451–9457

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., and 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

Gan, P., Hiroyama, R., Tsushima, A., Masuda, S., Shibata, A., Ueno, A., Kumakura, N., Narusaka, M., Hoat, T. X., Narusaka, Y., and others. 2020. Subtelomeric regions and a repeat-rich chromosome harbor multicopy effector gene clusters with variable conservation in multiple plant pathogenic *Colletotrichum* species. bioRxiv. :2020.04.28.061093

Gladyshev, E. 2017. Repeat-induced point mutation and other genome defense mechanisms in fungi. Pages 687–699 in: The Fungal Kingdom, John Wiley & Sons, Ltd.

Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J., and Thordal-Christensen, H. 2010. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. BMC Genomics. 11:317

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., and others. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29:644

Grell, M. N., Mouritzen, P., and Giese, H. 2003. A *Blumeria graminis* gene family encoding proteins with a C-terminal variable region with homologues in pathogenic fungi. Gene. 311:181–192

Hage, H., and Rosso, M.-N. 2021. Evolution of fungal carbohydrate-active enzyme portfolios and adaptation to plant cell-wall polymers. J. Fungi. 7:185

Harel, M., Kryger, G., Rosenberry, T. L., Mallender, W. D., Lewis, T., Fletcher, R. J., Guss, J. M., Silman, I., and Sussman, J. L. 2000. Three-dimensional structures of *Drosophila melanogaster* acetylcholinesterase and of its complexes with two potent inhibitors. Protein Sci. 9:1063–1072

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522

Hu, Z., He, B., Ma, L., Sun, Y., Niu, Y., and Zeng, B. 2017. Recent advances in ergosterol biosynthesis and regulation mechanisms in *Saccharomyces cerevisiae*. Indian J. Microbiol. 57:270–277

Iqbal, M., Dubey, M., Gudmundsson, M., Viketoft, M., Jensen, D. F., and Karlsson, M. 2018. Comparative evolutionary histories of fungal proteases reveal gene gains in the mycoparasitic and nematode-parasitic fungus *Clonostachys rosea*. BMC Evol. Biol. 18:1–17

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K. C. H., McGuire, B., Gubler, W. D., Walker, M. A., and Cantu, D. 2014a. Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. BMC Genomics. 15:1081

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., and others. 2014b. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236–1240

Justesen, A., Somerville, S., Christiansen, S., and Giese, H. 1996. Isolation and characterization of two novel genes expressed in germinating conidia of the obligate biotroph *Erysiphe graminis* f.sp. *hordei*. Gene. 170:131–135

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermiin, L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589

Kanehisa, M., Sato, Y., and Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428:726–731

Kato, S., Motoyama, T., Uramoto, M., Nogawa, T., Kamakura, T., and Osada, H. 2020. Induction of secondary metabolite production by hygromycin B and identification of the 1233A biosynthetic gene cluster with a self-resistance gene. J. Antibiot. (Tokyo). 73:475–479

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

Keilwagen, J., Hartung, F., and Grau, J. 2019. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol. Biol. Clifton NJ. 1962:161–177

Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 12:357–360

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736

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

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

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645

Letunic, I., and Bork, P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49:W293–W296

Levan, A., Fredga, K., and Sandberg, A. A. 1964. Nomenclature for centromeric position on chromosomes. Hereditas. 52:201–220

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

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25:1754–1760

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

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

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., and others. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. science. 326:289–293

Lin, D. H., and Hoelz, A. 2019. The structure of the nuclear pore complex (an update). Annu. Rev. Biochem. 88:725–783

Lubbers, R. J., Dilokpimol, A., Peng, M., Visser, J., Mäkelä, M. R., Hildén, K. S., and De Vries, R. P. 2019. Discovery of novel *p*-hydroxybenzoate-*m*-hydroxylase, protocatechuate 3, 4 ring-cleavage dioxygenase, and hydroxyquinol 1, 2 ring-cleavage dioxygenase from the filamentous fungus *Aspergillus niger*. Acs Sustain. Chem. Eng. 7:19081–19089

Lukashin, A. V., and Borodovsky, M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26:1107–1115

Lynch, M., and Conery, J. S. 2003. The evolutionary demography of duplicate genes. Pages 35–44 in: Genome Evolution: Gene and Genome Duplications and the Origin of Novel Gene Functions, A. Meyer and Y. Van de Peer, eds. Springer Netherlands, Dordrecht.

Min, B., and Choi, I.-G. 2019. Practical Guide for Fungal Gene Prediction from Genome Assembly and RNA-Seq Reads by FunGAP. Pages 53–64 in: Gene Prediction, Springer.

Müller, M. C., Kunz, L., Graf, J., Schudel, S., and Keller, B. 2021. Host adaptation through hybridization: genome analysis of triticale powdery mildew reveals unique combination of lineage-specific effectors. Mol. Plant. Microbe Interact. 34:1350–1357

Müller, M. C., Praz, C. R., Sotiropoulos, A. G., Menardo, F., Kunz, L., Schudel, S., Oberhänsli, S., Poretti, M., Wehrli, A., Bourras, S., Keller, B., and Wicker, T. 2019. A chromosome-scale genome assembly reveals a highly dynamic effector repertoire of wheat powdery mildew. New Phytol. 221:2176–2189

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274

Oakeshott, J., Claudianos, C., Campbell, P., Newcomb, R., and Russell, R. 2005. Biochemical genetics and genomics of insect esterases. Pages 309–381 in: Biochemical genetics and genomics of insect esterases, Elsevier, Oxford, Amsterdam.

Omura, S., Tomoda, H., Kumagai, H., Greenspan, M. D., Yodkovitz, J. B., Chen, J. S., Alberts, A. W., Martin, I., Mochales, S., Monaghan, R. L., and others. 1987. Potent inhibitory effect of antibiotic 1233A on cholesterol biosynthesis which specifically blocks 3-hydroxy-3-methylglutaryl coenzyme A synthase. J. Antibiot. (Tokyo). 40:1356–1357

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

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. 2022. Biostrings: Efficient manipulation of biological strings. R Package Version 2640.

Påhlman, A.-K., Granath, K., Ansell, R., Hohmann, S., and Adler, L. 2001. The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress. J. Biol. Chem. 276:3555–3563

Pedersen, B. S., and Quinlan, A. R. 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 34:867–868

Pedersen, C., van Themaat, E. V. L., McGuffin, L. J., Abbott, J. C., Burgis, T. A., Barton, G., Bindschedler, L. V., Lu, X., Maekawa, T., Weßling, R., and others. 2012. Structure and evolution of barley powdery mildew effector candidates. BMC Genomics. 13:694

Pennington, H. G., Jones, R., Kwon, S., Bonciani, G., Thieron, H., Chandler, T., Luong, P., Morgan, S. N., Przydacz, M., Bozkurt, T., and others. 2019. The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA. PLoS Pathog. 15:e1007620

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290–295

Pierleoni, A., Martelli, P. L., and Casadio, R. 2008. PredGPI: a GPI-anchor predictor. BMC Bioinformatics. 9:392

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

Putterill, J. J., Plummer, K. M., Newcomb, R. D., and Marshall, S. D. G. 2003. The carboxylesterase gene family from *Arabidopsis thaliana*. J. Mol. Evol. 57:487–500

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

Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 42:D503–D509

Saier Jr, M. H., Reddy, V. S., Tamang, D. G., and Västermark, Å. 2014. The transporter classification database. Nucleic Acids Res. 42:D251–D258

Sandmann, G., Misawa, N., Wiedemann, M., Vittorioso, P., Carattoli, A., Morelli, G., and Macino, G. 1993. Functional identification of *al-3* from *Neurospora crassa* as the gene for geranylgeranyl pyrophosphate synthase by complementation with crt genes, *in vitro* characterization of the gene product and mutant analysis. J. Photochem. Photobiol. B. 18:245–251

Schmidhauser, T. J., Lauter, F.-R., Schumacher, M., Zhou, W., Russo, V., and Yanofsky, C. 1994. Characterization of *al-2*, the phytoene synthase gene of *Neurospora crassa*. Cloning, sequence analysis, and photoregulation. J. Biol. Chem. 269:12060–12066

Sharma, G., Aminedi, R., Saxena, D., Gupta, A., Banerjee, P., Jain, D., and Chandran, D. 2019. Effector mining from the *Erysiphe pisi* haustorial transcriptome identifies novel candidates involved in pea powdery mildew pathogenesis. Mol Plant Pathol. 20:1506–1522

Shumate, A., and Salzberg, S. L. 2020. Liftoff: accurate mapping of gene annotations. Bioinformatics.

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

Slater, G., and Birney, E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6:1–11

Sood, S., Sharma, A., Sharma, N., and Kanwar, S. S. 2018. Carboxylesterases: sources, characterization and broader applications. Insights Enzyme Res. 01:2

Spanu, P. D. 2017. Cereal immunity against powdery mildews targets RNase-Like Proteins associated with Haustoria (RALPH) effectors evolved from a common ancestral gene. New Phytol. 213:969–971

Spanu, P. D. 2012. The genomics of obligate (and nonobligate) biotrophs. Annu. Rev. Phytopathol. 50:91–109

Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., van Themaat, E. V. L., Brown, J. K., Butcher, S. A., Gurr, S. J., and others. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 330:1543–1546

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743–761

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439

Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., and Smit, A. F. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob. DNA. 12:1–14

Tiley, G. P., Barker, M. S., and Burleigh, J. G. 2018. Assessing the performance of *Ks* plots for detecting ancient whole genome duplications. Genome Biol. Evol. 10:2882–2898

Tobiasen, C., Aahman, J., Ravnholt, K. S., Bjerrum, M. J., Grell, M. N., and Giese, H. 2007. Nonribosomal peptide synthetase (NPS) genes in *Fusarium graminearum*, *F. culmorum* and *F. pseudograminearium* and identification of NPS2 as the producer of ferricrocin. Curr. Genet. 51:43–58

Tomoda, H., Kumagai, H., Takahashi, Y., Tanaka, Y., Iwai, Y., and Omura, S. 1988. F-244 (1233A), a specific inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A synthase: taxonomy of producing strain, fermentation, isolation and biological properties. J. Antibiot. (Tokyo). 41:247–249

Vanneste, K., Van de Peer, Y., and Maere, S. 2013. Inference of genome duplications from age distributions revisited. Mol. Biol. Evol. 30:177–190

Víglaš, J., and Olejníková, P. 2021. An update on ABC transporters of filamentous fungi–from physiological substrates to xenobiotics. Microbiol. Res. 246:126684

Wang, D., Zou, L., Jin, Q., Hou, J., Ge, G., and Yang, L. 2018. Human carboxylesterases: a comprehensive review. Acta Pharm. Sin. B. 8:699–712

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J. C., and Paterson, A. H. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40:e49–e49

Wang, Y., Wang, X., Tang, H., Tan, X., Ficklin, S. P., Feltus, F. A., and Paterson, A. H. 2011. Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms S.R. Proulx, ed. PLoS ONE. 6:e28150

Widhalm, J. R., and Dudareva, N. 2015. A Familiar Ring to It: Biosynthesis of Plant Benzoic Acids. Mol. Plant. 8:83–97

Winnenburg, R. 2006. PHI-base: a new database for pathogen host interactions. Nucleic Acids Res. 34:D459–D464

Wu, T. D., and Watanabe, C. K. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 21:1859–1875

Wu, Y., Ma, X., Pan, Z., Kale, S. D., Song, Y., King, H., Zhang, Q., Presley, C., Deng, X., Wei, C.-I., and Xiao, S. 2018. Comparative genome analyses reveal sequence features reflecting distinct modes of host-adaptation between dicot and monocot powdery mildew. BMC Genomics. 19:705

Yan, N. 2015. Structural biology of the major facilitator superfamily transporters. Annu. Rev. Biophys. 44:257–283

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J. Integr. Biol. 16:284–287

Zaccaron, A. Z., Chen, L.-H., Samaras, A., and Stergiopoulos, I. 2022. A chromosome-scale genome assembly of the tomato pathogen *Cladosporium fulvum* reveals a compartmentalized genome architecture and the presence of a dispensable chromosome. Microb. Genomics. 8:000819

Zaccaron, A. Z., De Souza, J. T., and Stergiopoulos, I. 2021. The mitochondrial genome of the grape powdery mildew pathogen *Erysiphe necator* is intron rich and exhibits a distinct gene organization. Sci. Rep. 11:13924

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46:W95–W101

Zhang, Y., Xu, K., Yu, D., Liu, Z., Peng, C., Li, X., Zhang, J., Dong, Y., Zhang, Y., Tian, P., and others. 2019. The Highly Conserved Barley Powdery Mildew Effector *BEC1019* Confers Susceptibility to Biotrophic and Necrotrophic Pathogens in Wheat. Int. J. Mol. Sci. 20:4376

Chapter 6

The mitochondrial genome of the grape powdery mildew pathogen *Erysiphe necator* is intron rich and exhibits a distinct gene organization

Alex Z. Zaccaron
Jorge T. De Souza
Ioannis Stergiopoulos

Leading author statement

I contributed to the conceptualization of this chapter, performed most of the analyses, generated most of the figures, contributed to the investigation and interpretation, wrote the original draft, and contributed to the review and editing of the final draft.

This chapter was published in: *Scientific reports*, July 2021, volume 11, number 1 (https://doi.org/10.1038/s41598-021-93481-5).

Abstract

Powdery mildews are notorious fungal plant pathogens but only limited information exists on their genomes. Here we present the mitochondrial genome of the grape powdery mildew fungus *Erysiphe necator* and a high-quality mitochondrial gene annotation generated through cloning and Sanger sequencing of full-length cDNA clones. The *E. necator* mitochondrial genome consists of a circular DNA sequence of 188,577 bp that harbors a core set of 14 protein-coding genes that are typically present in fungal mitochondrial genomes, along with genes encoding the small and large ribosomal subunits, a ribosomal protein S3, and 25 mitochondrial-encoded transfer RNAs (mt-tRNAs). Interestingly, it also exhibits a distinct gene organization with atypical bicistronic-like expression of the *nad4L*/*nad5* and *atp6*/*nad3* gene pairs, and contains a large number of 70 introns, making it one of the richest in introns mitochondrial genomes among fungi. Sixty-four intronic ORFs were also found, most of which encoded homing endonucleases of the LAGLIDADG or GIY-YIG families. Further comparative analysis of five *E. necator* isolates revealed 203 polymorphic sites, but only five were located within exons of the core mitochondrial genes. These results provide insights into the organization of mitochondrial genomes of powdery mildews and represent valuable resources for population genetic and evolutionary studies.

6.1 Introduction

Erysiphe necator (syn. *Uncinula necator*) is an obligate biotrophic ascomycete fungus that belongs to the Erysiphaceae family (Leotiomycetes; Erysiphales) and causes grape powdery mildew, one of the most widespread and destructive fungal diseases in vineyards across the world (Gadoury et al. 2012). The predicted 126 ± 18 Mb nuclear genome of *E. necator* was sequenced before from five isolates of the fungus that originated from organic vineyards (i.e., isolates Branching and e1-101) or fields that received regular fungicide applications for control of the pathogen (i.e., isolates C-strain, Lodi, and Ranch9) (Jones et al. 2014). The analysis revealed a highly repetitive genome with frequent structural variations among the isolates that likely play a role in the adaptive responses of the fungus to fungicide stress. However, apart from this study, genomic resources for *E. necator* are to this date relatively scarce and there is no public reference mitochondrial (mt) genome available for this pathogen.

Mitochondria are double-membrane bound organelles commonly recognized as the power factories of eukaryotic cells, due to their ability to produce energy through oxidative phosphorylation (Chan 2006; Richardson et al. 2010). They carry their own genomes that are contained within single circular chromosomes. In mammals, mt genomes are approximately 16.6 kb in length and contain genes that typically lack introns (Gustafsson et al. 2016; Taanman 1999). In contrast, fungal mt genomes vary remarkably in size, ranging from 12 kb in *Rozella allomycis* (James et al. 2013) to 272 kb in *Morchella importuna* (Liu et al. 2020), and harbor genes that may too show extensive variation in intron content.

In fungi, mt genomes contain a standard set of 14 core genes (i.e., *atp6*, *atp8*, *atp9*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *cob*, *cox1*, *cox2*, and *cox3*) that encode proteins involved in the electron transport chain (ETC) and oxidative phosphorylation (Seif et al. 2005). They also harbor two genes encoding the small and large ribosomal subunits (*rns* and *rnl*, respectively) and a set of mt-encoded transfer RNAs (mt-tRNAs). Two additional genes, *rps3* and *rnpB*, that code for the 40S ribosomal protein S3 and the RNA subunit of the mitochondrial RNase P, respectively, are also sporadically found in fungal mt genomes (Seif et al. 2005).

Although most fungi exhibit a relatively similar repertoire of mt genes, in contrast the order of these genes is usually not well conserved, even among species of the same genus (Li et al. 2018). Nonetheless, some commonalities in gene arrangements exist as well, as for example is the case for the gene pairs *nad4L/nad5* and *nad2/nad3*, which appear next to each other in the mt genomes of most fungal species (Aguileta et al. 2014).

Fungal mt genes also exhibit large variation in their intron numbers, which in some species may be completely absent, as for example in the wheat pathogen *Zymoseptoria tritici* (Torriani et al. 2008), while in others there might be as many as 80 introns, as for example in the 'blue-stain' fungus of conifers *Endoconidiophora resinifera* (Zubaer et al. 2018). In general, fungal mt introns are typically classified into group I and group II (Saldanha et al. 1993), with group I introns further being classified into seven subgroups, i.e., IA, IA3, IB, IC1, IC2, ID and I derived (I*) (Lang et al. 2007). In contrast to spliceosomal introns, group I and group II introns resemble mobile genetic elements and often harbor open reading frames (ORFs) encoding catalytic enzymes that enable intron self-splicing and transposition to an intronless cognate allele. In particular, group I mt introns typically contain ORFs that encode homing endonucleases (HEs) of the LAGLIDADG or GIY-YIG families, whereas group II introns typically encode reverse transcriptases (RTs). Although both group I and group II introns can be found in fungi, the majority of fungal mt introns are of group I and have been shown to exhibit extensive presence/absence variation, owing to their mobility and horizontal mode of transmission (Lang et al. 2007).

In this study, we present a comprehensive assembly and annotation of the mt genome of *E. necator* isolate C-strain. All core protein-coding genes had their annotation manually curated by cDNA cloning and sequencing, which rectified spurious mitochondrial gene annotations that were not resolved by RNA-seq data alone. The results herein provide further insights into mt genome organization within Erysiphales and constitute valuable genomic resources for powdery mildew pathogens.

6.2 Results

6.2.1 Assembly and general features of the *E. necator* mt genome

BLASTn searches with the mt genome of *Sclerotinia borealis* (NC_025200.1) against the nuclear genome assembly of *E. necator* C-strain returned a 188,576 bp long scaffold (JNVN01000008.1) that represented the mt genome of *E. necator*. The scaffold contained homologs of all core mt genes, whereas the first and last 56 bp overlapped 100%, suggesting circularity. One 153 bp gap was present at 282 bp from one of the ends of the scaffold, but it was patched with whole-genome sequencing reads of *E. necator* C-strain (SRR1448449), thus generating a gapless mt genome assembly.

The resulting mt genome of *E. necator* C-strain corresponded to a long, circular, gapless DNA sequence containing 188,577 bp (Fig 6.1 and Table 6.1). The overall GC content was 33.8%, which is on the high-end for a fungal mt genome (Fig 6.S1 and Table 6.S1). The GC content of the protein-coding mt genes was 29.4%, reflecting preference for AT-rich codons (Table 6.S2), whereas the GC content of intergenic regions and introns was 38.9% and 33.1%, respectively, indicating that they largely contribute to the overall high GC content of the *E. necator* mt genome. GC-skew [(G-C)/(G+C)] and AT-skew [(A-T)/(A+T)] values were both positive (0.101 and 0.031, respectively), which is highly unusual as a positive AT-skew is rather rare in fungal mt genomes and thus far has been reported only in *Scytalidium auriculariicola* among 16 members of the Leotiomycetes (Chen et al. 2019).

A total of 106 genes and other ORFs were predicted in the mt genome of *E. necator*, all of which are transcribed from the sense strand. Coding-sequences of the mt core genes accounted for 8.3% (15.8 kb) of the genome, whereas intergenic regions and introns covered 15.0% (28.3 kb) and 73.9% (139.5 kb) of the genome, respectively, thus contributing to its enlargement. A self-blast search further revealed a considerable amount of repetitive DNA, which accounted for 8.0% of the mt genome (Table 6.1). In total, 104 forward, 23 palindromic, and 11 reverse short exact repeats were identified (Table 6.S3). Forty-five short tandem repeats were also identified, most of which were concentrated within intergenic or intronic regions (Table 6.S4). A notable exception was the tandem repeat ATCCGTAGG, which encoded for Ser-Val-Gly (SVG) and was inserted seven consecutive times in-frame with the last exon of *nad2*. This indicates that next to their potential role in genome rearrangements (Burger et al. 1999; Nedelcu 1997; Ogata et al. 2002; Beaudet et al. 2013), tandem repeats in the mt genome of *E. necator* actively contribute to the modification of protein sequences.

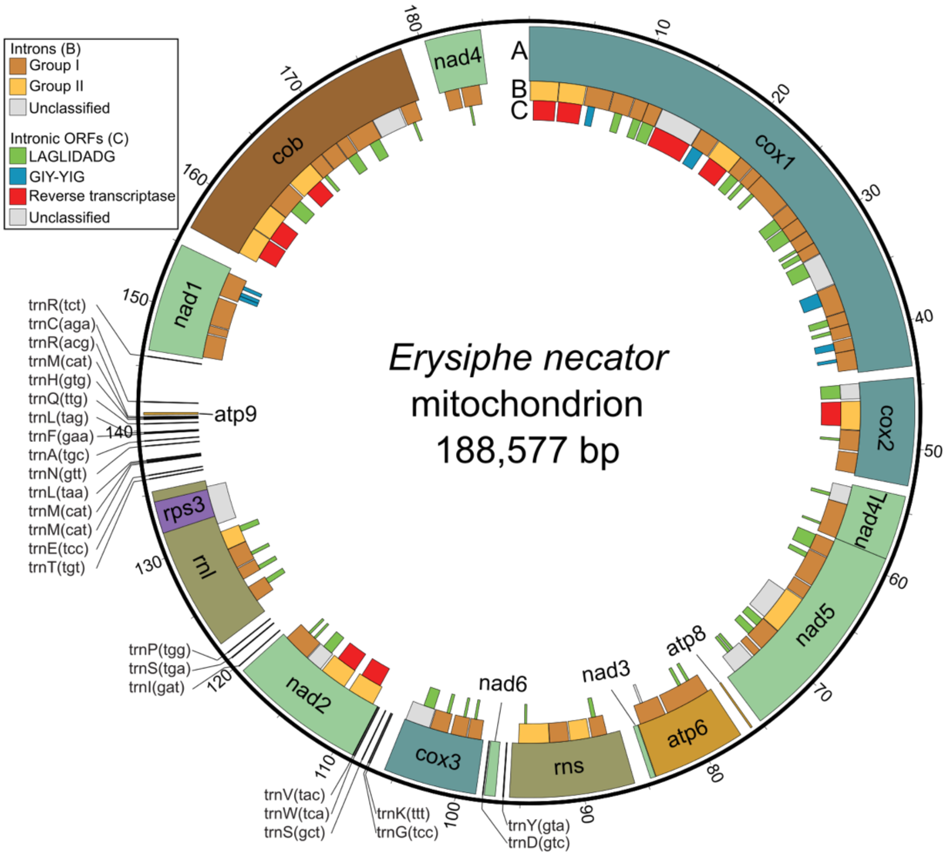


Figure 6.12: Organization of the mitochondrial (mt) genome of the grape powdery mildew fungus *Erysiphe necator*. The *E. necator* mt genome is a long and circular DNA molecule of 188,577 bp in size. Tracks: (A) Core protein-coding and other conserved genes present in the mt genome of *E. necator*. These include genes encoding the subunits of complex I (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*), complex III (*cob*), complex IV (*cox1*, *cox2* and *cox3*), the ATP-synthase complex (*atp6*, *atp8* and *atp9*), the small and large ribosomal subunits (*rns* and *rnl*), the ribosomal protein S3 (*rps3*), and a set of mt-tRNAs. (B) Introns present in the mt genes of *E. necator*. The introns are classified as group I and group II, or as unclassified. (C) Open reading frames (ORFs) present within introns, encoding homing endonucleases of the LAGLIDADG or GIY-YIG families, or reverse transcriptase. The figure was created with Circos v0.69-8 (Krzywinski et al. 2009) (http://www.circos.ca) and further edited with Inkscape v1.0.2 (https://inkscape.org).

Table 6.11: Assembly and gene annotation statistics of the mitochondrial genome of *Erysiphe necator*.

Feature	Value
Total size (bp)	188,577
Intergenic regions size (bp)	28,343
Intronic regions size (bp)	139,477
Overall GC (%)	33.8
Core coding sequences GC (%)	29.4
Intergenic regions GC (%)	38.9
Intronic regions GC (%)	33.1
GC-skew (G-C)/(G+C)	0.101
AT-skew (A-T)/(A+T)	0.031
Repetitive DNA (%)	8.0
Genes	106
Introns	70
Intronic ORFs	64
LAGLIDADG ORFs	44
GIY-YIG ORFS	9
Reverse transcriptase ORFs	10

6.2.2 Gene content and organization of the *E. necator* mt genome

The *ab initio* gene predictions performed with MFannot revealed that all the 14 core mt protein-coding genes are single-copy. The rRNA genes, *rns* and *rnl,* were also predicted within the mt genome, whereas *rps3*, which codes for the ribosomal protein S3, was detected within the fifth intron of *rnl* (*rnl*-i5). However, *mpB*, which encodes the RNA subunit of the mt RNase P, was absent (Fig 6.1 and Table 6.2). As frequently observed in fungal mt genomes, *nad5* and *nad4L* were located next to each other in the mt genome of *E. necator*. In contrast, *nad2* and *nad3*, which are commonly arranged side-by-side (Fig 6.2), were 29.9 kb apart and separated by the presence of three genes between them, namely *rns*, *nad6,* and *cox3* (Fig 6.1). Moreover, instead of clustering with *nad2*, *nad3* clustered with *atp6*, from which it was separated by a short intergenic region of 44 bp. Collectively, these observations indicate that *E. necator* has a unique arrangement of mt genes compared to non-powdery mildew fungal species.

Next to the core set of 14 protein-coding genes, 25 mt-tRNA genes, whose products are able to recognize the standard set of 20 amino acids required for the synthesis of the mt-encoded proteins, were also predicted within the *E. necator* mitogenome (Fig 6.1). All predicted mt-tRNAs also fold into common cloverleaf-like secondary structures (Fig 6.S2). Most mt-tRNA genes were single-copy, except for those encoding mt-tRNAs that decode arginine [*trnR(tct)* and *trnR(acg)*], leucine [*trnL(tag)* and *trnL(taa)*], and serine [*trnS(tga)* and *trnS(gct)*] that each had two copies, and the mt-tRNA gene for methionine [*trnM(cat)*] that was present in three copies. Almost all the mt-tRNA genes of *E. necator* were clustered in the vicinity of the *rnl* gene, a pattern that has been previously observed in fungal mt genomes and further positively correlates to the conservation of gene order (Aguileta et al. 2014; Chen et al. 2019; Mardanov et al. 2014; Zubaer et al. 2018). Specifically, of the 25 mt-tRNA genes, 15 were located within a 9.6 kb mt-tRNA-rich region between *rnl* and *nad1,* three were present between *rnl* and *nad2,* whereas the remaining seven were located upstream of *rnl,* between *nad2* and *cox3* (*n*=5), *cox3* and *nad6* (*n*=1), and *nad6* and *rns* (*n*=1) (Fig 6.1).

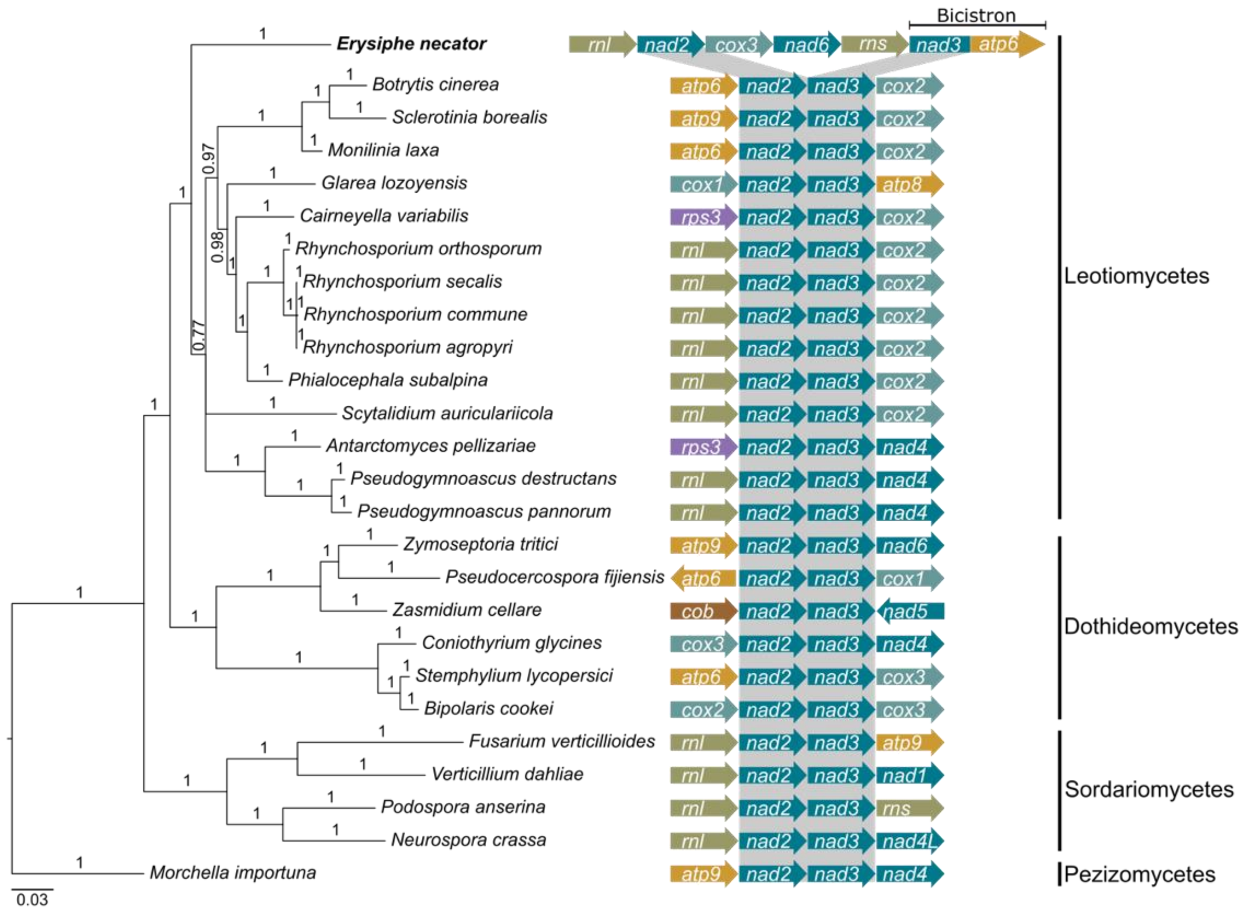


Figure 6.13: The mitochondrial (mt) genome of *Erysiphe necator* has an atypical organization of the *nad2* and *nad3* genes. A Bayesian phylogenetic tree of the mt genomes of *E. necator* and other 25 Ascomycetes is shown on the left-hand side of the image. The phylogenetic tree was inferred with MrBayes based on the concatenated alignment of the protein sequences of 12 mt genes (*atp6*, *nad1-6*, *nad4L*, *cox1-3,* and *cob*). For comparison, a phylogenetic tree based on the nuclear genomes is shown in Fig 6.8. Supporting values of branches are indicated as Bayesian posterior probabilities. *Morchella importuna* was used as outgroup. On the right-hand side of the image, the gene organization on each side of the *nad2* and *nad3* genes is indicted for the species shown in the tree. Genes are represented as arrows and are shown in the same order as they appear in the mt genomes. The *nad2* and *nad3* genes are typically next to each other in fungal mt genomes, but not in the mt genome of *E. necator* in which instead *nad3* and *atp6* are expressed bicistronically in the same RNA transcript. Accession numbers of the sequences utilized to construct the tree are available in Table 6.S12. The phylogenetic tree was edited with FigTree v1.4.2 (Rambaut 2007) (http://tree.bio.ed.ac.uk/software/figtree/) and the figure was created with Inkscape v1.0.2 (https://inkscape.org).

Table 6.12: Overall statistics of the core mitochondrial genes of *Erysiphe necator*. Intron density represents the number of introns per kb of exonic sequence.

Gene	Length (bp)	No. of introns	Exonic region (bp)	Intronic region (bp)	Intronic region (%)	Intron density	Start codon	Stop codon
atp6	7,339	2	756	6,583	89.7	2.6	ATG	TAA
atp8	147	0	147	0	0	0	ATG	TAG
atp9	180	0	180	0	0	0	TTA	TAG
nad1	8,585	4	1,062	7,523	87.6	3.8	ATG	TAA
nad2	10,819	4	1,755	9,064	83.8	2.3	ATG	TAG
nad3	417	0	417	0	0	0	ATG	TAG
nad4	4,420	2	1,443	2,977	67.4	1.4	ATG	TAG
nad4L	5,097	2	270	4,827	94.7	7.4	ATG	TAA
nad5	16,147	7	1,959	14,188	87.9	3.6	ATG	TAG
nad6	852	0	852	0	0	0	ATG	TAG
cox1	43,512	22	1,608	41,904	96.3	13.7	ATG	TAG
cox2	8,485	4	756	7,729	91.1	5.3	ATG	TAA
cox3	7364	4	816	6,548	88.9	4.9	ATG	TAA
cob	21,964	10	1,170	20,794	94.7	8.5	ATG	TAG
rps3	2,526	0	2,526	0	0	0	ATG	TAA
rnl	13,111	5	3,497	9,614	73.3	1.4	-	-
rns	9,865	4	2,139	7,726	78.3	1.9	-	-

6.2.3 Sanger sequencing of full-length cDNA clones and identification of bicistronic genes

To annotate the mt genes of *E. necator*, the publicly available RNA-seq data (SRR1502871 to SRR1502882) that were previously used to assist the gene annotation of the *E. necator* nuclear genome (Jones et al. 2014), was mapped to its mt genome. However, of the 393.3 million reads processed, only 2,619 reads (0.0007%) mapped to the mt genome, with the majority (1,604; 61.2%) aligning to the large or small ribosomal subunits. The rest of the genes sustained variable coverage with *nad3*, *nad4L*, *nad6*, *atp6*, and *atp8* having five or less reads mapping to their exons, thus prohibiting their accurate annotation. Therefore, an alternative approach was followed in order to properly curate the automatically inferred mt gene structures. Specifically, all 14 core protein-coding genes and the two ribosomal subunits were PCR amplified from a cDNA template of isolate C-strain and Sanger sequenced (Fig 6.S3 and Table 6.S5). In this way, all mt genes of *E. necator* had their structures manually inspected and successfully verified.

These experiments showed that all 14 protein-coding genes, the two rRNA genes, and the *rps3* gene present in the mt genome of *E. necator* were expressed. Of the 17 mt genes, eight had their *in silico* annotation confirmed by Sanger sequencing of their corresponding cDNA clones. These included the genes that were *in silico* annotated as intronless (i.e., *nad3*, *nad6*, *atp8*, *atp9,* and *rps3*) as well as *nad1*, *nad2,* and *nad5*. Interestingly, among the intronless genes whose ORFs were verified was *atp9,* even though it was predicted to encode a 59 amino acid long protein instead of the 74 amino acid atp9 protein typically found in other fungal species (Table A1S6). A multiple sequence alignment of fungal atp9 proteins showed that the *E. necator* atp9 was missing 11 amino acids at its N-terminus and a few other amino acids in relatively well-conserved regions, suggesting that it might not be functional (Fig 6.S4). However, a BLAST search against the nuclear genome of *E. necator* using the mt *atp9* as query revealed the presence of a 530 bp nuclear counterpart (KHJ33827), which encodes a 156 amino acid protein with a 76 amino acid long N-terminal mt targeting sequence. Evidence of allotropic expression of the nuclear *atp9* was observed based on RNA-seq data (Fig 6.S5), indicating that, as in other fungi (Zubaer et al. 2018; Torriani et al. 2014), the nuclear *atp9* of *E. necator* could be a functional substitute of its truncated mt one.

Although eight mt genes had their *in silico* annotation verified, the remaining (i.e., *cob*, *cox1*, *cox2*, *cox3*, *atp6*, *nad4*, *nad4L*, *rns,* and *rnl*) needed to have their predicted gene models manually adjusted. A total of seven exons were missed by the *in silico* annotations, including four in *cox1*, two in *nad4L,* and one in *atp6*. Also, three predicted exons in *cob* were absent in the sequenced cDNA of this gene. Curation of *nad4L* extended its coding sequence until it overlapped with *nad5* by one base pair, in that the last nucleotide of *nad4L* stop codon (TAA) was also the first nucleotide of *nad5* start codon (ATG). By using primers located at the start codon of *nad4L* and at the stop codon of *nad5* (Fig 6.S3 and Table 6.S5), an RT-PCR assay showed that the ORFs of these two genes were co-transcribed as a single RNA transcript (Fig 6.S6). Similar to the *nad4L*/*nad5* bicistron, *atp6* and *nad3* were also physically close to each other, and were co-transcribed in the same RNA transcript (Fig 6.S6). The only gene present between the gene pairs *nad4L*/*nad5* and *atp6*/*nad3* was *atp8*, and thus transcription of *atp8* as a polycistronic unit was investigated. However, there was no evidence suggestive of co-transcription of *nad4L/nad5/atp8* or *atp8/atp6/nad3*.

6.2.4 The mt genes of *E. necator* hold a large repertoire of introns and intron-encoded ORFs

The automatic gene annotations and subsequent manual curations revealed an unusually large number of 70 introns within the core mt genes of *E. necator,* with lengths varying from 714 bp (*nad1*-i2) to 4,142 bp (*atp6*-i1). Among the core protein-coding and rRNA mt genes, five genes (i.e., *nad6*, *nad3*, *atp8*, *atp9*, and *rps3*) and all the mt-tRNAs were intronless, whereas the rest harbored from as few as two introns in *atp6* to as many as 22 introns in *cox1* (Fig 6.1 and Table 6.2). The large number of introns present in *cox1*, which accounted for 96.3% of its sequence, expanded the size of this gene to 43.5 kb, which is comparable to the 47.5 kb long *cox1* from *Endoconidiophora rosinifera*, the longest *cox1* reported to date among members of the Ascomycetes (Zubaer et al. 2018). Intron density (i.e., the number of introns per kb of coding sequence) was also highest for *cox1* (13.7), followed by *cob* (8.5), and *nad4L* (7.4) (Table 6.2). This is perhaps not surprising, as *cox1* and *cob* are known to possess large intron numbers as compared to other fungal mt genes, and to exhibit frequent intron gain-and-loss events (Férandon et al. 2013; Lang et al. 2007; Yin et al. 2012).

As for most fungi, the majority (*n*=63) of mt introns in *E. necator* resembled self-splicing introns, which based on their putative secondary structure could be classified as group I (*n*=48) and group II (*n*=13) introns. Group I introns were further classified into subgroups IB (*n*=27), IC2 (*n*=9), ID (*n*=6), I derived (*n*=4), IC1 (*n*=1), and IA (*n*=1) (Fig 6.1). Notably, a set of 64 ORFs were found residing within the group I and II introns of the *E. necator* mt genes, of which 52 encoded HEs of the LAGLIDADG (*n*=44) and GIY-YIG (*n*=8) families, and ten encoded proteins with a domain architecture composed of an RT and an intron maturase. As expected, predicted HEs of the LAGLIDADG and GIY-YIG families were usually contained within group I introns, whereas RT-encoding ORFs were associated with group II introns (Table 6.S7). Specifically, of the

52 ORFs encoding HEs, 43 were located within group I introns, and nine RT-encoding ORFs were located within group II introns. However, exceptions were identified in the two ribosomal genes as, for example, *rns*-i4 and *rnl*-i4 contained ORFs with the LAGLIDADG nuclease motif, although they were classified as group II introns. Finally, of the 44 LAGLIDADG and eight GIY-YIG family HEs, 25 and five, respectively, appeared to have truncated domains, thus corresponding to likely degenerated HEs (Table 6.S8). The remaining two ORFs residing within the group I and II introns of the *E. necator* mt genes encoded a hybrid GIY-YIG/RT protein and a nuclease-associated modular DNA-binding domain 1 (NUMOD1), and were present within *nad5*-i4 and *atp6*-i2, respectively.

A notable feature of intronic HEs is that they can be inserted in-frame and thus translated as a fusion protein with their upstream exon. This, consequently, enhances their expression and their chances of fixation within a population (Emblem et al. 2014). In *E. necator*, a total of 58 ORFs encoding HEs or RTs were located within introns of protein-coding genes, of which 41 were in-frame with the upstream exon (Table 6.S8) and 25 further lacked stop codons in the region between the upstream exon and the predicted ORF start. Of these 25 ORFs, 14 encoded HEs of the LAGLIDADG family and four of the GIY-YIG family, whereas the remaining seven encoded RTs. Moreover, the 25 ORFs were overall closer to their upstream in-frame exons (average of 188 bp) compared to all the 64 intronic ORFs (average of 526 bp) found within the mt genes of *E. necator*. Collectively, these findings suggest that these 25 ORFs are likely capable of fusing with their in-frame upstream exons as a means of promoting their expression and fixation in the mt genome.

6.2.5 Mt genomes are highly conserved among *E. necator* isolates

By querying the mt genome of *E. necator* C-strain with BLASTn against the NCBI genomes database, scaffolds were identified that contained the mt genomes of isolates Branching (JNUS01000009.1), Ranch9 (JNUT01000020.1), and e1-101 (JOKO01000016.1). The mt genome of isolate Lodi was also identified but it parted into two scaffolds (JNUU01000038.1 and JNUU01000071.1). The size of the scaffolds containing the mt genome of the four isolates was comparable to that of isolate C-strain (188,577 bp), and ranged from

185,650 bp in Lodi, to 188,575 bp in Branching, 188,647 bp in e1-101, and 188,770 bp in Ranch9. Alignments with the mt genome of C-strain revealed a high level of conservation among the mt genomes of the five isolates. Specifically, all of the 106 genes and ORFs identified in the mt genome of C-strain were present in the mt genomes of the other four *E. necator* isolates. Also conserved was the order and orientation of these elements in the genome as well as the size and positions of intergenic regions and introns. The only exception appeared to be intron *nad5*-i4, which seemed absent in isolate Lodi (Fig 6.S7). However, this intron was located at the breakpoint between the two scaffolds that contained the mt genome of this isolate, whereas by mapping the whole genome sequencing (WGS) reads from isolate Lodi to the mt genome of C-strain, *nad5*-i4 had normal coverage (Fig 6.3), suggesting that it is also conserved in Lodi. Although mt gene content and genome organization were fully conserved among the five isolates of *E. necator*, 203 polymorphic sites were identified (Table 6.S9). Of the 203 polymorphic sites, 197 were short insertions or deletions (INDELs) and only six were single-nucleotide polymorphisms (SNPs). INDELs were abundant within intergenic regions and introns, particularly in the mt-tRNA-rich region between *rns* and *nad1* (Fig 6.3). From the 203 polymorphic sites, 126 were located within intergenic regions, 65 were present within introns, while 12 were located within exons of mt genes or intronic ORFs (Table 6.3 and Table 6.S9). Among these 12 polymorphic sites, eight were short INDELs corresponding to microsatellite-like homopolymeric regions of eight or more consecutive nucleotides. Five of the INDELs caused frameshifts of LAGLIDADG-encoding ORFs within *cox1*-i11 [$(C)_{9-11}$], *nad5*-i2 [$(T)_{10-11}$ and $(G)_{8-11}$], *cox3*-i2 [$(G)_{12-14}$] and *cob*-i5 [$(G)_{8-12}$], and two were located within *rns* [$(G)_{11-12}$] and *rnl* [$(G)_{9-11}$]. The remaining INDEL was identified within the last exon of *nad2* and corresponded to the tandem repeat ATCCGTAGG, which encoded for Ser-Val-Gly. This repeat was present seven times in isolates C-strain, Ranch9 and Lodi, and six times in isolates e1-101 and Branching (Table 6.3). Finally, of the remaining four polymorphic sites present within functional regions, two were located within intronic ORFs and two within coding sequences of conserved mt genes. Of the later ones, one induced a synonymous change at codon 151 of *nad4* (c.453C>A; p.V151V) and was found only in isolate Branching. However, the other one triggered a missense mutation at codon 143 (c.428G>C) of *cob*, which produces the notorious p.G143A amino acid substitution in cytochrome b that confers high levels of resistance to QoI fungicides (Fernández-Ortuño et al. 2008). This mutation was absent in C-strain, e1-101, and Branching but was present in isolates Ranch9 and Lodi (Table 6.3).

Interestingly, among the five *E. necator* isolates, 26% to 44% of the WGS reads mapped to the mt genome instead of the nuclear genome (Fig 6.4A and Table 6.S10). This indicates that the mt genome is overrepresented in the sequenced reads, most likely as a result of the multi-copy nature of mitochondria in cells. Based on the mt genome coverage to nuclear genome coverage ratio, the estimated mitochondria copy number per cell varied from 124 to 322 (Fig 6.4B and Table 6.S11). Finally, by dividing the sequenced base pairs from reads that did not map to the mt genome by the calculated coverage of the nuclear genome, then the estimated size of the nuclear genome of *E. necator* is between 78 Mbp and 95 Mbp (Table 6.S11). This estimate is considerably lower than the 126 ± 18 Mb genome size reckoned before using *k*-mer analysis (Jones et al. 2014), indicating that the later approach might have overestimated the genome size of *E. necator*.

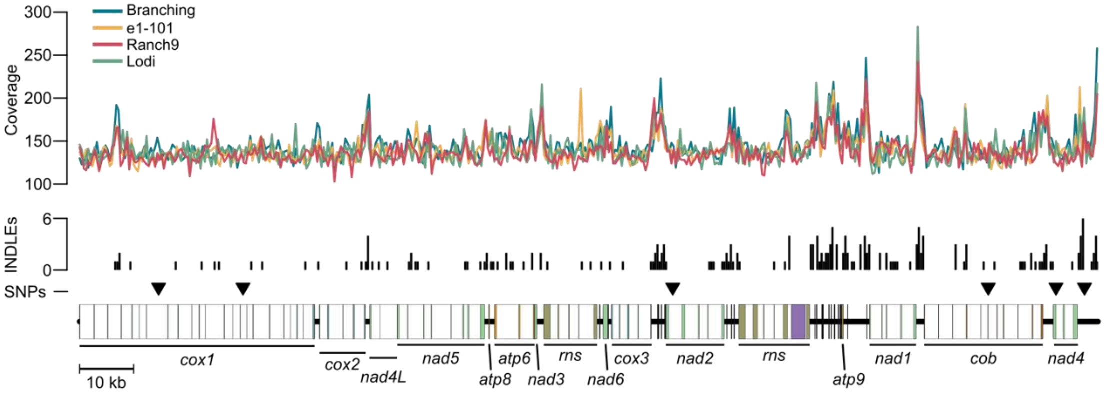


Figure 6.14: Comparative analysis of the mitochondrial (mt) genomes of five isolates of *E. necator* shows no presence/absence of introns and low genetic variability within functional regions. Whole-genome sequencing (WGS) coverage of four *E. necator* isolates across the reference mt genome of isolate C-strain is shown at the top. The histogram shows the number of insertions or deletions (INDELs) identified in different regions of the mt genome. Location of identified single nucleotide polymorphisms (SNPs) are indicated with triangles (total of six). The WGS coverage suggests that all regions of the reference mt genome are conserved in the other four isolates analyzed. Low number of SNPs and presence of almost all INDELs within intergenic or intronic regions indicate low genetic variability within functional regions of the mt genome of *E. necator*. Coverage of WGS reads and the INDEL histogram were generated with a non-overlapping sliding window of 400 bp. Coverage of WGS reads was normalized to 100x prior alignment. The figure was generated with R v4.0.3 (https://www.r-project.org) and further edited with Inkscape v1.0.2 (https://inkscape.org).

Table 6.13: Polymorphic sites within functional regions of the mitochondrial genome of *Erysiphe necator*. For each polymorphic site, its position in the genome is shown followed by the gene or intronic ORF affected by the polymorphism. The alleles for the respective isolates are shown. Variants at the DNA and protein levels are described according to the Human Genome Variation Society (HGVS) recommendations. Intronic ORFs encoding LAGLIDADG or reverse transcriptase domains are indicated with LD and RT, respectively.

Position (bp)	Gene/ORF	Variant (DNA)	Variant (protein)	C-strain	e1-101	Branching	Ranch9	Lodi
14,629	cox1-i7-RT	c.2094A>C	p.L698F	A	C	C	C	C
22,554	cox1-i11-LD	c.134C[11];[10];[9]	p.P48Lfs*10;p.P49Lfs*18	C[11]	C[9]	C[10]	C[10]	C[10]
61,221	nad5-i2-LD	c.34T[11];[10]	p.L15Yfs*5	T[11]	T[11]	T[11]	T[10]	T[10]
61,250	nad5-i2-LD	c.63G[10];[11];[8]	p.T25Dfs*16; p.G24Dfs*16	G[10]	G[8]	G[11]	G[10]	G[10]
86,443	*rns*	n.470G[11];[12]	-	G[11]	G[11]	G[11]	G[12]	G[12]
100,647	cox3-i2-LD	c.217G[14];[12]	p.G77*fs	G[14]	G[12]	G[14]	G[12]	G[12]
109,826	nad2-i1-RT	c.555C>T	p.L185L	C	C	C	T	T
119,171	*nad2*	c.1534TCCGTAGGA[7];[6]	p.SVG512[7];[6]	ATCCGTAGG[7]	ATCCGTAGG[6]	ATCCGTAGG[6]	ATCCGTAGG[7]	ATCCGTAGG[7]
130,655	*rnl*	n.2481G[9];[10];[11]	-	G[11]	G[10]	G[11]	G[9]	G[9]
167,386	cob-i5-LD	c.321G[12];[10];[11];[8]	p.G111Rfs*7;p.G111Afs*70;p.G110Afs*70	G[12]	G[11]	G[8]	G[10]	G[10]
168,253	*cob*	c.428G>C	p.G143A	G	G	G	C	C
180,789	*nad4*	c.453A>C	p.V151V	A	A	C	A	A

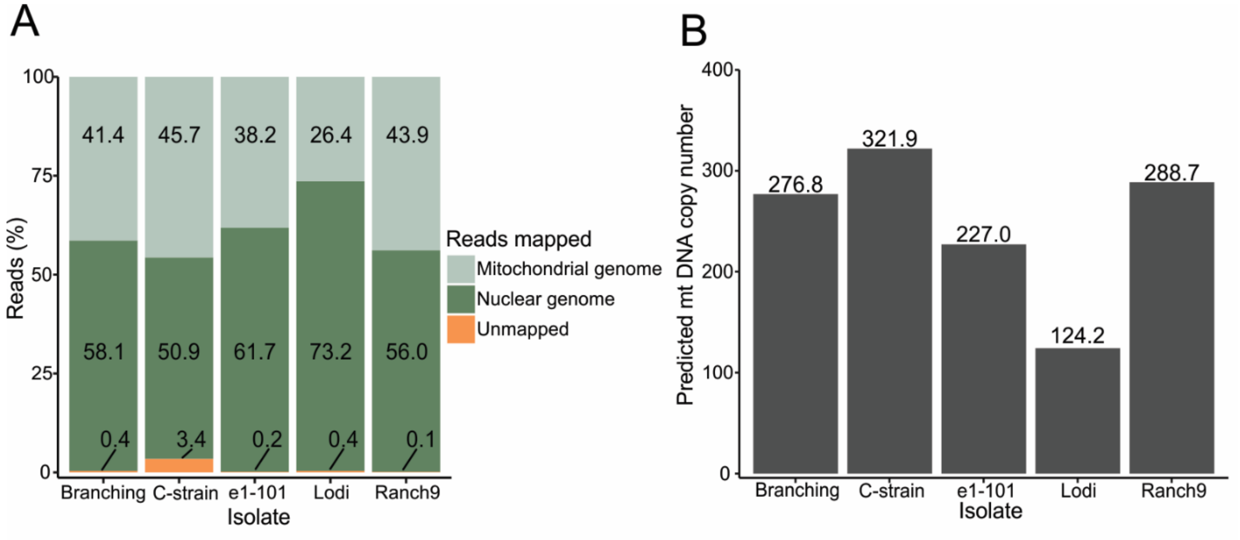


Figure 6.15: Estimated mitochondrial (mt) DNA copy numbers among isolates of *E. necator*. (A) Box-plot showing the percentage of whole-genome sequencing (WGS) reads from five *E. necator* isolates mapped to the nuclear and mt genomes of *E. necator* C-strain. The large percentage of WGS reads mapped to the mt genome indicates high abundance of mt DNA compared to nuclear DNA. (B) Box-plot showing the estimated mt genome copy number per cell for the analyzed isolates. Mt genome copy number was calculated based on the ratio between the mt genome coverage and the nuclear genome coverage.

6.3 Discussion

In this study, we present a high-quality mt genome for *E. necator*, an economically important powdery mildew pathogen, and thus provide further insights into the mt genome organization of members of the Erysiphales. Our analysis showed that the mt genome of *E. necator* is large but compact with dozens of group I introns encoding mostly HEs from the LAGLIDADG and GIY-YIG families. Moreover, the gene pairs *nad4L/nad5* and *atp6/nad3* exhibited bicistronic expression, which is exceptional among fungi. Further analysis of the mt genomes of five *E. necator* isolates revealed a high level of conservation of gene content and order but large variations in predicted mt DNA copy-numbers per isolate. Overall, the genomic resources presented herein will be of great value for future studies of population and evolutionary genomics of powdery mildews.

Identification of short exons is challenging as they can be easily mis-predicted by *ab initio* predictors. In this study, we followed a systematic approach to assemble the mt genome of *E. necator*, which consisted of validating the *ab initio* predictions of the mt genes by mapping of RNA-seq reads to the assembly and subsequent Sanger sequencing of full-length cDNA clones. This approach allowed us to correct several erroneous predictions that were not resolved by RNA-seq alone, and thus accurately adjust the gene annotations. For example, the 6 bp long *nad4L*-exon2 and the 11 bp long *cox1*-exon12 were not predicted by MFannot but they were resolved by Sanger sequencing of the cDNA clones. Such examples highlight the importance of verifying the structure of mt genes with cDNA sequences.

Manual curation of *nad4L* showed that its coding sequence overlaps by one base pair with the coding sequence of *nad5*, which is present just immediately downstream of *nad4L*. This is not unique, as the overlap of these two genes is found in other Leotiomycetes as well, including in *S. borealis*, *S. auriculariicola*, and *Antarctomyces pellizariae*. However, by using primers next to the start and stop codons of *nad4L* and *nad5*, we established that these two genes are expressed in *E. necator* from a single bicistronic transcript. Similarly, *atp6* and *nad3* are also side-by-side and are co-expressed in a bicistronic-like manner. To the best of our knowledge, co-transcription of the genes *nad4L*/*nad5* and *atp6*/*nad3* is rather exceptional among fungi (Kolondra et al. 2015). Moreover, while in most fungal mt genomes, including those of several phylogenetically distant fungal species such as of members of the Sordariomycetes (Aguileta et al. 2014; Zaccaron et al. 2017; Zubaer et al. 2018), Leotiomycetes (Chen et al. 2019; Mardanov et al. 2014), and Dothideomycetes (Franco et al. 2017; Torriani et al. 2008; Zaccaron and Bluhm 2017), the gene pairs *nad4L*/*nad5* and *nad2*/*nad3* are usually located close and next to each other, in *E. necator nad3* is paired with *atp6* instead of *nad2*. This is most likely the result of a gene rearrangement after divergence of the Erysiphales and a potential marker for powdery mildew pathogens. Furthermore, the widespread pairing among fungal mt genomes of *nad4L* and *nad3* to physically close genes raises the possibility that these genes require bicistronic-like behavior. One possible explanation for co-transcription of mt genes is that short mRNAs could be unstable or unable to interact effectively with the ribosomal unit (Kleidon et al. 2003; Kouvelis et al. 2004). For example, the coding sequences of *nad4L* and *nad3* in *E. necator* are relatively short, consisting of only 273 bp and 417 bp, respectively. However, this hypothesis does not account for *atp8*, which was not co-expressed with neighboring genes and has a coding sequence of only 147 bp. Nevertheless, future studies can shed light into the potential benefits and widespread behavior of bicistronic genes in mt genomes.

Sequencing of cDNA also revealed the presence of an *atp9* gene in the mt genome of *E. necator*. However, although transcribed, this gene is likely no longer functional because the encoded protein has several amino acids missing near the N-terminus, due to an in-frame stop codon present in its coding sequence. However, an *atp9* allele whose product can be translated to a full-length atp9 protein with an mt-targeting signal peptide at its N-terminus was identified in the nuclear genome of *E. necator*. This gene could be compensating for the inactive mt *atp9* allele, through allotopic expression in the nucleus and subsequent relocation of the produced protein into mitochondria. Indeed, allotopic expression of mtDNA-encoded genes that have migrated to the nucleus has been demonstrated in yeast (e.g., *atp8, bl4,* and *Var1p*) (Banroques et al. 1986; Nagley et al. 1988; Sanchirico et al. 1995) and human cell lines (e.g., *atp6* and *atp8*) (Boominathan et al. 2016; Kaltimbacher et al. 2006). A study has also demonstrated the successful allotopic expression of the *Podospora anserina atp9* gene in the nucleus of *Saccharomyces cerevisiae* (Bietenhader et al. 2012), indicating that, as with other genes encoding ATP synthase subunits, *atp9* can at least in principle also be functionally expressed from nuclear DNA and its product is translocated in mitochondria. However, the same study also showed that an engineered nuclear version of the yeast *atp9* gene that contained an mt-targeting sequence was unable to compensate the function of the yeast mt *atp9* gene, indicating that there are barriers to the mt import of allotopically expressed proteins. Nonetheless, the presence of nuclear copies of *atp9*, which may or may not be accompanied by a parallel loss of the mt allele, have been reported in a number of fungal species (Déquard-Chablat et al. 2011; Franco et al. 2017; Zubaer et al. 2018), including in the Leotiomycetes *Rhynchosporium* spp. (Torriani et al. 2014).

The 104 kb and 139 kb mt genomes of the barley powdery mildew *Blumeria graminis* f. sp. *hordei* isolates DH14 and RACE1, respectively, have been previously reported (Frantzeskakis et al. 2018). Compared to *E. necator*, the mt genome of *B. graminis* f. sp. *hordei* is considerably smaller but harbors all core mt genes in the same order and orientation as in *E. necator*. This indicates that their difference in mt genome size is likely due to presence/absence of nonfunctional regions, whereas no major mt gene rearrangement is present between them. Similar to *E. necator*, *B. graminis* f. sp. *hordei* also contains a nuclear-encoded *atp9* homolog that likely compensates for the absence of a functional mt-encoded *atp9* (Frantzeskakis et al. 2018). Notably, as in *E. necator*, the mt genome of *B. graminis* f. sp. *hordei* also contains the gene pair *atp6*/*nad3* next and physically close to each other, which indicates that this atypical gene pairing is common among powdery mildews.

Similar to previous reports of mt genomes of other members of the Leotiomycetes, such as *S. borealis* (Mardanov et al. 2014) and *Monilinia laxa* (Yildiz and Ozkilinc 2020), the mt genome of *E. necator* is also enriched with ORFs encoding HEs and RTs. HE genes are selfish genetic elements that spread at a super-mendelian rate within a population. They are believed to have no effect on the fitness of the host organism, and therefore are not subject to natural selection. Once fixed in a population, these elements accumulate mutations that eventually disrupt their ability to spread (Burt et al. 2006; Burt and Koufopanou 2004). However, comparative analysis among five isolates revealed that only six out of the 64 ORFs encoding HEs or RTs contained polymorphic sites, indicating that these ORFs have little genetic variability in the mt genome of *E. necator*. One possible explanation is that these enzymes could function as maturases required for proper intron splicing (Belfort 2003; Bonen and Vogel 2001). Mutations could disrupt their function, causing retention of introns in mature transcripts and interfering with the function of core mt genes. Future studies can reveal how active these enzymes are, their importance for the proper function of the mt genome of *E. necator*, as well as their distribution among different populations of powdery mildew pathogens.

6.4 Materials and methods

6.4.1 Mt genome assembly

A scaffold (JNVN01000008.1) containing the mt genome of *E. necator* was initially identified by querying with BLASTn (e-value<1e-5) the mt genome of the phylogenetically close-related species *Sclerotinia borealis* (NC_025200.1) (Mardanov et al. 2014) against the nuclear genome assembly of *E. necator* C-strain (ASM79871v1) (Jones et al. 2014). To produce the final mt genome of *E. necator*, a new assembly was generated in order to patch a 153 bp gap that was present within scaffold JNVN01000008.1. For this purpose, whole-genome sequencing (WGS) reads from *E. necator* isolate C-strain (SRR1448449) were obtained from NCBI. Reads were then trimmed with fastp v0.20 (Chen et al. 2018), and those with a *k*-mer matching to scaffold JNVN01000008.1 were extracted with the *bbduk.sh* script of the BBMap v38-72 software package (Bushnell 2014), using the parameters *k=31* and *hdist=1*. Extracted reads were then processed with the *bbnorm.sh* script of BBMap to normalize the depth of coverage to 100x, and the normalized reads were assembled into contigs with SPAdes v3.14 (Bankevich et al. 2012) utilizing *k*-mer values of 21, 33, 55, 77, 99, and 127. The assembled contigs were finally mapped to scaffold JNVN01000008.1 with the Burrows-Wheeler Aligner – Maximal Exact Matches (BWA-MEM; v0.7.17) algorithm (Li and Durbin 2009) and the gap was manually patched. Finally, the same reads utilized in the assembly step were mapped to the gap-filled contig with BWA-MEM v0.7.17 followed by two rounds of polishing with Pilon v1.23 (Walker et al. 2014).

6.4.2 Annotation of mt genes

The assembled mt genome was initially annotated with MFannot, using the genetic code 4 (Mold, Protozoan and Coelenterate Mt Code) (Valach et al. 2014). Genes encoding mt-tRNAs and their secondary structures were obtained with MITOS2 (Bernt et al. 2013). Introns were classified into group I or group II with RNAweasel (Lang et al. 2007). Intronic ORFs were identified with ORFfinder v0.4.3 (Wheeler et al. 2007), using as a minimum ORF length 200 bp and genetic code 4. ORFs encoding homing endonucleases (HEs) or reverse transcriptases (RTs) were identified and classified based on their conserved domains identified by querying (e-value<1e-3) the encoded peptide sequences against the NCBI conserved domain database (CDD) (Marchler-Bauer et al. 2017). Conserved domains within introns were identified by translating the entire intronic sequences in six frames with the *transeq* script of the EMBOSS software package v6.6.0 (Rice et al. 2000), utilizing the genetic code 4 and querying (e-value<1e-3) the peptide sequences against the NCBI CDD. Codon usage was determined with the *cusp* script of the EMBOSS software package v6.6.0 (Rice et al. 2000). Short exact repeats were identified with REPuter (Kurtz and Schleiermacher 1999) using minimum repeat length of 8 bp and e-value<1e-5. Tandem repeats were identified with Tandem Repeat Finder v4.09 (Benson 1999) and the overall percentage of repeats in the mt genome was calculated based on self BLASTn searches (e-value<1e-10), utilizing the parameter -*task blastn*. Circular representations of the mt genome was created with Circos v0.69-8 (Krzywinski et al. 2009).

6.4.3 Confirmation of mt genes by full length cDNA clones and Sanger sequencing

To validate the *in silico* annotations of protein-coding mt genes, RNA-seq reads of *E. necator* isolate C-strain were obtained from NCBI (accessions SRR1502871 to SRR1502882) and mapped to the mt genome with HISAT2 v2.2.1 (Kim et al. 2015). However, the overall low number of RNA-seq reads mapped prohibited curation of most of the *E. necator* genes. Therefore, a different approach was followed to adjust the annotations of the mt genes based on cDNA sequencing. To do so, *E. necator* C-strain was maintained on detached leaves of *Vitis vinifera* cv. Carignan in the laboratory as described before (Jones et al. 2014). RNA was extracted from spores collected from colonized leaves by using the TRIzol reagent (Invitrogen, Carlsbad, CA). Complementary DNA was synthesized with the SuperScript First-Strand Synthesis Kit, (Invitrogen, Cat no. 12371-019) according to the manufacturer’s protocol. Primers were designed to capture the entire ORF of genes (Fig 6.S3 and Table 6.S5) and utilized to PCR-amplify them from the cDNA template. For cloning, PCR products were separated on 2% agarose gel, bands were excised and subjected to column purification using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Cat no. D4001). Purified cDNA fragments were ligated into pGEM-T Easy vector (Promega, Madison, WI, United States) according to the manufacturer's instructions and transformed into *Escherichia coli* strain DH5α, using the heat-shock method. Either PCR products or fragments cloned in the pGEM-T Easy vector were Sanger-sequenced. Generated ABI sequence files were mapped to the *in silico* predicted mt genes of *E. necator* with SnapGene v5.0.7 (GSL Biotech; available at snapgene.com). Alignments were visualized with SnapGene and exon-intron boundaries were manually adjusted. Start and stop codons were adjusted based on homology with reviewed fungal mt proteins from UniProt/Swiss-Prot database (The UniProt Consortium 2019).

6.4.4 Phylogenetic analysis

To construct a phylogenetic tree of mt genomes, protein sequences were obtained from NCBI using the *efetch* module of the Entrez Direct package v13.9 (Kans 2013). To construct a phylogenetic tree of nuclear genomes, universal single-copy genes were identified with BUSCO v4.0.6 (Simão et al. 2015) using the Eukaryote data set v10. Protein sequences were aligned with MAFFT v7.475 (Katoh et al. 2002) and positions containing gaps in the alignments were removed with trimAl v1.4.1 (Capella-Gutiérrez et al. 2009). The resulting alignments were concatenated, thus producing alignments of 2,651 and 51,776 amino acids for the mt and nuclear genomes, respectively that were subsequently used to construct trees with the Bayesian Inference method implemented in MrBayes v3.2.6 (Huelsenbeck and Ronquist 2001). Four chains were run with one cold and three hot for 500,000 generations, and sampling every 200 generations. The first 25% of samples were discarded as burn-in. The amino acid substitution model was set to *mixed*, which leveraged over 10 different models implemented in MrBayes, and subsequently selected *Cprev* and *Wag* as the best-fitted models for the mt and nuclear genomes, respectively. Stationarity was observed based on the average standard deviation of split frequencies, which was less than 0.005 at th e end of the run. The trees were visualized and edited with FigTree v1.4.2 (Rambaut 2007). Accession numbers of the proteins utilized to construct the phylogenetic trees are shown in Table 6.S12 and Table 6.S13.

6.4.5 Comparison of mt genomes from different isolates of *E. necator*

Whole-genome sequencing reads of *E. necator* were downloaded from NCBI database for isolates e1-101 (SRR1448468), Branching (SRR1448453), Ranch9 (SRR1448454), Lodi (SRR1448470), and C-strain (SRR1448450). Reads were trimmed with fastp v0.20 (Chen et al. 2018) with default settings, except of the required read length that was set to 40 bp. Reads were then mapped simultaneously to the nuclear (GCA_000798715.1) and mt genomes of *E. necator* C-strain, using the BWA-MEM v0.7.17 software package (Li and Durbin 2009). Mapped reads considered as PCR duplicates were marked with samblaster 0.1.26 (Faust and Hall 2014). The overall alignment rate and the number of reads that mapped to the mt and nuclear genomes were determined with SAMTools v1.9 (Li et al. 2009). Coverage of the nuclear and mt genomes was determined with mosdepth v0.3.1 (Pedersen and Quinlan 2018) with parameters adjusted to calculate the median coverage (option *--use-median*) and to ignore unmapped reads, secondary alignments and PCR duplicates (option *--flag 1796*). To avoid repetitive regions in the nuclear genome that can skew coverage values, mosdepth calculated the median coverage of all predicted exons in the nuclear genome. Subsequently, the nuclear genome coverage was estimated as the median coverage of all exons. The nuclear genome size of *E. necator* was estimated as the total number of sequenced bases from reads not marked as PCR duplicates that did not map to the mt genome divided by the nuclear genome coverage. To identify polymorphic sites in the mt genome, reads had their coverage normalized to 100x with the *bbnorm.sh* script of the BBMap v38.18 software package (Bushnell 2014), and they were then mapped to the reference *E. necator* mt genome using the BWA-MEM v0.7.17 software package (Li and Durbin 2009). Short INDELs and SNPs were identified with FreeBayes v1.3.5 with ploidy set to 1 (Garrison and Marth 2012), and subsequently filtered with VCFtools v0.1.16 with minimum quality of 30 (Danecek et al. 2011). Filtered INDELs and SNPs were annotated with SnpEff v5.0 (Cingolani et al. 2012) based on a database constructed from the GenBank file of the mt genome (MT880588). Polymorphic sites that overlapped with exons, introns or intergenic regions were identified with the subcommand *intersect* from BEDTools v2.29.0 (Quinlan and Hall 2010).

6.5 Data availability

The annotated mt genome of *E. necator* has been submitted to GenBank under the accession number MT880588. Scripts used in the analysis were designed with the Snakemake workflow manager (Köster and Rahmann 2012), and are available at https://github.com/alexzaccaron/2021_enec_mt.

Acknowledgments

We thank Dr. Dario Cantu for providing *E. necator* isolate C-strain and Carolina Nogueira for help with maintaining *E. necator* cultures in the lab. Alex Z. Zaccaron acknowledges University of California, Davis, for providing funding through the Dean's Distinguished Graduate (DDGF) Fellowship program. Jorge T. De Souza acknowledges funding from the CAPES-PrInt (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) program, Brazil. Finally, we are thankful to the United States Department of Agriculture – National Institute for Food and Agriculture – Specialty Crop Research Initiative FRAME project (number 2018-03375) for partially supporting this study.

Author Contributions

A.Z.Z.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review and editing, Visualization. **J.T.S.:** Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **I.S.** Conceptualization, Methodology, Writing - original draft, Writing - review and editing, Visualization, Supervision, Project administration, Funding acquisition.

6.6 References

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., and Gabaldón, T. 2014. High variability of mitochondrial gene order among fungi. Genome Biol Evol. 6:451–465

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477

Banroques, J., Delahodde, A., and Jacq, C. 1986. A mitochondrial RNA maturase gene transferred to the yeast nucleus can control mitochondrial mRNA splicing. Cell. 46:837–844

Beaudet, D., Terrat, Y., Halary, S., de la Providencia, I. E., and Hijri, M. 2013. Mitochondrial genome rearrangements in *Glomus* species triggered by homologous recombination between distinct mtDNA haplotypes. Genome Biol. Evol. 5:1628–1643

Belfort, M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes Dev. 17:2860–2863

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P. F. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69:313–319

Bietenhader, M., Martos, A., Tetaud, E., Aiyar, R. S., Sellem, C. H., Kucharczyk, R., Clauder-Münster, S., Giraud, M.-F., Godard, F., Salin, B., Sagot, I., Gagneur, J., Déquard-Chablat, M., Contamine, V., Denmat, S. H.-L., Sainsard-Chanet, A., Steinmetz, L. M., and Rago, J.-P. di. 2012. Experimental relocation of the mitochondrial *ATP9* gene to the nucleus reveals forces underlying mitochondrial genome evolution. PLoS Genet. 8:e1002876

Bonen, L., and Vogel, J. 2001. The ins and outs of group II introns. TRENDS Genet. 17:322–331

Boominathan, A., Vanhoozer, S., Basisty, N., Powers, K., Crampton, A. L., Wang, X., Friedricks, N., Schilling, B., Brand, M. D., and O'Connor, M. S. 2016. Stable nuclear expression of *ATP8* and *ATP6* genes rescues a mtDNA Complex V *null* mutant. Nucleic Acids Res. 44:9342–9357

Burger, G., Saint-Louis, D., Gray, M. W., and Lang, B. F. 1999. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: cyanobacterial introns and shared ancestry of red and green algae. Plant Cell. 11:1675–1694

Burt, A., and Koufopanou, V. 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr. Opin. Genet. Dev. 14:609–615

Burt, A., Trivers, R., and Burt, A. 2006. *Genes in conflict: the biology of selfish genetic elements*. Harvard University Press.

Bushnell, B. 2014. BBMap: a fast, accurate, splice-aware aligner. Available at: Available at http://sourceforge.net/projects/bbmap [Accessed December 2, 2020].

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

Chan, D. C. 2006. Mitochondria: dynamic organelles in disease, aging, and development. Cell. 125:1241–1252

Chen, C., Li, Q., Fu, R., Wang, J., Xiong, C., Fan, Z., Hu, R., Zhang, H., and Lu, D. 2019. Characterization of the mitochondrial genome of the pathogenic fungus *Scytalidium auriculariicola* (Leotiomycetes) and insights into its phylogenetics. Sci. Rep. 9:17447

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 6:80–92

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., and others. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158

Déquard-Chablat, M., Sellem, C. H., Golik, P., Bidard, F., Martos, A., Bietenhader, M., di Rago, J.-P., Sainsard-Chanet, A., Hermann-Le Denmat, S., and Contamine, V. 2011. Two nuclear life cycle–regulated genes encode interchangeable subunits *c* of mitochondrial ATP synthase in *Podospora anserina*. Mol. Biol. Evol. 28:2063–2075

Emblem, Å., Okkenhaug, S., Weiss, E. S., Denver, D. R., Karlsen, B. O., Moum, T., and Johansen, S. D. 2014. Sea anemones possess dynamic mitogenome structures. Mol. Phylogenet. Evol. 75:184–193

Faust, G. G., and Hall, I. M. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 30:2503–2505

Férandon, C., Xu, J., and Barroso, G. 2013. The 135 kbp mitochondrial genome of *Agaricus bisporus* is the largest known eukaryotic reservoir of group I introns and plasmid-related sequences. Fungal Genet. Biol. 55:85–91

Fernández-Ortuño, D., Torés, J. A., de Vicente, A., and Pérez-García, A. 2008. Mechanisms of resistance to QoI fungicides in phytopathogenic fungi. Int. Microbiol. 11:1–9

Franco, M. E. E., López, S. M. Y., Medina, R., Lucentini, C. G., Troncozo, M. I., Pastorino, G. N., Saparrat, M. C. N., and Balatti, P. A. 2017. The mitochondrial genome of the plant-pathogenic fungus *Stemphylium lycopersici* uncovers a dynamic structure due to repetitive and mobile elements. PLoS One. 12:e0185545

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., and 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

Gadoury, D. M., Cadle-Davidson, L., Wilcox, W. F., Dry, I. B., Seem, R. C., and Milgroom, M. G. 2012. Grapevine powdery mildew (*Erysiphe necator*): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. Mol. Plant Pathol. 13:1–16

Garrison, E., and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. Available at: Preprint at https://arxiv.org/abs/1207.3907.

Gustafsson, C. M., Falkenberg, M., and Larsson, N.-G. 2016. Maintenance and expression of mammalian mitochondrial DNA. Annu. Rev. Biochem. 85:133–160

Huelsenbeck, J. P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755

James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J. E. 2013. Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. Curr. Biol. 23:1548–1553

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K. C. H., McGuire, B., Gubler, W. D., Walker, M. A., and Cantu, D. 2014. Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. BMC Genomics. 15:1081

Kaltimbacher, V., Bonnet, C., Lecoeuvre, G., Forster, V., Sahel, J.-A., and Corral-Debrinski, M. 2006. mRNA localization to the mitochondrial surface allows the efficient translocation inside the organelle of a nuclear recoded ATP6 protein. RNA. 12:1408–1417

Kans, J. 2013. Entrez Direct: E-utilities on the Unix command line. Available at: https://www.ncbi.nlm.nih.gov/books/NBK179288/ [Accessed March 20, 2021].

Katoh, K., Misawa, K., Kuma, K.-I., and 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., Langmead, B., and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 12:357–360

Kleidon, J., Plesofsky, N., and Brambl, R. 2003. Transcripts and transcript-binding proteins in mitochondria of *Neurospora crassa*. Mitochondrion. 2:345–360

Kolondra, A., Labedzka-Dmoch, K., Wenda, J. M., Drzewicka, K., and Golik, P. 2015. The transcriptome of *Candida albicans* mitochondria and the evolution of organellar transcription units in yeasts. BMC Genomics. 16:1–22

Köster, J., and Rahmann, S. 2012. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 28:2520–2522

Kouvelis, V. N., Ghikas, D. V., and Typas, M. A. 2004. The analysis of the complete mitochondrial genome of *Lecanicillium muscarium* (synonym *Verticillium lecanii*) suggests a minimum common gene organization in mtDNAs of Sordariomycetes: phylogenetic implications. Fungal Genet. Biol. 41:930–940

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645

Kurtz, S., and Schleiermacher, C. 1999. REPuter: fast computation of maximal repeats in complete genomes. Bioinforma. Oxf. Engl. 15:426–427

Lang, B. F., Laforest, M.-J., and Burger, G. 2007. Mitochondrial introns: a critical view. Trends Genet. 23:119–125

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760

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

Li, Q., Wang, Q., Chen, C., Jin, X., Chen, Z., Xiong, C., Li, P., Zhao, J., and Huang, W. 2018. Characterization and comparative mitogenomic analysis of six newly sequenced mitochondrial genomes from ectomycorrhizal fungi (*Russula*) and phylogenetic analysis of the Agaricomycetes. Int. J. Biol. Macromol. 119:792–802

Liu, W., Cai, Y., Zhang, Q., Chen, L., Shu, F., Ma, X., and Bian, Y. 2020. The mitochondrial genome of *Morchella importuna* (272.2 kb) is the largest among fungi and contains numerous introns, mitochondrial non-conserved open reading frames and repetitive sequences. Int. J. Biol. Macromol. 143:373–381

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., and Bryant, S. H. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45:D200–D203

Mardanov, A. V., Beletsky, A. V., Kadnikov, V. V., Ignatov, A. N., and Ravin, N. V. 2014. The 203 kbp mitochondrial genome of the phytopathogenic fungus *Sclerotinia borealis* reveals multiple invasions of introns and genomic duplications. PLoS One. 9:e107536

Nagley, P., Farrell, L. B., Gearing, D. P., Nero, D., Meltzer, S., and Devenish, R. J. 1988. Assembly of functional proton-translocating ATPase complex in yeast mitochondria with cytoplasmically synthesized subunit 8, a polypeptide normally encoded within the organelle. Proc. Natl. Acad. Sci. 85:2091–2095

Nedelcu, A. M. 1997. Fragmented and scrambled mitochondrial ribosomal RNA coding regions among green algae: a model for their origin and evolution. Mol. Biol. Evol. 14:506–517

Ogata, H., Audic, S., Abergel, C., Fournier, P.-E., and Claverie, J.-M. 2002. Protein coding palindromes are a unique but recurrent feature in *Rickettsia*. Genome Res. 12:808–816

Pedersen, B. S., and Quinlan, A. R. 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 34:867–868

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

Rambaut, A. 2007. FigTree. Available at: Available at http://tree.bio.ed.ac.uk/software/figtree/ [Accessed March 20, 2021].

Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16:276–277

Richardson, D. R., Lane, D. J. R., Becker, E. M., Huang, M. L.-H., Whitnall, M., Suryo Rahmanto, Y., Sheftel, A. D., and Ponka, P. 2010. Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. Proc. Natl. Acad. Sci. 107:10775–10782

Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A. M. 1993. Group I and group II introns. FASEB J. 7:15–24

Sanchirico, M., Tzellas, A., Mason, T. L., Fox, T. D., Conrad-Webb, H., and Perlman, P. S. 1995. Relocation of the unusual *VAR*1 gene from the mitochondrion to the nucleus. Biochem. Cell Biol. 73:987–995

Seif, E., Leigh, J., Liu, Y., Roewer, I., Forget, L., and Lang, B. F. 2005. Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. Nucleic Acids Res. 33:734–744

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

Taanman, J.-W. 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta BBA-Bioenerg. 1410:103–123

The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47:D506–D515

Torriani, S. F. F., Goodwin, S. B., Kema, G. H. J., Pangilinan, J. L., and McDonald, B. A. 2008. Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. Fungal Genet. Biol. 45:628–637

Torriani, S. F. F., Penselin, D., Knogge, W., Felder, M., Taudien, S., Platzer, M., McDonald, B. A., and Brunner, P. C. 2014. Comparative analysis of mitochondrial genomes from closely related *Rhynchosporium* species reveals extensive intron invasion. Fungal Genet. Biol. 62:34–42

Valach, M., Burger, G., Gray, M. W., and Lang, B. F. 2014. Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. Nucleic Acids Res. 42:13764–13777

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., and others. 2007. Database resources of the national center for biotechnology information. Nucleic Acids Res. 36:D13–D21

Yildiz, G., and Ozkilinc, H. 2020. First characterization of the complete mitochondrial genome of fungal plant-pathogen *Monilinia laxa* which represents the mobile intron rich structure. Sci. Rep. 10:1–14

Yin, L.-F., Hu, M.-J., Wang, F., Kuang, H., Zhang, Y., Schnabel, G., Li, G.-Q., and Luo, C.-X. 2012. Frequent gain and loss of introns in fungal cytochrome b genes. PLoS One. 7:e49096

Zaccaron, A. Z., and Bluhm, B. H. 2017. The genome sequence of *Bipolaris cookei* reveals mechanisms of pathogenesis underlying target leaf spot of sorghum. Sci. Rep. 7:17217

Zaccaron, A. Z., Woloshuk, C. P., and Bluhm, B. H. 2017. Comparative genomics of maize ear rot pathogens reveals expansion of carbohydrate-active enzymes and secondary metabolism backbone genes in *Stenocarpella maydis*. Fungal Biol. 121:966–983

Zubaer, A., Wai, A., and Hausner, G. 2018. The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. Sci. Rep. 8:17591

6.7 Supplementary materials

6.7.1 Supplementary figures

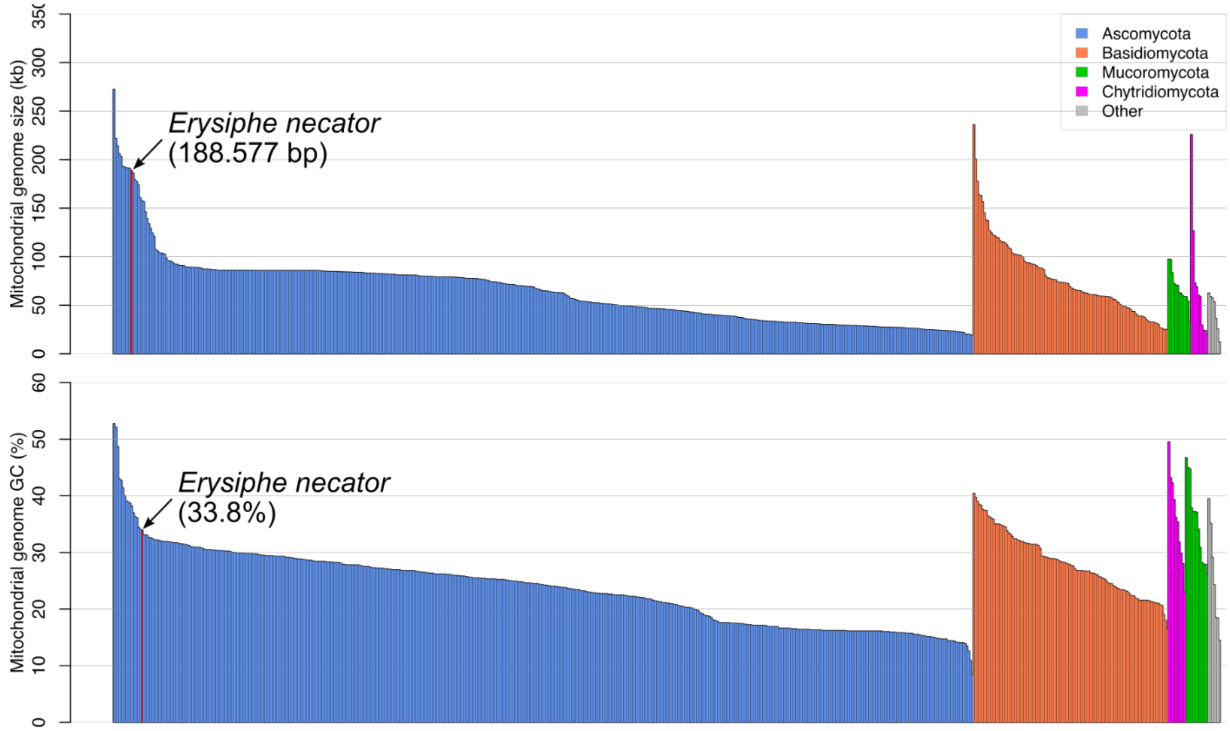


Figure 6.S1: Comparison of size and GC content of 636 fungal mitochondrial (mt) genomes. Each mt genome is represented by a single bar in the graph and is color-coded based on the phylum that the respective species belongs; blue: Ascomycota, orange: Basidiomycota; green: Mucoromycota, purple: Chytridiomycota, grey: Zoopagomycota, Blastocladiomycota and Cryptomycota. The bar representing the mt genome of the powdery mildew pathogen *Erysiphe necator* is highlighted in red. Mt genomes were organized by phylum and then by size (top bar chart) and GC content (bottom bar chart). The complete list of the fungal mt genomes included in these bar graphs can be found in Table 6.S1.

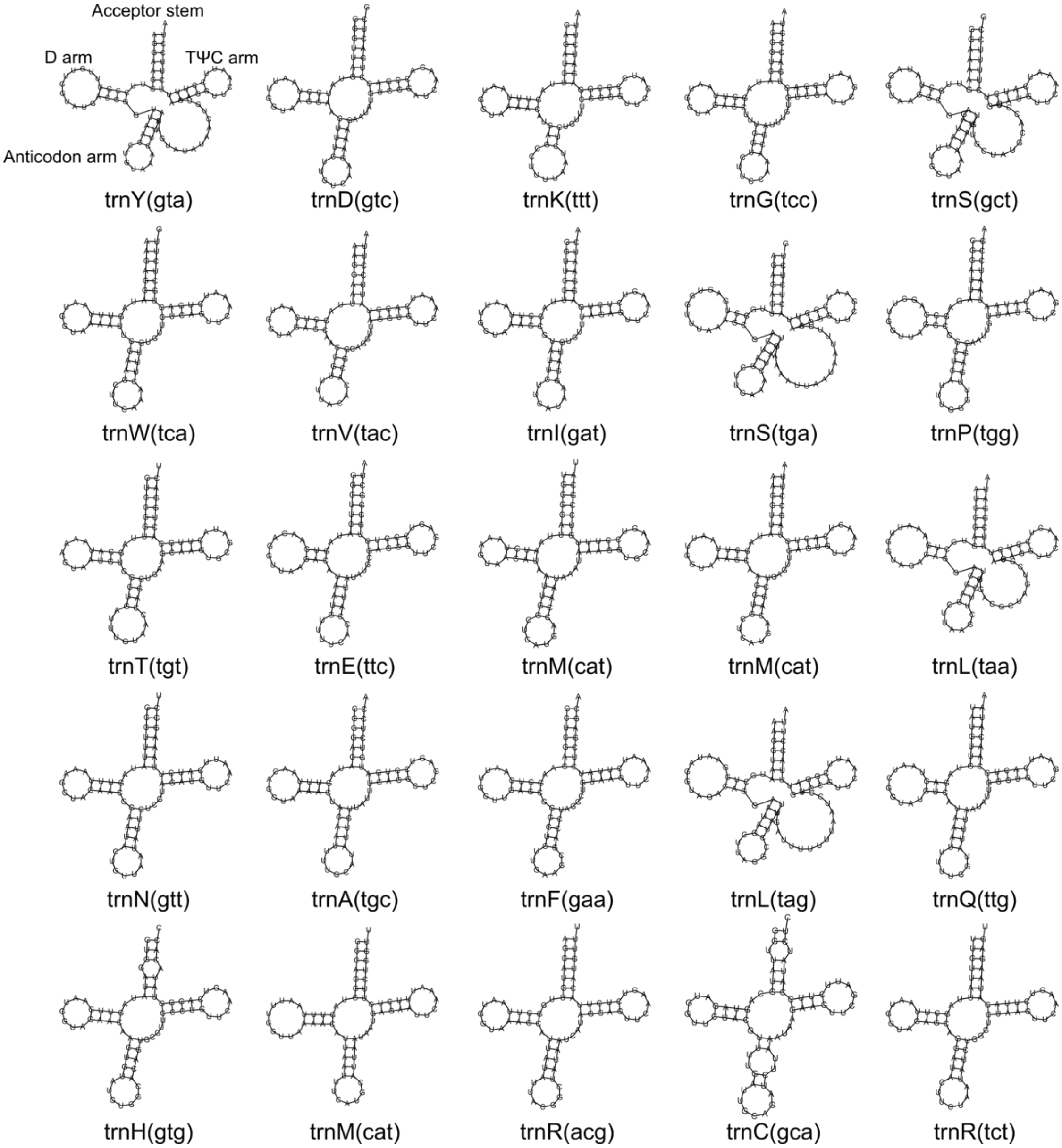


Figure 6.S2: Predicted secondary structures of mitochondrial tRNAs of *Erysiphe necator*. The 25 tRNAs are arranged in the figure top to bottom, left to right, according to their order of occurrence in the mitochondrial genome. The tRNA arms are illustrated for trnY(gta). Respective tRNA-anticodons are shown between parentheses. Structures of tRNAs were predicted with MITOS2 web server (http://mitos2.bioinf.uni-leipzig.de/index.py).

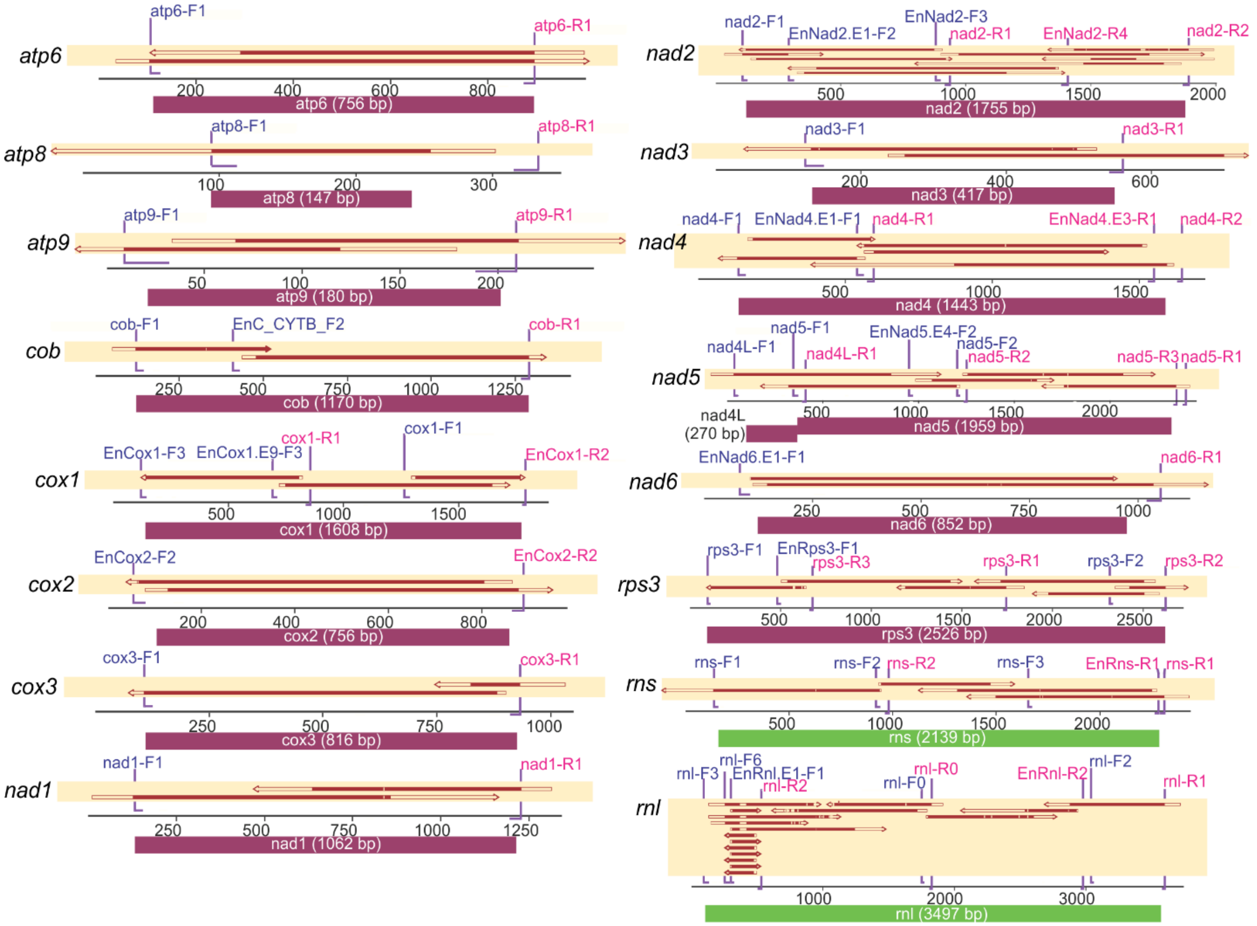
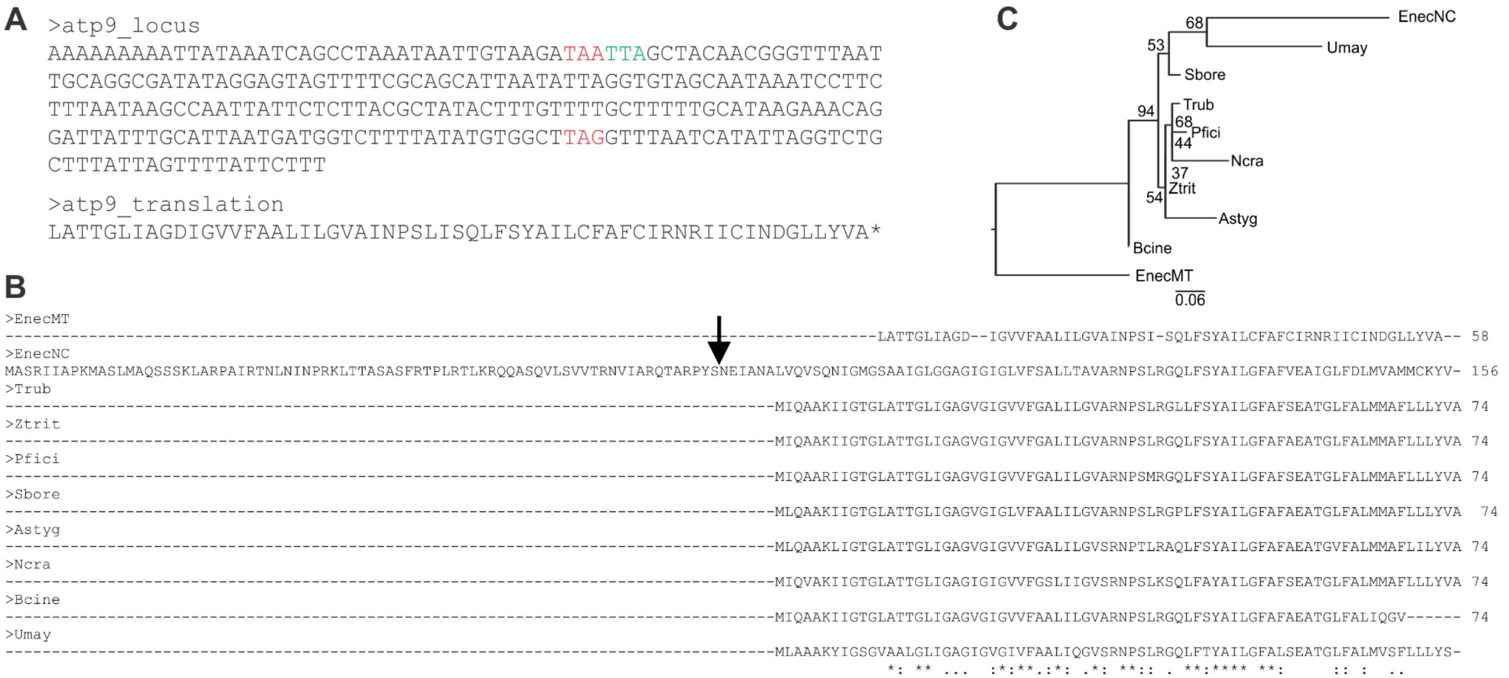


Figure 6.S3: Manual verification and confirmation of *Erysiphe necator* mitochondrial genes. Thick dark red and green bars represent coding sequences (exons for rRNAs). Thinner arrows indicate coverage of sequenced PCR products. Within these thin arrows, filled regions represent matches to the coding (or exonic) sequence, while transparent regions indicate mismatches or gaps in the sequenced PCR product. Deletions in the reference sequence are represented by red triangles (no occurrences). Location of primers are indicated with thin horizontal purple lines above (forward primers, in blue) or below (reverse primers, in pink) the scale bar. Primer sequences are shown in Table 6.S5. Gene names are indicated. Figures were generated with SnapGene v5.0.7 (https://www.snapgene.com).



```
A >atp9_locus
  AAAAAAAAATTATAAATCAGCCTAAATAATTGTAAGATAATTAGCTACAACGGGTTTAAT
  TGCAGGCGATATAGGAGTAGTTTTCGCAGCATTAATATTAGGTGTAGCAATAAATCCTTC
  TTTAATAAGCCAATTATTCTCTTACGCTATACTTTGTTTTGCTTTTTGCATAAGAAACAG
  GATTATTTGCATTAATGATGGTCTTTTATATGTGGCTTAGGTTTAATCATATTAGGTCTG
  CTTTATTAGTTTTATTCTTT

  >atp9_translation
  LATTGLIAGDIGVVFAALILGVAINPSLISQLFSYAILCFAFCIRNRIICINDGLLYVA*
```

```
B
>EnecMT
----------------------------------------------------------------------------------------------LATTGLIAGD--IGVVFAALILGVAINPSI-SQLFSYAILCFAFCIRNRIICINDGLLYVA-- 58
>EnecNC
MASRIIAPKMASLMAQSSSKLARPAIRTNLNINPRKLTTASASFRTPLRTLKRQQASQVLSVVTRNVIARQTARPYSNEIANALVQVSQNIGMGSAAIGLGGAGIGIGLVFSALLTAVARNPSLRGQLFSYAILGFAFVEAIGLFDLMVAMMCKYV- 156
>Trub
-----------------------------------------------------------------------------------MIQAAKIIGTGLATTGLIGAGVGIGVVFGALILGVARNPSLRGLLFSYAILGFAFSEATGLFALMMAFLLLYVA 74
>Ztrit
-----------------------------------------------------------------------------------MIQAAKIIGTGLATTGLIGAGVGIGVVFGALILGVARNPSLRGQLFSYAILGFAFAEATGLFALMMAFLLLYVA 74
>Pfici
-----------------------------------------------------------------------------------MIQAARIIGTGLATTGLIGAGVGIGVVFGALILGVARNPSMRGQLFSYAILGFAFSEATGLFALMMAFLLLYVA 74
>Sbore
-----------------------------------------------------------------------------------MLQAAKIIGTGLATTGLIGAGVGIGLVFAALILGVARNPSLRGPLFSYAILGFAFAEATGLFALMMAFLLLYVA  74
>Astyg
-----------------------------------------------------------------------------------MLQAAKLIGTGLATTGLIGAGVGIGVVFGALILGVSRNPTLRAQLFSYAILGFAFAEATGVFALMMAFLILYVA 74
>Ncra
-----------------------------------------------------------------------------------MIQVAKIIGTGLATTGLIGAGIGIGVVFGSLIIGVSRNPSLKSQLFAYAILGFAFSEATGLFALMMAFLLLYVA 74
>Bcine
-----------------------------------------------------------------------------------MIQAAKIIGTGLATTGLIGAGVGIGVVFAALILGVARNPSLRGQLFSYAILGFAFAEATGLFALIQGV------ 74
>Umay
-----------------------------------------------------------------------------------MLAAAKYIGSGVAALGLIGAGIGVGIVFAALIQGVSRNPSLRGQLFTYAILGFALSEATGLFALMVSFLLLYS-
                                                                                               *: ** ...  :*:**.:*: .*: **:: . **:**** **:     :: :  ..
```

Figure 6.S4: The mitochondrial ATP synthase subunit 9 (*atp9*) gene is likely a pseudogene in *Erysiphe necator*. (A) Nucleotide sequence of the truncated mitochondrial (mt) *atp9* gene of *E. necator* and its translated 59 amino acid product based on the Mold, Protozoan, and Coelenterate mt genetic code (genetic code = 4). Predicted start and stop codons are highlighted in green and red, respectively. The three nu cleotides immediately before the predicted start codon are also highlighted and encode a different stop codon. (B) Amino acid alignment of the mt ATP synthase subunit 9 encoded by the mt (EnecMT) and nuclear (EnecNC) *atp9* genes of *E. necator*. Included in the alignment are also homologous atp9 proteins from *Botrytis cinerea* (Bcine; AGN49018), *Zymoseptoria tritici* (Ztrit; YP_001648752), *Trichophyton rubrum* (Trub; Q01554), *Pestalotiopsis fici* (Pfici; AOW71154), *Neurospora crassa* (Ncra; Q12635), *Annulohypoxylon stygium* (Astyg; AYE67549), *Sclerotinia borealis* (Sbore; YP_009072368), and *Ustilago maydis* (Umay; Q0H8W9). The predicted cleavage site of the mt signal present in the ATP synthase subunit 9 encoded by the nuclear *atp9* gene of *E. necator* (KHJ33827) is indicated by an arrow. The alignment shows that the EnecMT atp9 misses a few amino acids in its N-terminus and a few others that are fairly conserved among fungal atp9 proteins. (C) Maximum likelihood phylogenetic tree constructed based on the amino acid alignment of selected atp9 proteins presented in Panel B. The tree was inferred with IQ-TREE v1.6.11 utilizing the cpREV amino acid substitution model, selected automatically by IQ-TREE as the best model, and 1,000 ultra-fast bootstrap replicates. The tree shows that the atp9 protein encoded by the mt *atp9* gene of *E. necator* strains apart from other fungal atp9 proteins.

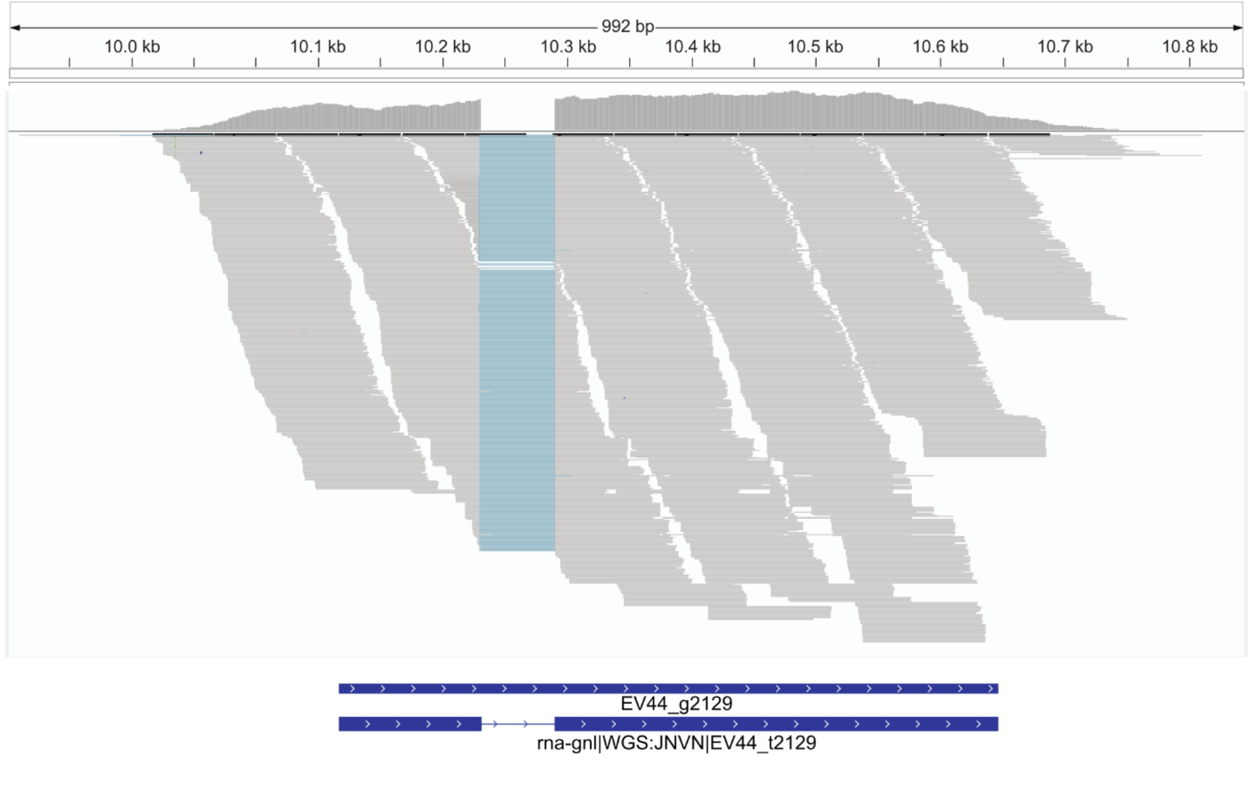


Figure 6.S5: The ATP synthase subunit 9 (*atp9*) gene encoded in the nuclear genome of *Erysiphe necator* is expressed. RNA-seq coverage at the region of the *E. necator* C-strain nuclear genome encoding a putative ATP synthase subunit 9 (GenBank accession KHJ33827). Gene structure is shown in blue at the bottom. A total of 3,317 RNA-seq reads mapped to the region and are shown in grey (GenBank SRA accession SRR1502880). Region was visualized with IGV v2.6.1 (https://igv.org).

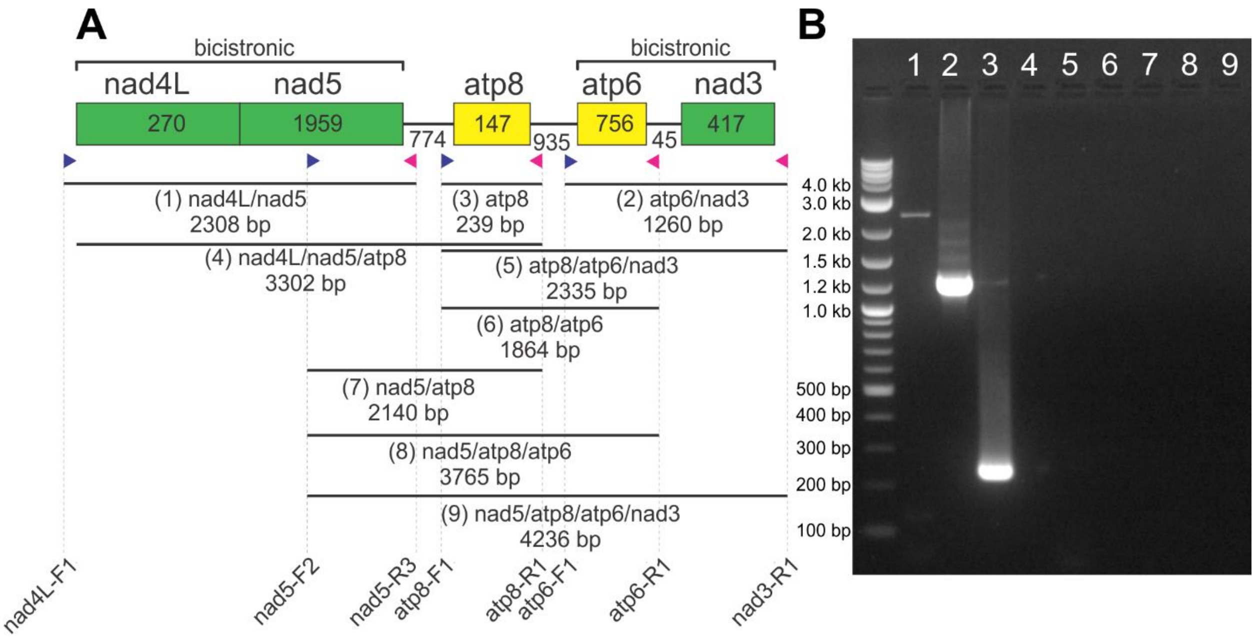


Figure 6.S6: Co-transcription of the mitochondrial gene pairs *nad4L/nad5* and *atp6/nad3* in *Erysiphe necator*. (A) Schematic representation of the organization of the *nad4L*, *nad5*, *atp8*, *apt6*, and *nad3* genes in the mitochondrial (mt) genome of *E. necator*. Length (bp) of coding sequences (i.e., with introns spliced out) are shown inside boxes that represent each gene. Lengths in base pairs of intergenic regions are shown in between boxes. Location of primers in the coding sequences are indicated with triangles. Horizontal lines indicate distances (bp) between primers. Introns were not considered when calculating these distances. Primer sequences are shown in Table 6.S5. (B) Agarose gel electrophoresis of PCR products depicting co-transcription of the mitochondrial gene-pairs *nad4L/nad5* and *atp6/nad3* in *E. necator*. cDNA of isolate C-strain was used as template in PCRs with primers aimed to detect bicistronic or polycistronic expression among the *nad4L*, *nad5*, *atp8*, *apt6*, and *nad3* genes. The nine primer combinations used (1-to-9) and the expected size of the products are shown in panel A. The exact location of the primers is shown in Fig 6.S3. Amplification of *atp8* was used as a positive control for the PCRs.

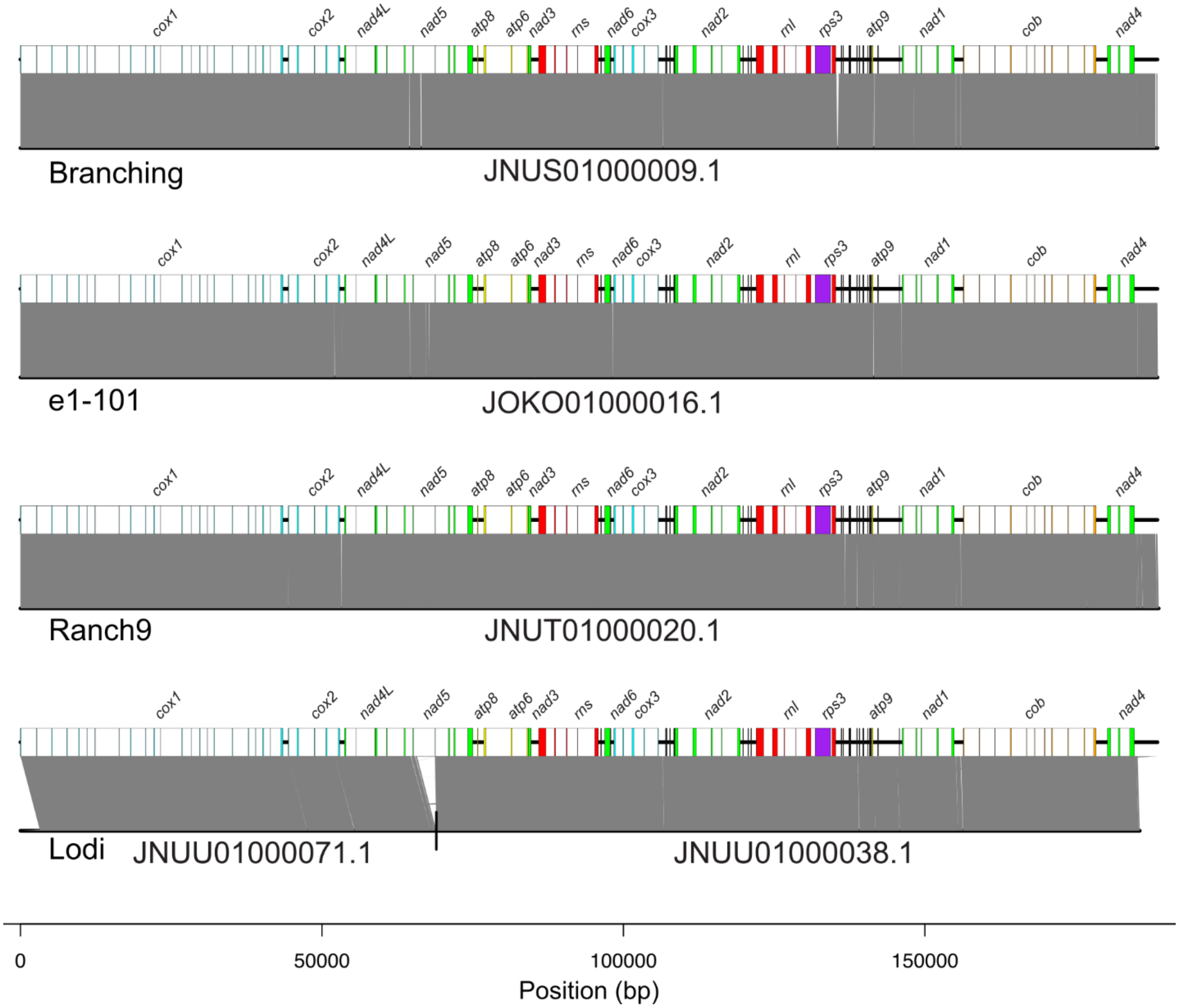


Figure 6.S7: The mitochondrial genomes of different isolates of *Erysiphe necator* are highly conserved. (A) Homologous regions between the mitochondrial (mt) genome of the *E. necator* C-strain (top strand) and the assembled scaffolds (bottom strand) containing the mt genomes of isolates Branching, Ranch9, e1-101, and Lodi. GenBank accession numbers of the contigs are show at the bottom. Mt genes of *E. necator* C-strain are represented as rectangles with intronic regions shown in white. Gray ribbons, seen as grey blocks in the figure due to their density, represent homologous regions identified with BLASTn searches with e-value < 1e-20. Scaffolds were rotated in order to match the start position of the mt genome of C-strain. The mt genome of strain Lodi is fragmented into two scaffolds, indicated by a vertical line at position 68,980. Intron 4 of gene *nad5* in isolate Lodi is missing in the assembled scaffolds. However, the region of this intron had normal coverage when whole-genome sequencing reads from isolate Lodi were mapped to the reference mt genome (Fig 6.3), indicating that this intron is conserved in isolate Lodi.

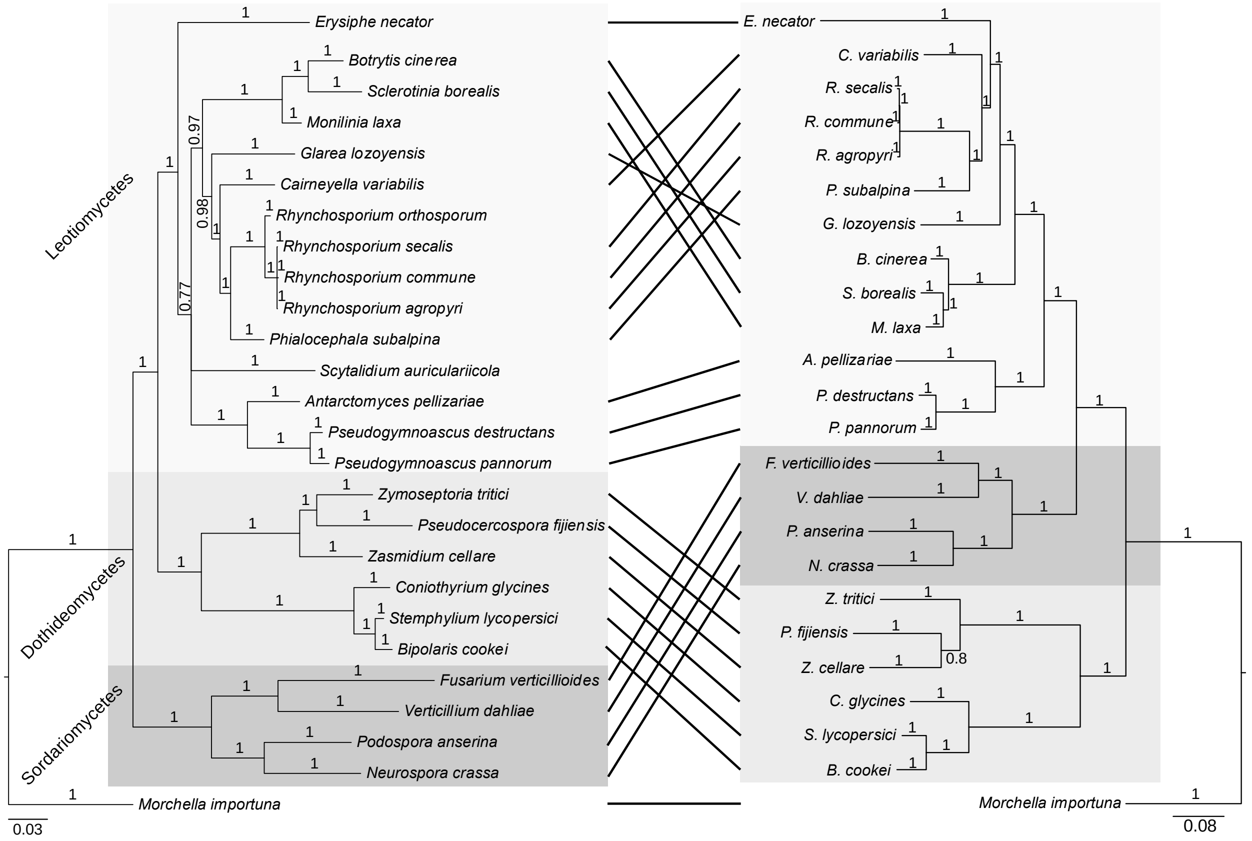


Figure 6.S8: Comparison of the mitochondrial (mt) and nuclear genome phylogeny of different fungi. Bayesian phylogenetic trees were constructed based on the mt and nuclear genomes, and are shown on the left-hand and right-hand side, respectively. The mt genome tree was constructed based on the concatenated alignment of the protein sequences of 12 core mt genes (*atp6*, *nad1-6*, *nad4L*, *cox1-3,* and *cob*). The nuclear genome tree was constructed based on the concatenated alignment of the protein sequences of 113 universally conserved genes in Eukaryotes. Supporting values of branches are indicated as Bayesian posterior probabilities. *Morchella importuna* (Pezizomycete) was used as outgroup in both trees. To construct the nuclear genome tree, universally conserved genes were identified with BUSCO v4.0.6 using the Eukaryote data set v10, which contains a total of 255 genes. For species with gene annotation available at NCBI, BUSCO was executed in proteome mode (*--mode proteins*). For species with no gene annotation available at NCBI, BUSCO was executed in genome mode (*--mode genome*). Species with no nuclear genome assembly available at NCBI were not included in the nuclear genome tree. All complete BUSCO genes (n=113), i.e., not duplicated or fragmented, in all analyzed species were utilized to infer the tree. Both trees were constructed with MrBayes v3.2.6 using the same parameters (see Methods in the main manuscript). Accession numbers of the proteins used to construct the trees are shown in Table 6.S12 and Table 6.S13 for the mt and nuclear genome trees, respectively.

6.7.2 Supplementary tables

Table 6.S1: Statistics of fungal mitochondrial genomes publicly available. Genomes were sorted by length. *Erysiphe necator* is highlighted. These mitochondrial genomes were found in the NCBI Organelle database (https://www.ncbi.nlm.nih.gov/genome/organelle/) as of January 2020. This table is available at https://zenodo.org/records/11211529.

Table 6.S2: Codon usage of the core mitochondrial genes from *Erysiphe necator*. The table shows the total number of codons for each gene. The last two columns show the number of GC bases for each codon and usage fraction among synonymous codons. This table is available at https://zenodo.org/records/11211529.

Table 6.S3: Repeats identified with REPuter in the mitochondrial genome of *Erysiphe necator*. Repeats were identified with REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) at E-value < 1e-5 and minimum repeat length of 8 bp. This table is available at https://zenodo.org/records/11211529.

Table 6.S4: Tandem repeats identified in the *Erysiphe necator* mitochondrial genome. Start and end coordinates in the genome are shown as well as the predicted copy number. The repeats were identified with the Tandem Repeat Finder program (https://tandem.bu.edu/trf/trf.html).

Start (bp)	End (bp)	Consensus	Copy number
9324	9363	AGCACTAAATAAATAAG	2.4
9565	9607	GGGTCCGTA	4.9
9565	9617	GGGTCCGTTGGTCCGTA	3.1
16596	16624	GCGTGCGGA	3.2
28629	28664	GGGTCCCTTTATAGCT	2.3
42055	42126	TAATATTGA	8
47399	47494	ACGGACCA	11.9
47394	47521	ACCCAACGGACCAACGGACCAACGGACCAACGGACCAACGGACCCTACGGACCCAACGG	2.2
53272	53318	GGGGGGGGGAGGAAGGAGGAA	2.1
75391	75445	GGTCCTCTTTTGTAAAAATAGGGTGAG	2
76418	76496	CAGCTCTTATTTATATGGGCTACGCATCCGTAGCAGGAG	2
85277	85389	CCCTTATTTATGGGCTCGCATCCGTAGAGGAGGACGCATTAC	2.7
85277	85409	CCCTTATTTATGGGCTCGCATCCGTAGGAGGACGCATCC	3.3
85751	85829	AGCAGCTC	9.9
92957	92985	GAAGGGGGGAGGG	2.2
96090	96120	CTCCGCATG	3.6
96765	96791	AGCAGCTG	3.4
102186	102228	AATTATTTAATTAATGATATT	2
104985	105009	AGGATCCGT	2.8
105158	105196	CCTTCGGAT	4.3
106459	106488	TAAGGGGGGGGGGG	2.1
119171	119233	ATCCGTAGG	7
119436	119476	CTTCTGGTCCGAAGGTTTT	2.2
120954	120988	GCCCTAAATAAATAAGA	2.1
121976	122019	GGGGGGGGGGGAACAAGTCCCGA	1.9
124601	124628	GCGGGAGC	3.5
133466	133511	AGCAGCTC	5.8
133466	133511	AGCAGCTCAGCAGCTAACCAGCTC	1.9
135593	135626	GGTCGAAG	4.1
140217	140264	GGGGGGGGTTAAGTAGGGCA	2.4
143993	144017	AAAATAAAAAA	2.3
145245	145299	TAGGCAGCCAGCAGGCCTTG	2.8
151282	151310	AGATAGATAAAT	2.4
162163	162192	CCTACGGAT	3.3
170733	170762	GGGAATCCCTGAC	2.3
178817	178859	GAGGGAGGCTGCATGTGG	2.4
178828	178863	ATGTGGGAGGGAGGGTGG	2
178862	178935	GGCA	18.5
178857	178931	AGGGAGGCAGGCAGGCAGGCAGGC	3.1
178926	178971	GGCAGGCAGGTGGGAGGGAGGGT	2
178999	179129	ACCAGTACCTGGCCGTCAGGAACTAGTTAAAAAAAATGTACCCTCCGA	2.6
185000	185028	CAGCTCCCTTTGAAC	1.9
185036	185063	G	28
187014	187053	TATTTGGTGATTAGTA	2.5
187692	187717	AGGATCCGT	2.9

Table 6.S5: Primers utilized to amplify and sequence the core mitochondrial genes of *Erysiphe necator*. The location of the primers in the mature gene transcripts are shown in Supplementary Fig 6.S3. This table is available at https://zenodo.org/records/11211529.

Table 6.S6: Mitochondrial *atp9* genes in members of Leotiomycetes and other fungal species. In *Erysiphe necator* as well in *Scytalidium auriculariicola* and *Botrytis cinerea* (highlighted), *atp9* is present but encodes a truncated protein.

Organism	Class	atp9 status	GenBank accession	Protein length (aa)
Erysiphe necator	Leotiomycetes	Truncated	NA	59
Sclerotinia borealis	Leotiomycetes	Present	YP_009072368.1	74
Monilinia fructicola	Leotiomycetes	Present	QGA74362.1	74
Scytalidium auriculariicola	Leotiomycetes	Present	QDG01211.1	63
Botrytis cinerea	Leotiomycetes	Present	AGN49018.1	68
Marssonina brunnea f. sp. 'multigermtubi'	Leotiomycetes	Present	YP_004842021.1	108
Glarea lozoyensis	Leotiomycetes	Present	YP_009306738.1	74
Phialocephala subalpina	Leotiomycetes	Present	YP_004733037.1	74
Gamarada debralockiae	Leotiomycetes	Present	AWL21288.1	74
Cairneyella variabilis	Leotiomycetes	Present	YP_009240971.1	74
Zymoseptoria tritici	Dothideomycetes	Present	YP_001648752	74
Trichophyton rubrum	Eurotiomycetes	Present	Q01554	74
Pestalotiopsis fici	Sordariomycetes	Present	AOW71154	74
Neurospora crassa	Sordariomycetes	Present	Q12635	74
Annulohypoxylon stygium	Sordariomycetes	Present	AYE67549	74
Ustilago maydis	Ustilaginomycetes	Present	Q0H8W9	73

Table 6.S7: Classification of *Erysiphe necator* mitochondrial introns and of the conserved domains encoded by their sequences. Introns start and end coordinates in the mt genome of *E. necator* are shown. Conserved domains found within the introns are shown. Conserved domains with more than one copy are indicated in the last column. This table is available at https://zenodo.org/records/11211529.

Table 6.S8: Intronic open reading frames (ORFs) encoding homing endonucleases (HEs) and reverse transcriptases (RTs) in the mitochondrial genome of *Erysiphe necator*. ORF IDs are composed of the gene name and the intron number (5' to 3' direction) where the ORFs are inserted. ORFs encoding proteins containing conserved domains related to RT (pfam00078) and HEs of the families LAGLIDADG_1 (pfam00961), LAGLIDADG_2 (pfam03161) and GIY-YIG (cd10445) are indicated. Combination of more than one domain is indicted with a plus sign. The length of the domains (in amino acids) is also indicated, when present. ORFs were considered degenerated if the length of the conserved domains is shorter than the expected. For each intronic ORF, the phase of the preceding, i.e., upstream exon is given in relation to the start codon (phase = 0) of the respective gene. The phase of the ORF is in relation to the previous exon, i.e., ORFs in-frame with the previous exon have phase = 0. Finally, the distance in base pairs between the start of the ORF and the end of the previous exon is shown as well as the number of in -frame stop codons (TAA and TAG) in these regions. This table is available at https://zenodo.org/records/11211529.

Table 6.S9: Polymorphic sites identified in the mitochondrial genes of *Erysiphe necator*. For each polymorphic site, its position in the genome is shown (C-strain as reference; GenBank MT880588.1) as well as the alleles of the reference genome (referred to as allele 0) and alternative alleles (referred to as alleles 1, 2 or 3, separated by commas). Alleles of four other *E. necator* isolates (Lodi, E1-101, Branching and Ranch9) are indicated. Polymorphic sites located within functional regions of the genome are highlighted. Further information of these sites is given in the form of the gene affected, which can be an intron encoded ORF (LD: LAGLIDADG, RT: reverse transcriptase) or a conserved mitochondrial gene; variants at the DNA and protein levels; and a description of the mutation. This table is available at https://zenodo.org/records/11211529.

Table 6.S10: Overrepresentation of the mitochondrial (mt) genome of different *Erysiphe necator* isolates in whole-genome sequencing (WGS) reads. Reads were mapped to the mt and nuclear genomes simultaneously.

Isolate	SRA accession	Total reads	Read length (bp)	Total reads after QC	Read length after QC (bp)	Total reads mapped	Total reads mapped (%)	Total reads mapped to mt genome	Total reads mapped to mt genome (%)	Total reads mapped to nuclear genome	Total reads mapped to nuclear genome (%)
C-strain	SRR1448450	33572038	260	32573584	257	31453034	96.6	14875950	45.7	16577084	50.9
Branching	SRR1448453	27947990	150	26836458	149	26732950	99.6	11102719	41.4	15630231	58.2
Ranch-9	SRR1448454	28550362	150	27506730	149	27468245	99.9	12066553	43.9	15401692	56.0
e1-101	SRR1448468	37319124	150	35866028	149	35805050	99.8	13683155	38.2	22121895	61.7
Lodi	SRR1448470	30404820	150	29208704	149	29097542	99.6	7710106	26.4	21387436	73.2

Table 6.S11: Estimation of mtDNA copy number and size of the nuclear genome of five isolates of *Erysiphe necator*. Only reads that passed quality control (QC) and were not flagged as PCR duplicates (SAM Flag 1024) were considered for the calculations. The coverage of the nuclear genome was determined as the median coverage of exons from nuclear genes (GCA_000798715.1), whereas coverage of the mt genome was calculated as the median coverage of all bases in the mt genome. The estimated mtDNA copy number was calculated as the ratio between the mt genome coverage and the nuclear genome coverage. The estimated nuclear genome size was calculated as the total base pairs from reads after QC, not flagged as PCR duplicates, that did not map to the mt genome, divided by the nuclear genome coverage.

Isolate	Total reads	Read length (bp)	Total bases	Total reads mapped to mt genome	Total bases mapped to mt genome	Total reads not mapped to mt genome	Total bases not mapped to mt genome	Nuclear genome coverage	Mt genome coverage	Estimated mtDNA copy number	Estimated nuclear genome size
C-strain	28359826	257	7288475282	11285250	2900309250	17074576	4388166032	46	14809	322	95394914
Branching	23907590	149	3562230910	9086345	1353865405	14821245	2208365505	28	7751	277	78870197
Ranch-9	19897648	149	2964749552	7999140	1191871860	11898508	1772877692	22	6351	289	80585350
e1-101	32210299	149	4799334551	11079774	1650886326	21130525	3148448225	38	8627	227	82853901
Lodi	24447869	149	3642732481	5926614	883065486	18521255	2759666995	34	4222	124	81166676

Table 6.S12: GenBank accession numbers of genes utilized to construct a phylogenetic tree of mitochondrial genomes. This table is available at https://zenodo.org/records/11211529.

Table 6.S13: GenBank accession numbers and sequences of genes utilized to construct a phylogenetic tree of nuclear genomes. Protein sequences from a total of 113 genes universally conserved in Eukaryotes were used to infer a tree. Genes were identified with BUSCO v4.0.6 using eukaryota_odb10. BUSCO was executed in genome mode for genomes without gene annotation. Protein accession number or sequence is shown for each BUSCO gene. BUSCO gene IDs are shown in the headers of the last 113 columns. BUSCO genes for *Scytalidium auriculariicola* and *Rhynchosporium orthosporum* were not predicted because they have no reference nuclear genome assemblies available at NCBI. This table is available at https://zenodo.org/records/11211529.

Chapter 7

Characterization of the mitochondrial genomes of three powdery mildew pathogens reveals remarkable variation in size and nucleotide composition.

Alex Z. Zaccaron
Ioannis Stergiopoulos

Leading author statement

I contributed to the conceptualization of this chapter, performed all analyses, generated all figures, contributed to the investigation and interpretation, wrote the original draft, and contributed to the review and editing of the final draft.

This chapter was published in: *Microbial Genomics*, December 2021, volume 7, number 12 (https://doi.org/10.1099/mgen.0.000720).

Abstract

Powdery mildews comprise a large group of economically important phytopathogenic fungi. However, limited information exists on their mitochondrial genomes. Here, we assembled and compared the mitochondrial genomes of the powdery mildew pathogens *Blumeria graminis* f. sp. *tritici*, *Erysiphe pisi*, and *Golovinomyces cichoracearum*. Included in the comparative analysis was also the mitochondrial genome of *Erysiphe necator* that was previously analyzed. The mitochondrial genomes of the four Erysiphales exhibit a similar gene content and organization but a large variation in size, with sizes ranging from 109,800 bp in *B. graminis* f. sp. *tritici* to 332,165 bp in *G. cichoracearum*, which is the largest mitochondrial genome of a fungal pathogen reported to date. Further comparative analysis revealed an unusual bimodal GC distribution in the mitochondrial genomes of *B. graminis* f. sp. *tritici* and *G. cichoracearum* that was not previously observed in fungi. The cytochrome *b* (*cob*) genes of *E. necator, E. pisi*, and *G. cichoracearum* were also exceptionally rich in introns, which in turn harbored rare open reading frames encoding reverse transcriptases that were likely acquired horizontally. *Golovinomyces cichoracearum* had also the longest *cob* gene (45 kb) among 703 fungal *cob* genes analyzed. Collectively, these results provide novel insights into the organization of mitochondrial genomes of powdery mildew pathogens and represent valuable resources for population genetics and evolutionary studies.

7.1 Introduction

Mitochondria are organelles that perform critical functions in most eukaryotic cells. They are responsible for energy production through oxidative phosphorylation, and are further involved in iron metabolism, aging, and programmed cell death (Chan 2006; Richardson et al. 2010). According to the endosymbiotic hypothesis, mitochondria are the descendants of an alpha-proteobacterium that was engulfed by an ancient eukaryotic cell, thus giving rise to a symbiotic relationship (Lang et al. 1999). This hypothesis is supported by the fact that mitochondria have their own, typically circular, genomes. Studies suggest that most of the ancient endosymbiont genes migrated to the nuclear genome or were lost throughout evolution, resulting in a substantial size reduction of the mitochondrial (mt) genome . Nonetheless, remarkable mt genome size variation has been observed, particularly among fungi in which it ranges from 12 kb in *Rozella allomycis* (James et al. 2013) to 531 kb in *Morchella crassipes* (Liu et al. 2020b).

Size variation among fungal mt genomes can often be explained by the number and size of introns present in their genes (Deng et al. 2018), with smaller mt genomes typically containing no or few introns and larger mt genomes being enriched in mt introns (Aguileta et al. 2014; Mardanov et al. 2014; Zubaer et al. 2018). Based on their secondary RNA structures and splicing mechanism, mt introns are typically classified into group I and group II introns, with group I introns further being classified into seven subgroups, i.e., IA, IA3, IB, IC1, IC2, ID, and I derived (I*) (Saldanha et al. 1993; Lang et al. 2007). Fungal mt introns also frequently host open reading frames (ORFs) encoding homing endonucleases (HEs) of the LAGLIDADG or GIY-YIG families, or reverse transcriptases (RTs), with the former most frequently found in group I introns and the latter in group II introns (Burt and Koufopanou 2004; Lambowitz and Zimmerly 2011). HEs enable intron self-splicing and transposition to an intronless cognate allele by cleaving DNA at specific target sequences of 13 to 40 bp long. This, in turn, activates the double-strand break repair mechanism of the cell that converts the intron$^-$ to an intron$^+$ allele, by using the intron$^+$ allele carrying the HE gene as template (Goddard and Burt 1999). In contrast, the mobility of group II introns carrying RT-coding ORFs is similar to that of retroelements in which RNA reverse splice into a DNA target, followed by reverse transcription (Lambowitz and Zimmerly 2011).

Despite their differences in size and intron content, the majority of fungal mt genomes typically contain a set of 14 core genes encoding proteins involved in the electron transport chain (ETC) and oxidative phosphorylation (Hausner 2003; Aguileta et al. 2014). These include *atp6*, *atp8,* and *atp9* that code for subunits 6, 8, and 9 of the ATP-synthase complex; *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5,* and *nad6* that code for seven Type I NADH dehydrogenases subunits of complex I; *cob* that codes for cytochrome b, one of the three catalytic subunits of complex III; and finally *cox1*, *cox2,* and *cox3* that code for the cytochrome c oxidase subunits 1, 2, and 3, respectively, of complex IV (Hausner 2003). In addition to protein-coding genes, fungal mt genomes also contain a set of mt-tRNAs as well as two genes, *rns* and *rnl*, encoding the small and large ribosomal subunits, respectively (Hausner 2003; Aguileta et al. 2014).

Mt genomes are an important genomic resource for studies in population genetics and evolutionary genomics. For phytopathogenic fungi, they are also of particular significance because mutations in mt proteins are often associated with resistance to fungicides that block the electron transport chain (ETC) and the production of energy in cells. The quinone outside inhibitor (QoI) fungicides, for instance, inhibit mt respiration by binding to the outer quinol-oxidation (Q_o) site of mt complex III, thus halting the production of ATP (Bartlett et al. 2002). Resistance to QoIs in fungi has been predominately associated with point mutations in the cytochrome *b* (*cob*) gene, with the most common mutations leading to the amino acid (aa) substitutions p.F129L, p.G137R, and p.G143A (Fernández-Ortuño et al. 2008). Of these, the p.G143A mutation has been shown to confer high levels of resistance to QoIs, often leading to the collapse of pathogen control in the fields. However, the formation of this amino acid substitution is blocked in some species by the presence of an intron at codon 143, as on the nucleotide level the substitution leads to the disruption of the intron splicing site that prevents intron splicing and leads to a frameshift. Thus, the presence or absence of this intron in cytochrome *b* is sometimes used as a predictor of resistance development to QoIs by the p.G143A mutation, although it is reported that in some species, such as *Botrytis cinerea* (Banno et al. 2009) and *Monilinia* spp. (Luo et al. 2010), different isolates carry different alleles of the cytochrome *b* gene that differ in the presence or absence of this intron.

Powdery mildew (PM) fungi represent a large and diverse group of obligate biotrophic Ascomycetes (Leotiomycetes, Erysiphales) that cause diseases in a wide range of monocots and dicots. The Erysiphales consists of approximately 900 species from 18 genera that infect nearly 10,000 species of angiosperms, including fruit trees, cereals, vegetables, and flowering plants, which highlights the economic importance of these pathogens (Braun and Cook 2012). Notably, the wheat PM fungus *Blumeria graminis* f. sp. *tritici* was recently ranked number eight among pests and pathogens that are causing the highest yield losses in wheat worldwide (Savary et al. 2019). In contrast to the monocot-adapted PM fungus *B. graminis* f. sp. *tritici* that only infects wheat, *Golovinomyces cichoracearum, Erysiphe pisi,* and *Erysiphe necator* are dicot-adapted fungi. Specifically, *Erysiphe pisi* is the causal agent of pea PM. It can colonize almost all parts of the plant and under severe infections it may induce yield losses of up to 50% (Fondevilla and Rubiales 2012; Warkentin et al. 1996). *Erysiphe necator* causes PM on grapes. It is distributed globally in grape producing areas and is considered the major disease threat to grape fruit production and quality (Gadoury et al. 2012). Finally, *Golovinomyces cichoracearum* is another dicot-adapted PM pathogen that has a broad host range that encompasses many plant species within members of the Asteraceae and Cucurbitaceae families, including important economic crops like lettuce, potato, melon, and several others (Braun 1987; Adam and Somerville 1996).

Despite the importance of PM fungi, genomic resources, including of mt genomes, for these plant pathogens are still scarce. Recently, the mt genome of *E. necator* was annotated and analyzed (Zaccaron et al. 2021). It revealed an atypical mt gene organization that differs from other fungal species and has one of the highest among fungi number of mt introns. To further gain an insight into the mt genomes of PM fungi, here we assembled and annotated the complete mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi,* and *G.*

cichoracearum. We next performed a comparative genomic analysis among the four PM fungi. The results herein provide novel insights into mt genome organization within Erysiphales and constitute valuable genomic resources for population genetic studies of PM pathogens.

7.2 Results

7.2.1 The mt genomes of PMs vary in size but have a similar complement of core mt genes and mt-tRNAs

To assemble the mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi*, and *G. cichoracearum*, whole-genome sequencing reads of the three species were obtained from NCBI and the reads corresponding to their mt genomes were extracted and assembled (Fig 7.S1). The resulting mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi*, and *G. cichoracearum* corresponded to single circular DNA sequences of 109,800 bp, 188,623 bp, and 332,165 bp, respectively (Fig 7.1). Notably, *G. cichoracearum*, *E. pisi*, and *E. necator* along with the non-PM species *Sclerotinia borealis* (mt genome size of 203,051 bp) have the largest mt genomes to date among Leotiomyecetes (Table 7.S1), whereas *G. cichoracearum* further has the second largest fungal mt genome currently available at the NCBI database (Fig 7.S2).

A total of 44, 111, and 58 genes and other ORFs were predicted in the mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi*, and *G. cichoracearum*, respectively, with all genes and ORFs transcribed from the sense strand (Fig 7.1). Despite the contrasting number of gene content, all three PM fungi contained single copies of the 13 core mt protein-coding genes involved in the ETC and oxidative phosphorylation, (i.e., *atp6*, *atp8*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *cob*, *cox1*, *cox2*, and *cox3*), as well as of the two mt rRNA genes (*rns* and *rnl*), and the putative ribosomal protein S3 (*rps3*) gene, located within an intron of *rnl*. The remaining genes corresponded to intronic ORFs predicted to encode homing endonucleases (HEs) of the LAGLIDADG or GIY-YIG families, or reverse transcriptases (RTs), and free-standing ORFs encoding type-B DNA polymerases, present only in *B. graminis* f. sp. *tritici* and *E. pisi* (Table 7.1). The total size of the exonic sequences of all mt genes, excluding intronic and free-standing ORFs, was also comparable among the three PM species and accounted for 22.1 kb (20.1%) of the mt genome in *B. graminis* f. sp. *tritici*, 22.6 kb (12.0%) in *E. pisi*, and 24.5 kb (7.4%) in *G. cichoracearum*. In *E. necator*, the total size of exonic sequences of all mt genes, excluding intronic and free-standing ORFs, was previously determined to account for 23.3 kb (12.3%) of its mt genome (Zaccaron et al. 2021), thus agreeing with the other three PM species (Fig 7.2A). Finally, also conserved among the three PM species and *E. necator* was the arrangement and orientation of the 13 core mt protein-coding genes in their mt genomes (Fig 7.2B and Fig 7.2C). This included *nad4L* and *nad5*, which overlapped by one base in that the last base pair of the stop codon of *nad4L* was the first base pair of the start codon of *nad5,* as well as *atp6* and *nad3*, which were positioned adjacent to each other, an atypical arrangement among fungal mt genomes (Zaccaron et al. 2021), and were separated by very short intergenic spacers of 46 bp in *B. graminis* f. sp. *titici,* 44 bp in *E. necator* and *E. pisi,* and 40 bp in *G. cichoracearum*.

Despite the overall conservation of the core gene complement among the mt genomes of the four PM species, some differences were noted as well. Specifically, while a transcriptionally active vestigial *atp9* gene that encodes a truncated protein was previously found in *E. necator* (Zaccaron et al. 2021), no ORF encoding a putative atp9 protein was identified in the other three PM species. Indeed, by q uerying the *E. necator atp9* coding sequence with BLASTn, homologous sequences were identified in the mt genomes of *E. pisi* and *G. cichoracearum*, but not in *B. graminis* f. sp. *tritici*. However, when translated, the homologous to *atp9* sequence in *E. pisi* contained a premature stop codon, whereas the one in *G. cichoracearum* contained insertions and deletions that caused frameshifts (Fig 7.S3). Another difference among the four PM species is the presence of an ORF upstream of the *rnl* gene in the mt genomes of *E. pisi* and *B. graminis* f. sp. *tritici* that encodes a putative DNA polymerase type-B (*dpo*), which is absent in *G. cichoracearum* as well as in *E. necator* (Zaccaron et al. 2021). Finally, whereas in *B. graminis* f. sp. *tritici*, *E. necator*, and *E. pisi* *rps3* is located within the last intron of *rnl*, in *G. cichoracearum rps3* is present within the penultimate intron of *rnl*.

In addition to the core mt protein-coding genes, the mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi,* and *G. cichoracearum* contained 25, 25, and 26 mt-tRNA encoding genes, respectively, capable of recognizing the standard set of 20 amino acids. These numbers are once again comparable to the 25 mt-tRNA genes previously found in *E. necator* (Zaccaron et al. 2021). As for the core mt genes, the order and orientation of the mt-tRNAs was conserved among the four mt genomes (Fig 7.2C), whereas the predicted secondary structures of the identified mt-tRNAs revealed that they all fold into common cloverleaf-like structures (Fig 7.S4). Most mt-tRNA genes identified in the mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi,* and *G. cichoracearum* were single-copy. Exceptions were the mt-tRNAs that decode arginine (*trnR*), leucine (*trnL*), methionine (*trnM*), and serine (*trnS*) that were present in two, two, three, and two copies, respectively, in the mt genomes of *B. graminis* f. sp. *tritici* and *E. pisi*, whereas *G. cichoracearum* also contained the same number of copies of *trnR*, *trnL* and *trnM*, but had a third copy of *trnS* between *nad1* and *cob*.

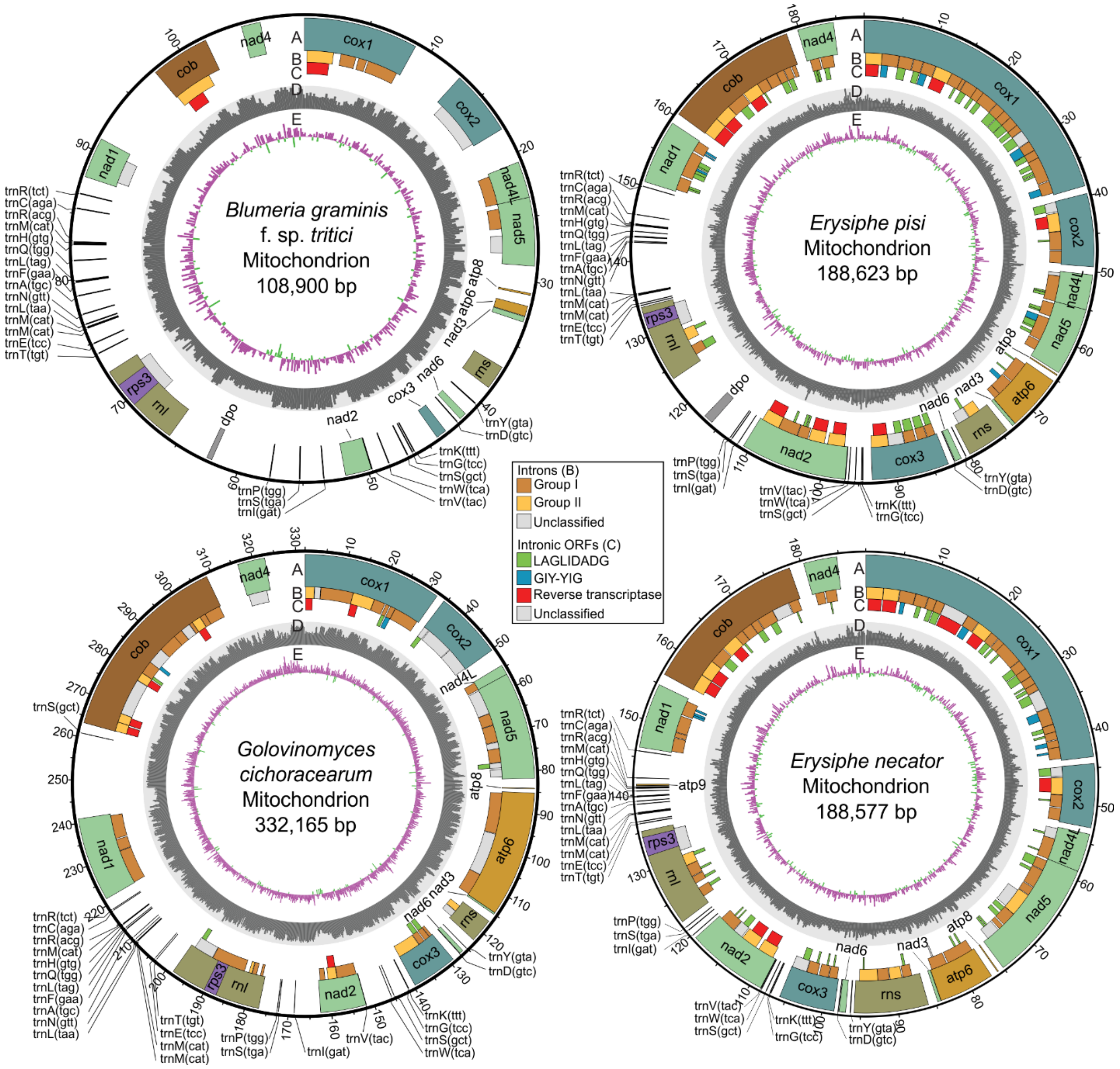


Figure 7.16: Organization of the mitochondrial genomes of the powdery mildew fungi *Blumeria graminis* f. sp. *tritici*, *Erysiphe necator*, *Erysiphe pisi*, and *Golovinomyces cichoracearum*. Tracks: (A) Core protein-coding and other conserved genes present in the mitochondrial (mt) genomes of the four powdery mildew species. These include genes encoding the subunits of complex I (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*), complex III (*cob*), complex IV (*cox1*, *cox2* and *cox3*), the ATP-synthase complex (*atp6*, *atp8* and *atp9*), the small and large ribosomal subunits (*rns* and *rnl*), the ribosomal protein S3 (*rps3*), and a set of mt-tRNAs. Open reading frames (ORFs) encoding DNA polymerase type-B are indicated as *dpo*. (B) Introns present in the mt genes of the four powdery mildew species. The introns are classified as group I and group II introns, or as unclassified based on their secondary RNA structures and splicing mechanism. (C) ORFs present within introns, encoding homing endonucleases of the LAGLIDADG and GIY-YIG families, or reverse transcriptases. (D) GC content scaled from 0% to 70%. (E) GC skew calculated as (G-C)/(G+C), with positive and negative values in purple and green, respectively. Both GC content and GC skew were calculated using a non-overlapping sliding window of 200 bp. Data for the mt genome of *E. necator* were obtained from (Zaccaron et al. 2021) and the circular representation of its mt genome is accordingly adjusted.

Table 7.14: Mitochondrial genome statistics of the powdery mildew fungi *Blumeria graminis* f. sp. *tritici* (Bgt), *Erysiphe necator* (En), *Erysiphe pisi* (Ep), and *Golovinomyces cichoracearum* (Gc).

Genome features	*Bgt*	*En*[1]	*Ep*	*Gc*
Total size (bp)	109,800	188,577	188,623	332,165
Overall GC (%)	48.3	33.9	34	45.1
GC-skew (G-C)/(G+C)	0.173	0.101	0.117	0.16
AT-skew (A-T)/(A+T)	0.015	0.031	0.032	0.071
Repetitive DNA (%)	13.6	8	12.7	11.3
Genes	44	107	111	58
Mt-tRNAs	25	25	26	25
Introns	11	70	61	53
Group I introns	5	48	42	28
Group II introns	2	13	12	10
Unclassified introns	4	9	7	15
Average intron size (bp)	2,359	1,992	2,025	3,900
Intergenic regions size (bp)	62,989	28,343	43,247	103,902
Intronic regions size (bp)	25,956	139,477	123,556	206,694
Core exonic sequences GC (%)	32.6	32	31.8	33.6
Intergenic regions GC (%)	54.5	38.9	35.9	47.6
Intronic regions GC (%)	45.5	33.2	33.7	45.1
Intronic ORFs	2	65	69	16
LAGLIDADG ORFs	0	44	50	7
GIY-YIG ORFS	0	9	9	2
Reverse transcriptase ORFs	2	11	10	7
Accession	MT880591	MT880588	MT880589	MT880590

[1] Data for the mitochondrial genome of *E. necator* were obtained from (Zaccaron et al. 2021).

7.2.2 Repetitive DNA content and mt introns are poorly conserved among PMs

Although coding sequences of the core mt genes were comparable in size among the three PM species analyzed in this study and *E. necator* (Zaccaron et al. 2021), large differences were found in their repetitive DNA content as well as their intronic and intergenic regions (Fig 7.2A and Table 7.1). Specifically, repetitive DNA accounted for 13.6%, 12.7%, and 11.3% of the mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi*, and *G. cichoracearum,* which also harbored 142, 97, and 206 short tandem repeats, respectively (Table 7.S2). This contrasts with the mt genome of *E. necator*, which has a repetitive DNA content of 8% and 45 short tandem repeats (Zaccaron et al. 2021), and suggests that repetitive DNA contributes to a greater extend to the organization of the mt genomes of *B. graminis* f. sp. *tritici*, *E. pisi*, and *G. cichoracearum* as compared to *E. necator*.

Large differences in intron content were observed as well among the mt genomes of the four PM species, with variations extending to the number, density, size, and distribution of introns hosted within orthologous mt genes as well as conserved domains present in these introns (Fig 7.3A). Specifically, a total of 11, 61, and 53 introns were identified in the mt genomes of *B. graminis* f. sp. *tritici, E. pisi,* and *G. cichoracearum* (Table 7.S3). Intron density was also lowest for *B. graminis* f. sp. *tritici* (0.7 intron per kb of cds), followed by *G. cichoracearum* (3.3 introns per kb of cds), and *E. pisi* (4.0 introns per kb of cds). The number of introns previously identified in the mt genome of *E. necator* was 70, with an intron density of 4.4 introns per kb of cds (Zaccaron et al. 2021). This indicates that intron features vary more homogeneously between *E. necator* and *E. pisi* rather than between the other species-pairs (Fig 7.3A). A weak correlation (*R* = 0.49, *p*-value = 0.51) existed between intron numbers and mt genome size among the four PM species (Fig 7.S5A), whereas a stronger correlation (*R* = 0.95, *p*-value = 0.055) was observed between the total size of introns and mt genome size (Fig 7.S5B), indicating that intron size, rather than intron numbers, dictated the variation in mt genome size among the four PM species. This conclusion is also supported by the differences in size among orthologous genes in the four PM species (Table 7.S4), which are largely explained by the size of their introns (*R* = 1, *p*-value < 2.2e-16) (Fig 7.S5C).

Next to their differences in size, introns inserted at the same sites within orthologous genes often shared limited primary sequence identity as well, indicating poor intron conservation among the four Erysiphales (Fig 7.3B and Table 7.S5). The most conserved intron among the four PM species was the intron inserted at position 99 of the cytochrome *b* gene (*cob*) cds (cob-99), with an average pairwise identity of 63.0%, indicating that there are no introns that are highly conserved among all four species. The poor conservation of introns inserted at the same sites within orthologous genes further extended to the presence/absence of ORFs embedded in them. Indeed, ORFs encoding HEs or RTs were found in 69 (85.2%) and 65 (82.8%) of the introns in *E. pisi* and *E. necator*, respectively, but only in 16 (30.2%) and two (18.2%) of the introns in *G. cichoracearum* and *B. graminis* f. sp. *tritici,* respectively (Fig 7.3A). These results indicate that *E. pisi* (1.13 ORFs/intron) and *E. necator* (0.93 ORFs/intron) were richer in intronic ORFs as compared to *B. graminis* f. sp. *tritici* (0.18 ORFs/intron) and *G. cichoracearum* (0.30 ORFs/intron). Overall, of the 61 introns (42 group I, 12 group II, and 7 unclassified) detected in the mt genome of *E. pisi*, 42 contained HEs of the LAGLIDADG (n=36) or GIY-YIG (n=8) families, and 10 contained RT-encoding ORFs. In an analogous way, *B. graminis* f. sp. *tritici* contained two group II introns containing RT-encoding ORFs, but none HE-encoding ORF, whereas *G. cichoracearum* had seven RT-encoding ORFs within its group II introns and five group I introns with HE-encoding ORFs of the LAGLIDADG (n=4) or GIY-YIG (n=1) families (Fig 7.3A).

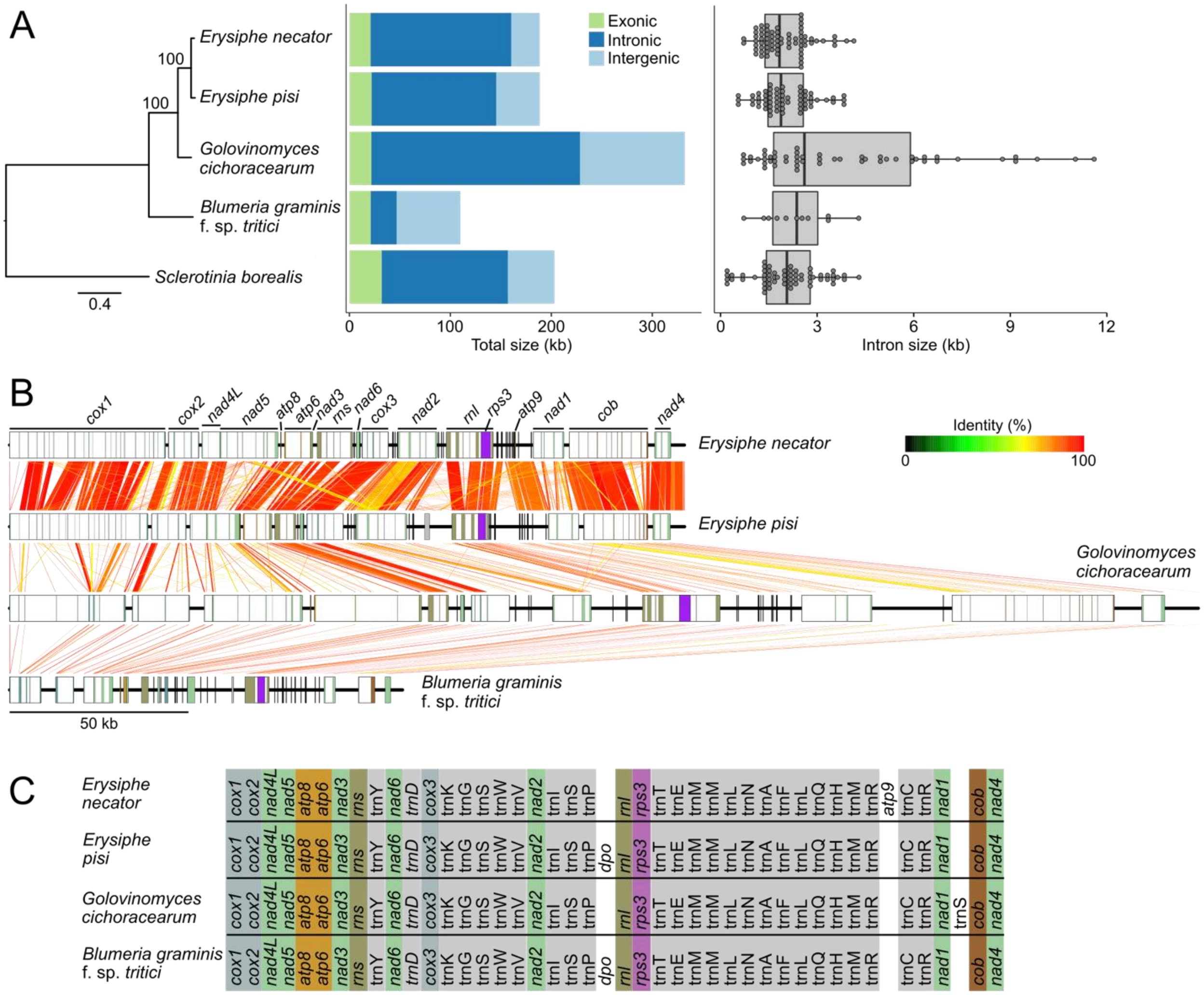


Figure 7.17: Comparison of the mitochondrial (mt) genomes of the powdery mildew (PM) fungi *Blumeria graminis* f. sp. *tritici*, *Erysiphe necator*, *Erysiphe pisi*, and *Golovinomyces cichoraceaum*. (A) Maximum likelihood phylogenetic tree constructed based on the amino acid sequences encoded by 14 core mt genes. The tree was rooted on *Sclerotinia borealis,* which is also a species of Leotiomycetes. The bar plot shows the total size of the mt genomes of the four PM species. Bars show the total size of exonic (excluding ORFs encoding homing endonucleases, reverse transcriptases, or DNA polymerases), intronic, and intergenic regions. Differences in mt genome size are due to differences in the size of the intronic and intergenic regions. The box plot depicts the number and length of the introns present in the mt genes of the four PM species. Individual introns are represented by points and are grouped into bins of 150 bp in width. (B) The mt genomes of the four PM species are syntenic. Genes are represented as rectangles with intronic regions in white and exonic regions color-coded as in Fig 7.1. Ribbons connect homologous regions identified with BLASTn (e-value < 1e-10) and are color-coded based on the percent of nucleotide identity. (C) Mitochondrial gene order in the four PM species is highly conserved. The *rps3* gene is encoded within an intron of *rnl*. The *atp9* gene was predicted in *E. necator* but not in the other three PM species. ORFs encoding putative DNA polymerases type-B are represented as *dpo*. Data for the mt genome of *E. necator* were obtained from (Zaccaron et al. 2021).

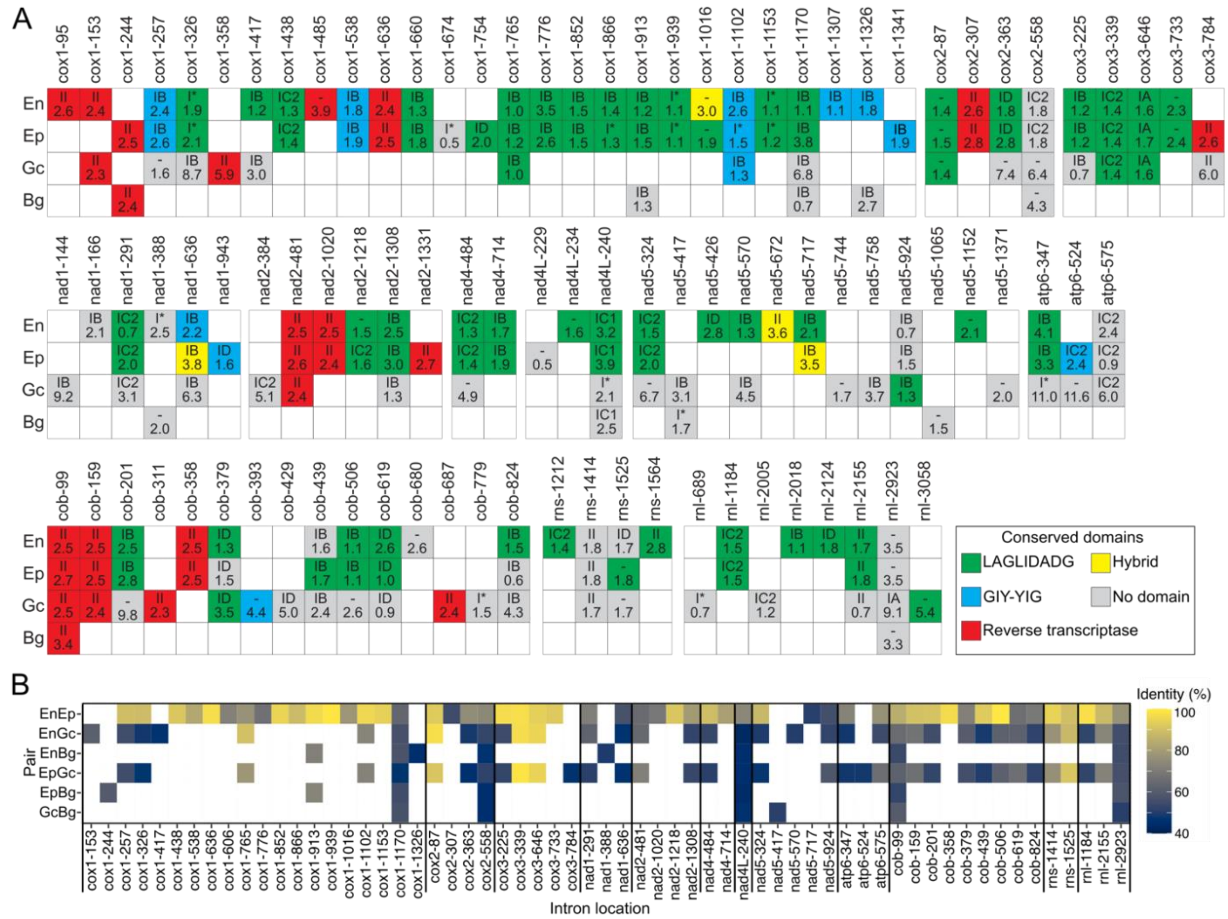


Figure 7.18: Intron insertion sites in mitochondrial (mt) genes of the powdery mildew (PM) fungi *Blumeria graminis* f. sp. *tritici*, *Erysiphe necator*, *Erysiphe pisi*, and *Golovinomyces cichoracearum*. (A) Introns are depicted as color-coded square boxes and their insertion sites within respective genes are shown as columns. Base-pair coordinates are given at the top and are in reference to the coding sequences of the *E. necator* mt genes. Intron type and length (kb) are shown inside the boxes, at the top and bottom, respectively. Colors indicate the absence (grey boxes) or presence (colored boxes) of conserved domains in the introns. Conserved domains include homing endonucleases (HEs) from the LAGLIDADG (green) or GIY-YIG families (blue boxes), reverse transcriptases (RTs) (red boxes), and hybrid LAGLIDADG/GIY-YIG or GIY-YIG/RT domains (yellow boxes). The introns inserted at *cox1*-1016, *nad1*-636, and *nad5*-717 positions contain a hybrid LAGLIDADG/GIY-YIG domain, and the intron inserted at the *nad5*-672 position contains a hybrid GIY-YIG/RT domain. Absence of introns is indicated with blank white boxes. Genes without introns in all four species are not shown. The figure shows that intron content and intron insertion sites within orthologous mt genes vary widely among the four PM species, but are correlated to the phylogenetic relatedness of the species. (B) Heat-map showing the pairwise nucleotide identity between introns inserted in the same site within orthologous mt genes of the four PM species. En: *E. necator*; Ep: *E. pisi*; Gc: *G. cichoracearum*; Bg: *B. graminis* f. sp. *tritici*. As for other intron features, introns inserted in the same site share limited primary sequence identity that is largely defined by the phylogenetic relatedness of the species. The data for the mt genome of *E. necator* were obtained from (Zaccaron et al. 2021).

7.2.3 The mt genomes of *B. graminis* f. sp. *tritici* and *G. cichoracearum* show a bimodal distribution of GC content

High variation in GC content was observed among the mt genomes of the three PM species, ranging from 34.0% in *E. pisi*, 45.1% in *G. cichoracearum*, and 48.3% in *B. graminis* f. sp. *tritici*, which indicates substantial differences in nucleotide composition (Table 7.1). Assessment of GC content separately in gene coding and non-coding sequences indicated that exonic sequences in *B. graminis* f. sp. *tritici, E. pisi,* and *G. cichoracearum* exhibited an average GC content of 32.6%, 31.8%, and 33.6%, respectively, whereas intergenic and intronic sequences exhibited an average GC content of 51.9%, 34.3%, and 45.9%, respectively (Table 7.1). The overall GC content of the *E. necator* mt genome was previously determined as 33.8%, with exonic and intronic sequences exhibiting an average GC content of 32.0% and 34.1%, respectively (Zaccaron et al. 2021). Collectively, these results indicate that the differences in GC content among the four Erysiphales are mainly driven by nucleotide compositional differences between functional and non-functional sequences.

An inspection using a sliding window analysis of the variation in the GC content across the four mt genomes further showed that GC content was unimodally distributed in the mt genomes of *E. necator* and *E. pisi*. In contrast, the mt genomes of *B. graminis* f. sp. *tritici* and *G. cichoracearum* exhibited a bimodal distribution, characterized by interspersed isochore-like regions of GC-poor (i.e., GC of 20-40%) and GC-rich (i.e., GC of 50-65%) content (Fig 7.4). The GC-rich regions were longer in *B. graminis* f. sp. *tritici* (average of 837 bp) and covered a higher fraction of its mt genome (52.6%) compared to *G. cichoracearum* (average length of 450 bp, covering 45.9% of its mt genome). In contrast, only 2.9% and 3.7% of the mt genomes of *E. necator* and *E. pisi*, respectively, had GC content higher than 50%. As expected, almost all of the GC-rich regions in the mt genomes of *B. graminis* f. sp. *tritici* (66 out of 69 regions) and *G. cichoracearum* (230 out of 236 regions) were located within intergenic regions or introns and did not appear to affect neighboring coding sequences. Repetitive DNA identification performed with RepeatMasker showed that none of the GC-rich regions contained interspersed repeats, whereas simple repeats and low complexity regions accounted for only 7.5% of the GC-rich regions in *B. graminis* f. sp. *tritici* and 3.3% of the regions in *G. cichoracearum*, suggesting that the GC-rich regions are not composed of repetitive DNA. Finally, homology searches performed with BLASTn (e-value < 1e-5) further revealed that none of the GC-rich regions was conserved between *G. cichoracearum* and *B. graminis* f. sp. *tritici*, whereas no evidence of a possible migration of these sequences from their respective nuclear genome were obtained as well.

To further inquire on the origin and evolution of the GC-rich regions in *B. graminis* f. sp. *tritici* and *G. cichoracearum*, those that did not overlap with functional sequences were then queried in BLASTn searches against the NCBI nr database. From the 66 GC-rich regions present in *B. graminis* f. sp. *tritici* queried, 58 had only a single hit on one of the scaffolds of the *B. graminis* f. sp. *tritici* assembly containing its mt genome. The other eight regions had somewhat significant hits (e-value < 1e-30) to sequences from a wide range of species, including fungi and animals (Table 7.S6). When querying the 230 GC-rich regions present in *G. cichoracearum*, then 212 had hits solely to scaffolds from a genome assembly of the oomycete *Albugo laibachii*. However, further BLAST searches revealed that this same *Albugo laibachii* genome assembly contained a contig (FR825110.1) highly homologous (99% nucleotide identity) to the internal transcribed spacer (ITS) sequences from *Golovinomyces* spp. Taken together, these data suggest that the GC-rich isochore-like regions present in the genomes of *B. graminis* f. sp. *tritici* and *G. cichoracearum* are likely to have been formed *de novo* in each of the two species or to have been transmitted vertically rather than horizontally.

Bimodal distribution of GC content has been previously described in nuclear genomes of fungi, but it is hardly ever reported in fungal mt genomes. Therefore, we next examined the patterns of GC content variation in 22 mt genomes from phylogenetically related species of Leotiomycetes. However, only the genome of *Blumeria graminis* f. sp. *hordei* was organized into isochore-like regions of high-GC and low-GC regions, as were the mt genomes of *B. graminis* f. sp. *tritici* and *G. cichoracearum* (Fig 7.S6). We then extended our analysis to 949 mt genomes from phylogenetically diverse fungi, with the stipulation that they should have an average GC content of 40% or higher. Surprisingly, from the 949 fungal mt genomes analyzed, only 23 from 21 different species, representing four phyla, had GC ≥40%, indicating that fungal mt genomes are predominately AT-rich (Fig 7.S7 and Table 7.S1). However, the majority (n=17) of these 22 mt genomes had a single GC peak between 40-60%, indicative of a homogeneous GC content and a unimodal distribution (Fig 7.S8). The mt genomes of only seven species (i.e. *B. graminis* f. sp. *hordei*, *Magnusiomyces tetraspermus*, *Morchella crassipes*, *Alternaria alternata*, *Pyrenema ompalodes*, *Synchytrium endobioticum*, and *Glomus cerebriforme*) exhibited GC content distribution patterns suggestive of bimodal organization but only in *B. graminis* f. sp. *hordei* the heterogeneity in GC content was as pronounced as in *B. graminis* f. sp. *tritici* and *G. cichoracearum* (Fig 7.S8). Taken together, these data indicate that *B. graminis* f. sp. *tritici* and *G. cichoracearum* display a rather idiosyncratic mode of bimodality in GC content in their mt genomes.

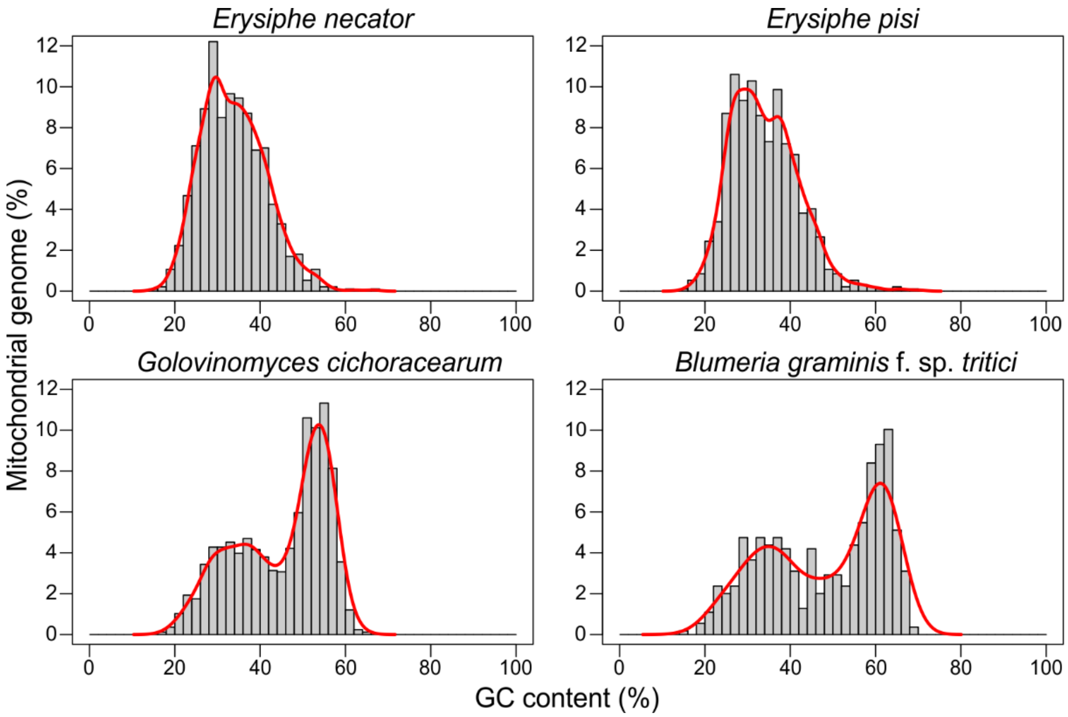


Figure 7.19: GC content comparison among the mitochondrial (mt) genomes of the powdery mildew fungi *Blumeria graminis* f. sp. *tritici*, *Erysiphe necator*, *Erysiphe pisi*, and *Golovinomyces cichoracearum*. Histograms depict the percentage of the respective mt genome (Y-axis) containing the indicated GC content (X-axis). Solid lines represent estimated distributions based on the kernel density of the histograms. The graphs show that the mt genomes of *E. necator* and *E. pisi* have a unimodal distribution of GC content with a peak at approximately 29%, whereas the mt genomes of *G. cichoracearum* and *B. graminis* f. sp. *tritici* exhibit a bimodal distribution of GC content, with two peaks at 36% and 53%, and 35% and 61%, respectively. Data for the mt genome of *E. necator* were obtained from (Zaccaron et al. 2021).

7.2.4 Erysiphales have large cytochrome *b* genes with rare introns embedded in them that contain RT-encoding ORFs

The cytochrome *b* (*cob*) gene receives particular attention in phytopathogenic fungi as point mutations in this gene are associated with resistance to QoI fungicides (Fernández-Ortuño et al. 2008). The *cob* genes of *B. graminis* f. sp. *tritici, E. necator*, *E. pisi,* and *G. cichoracearum* exhibited fairly conserved coding sequences, both in size (1,161 bp, 1,170 bp, 1,170 bp, and 1,170 bp, respectively) and nucleotide identity (89.4% to 97.7%) (Table 7.S7). In contrast, a large variation was seen in the size (4.5 kb to 45.3 kb) and number of introns (1 to 13) present in them (Zaccaron et al. 2021) (Fig 7.5A, Fig 7.5B and Table 7.S7), whereas pairwise alignments of homologous introns further showed that they exhibit a much broader range of nucleotide identities as compared to the *cob* coding sequences (Fig 7.3B). However, despite such differences in size and primary sequence identity, some notable similarities in the distribution and content of introns present in the *cob* genes of *E. necator*, *E. pisi,* and *G. cichoracearum* were observed as well. The *cob* gene in *B. graminis* f. sp. *tritici* stood apart from its orthologs in the other three PM species, as it contained only one intron (cob-99), which was conserved among all four PM species (Fig 7.5B). Intron insertion sites in the *cob* genes of *E. necator*, *E. pisi,* and *G. cichoracearum* were fairly conserved, with nine sites shared between *E. necator* and *E. pisi*, and eight sites being conserved in all three species (Fig 7.5C). Notably, some of the differences in intron insertions sites were found around codons 129, 137, and 143, which affect amino acids that when mutated confer resistance to QoI fungicides (Fernández-Ortuño et al. 2008). Specifically, the intron cob-429 that is inserted at codon 143 and its presence blocks the formation of the p.G143A substitution (Vallières et al. 2011), was present in *G. cichoracearum* but not in the other three PM species. One more intron, cob-393, was found six base pairs downstream of codon 129 and was present only in *G. cichoracearum,* whereas a third intron, cob-379, was inserted five base pairs upstream of codon 129 and was present in *E. necator*, *E. pisi,* and *G. cichoracearum* (Fig 7.5C). Although these introns do not directly affect codon 129, it is plausible that their mis-splicing may indirectly affect this codon and thus QoI resistance as well.

In order to further identify structural differences among the *cob* genes of the PM species examined in this study and of other phylogenetically closely related fungal species, we obtained this gene from 19 other members of Leotiomycetes outside the Erysiphales (Table 7.S8). The size of the *cob* gene varied greatly among these 19 species, ranging from 1.1 kb in *Rhynchosporium orthosporum* to 17.1 kb in *Sclerotinia borealis*. Thirteen species had no introns in their *cob* genes, whereas the remaining six species had three to six introns (Fig 7.5A and Fig 7.5B). Interestingly, intron insertion positions in *cob* were somewhat conserved within Leotiomycetes, as the 61 introns identified were inserted into 18 distinct sites between positions 99 and 824 of the *cob* cds (Fig 7.5C). Of these, insertion sites 99, 159, 311, 358, 379, 619, and 687 were unique to the Erysiphales, whereas from the remaining 11 insertion sites, eight were shared between at least one species of Erysiphales and one other species of Leotiomycetes. Among the shared intron insertion sites was cob-429, which was present in *G. cichoracearum*, *B. cinerea, Monilinia laxa*, and *M. fructicola*. Notably, group II introns with RT domains were present in five different sites within *cob* and were exclusive to the Erysiphales. Moreover, the intron cob-393 containing a GIY-YIG HE domain that was present in *G. cichoracearum* but absent in *B. graminis* f. sp. *tritici*, *E. necator*, and *E. pisi*, was conserved in all other analyzed members of the Leotiomycetes and in *Neurospora crassa* (Fig 7.5C). Taken, together, these results indicate that the *cob* genes of PM species are rather unique with respect to their intron content among Leotiomycetes.

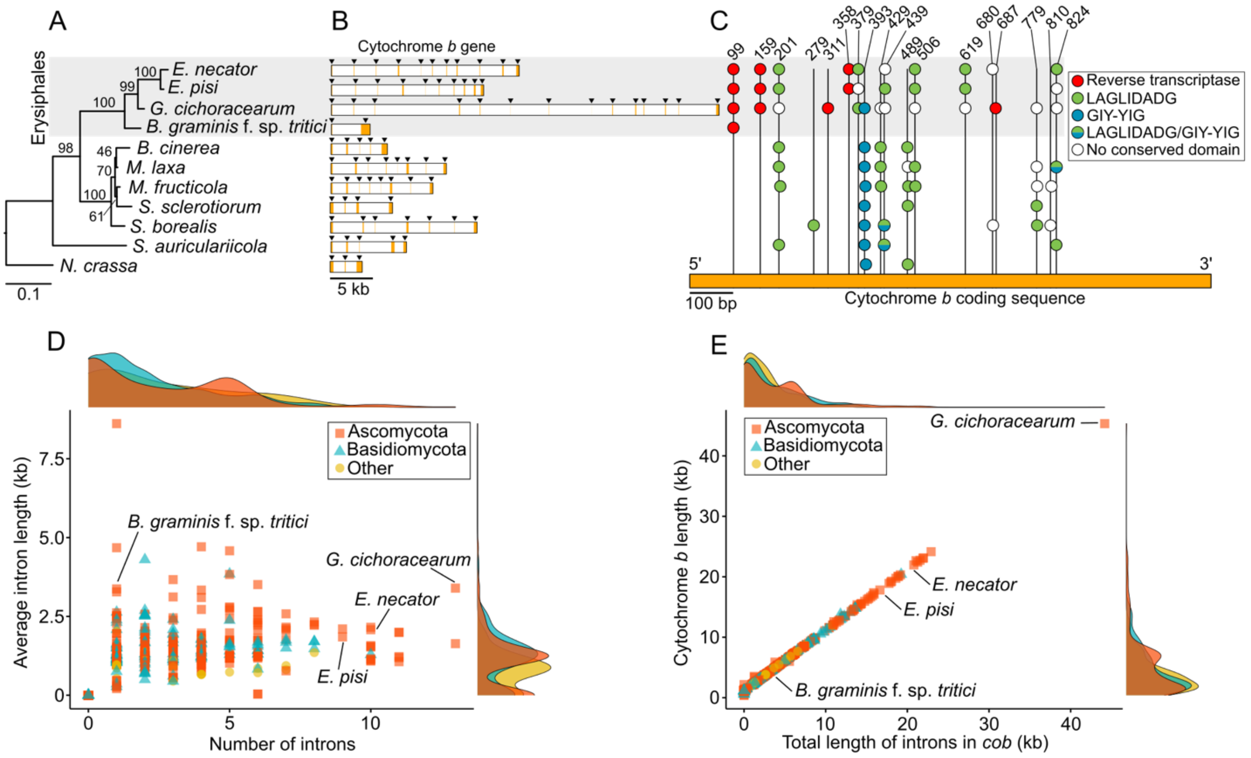


Figure 7.20: Comparative analysis of cytochrome b (cob) genes from the powdery mildew fungi *Blumeria graminis* f. sp. *tritici*, *Erysiphe necator*, *Erysiphe pisi*, and *Golovinomyces cichoracearum*, and other fungal species. (A) Maximum likelihood phylogenetic tree constructed based on *cob* coding sequences from ten species of Leotiomycetes and the outgroup species *Neurospora crassa*. (B) *Cob* gene structure from the eleven species used to construct the phylogenetic tree. Exons and introns are represented by orange and white bars, respectively. For clearer visualization, exons are also indicated with black triangles. The image shows that *cob* genes differ dramatically in size and intron content among

species of Leotiomycetes. (C) Insertion sites of the introns present in the *cob* genes of the eleven species used in our analysis and the conserved domains encoded within these introns. Insertion sites are represented by vertical lines and presence of introns is indicated by circles color-coded based on their conserved domain content. This figure shows that species of Erysiphales are the only ones among the Leotiomycetes that contain introns with reverse transcriptases (RTs) embedded in them, and that intron insertion sites are poorly conserved between them and the rest of the Leotiomycetes. (D, E) Scatterplots that show the average intron length in *cob*, number of introns in *cob*, total length of *cob*, and total length of introns in *cob*, among 703 *cob* genes of fungal species of Ascomycota, Basidiomycota, and 'other' fungal phyla (i.e., Blastocladiomycota, Chytridiomycota, Cryptomycota, Mucoromycota, and Zoopagomycota) included in the analysis. The scatterplots shows that with the exception of *B. graminis* f. sp. *tritici,* Erysiphales have unusually large and intron-rich *cob* genes. Data for the mt genome of *E. necator* were obtained from (Zaccaron et al. 2021).

7.2.5 Comparative analysis of fungal cytochrome *b* genes further highlights their idiosyncratic nature in *E. necator*, *E. pisi*, and *G. cichoracearum*

To probe further into the idiosyncrasy of the *cob* genes from the three Erysiphales, we then extracted this gene from fungal mt genomes available at NCBI. A total of 703 *cob* genes were retrieved, representing 254 fungal genera from seven phyla (Table 7.S8). Overall, intron abundance among *cob* genes was low, with 526 (74.8%) of the genes having four or less introns, and 218 (31.0%) having no introns (Fig 7.5D). Average intron density (i.e., the number of introns per kb of cds) among the 485 *cob* genes with at least one intron was 3.1, which is considerably lower than the average intron density of *cob* genes from the three PM species (9.1) or from all four together (7.1). Overall, *G. cichoracearum* and the yeast *Metschnikowia amazonensis* possessed the highest number of introns (*n*=13) in *cob* among all 703 fungal mt genomes analyzed (Fig 7.5D). Seventeen other fungal mt genomes had a total of nine or more introns in their *cob* gene. Among them, there were 13 yeast species of the *Metschnikowia*/*Candida* genera (9-11 introns), *E. necator* (10 introns), *Agaricus bisporus* (10 introns), *Juglanconis juglandina* (10 introns), and *E. pisi* (9 introns). Next to having the largest number of introns in *cob*, *G. cichoracearum* also had the largest *cob* gene (45.3 kb) among the 703 fungal *cob* genes analyzed, followed by *Morchella importuna* (24.1 kb) (Fig 7.5E). These results indicate that, compared with other fungal species, *G. cichoracearum*, *E. necator*, and *E. pisi* have unusually large *cob* genes that are rich in introns.

The same dataset of 703 fungal mt genomes was next utilized to search for species that, similar to the four PMs analyzed in this study, have introns in their *cob* genes with RT-encoding ORFs (NCBI accession cd01651) embedded in them. Surprisingly, next to the four PMs, only 33 more fungal species had a *cob* gene with at least one intron containing an RT domain (Fig 7.6A). The fungal species were from taxonomically diverse classes and typically each taxonomic class will be represented by only one to four species. However, two notable exceptions for which the presence of an RT domain within *cob* introns is common are the PMs and the yeast *Metschnikowia* spp., as all four PM species and ten of the 17 of *Metschnikowia* spp. included in the analysis had such domains. Moreover, while PMs harbored up to four *cob* introns with an RT-encoding ORF (*G. cichoracearum*) and *Metschnikowia* spp. up to six introns, other fungal species typically harbored a single *cob* intron with an RT domain. Collectively, this data indicate that the two lineages differ from other fungi by containing RT-encoding ORFs in multiple introns within *cob*. Finally, a total of 73 RT-encoding ORFs were identified within the introns of the 37 *cob* genes (Table 7.S9). These introns were inserted into nine different sites within the *cob* cds, with insertion sites cob-99 and cob-687 being the most common ones, as they harbored 36 (49.3%) of the 73 RT-encoding ORFs identified (Fig 7.6B). Interestingly, a phylogenetic tree constructed using the translated RT-encoding ORFs revealed grouping based on the insertion site of the introns that contained them in the *cob* gene rather than their respective species phylogeny, suggesting that some of these domains could have been acquired through horizontal transfer (Fig 7.6C).

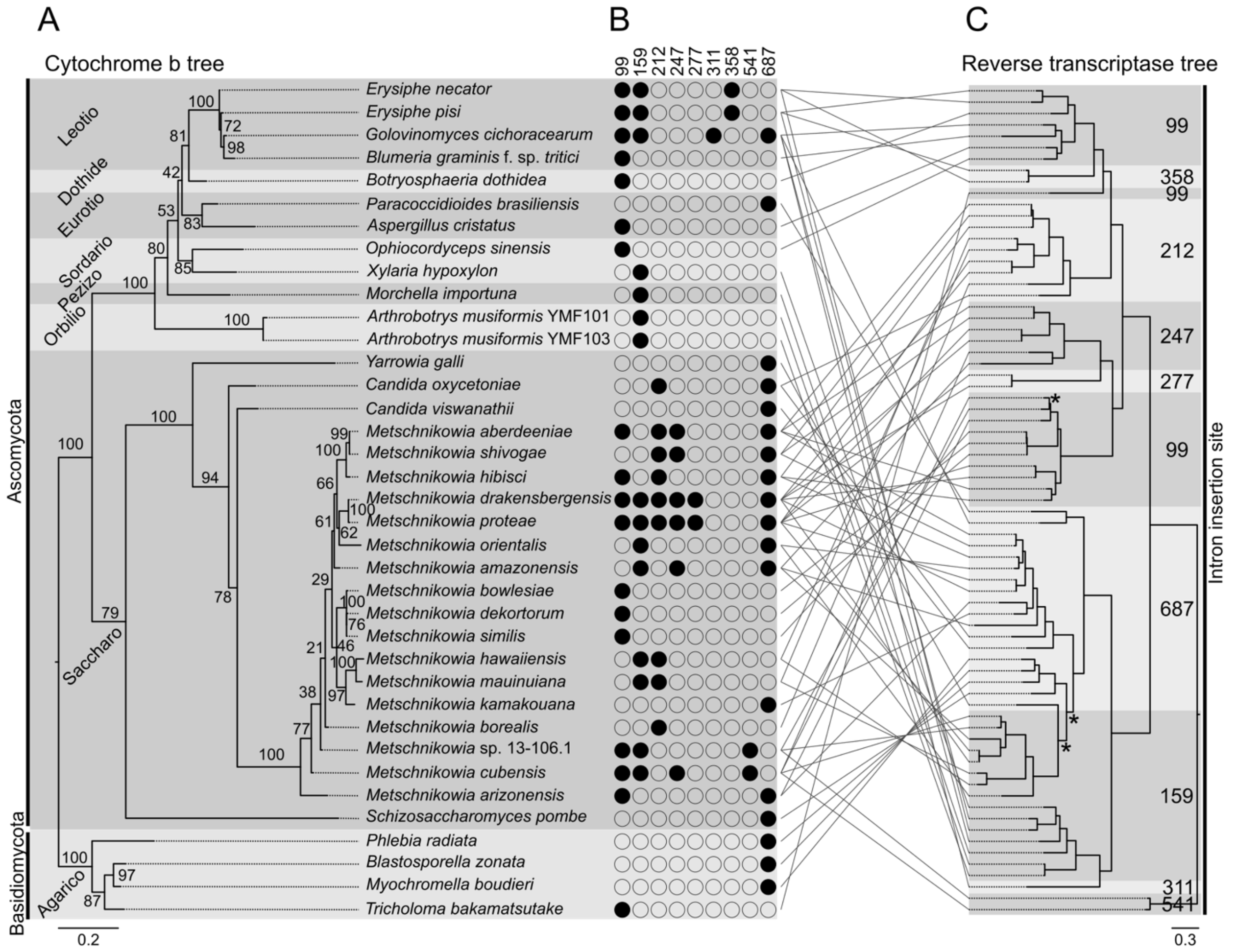

A
Cytochrome b tree
B
99 159 212 247 277 311 358 541 687
C
Reverse transcriptase tree
Leotio
Dothide
Eurotio
Sordario
Pezizo
Orbilio
Saccharo
Agarico
Ascomycota
Basidiomycota
Erysiphe necator ●●○○○○●○○
Erysiphe pisi ●●○○○○●○○
Golovinomyces cichoracearum ●●○○○●○○●
Blumeria graminis f. sp. tritici ●○○○○○○○○
Botryosphaeria dothidea ●○○○○○○○○
Paracoccidioides brasiliensis ○○○○○○○○●
Aspergillus cristatus ●○○○○○○○○
Ophiocordyceps sinensis ●○○○○○○○○
Xylaria hypoxylon ○●○○○○○○○
Morchella importuna ○●○○○○○○○
Arthrobotrys musiformis YMF101 ○●○○○○○○○
Arthrobotrys musiformis YMF103 ○●○○○○○○○
Yarrowia galli ○○○○○○○○●
Candida oxycetoniae ○○●○○○○○●
Candida viswanathii ○○○○○○○○●
Metschnikowia aberdeeniae ●○●●○○○○●
Metschnikowia shivogae ○○●●○○○○●
Metschnikowia hibisci ●○●○○○○○●
Metschnikowia drakensbergensis ●●●●●○○○●
Metschnikowia proteae ●●●●●○○○●
Metschnikowia orientalis ○●○○○○○○●
Metschnikowia amazonensis ○●○●○○○○●
Metschnikowia bowlesiae ●○○○○○○○○
Metschnikowia dekortorum ●○○○○○○○○
Metschnikowia similis ●○○○○○○○○
Metschnikowia hawaiiensis ○●●○○○○○○
Metschnikowia mauinuiana ○●●○○○○○○
Metschnikowia kamakouana ○○○○○○○○●
Metschnikowia borealis ○○●○○○○○○
Metschnikowia sp. 13-106.1 ●●○○○○○●○
Metschnikowia cubensis ●●○●○○○●○
Metschnikowia arizonensis ●○○○○○○○●
Schizosaccharomyces pombe ○○○○○○○○●
Phlebia radiata ○○○○○○○○●
Blastosporella zonata ○○○○○○○○●
Myochromella boudieri ○○○○○○○○●
Tricholoma bakamatsutake ●○○○○○○○○
100
72
98
81
42
53
83
80
85
100
100
100
100
94
99
100
66
100
61
62
29
78
100
76
46
21
100
38
97
77
100
79
100
97
87
0.2
99
358
99
212
247
277
99
687
159
311
541
Intron insertion site
*
*
*
0.3

Figure 7.21: The cytochrome b (*cob*) genes from the powdery mildew fungi *Blumeria graminis* f. sp. *tritici*, *Erysiphe necator*, *Erysiphe pisi*, and *Golovinomyces cichoracearum* contain rare intronic ORFs encoding reverse transcriptases (RTs). (A) Maximum likelihood phylogenetic tree obtained based on the translated amino acid sequences of all fungal *cob* genes containing RT-encoding ORFs among 703 fungal mitochondrial genomes analyzed. The taxonomic classes of the fungal species are shown on the left. (B) Insertion sites in their respective *cob* cds of introns harboring RT-encoding ORFs. Presence and absence of RT-encoding ORFs are indicated with filled and blank circles, respectively. (C) Maximum likelihood phylogenetic tree obtained based on the translated amino acid sequences of RT-encoding ORFs present within *cob* introns. Insertion sites of the introns harboring the RT-encoding ORFs are shown on the right. For clearer visualization, support bootstrap values were omitted. Branches with low bootstrap support (<75%) are labeled with an asterisk (*). Pairwise comparison of the two trees shows multiple cases of phylogenetic incongruences that could be indicative of horizontal gene transfer events. All intron insertion sites are reported using *E. necator* as reference. Data for *E. necator* were obtained from (Zaccaron et al. 2021).

7.3 Discussion

In this study we assembled the mt genomes of three PM species and further compared them with the mt genome of a fourth species, in order to reveal contrasting differences among mt genomes of PM species, and to show idiosyncratic features of their mt genomes as compared to those of other fungal species. Our comparative analysis revealed a markedly variability in mt genome sizes among PM species, which ranged from 109,800 bp to 332,165 bp. The 109.8 kb mt genome of *B. graminis* f. sp. *tritici* was the smallest among the four PM analyzed, and is comparable in size with the mt genomes of *B. graminis* f. sp. *hordei* isolates DH14 (106.3 kb) and RACE1 (139.2 kb) described previously (Frantzeskakis et al. 2018). The seemingly larger genome size of *B. graminis* f. sp. *hordei* RACE1 is caused by a 32.2 kb duplication at the immediate ends of the contig representing the mt genome of this isolate (Frantzeskakis et al. 2018). However, this duplication is likely the result of untrimmed overlapping ends, which often occur during assembly of circular genomes, and thus by removing this duplication then the size of the mt genome of the *B. graminis* f. sp. *hordei* RACE1 isolate is reduced to 107.1 kb, which is comparable to other mt genomes of *B. graminis* (Frantzeskakis et al. 2018). Recently, a 26.0 kb mt genome for the PM *Podosphaera xanthii* has been described (Kim et al. 2019), which is considerably smaller than the PM mt genomes presented herein. However, a phylogenetic analysis failed to group the *Podosphaera xanthii* mt genome with members of the Leotiomycetes (Kim et al. 2019), suggesting that this mt genome is not from a PM species. The largest fungal mt genomes reported to date are those of the so-called 'true morels' fungi *Morchella importuna, M. conica,* and *M. crassipes,* sized at 272,238 bp (Liu et al. 2020a), 280,763 bp (Li et al. 2020), and 531,195 bp (Liu et al. 2020b), respectively. Thus, at 332,165 bp, the mt genome of *G. cichoracearum* currently stand as the second largest among fungi and the largest among pathogenic fungi. With the advent of new cost-effective sequencing technologies, this pattern will likely continue and even larger fungal mt genomes will possibly be reported, thus highlighting the variable and dynamic nature of fungal mt genomes.

Our analyses indicated that the variation in size among the mt genomes of the four Erysiphales examined in this study is mainly due to differences in the number and length of their intergenic and intronic regions. In contrast, their core protein-coding mt genes were mostly conserved and showed no evidence of gene duplication. Also, no other genomic rearrangements were observed, although these can even occur among fungal species of the same genus (Li et al. 2018). A recent study, which analyzed gene rearrangements in the mt genomes of 16 species of Leotiomycetes, identified five distinct groups, indicating that Leotiomycetes have undergone major gene rearrangements during their evolutionary history (Chen et al. 2019). Interestingly, the gene arrangement present in the four Erysiphales analyzed in this study is not in agreement with any of those described in the other 16 Leotimycete species (Chen et al. 2019), and thus constitutes a novel type of mt gene arrangement within Leotiomycetes. A noteworthy discrepancy of the Erysiphales is that the *nad2* and *nad3* genes are not placed next to each other, as is the case in several fungal species (Aguileta et al. 2014; Zaccaron et al. 2021), including in the 16 species of Leotiomycetes analyzed before (Chen et al. 2019). This suggests that this unusual rearrangement occurred after divergence of the Erysiphales, and the gene order was maintained during subsequent speciation events.

Next to differences in mt intron size and numbers, intron content was also very different among the four PM species. Homing endonuclease genes (HEGs) are selfish genetic elements that spread at a super-Mendelian rate within a population (Burt and Koufopanou 2004). They are thought to have no effect on the fitness of the host organism, and therefore are not subject to natural selection. Once fixed in a population, these elements accumulate mutations that eventually disrupt their ability to spread. Their preservation during evolution is ensured by a cyclical model of acquisition, degeneration and loss, in which HEGs constantly move to new species by horizontal transfer before they degenerate, whereas the degenerated HEGs in the donor species are eventually lost from its genome (Burt and Koufopanou 2004; Goddard and Burt 1999). This loss is mainly through a mechanism of precise excision, mediated by reverse transcription of spliced RNA. As a result, it leads to the re-constitution of the sequence recognized by the HE, thus making the sequence susceptible again to re-invasion, and starting the process of acquisition, degeneration, and loss of HEGs. This cyclical model of HEG propagation could explain the marked intron content variability among Erysiphales, which suggests extensive intron gain and/or loss throughout their evolutionary history. Another explanation for the observed intron variability is the likelihood of a species to acquire horizontally transferred sequences. It is believed in this respect that HEGs are more commonly found in organelles of eukaryotes of simpler organization, such as fungi, algae and protists, because of the ease of horizontal transmission among these taxa, for which access to the germline is more effortless (Burt et al. 2006). However, this scenario would imply that the mt genome of *B. graminis* f. sp. *tritici* would be less susceptible to horizontal transmission compared to the mt genomes of *E. necator*, *E. pisi*, and *G. cichoracearum*. Although a definite answer cannot be given, both horizontal transmission of HEGs and their degeneration could be evoked to explain the different conservation levels observed among introns.

Another interesting feature observed in the mt genomes of *B. graminis* f. sp. *tritici* and *G. cichoracearum* was an unusual bimodal GC distribution that was absent in *E. necator*, *E. pisi,* and other Leotiomycetes. The high GC content regions had no apparent impact on coding sequences, because they were almost exclusively located within intronic or intergenic regions. Lack of homology between these high-GC regions of *B. graminis* f. sp. *tritici* and *G. cichoracearum* does not support inheritance from a common ancestor, but may alternatively suggest acquisition by horizontal transfer. Also, homology searches performed with BLASTn against the NCBI nr database indicated that these regions were not well conserved in other species, except in the oomycete *Albugo laibachii*, which contained sequences highly homologous to the high-GC regions from *G. cichoracearum*. As with PM pathogens, *A. laibachii* is also an obligate biotrophic pathogen that causes downy mildew on *Arabidopsis thaliana*. Previous studies have shown that both *G. cichoracearum* and *G. orontii* can also infect *A. thaliana* (Adam and Somerville 1996; Plotnikova et al. 1998), and thus sharing the same host could facilitate horizontal transmission between *G. cichoracearum* and *A. laibachii*. However, this hypothetical horizontal transmission might not be the case because in the *A.*

laibachii genome assembly we identified a scaffold that matches the ITS sequence from *Golovinomyces* spp. This raises the possibility of contamination by *Golovinomyces* spp. of the sequenced sample of *A. laibachii*. Nevertheless, the possible origin of these high-GC regions in *G. cichoracearum* and *B. graminis* f. sp. *tritici* remains unclear and future studies may shed light into their formation and impact on the mt genome.

A recent study identified 21 intron insertion sites within *cob* among 129 fungal species of the Ascomycota (Guha et al. 2018). In the present study, among only nine representatives of the Leotiomycetes, we identified 18 insertion sites, seven of which were unique to Erysiphales. Interestingly, the most commonly found insertion sites in fungal *cob* genes were at position 393 and 490 (Guha et al. 2018). However, while these insertion sites were common among other Leotiomycetes, they were rare among Erysiphales. Instead, the most common intron within Erysiphales was at position 99, which was absent in other Leotiomycetes and was not reported in the study presented in (Guha et al. 2018). Concomitant to rare intron insertion sites, RT-encoding ORFs were also common in Erysiphales. Although, the presence of RT-encoding ORFs in fungal mt introns has been frequently reported, we show that within fungal *cob* introns these ORFs are rare, except among the Erysiphales and *Metschnikowia* spp. In addition, a phylogenetic analysis revealed that the RT-encoding ORFs group based on their respective insertion site, which differs substantially from their respective host species phylogeny. This contrasting phylogenetic placement is indicative of horizontal transfer and is in accordance with the hypothesized cyclical model of HEG acquisition, degeneration, and loss.

The presence of an intron adjacent to codon 143 of the *cob* gene (i.e., intron *cob*-429) has been shown to prevent the p.G143A mutation in fungal pathogens. Mutation at this codon can interfere with the splicing of the intron, which consequently leads to a presumably deficient *cob*. This mechanism has been previously reported in the Leotiomycetes *B. cinerea* (Banno et al. 2009) and *M. fructicola* (Luo et al. 2010). Particularly, two different *cob* alleles were identified among isolates of *B. cinerea* based on the presence or absence of the cob-429 intron. This same study revealed that individuals carrying the p.G143A mutation did not have the cob-429 intron, supporting the hypothesis that this intron prevents resistance to QoI fungicides acquired by the p.G143A mutation. In accordance with previous studies, the presence of the cob-429 intron was also detected in *G. cichoracearum*, suggesting that isolates of the fungus bearing this intron are unable to acquire resistance to QoIs trough the p.G143A mutation. Further studies can reveal the frequency of this intron among populations of *G. cichoracearum*.

In summary, in this study we produced high-quality mt genomes for three economically important PM pathogens. By doing so, we provided novel insights into the mt genome organization of members of the Erysiphales and expanded the spectrum of mt genomic resources for these pathogens. Notably, the mt genomes of PM pathogens are highly syntenic but vary greatly in size. *Blumeria graminis* f. sp. *tritici* and *G. cichoracearum* also differed substantially from other fungal species by having an unusual bimodal GC content with low GC regions interspersed with high GC regions. In addition, mt genomes of the four PM pathogens differed from other fungi by having atypical RT-encoding ORFs within the *cob* gene. The analyzed mt genomes of Erysiphales also presented a dynamic architecture of presence/absence of introns and intronic ORFs encoding HEs and RTs. In this context, future studies can elucidate how active HEs and RTs are and their evolutionary impact in the mt genomes of the analyzed Erysiphales.

7.4 Materials and methods

7.4.1 Assembly of mt genomes

To assemble the mitochondrial (mt) genome of *B. graminis* f. sp. *tritici* isolate 96224, reads were obtained from NCBI (SRR7642212) and trimmed with fastp v0.20 (Chen et al. 2018). Reads representing the mt genome of *B. graminis* f. sp. *tritici* were extracted based on exact *k*-mer matching (*k*=29) performed with the *bbduk.sh* script of the BBMap v38-60 software package (Bushnell 2014). A scaffold (UNSH01000099.1) containing the mt genome of *B. graminis* f. sp. *hordei* isolate RACE1 (Frantzeskakis et al. 2018) was utilized as bait to extract the reads. Extracted reads were processed with the *bbnorm.sh* script of BBMap to normalize the coverage to 100x. The resulting reads (total of 179,166 paired-end reads) were then assembled with SPAdes v3.14 (Bankevich et al. 2012) with the flag *isolate* enabled and *k*-mer values of 21, 33, 55, and 77. With these parameters, SPAdes assembled a contig of 109,874 bp, which was further polished with Pilon v1.23 (Walker et al. 2014) . To assemble the mt genome of *E. pisi* isolate Palampur-1, PacBio reads were obtained from NCBI (SRR11059780). PacBio reads were processed with the *removesmartbell.sh* script of the BBMap v38-60 software package to remove remaining SMRTbell adapters and then mapped to the mt genome of *E. necator* (Zaccaron et al. 2021) with minimap2 v2.16 (Li 2018) with parameters to map PacBio reads (option *map-pb*) and output the alignment in SAM format (option *-a*). The output of minimap2 was converted to BAM format and reads that successfully mapped were extracted with SAMtools v1.9. From the extracted reads, the 27,238 longest reads estimated to cover the mt genome 100 times were selected and then assembled with Canu v1.9 (Koren et al. 2017) using parameters *genomeSize=200k* and *corOutCoverage=60*. Canu assembled a 215,191 bp contig, which was further polished with Flye v2.7 (Kolmogorov et al. 2019) based on three polishing rounds. To assemble the mt genome of *G. cichoracearum* isolate UCSC1, PacBio reads were downloaded from NCBI (SRR6829655) and processed with the *removesmartbell.sh* script of the BBMap v38-60 software package. First, coding sequences of mt genes from *E. necator* (Zaccaron et al. 2021) were queried with BLASTn (e-value < 1e-10) against the genome assembly of *G. magnicellulatus* isolate FPH2017-1 (Farinas et al. 2019). Three contigs (VCMJ01000001.1, VCMJ01000002.1, and VCMJ01000079.1) were identified to represent fragments of the mt genome of *G. magnicellulatus*. Subsequently, the PacBio reads of *G. cichoracearum* were mapped to these three contigs with minimap2 and successfully mapped reads were extracted with SAMtools v1.9. From the mapped reads, the longest 4,912 reads were estimated to cover the mt genome 100 times and were then assembled with Canu v1.9 with parameters *genomeSize=350k* and *corOutCoverage=60*. Canu assembled a 358,768 bp contig, which was further polished with Flye v2.7 based on three polishing rounds.

A final round of polishing was carried out with Pilon v1.23, by utilizing Illumina reads of *G. cichoracearum* isolate UCSC1 (SRR4017398) mapped to the assembled mt genome with BWA-MEM v0.7.17 (Li and Durbin 2009). Self BLASTn searches were performed to identify overlapping ends of the assembled contigs, suggesting circularity. For *B. graminis* f. sp. *tritici*, the first and last 77 bp of the assembled contig were identical. For *E. pisi* and *G. cichoracearum* overlapping ends of 26.7 kb and 26.6 kb were identified to overlap almost perfectly, with 99.6% and 99.8% identity, respectively. One of the overlapping ends were removed and the contigs were rotated so that the first position was the start coordinate of the *cox1* gene. The workflow of how the genomes were assembled is shown in Fig 7.S1.

7.4.2 Annotation of mt genomes

Assembled mt genomes were initially annotated with the MFannot webserver, using the genetic code 4 (Mold, Protozoan and Coelenterate Mt Code) (Valach et al. 2014). To validate the predicted annotations of protein-coding mt genes from *B. graminis* f. sp. *tritici*, *E. pisi,* and *G. cichoracearum*, public RNAseq datasets from these organisms were obtained from NCBI (accessions SRR6026494, SRR7066906 and SRR6232712, respectively). RNAseq reads were mapped to the respective mt genome with HISAT2 v2.1.0 (Kim et al. 2015) with default settings. Alignments were visualized with IGV v2.5.3 (Robinson et al. 2011) and gene annotations were inspected and adjusted manually.

Genes encoding mt-tRNAs and their secondary structures were obtained with MITOS2 (Bernt et al. 2013). Introns were classified into group I or group II with RNAweasel (Lang et al. 2007). Intronic ORFs were identified with ORFfinder v0.4.3 (Wheeler et al. 2007), using as a minimum ORF length 200 bp and genetic code 4. ORFs encoding homing endonucleases (HEs) or reverse transcriptases (RTs) were identified and classified based on their conserved domains identified by querying the encoded peptide sequences against the NCBI conserved domain database (CDD) (Marchler-Bauer et al. 2017) with an e-value < 1e-3. Conserved domains within introns were identified by translating the entire intronic sequences in 6 frames with the *transeq* script from EMBOSS v6.6.0 (Rice et al. 2000), utilizing the genetic code 4 and querying the peptide sequences against the NCBI CDD with an e-value < 1e-3. Circular representations of the mt genomes were created with Circos v0.69-8 (Krzywinski et al. 2009). Tandem repeats were identified with Tandem Repeat Finder v4.09 (Benson 1999) and overall percentage of repeats of the mt genomes was calculated based on self BLASTn searches, utilizing the parameter *-task blastn* and an e-value < 1e-10.

7.4.3 Identification of introns insertion sites and conservation of intronic sequences

To identify the insertion sites of introns in mt genes, mature transcripts of the genes were aligned with the *mafft-linsi* script from MAFFT v7.455 (Katoh et al. 2002) and the alignments were visualized with SnapGene v5.0.7 (GSL Biotech; available at snapgene.com). Intron insertion sites (i.e., last base-pair position of exons) were reported using *E. necator* C-strain as reference (Zaccaron et al. 2021). For better interpretation and visualization, intron insertion sites that differed by at most three base pairs were grouped into one single insertion site. Overall conservation of intronic sequences of the four PM pathogens was determined by pairwise nucleotide alignments produced by the *pairwiseAlignment* function from the R package Biostrings v2.57.2 (Pagès et al. 2022) within R v4.0.3 (R Core Team 2020). To minimize negative impact in the percent identity values due to differences in introns size, the alignment type option in *pairwiseAlignment* was set to *local-global*, which performs a local alignment of the pattern (sequence 1) with a global alignment of subject (sequence 2), where size of subject is at most the size of pattern. Pairwise alignments were processed with the *pid* function from Biostrings (Pagès et al. 2022) to calculate the percent sequence identity considering internal gap positions (option *type=PID1*).

7.4.4 Nucleotide composition and GC-rich regions of the mt genomes

Distribution of GC content, GC skew [(G-C)/(G+C)] and AT skew [(A-T)/(A+T)] of the mt genomes were determined by calculating the nucleotide composition of non-overlapping sliding windows of 200 bp with the command *comp* of the seqtk v1.3-r106 script (https://github.com/lh3/seqtk). Histograms of GC content were generated within the R software package v4.0.3 (R Core Team 2020) and the kernel density estimations were obtained with the *density* function within R, utilizing the Gaussian method. Consecutive 200 bp-windows with GC > 50% were merged to estimate length of isochore-like GC-rich regions. To investigate if the GC-rich regions (i.e., windows with GC > 50%) from the mt genomes of *G. cichoracearum* and *B. graminis* f. sp. *tritici* migrated from the nuclear genome, the GC-rich regions were queried with BLASTn (e-value < 1e-5) against the nuclear genome of *G. cichoracearum* (GCA_003611235.1) and *B. graminis* f. sp. *tritici* (GCA_900519115.1).

7.4.5 Phylogenetic trees

Phylogenetic trees were built based on multiple sequence alignments generated with the *mafft-linsi* script from MAFFT v7.455 (Katoh et al. 2002). Alignments were processed with trimAl v1.4 (Capella-Gutiérrez et al. 2009) to remove all sites containing gaps. Maximum likelihood phylogenetic trees were inferred with IQTREE v1.6.11 (Nguyen et al. 2015) utilizing ModelFinder**Error! Bookmark not defined.** (Kalyaanamoorthy et al. 2017) to automatically select the best substitution model. Support for branches was obtained based on 1,000 ultrafast bootstrap replicates (Hoang et al. 2018). The phylogenetic tree of the four PM pathogens was constructed based on the concatenated amino acid sequences of 14 protein-coding mt genes, i.e., *atp6*, *atp8*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *cox1*, *cox2*, *cox3*, *cob*, and *rps3*, utilizing the substitution model cpREV+F+R2. The tree was rooted on *Sclerotinia borealis*. The phylogenetic tree of the *cob* gene from Leotiomycetes was constructed based on coding sequences and utilizing the substitution model TIM+F+G4. The tree was rooted on *Neurospora crassa*. The phylogenetic trees of *cob* genes and ORFs encoding RTs within *cob* introns were constructed based on amino acid sequences and utilizing the substitution models mtZOA+G4 and WAG+F+I+G4, respectively. Trees were visualized and edited with FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

7.4.6 Public data acquisition

For comparative analyses, fungal mt genomes were obtained from NCBI Nucleotide and Organelle Genome Resources databases (Wolfsberg et al. 2001) as of May 6th, 2021. Nucleotide and protein sequences were obtained from NCBI database utilizing the *efetch* command from NCBI Entrez Direct E-utilities v11.0 (Kans 2020). Reads were obtained from NCBI Sequence Read Archive by generating downloadable links with SRA-Explorer online tool (http://sra-explorer.info) based on the accession numbers. Fungal *cob* genes were retrieved from the annotated mt genomes and utilized to mine for additional fungal *cob* genes by querying them against the NCBI nr database using BLASTp with e-value < 1e-5 and results restricted to Fungi (taxid 4751).

7.5 Data availability

The assembled mitochondrial genomes of *B. graminis* f. sp. *tritici*, *E. pisi*, and *G. cichoracearum* have been submitted to GenBank under the accession numbers MT880591, MT880589, and MT880590, respectively. Accession numbers of all other mitochondrial genomes and cytochrome *b* genes utilized in the comparative analyses are listed in Table 7.S1 and Table 7.S8, respectively, available at https://zenodo.org/records/11211529. Scripts utilized in this study were deposited in a public GitHub repository available at https://github.com/alexzaccaron/2021_pms_mt. The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

Author Contributions

Conceptualization: AZZ, IS; **Data Curation:** AZZ; **Formal Analysis:** AZZ; **Funding Acquisition:** IS; **Investigation:** AZZ; **Methodology:** AZZ, IS; **Project Administration:** IS; **Resources:** AZZ; **Software:** AZZ; **Supervision:** IS; **Validation:** AZZ; **Visualization:** AZZ, IS; **Writing – Original Draft Preparation:** AZZ, IS; **Writing – Review & Editing:** AZZ, IS.

Funding information

Alex Z. Zaccaron received funding from University of California, Davis, through the Dean's Distinguished Graduate Fellowship (DDGF) program. This study was partially funded by the United States Department of Agriculture – National Institute for Food and Agriculture – Specialty Crop Research Initiative FRAME project (number 2018-03375).

7.6 References

Adam, L., and Somerville, S. C. 1996. Genetic characterization of five powdery mildew disease resistance loci in *Arabidopsis thaliana*. Plant J. 9:341–356

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., and Gabaldón, T. 2014. High variability of mitochondrial gene order among fungi. Genome Biol. Evol. 6:451–465

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477

Banno, S., Yamashita, K., Fukumori, F., Okada, K., Uekusa, H., Takagaki, M., Kimura, M., and Fujimura, M. 2009. Characterization of QoI resistance in *Botrytis cinerea* and identification of two types of mitochondrial cytochrome b gene. Plant Pathol. 58:120–129

Bartlett, D. W., Clough, J. M., Godwin, J. R., Hall, A. A., Hamer, M., and Parr-Dobrzanski, B. 2002. The strobilurin fungicides. Pest Manag. Sci. 58:649–662

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P. F. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69:313–319

Braun, U. 1987. A monograph of the Erysiphales (powdery mildews). Beih. Zur Nova Hedwig. 89:1–700

Braun, U., and Cook, R. T., eds. 2012. *Taxonomic manual of Erysiphales (powdery mildews)*. CBS Biodiversity Series, Utrecht, the Netherlands.

Bullerwell, C. E., and Lang, B. F. 2005. Fungal evolution: the case of the vanishing mitochondrion. Curr Opin Microbiol. 8:362–369

Burt, A., and Koufopanou, V. 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr. Opin. Genet. Dev. 14:609–615

Burt, A., Trivers, R., and Burt, A. 2006. *Genes in conflict: the biology of selfish genetic elements*. Harvard University Press.

Bushnell, B. 2014. *BBMap: a fast, accurate, splice-aware aligner*. Lawrence Berkeley National Lab (LBNL), Berkeley, CA.

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

Chan, D. C. 2006. Mitochondria: dynamic organelles in disease, aging, and development. Cell. 125:1241–1252

Chen, C., Li, Q., Fu, R., Wang, J., Xiong, C., Fan, Z., Hu, R., Zhang, H., and Lu, D. 2019. Characterization of the mitochondrial genome of the pathogenic fungus *Scytalidium auriculariicola* (Leotiomycetes) and insights into its phylogenetics. Sci. Rep. 9:17447

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890

Deng, Y., Hsiang, T., Li, S., Lin, L., Wang, Q., Chen, Q., Xie, B., and Ming, R. 2018. Comparison of the mitochondrial genome sequences of six *Annulohypoxylon stygium* isolates suggests short fragment insertions as a potential factor leading to larger genomic size. Front. Microbiol. 9:2079

Farinas, C., Gluck-Thaler, E., Slot, J. C., and Peduto Hand, F. 2019. Whole-genome sequence of the phlox powdery mildew pathogen *Golovinomyces magnicellulatus* strain FPH2017-1. Microbiol. Resour. Announc. 8:e00852-19

Fernández-Ortuño, D., Torés, J. A., de Vicente, A., and Pérez-García, A. 2008. Mechanisms of resistance to QoI fungicides in phytopathogenic fungi. Int. Microbiol. 11:1–9

Fondevilla, S., and Rubiales, D. 2012. Powdery mildew control in pea. A review. Agron. Sustain. Dev. 32:401–409

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., and 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

Gadoury, D. M., Cadle-Davidson, L., Wilcox, W. F., Dry, I. B., Seem, R. C., and Milgroom, M. G. 2012. Grapevine powdery mildew (*Erysiphe necator*): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. Mol. Plant Pathol. 13:1–16

Goddard, M. R., and Burt, A. 1999. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci U A. 96:13880–13885

Guha, T. K., Wai, A., Mullineux, S.-T., and Hausner, G. 2018. The intron landscape of the mtDNA *cytb* gene among the Ascomycota: introns and intron-encoded open reading frames. Mitochondrial DNA Part A. 29:1015–1024

Hausner, G. 2003. Fungal Mitochondrial Genomes, Plasmids and Introns. Pages 101–131 in: Fungal Genomics, Applied Mycology and Biotechnology. D.K. Arora and G.G. Khachatourians, eds. Elsevier Science 2003;, New York.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522

James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J. E. 2013. Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. Curr. Biol. 23:1548–1553

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14:587–589

Kans, J. 2020. Entrez direct: E-utilities on the UNIX command line. in: Entrez Programming Utilities Help, National Center for Biotechnology Information, Bethesda, MD.

Katoh, K., Misawa, K., Kuma, K.-I., and 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., Langmead, B., and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 12:357–360

Kim, S., Jung, M., Oh, E. A., Ho Kim, T., and Kim, J.-G. 2019. Mitochondrial genome of the *Podosphaera xanthii*: a plant pathogen causes powdery mildew in cucurbits. Mitochondrial DNA Part B. 4:4172–4173

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37:540–546

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645

Lambowitz, A. M., and Zimmerly, S. 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb. Perspect. Biol. 3:a003616

Lang, B. F., Gray, M. W., and Burger, G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. Annu. Rev. Genet. 33:351–397

Lang, B. F., Laforest, M.-J., and Burger, G. 2007. Mitochondrial introns: a critical view. Trends Genet. 23:119–125

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

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25:1754–1760

Li, Q., Chen, C., Xiong, C., Jin, X., Chen, Z., and Huang, W. 2018. Comparative mitogenomics reveals large-scale gene rearrangements in the mitochondrial genome of two *Pleurotus* species. Appl. Microbiol. Biotechnol. 102:6143–6153

Li, W., Zhang, F., and Gao, L. 2020. SMRT-based mitochondrial genome of the edible mushroom *Morchella conica*. Mitochondrial DNA Part B. 5:3219–3220

Liu, W., Cai, Y., Zhang, Q., Chen, L., Shu, F., Ma, X., and Bian, Y. 2020a. The mitochondrial genome of *Morchella importuna* (272.2 kb) is the largest among fungi and contains numerous introns, mitochondrial non-conserved open reading frames and repetitive sequences. Int. J. Biol. Macromol. 143:373–381

Liu, W., Cai, Y., Zhang, Q., Shu, F., Chen, L., Ma, X., and Bian, Y. 2020b. Subchromosome-scale nuclear and complete mitochondrial genome characteristics of *Morchella crassipes*. Int. J. Mol. Sci. 21:483

Luo, C.-X., Hu, M.-J., Jin, X., Yin, L.-F., Bryson, P. K., and Schnabel, G. 2010. An intron in the cytochrome b gene of *Monilinia fructicola*. mitigates the risk of resistance development to QoI fungicides. Pest Manag Sci. 66:1308–1315

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., and Bryant, S. H. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45:D200–D203

Mardanov, A. V., Beletsky, A. V., Kadnikov, V. V., Ignatov, A. N., and Ravin, N. V. 2014. The 203 kbp mitochondrial genome of the phytopathogenic fungus *Sclerotinia borealis* reveals multiple invasions of introns and genomic duplications. PLoS One. 9:e107536

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. 2022. Biostrings: Efficient manipulation of biological strings.

Plotnikova, J. M., Reuber, T. L., Ausubel, F. M., and Pfister, D. H. 1998. Powdery mildew pathogenesis of *Arabidopsis thaliana*. Mycologia. 90:1009–1016

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020, Vienna, Austria.

Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16:276–277

Richardson, D. R., Lane, D. J. R., Becker, E. M., Huang, M. L.-H., Whitnall, M., Suryo Rahmanto, Y., Sheftel, A. D., and Ponka, P. 2010. Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. Proc. Natl. Acad. Sci. 107:10775–10782

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. 2011. Integrative genomics viewer. Nat. Biotechnol. 29:24–26

Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A. M. 1993. Group I and group II introns. FASEB J. 7:15–24

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. 2019. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. 3:430–439

Valach, M., Burger, G., Gray, M. W., and Lang, B. F. 2014. Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. Nucleic Acids Res. 42:13764–13777

Vallières, C., Trouillard, M., Dujardin, G., and Meunier, B. 2011. Deleterious effect of the Qo inhibitor compound resistance-conferring mutation G143A in the intron-containing cytochrome *b* gene and mechanisms for bypassing it. Appl. Environ. Microbiol. 77:2088–2093

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963

Warkentin, T., Rashid, K., and Xue, A. 1996. Fungicidal control of powdery mildew in field pea. Can. J. Plant Sci. 76:933–935

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., and others. 2007. Database resources of the national center for biotechnology information. Nucleic Acids Res. 36:D13–D21

Wolfsberg, T. G., Schafer, S., Tatusov, R. L., and Tatusova, T. A. 2001. Organelle genome resources at NCBI. Trends Biochem. Sci. 26:199–203

Zaccaron, A. Z., De Souza, J. T., and Stergiopoulos, I. 2021. The mitochondrial genome of the grape powdery mildew pathogen *Erysiphe necator* is intron rich and exhibits a distinct gene organization. Sci. Rep. 11:13924

Zubaer, A., Wai, A., and Hausner, G. 2018. The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. Sci. Rep. 8:17591

7.7 Supplementary materials

7.7.1 Supplementary figures

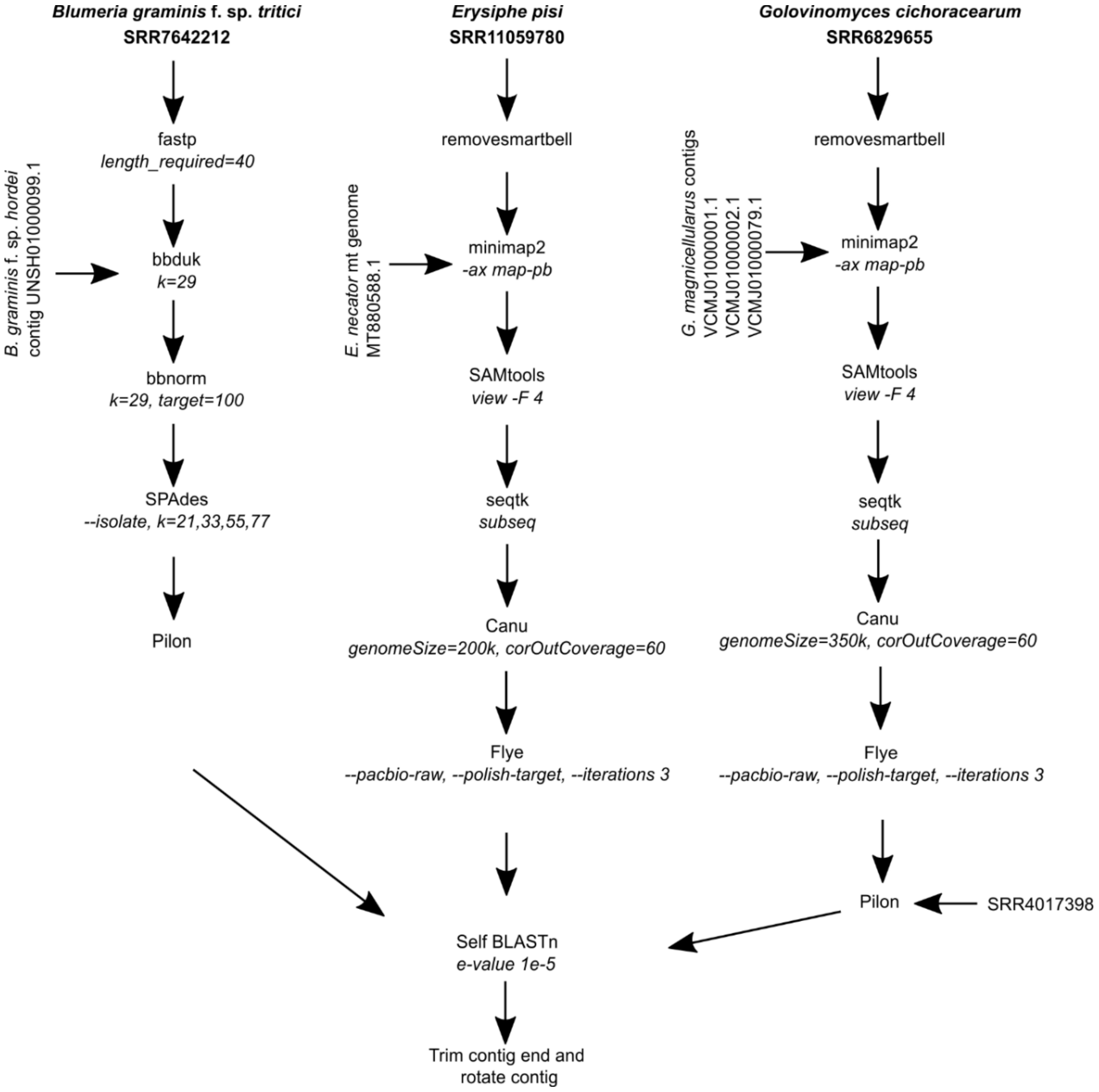


Figure 7.S1: Assembly workflow of the mitochondrial genomes of *Blumeria graminis* f. sp. *tritici*, *Erysiphe pisi*, and *Golovinomyces cichoracearum*. Arrows indicate flow of the tasks. Name of the software and parameters used are indicated. Accession numbers of assembled contigs and sequencing reads used are shown. Details are given in the Materials and Methods.

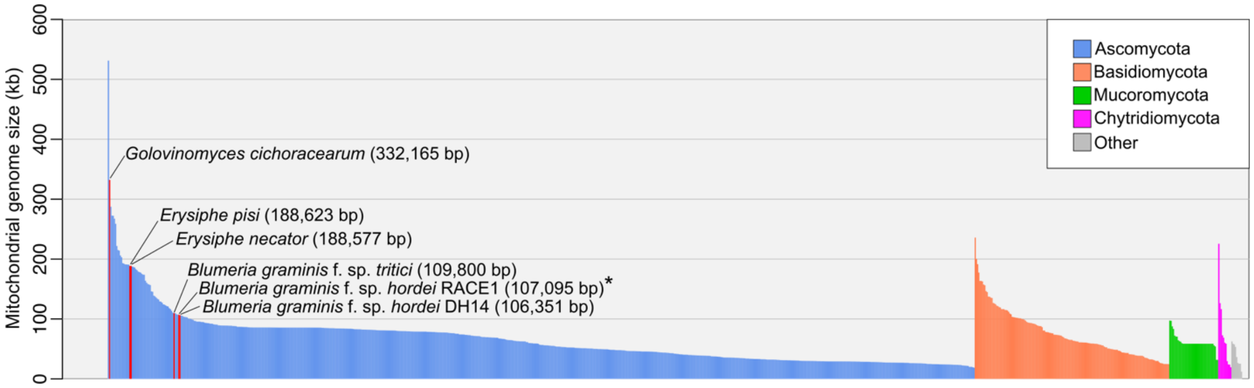
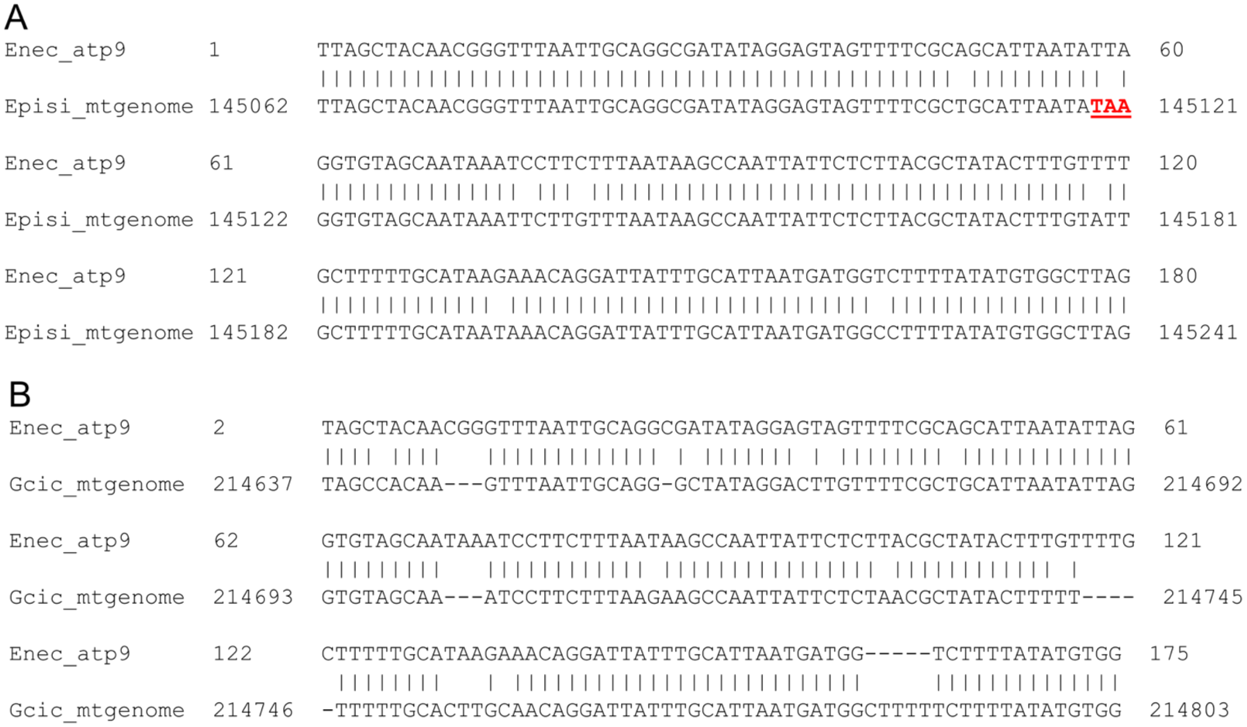


Figure 7.S2: Mitochondrial (mt) genome size comparison of 950 fungal mt genomes. Each mt genome is represented by a single bar in the graph and is color-coded based on the phylum of the fungal species that it is from. Blue: Ascomycota; orange: Basidiomycota; green: Mucoromycota; purple: Chytridiomycota; grey: Zoopagomycota, Blastocladiomycota, and Cryptomycota. The bars representing the mt genomes of the powdery mildew fungi *Blumeria graminis* f. sp. *tritici*, *B. graminis* f. sp. *hordei*, *Erysiphe necator, Erysiphe pisi,* and *Golovinomyces cichoracearum* are highlighted in red, even though they are all species of Ascomycota. The largest mt genome is that of *Morchella crassipes* (531.2 kb), whereas *Golovinomyces cichoracearum* has the second largest mt genome (332.2 kb). The complete list of the fungal species included in this bar graph can be found in Table 7.S1. The size of the mt genome of *B. graminis* f. sp. *hordei* RACE1 marked with an asterisk is shown after trimming one of the 32,179 bp overlapping ends of the original assembled contig (139,274 bp).



A

```
Enec_atp9      1        TTAGCTACAACGGGTTTAATTGCAGGCGATATAGGAGTAGTTTTCGCAGCATTAATATTA  60
                        ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||| |
Episi_mtgenome 145062   TTAGCTACAACGGGTTTAATTGCAGGCGATATAGGAGTAGTTTTCGCTGCATTAATATAA  145121

Enec_atp9      61       GGTGTAGCAATAAATCCTTCTTTAATAAGCCAATTATTCTCTTACGCTATACTTTGTTTT  120
                        ||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||| ||
Episi_mtgenome 145122   GGTGTAGCAATAAATTCTTGTTTAATAAGCCAATTATTCTCTTACGCTATACTTTGTATT  145181

Enec_atp9      121      GCTTTTTGCATAAGAAACAGGATTATTTGCATTAATGATGGTCTTTTATATGTGGCTTAG  180
                        ||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||
Episi_mtgenome 145182   GCTTTTTGCATAATAAACAGGATTATTTGCATTAATGATGGCCTTTTATATGTGGCTTAG  145241
```

B

```
Enec_atp9      2        TAGCTACAACGGGTTTAATTGCAGGCGATATAGGAGTAGTTTTCGCAGCATTAATATTAG  61
                        |||| ||||   ||||||||||||| | ||||||| | |||||||| |||||||||||||
Gcic_mtgenome  214637   TAGCCACAA---GTTTAATTGCAGG-GCTATAGGACTTGTTTTCGCTGCATTAATATTAG  214692

Enec_atp9      62       GTGTAGCAATAAATCCTTCTTTAATAAGCCAATTATTCTCTTACGCTATACTTTGTTTTG  121
                        |||||||||   |||||||||||| |||||||||||||||| |||||||||||| |
Gcic_mtgenome  214693   GTGTAGCAA---ATCCTTCTTTAAGAAGCCAATTATTCTCTAACGCTATACTTTTT----  214745

Enec_atp9      122      CTTTTTGCATAAGAAACAGGATTATTTGCATTAATGATGG-----TCTTTTATATGTGG  175
                         ||||||||   | ||||||||||||||||||||||||||     ||||||||||||||
Gcic_mtgenome  214746   -TTTTTGCACTTGCAACAGGATTATTTGCATTAATGATGGCTTTTTCTTTTATATGTGG  214803
```

Figure 7.S3: The mitochondrial (mt) genomes of the powdery mildew fungi *Erysiphe pisi* and *Golovinomyces cichoracearum* likely contain non-functional homologs of the *atp9* gene. Alignments obtained with BLASTn by querying the predicted coding sequence of the mt *atp9* gene of *E. necator* (QQY98148.1; 180 bp) against the mt genome of *B. graminis* f. sp. *tritici* (no hits; not shown), *E. pisi* (A), and *G. cichoracearum* (B). A premature stop codon in *E. pisi* and insertions/deletions in *G. cichoracearum* suggest that these two species contain non-functional *atp9* homologs, whereas *B. graminis* f. sp. *tritici* completely lost its *atp9* homolog. BLASTn search was performed with parameters *-evalue 1e-3* and *-task blastn*.

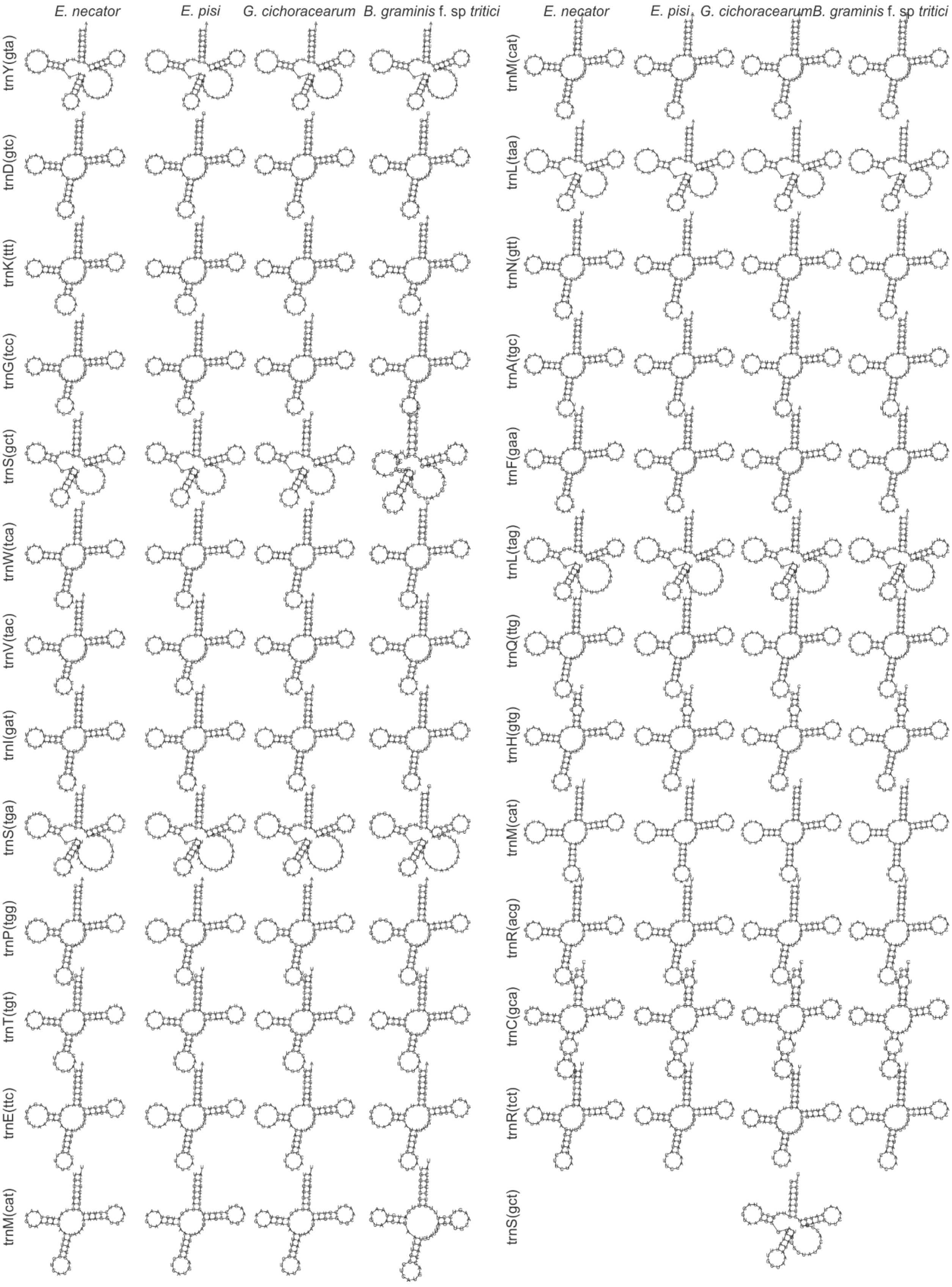

E. necator
E. pisi
G. cichoracearum
B. graminis f. sp tritici
trnY(gta)
trnD(gtc)
trnK(ttt)
trnG(tcc)
trnS(gct)
trnW(tca)
trnV(tac)
trnI(gat)
trnS(tga)
trnP(tgg)
trnT(tgt)
trnE(ttc)
trnM(cat)
E. necator
E. pisi
G. cichoracearum
B. graminis f. sp tritici
trnM(cat)
trnL(taa)
trnN(gtt)
trnA(tgc)
trnF(gaa)
trnL(tag)
trnQ(ttg)
trnH(gtg)
trnM(cat)
trnR(acg)
trnC(gca)
trnR(tct)
trnS(gct)

Figure 7.S4: Predicted secondary structures of mitochondrial (mt) tRNAs from the powdery mildew fungi *Blumeria graminis* f. sp. *tritici*, *Erysiphe necator*, *Erysiphe pisi*, and *Golovinomyces cichoracearum*. The mt-tRNAs are arranged from the top left to the bottom right of the figure according to their order in their respective mitochondrial genomes. Respective mt-tRNA-anticodons are shown between parentheses. Mt-tRNA and mt-tRNA-anticodon structures were predicted with MITOS2 web server (http://mitos2.bioinf.uni-leipzig.de/index.py). Data for the mitochondrial genome of *E. necator* were obtained from (Zaccaron et al. 2021).

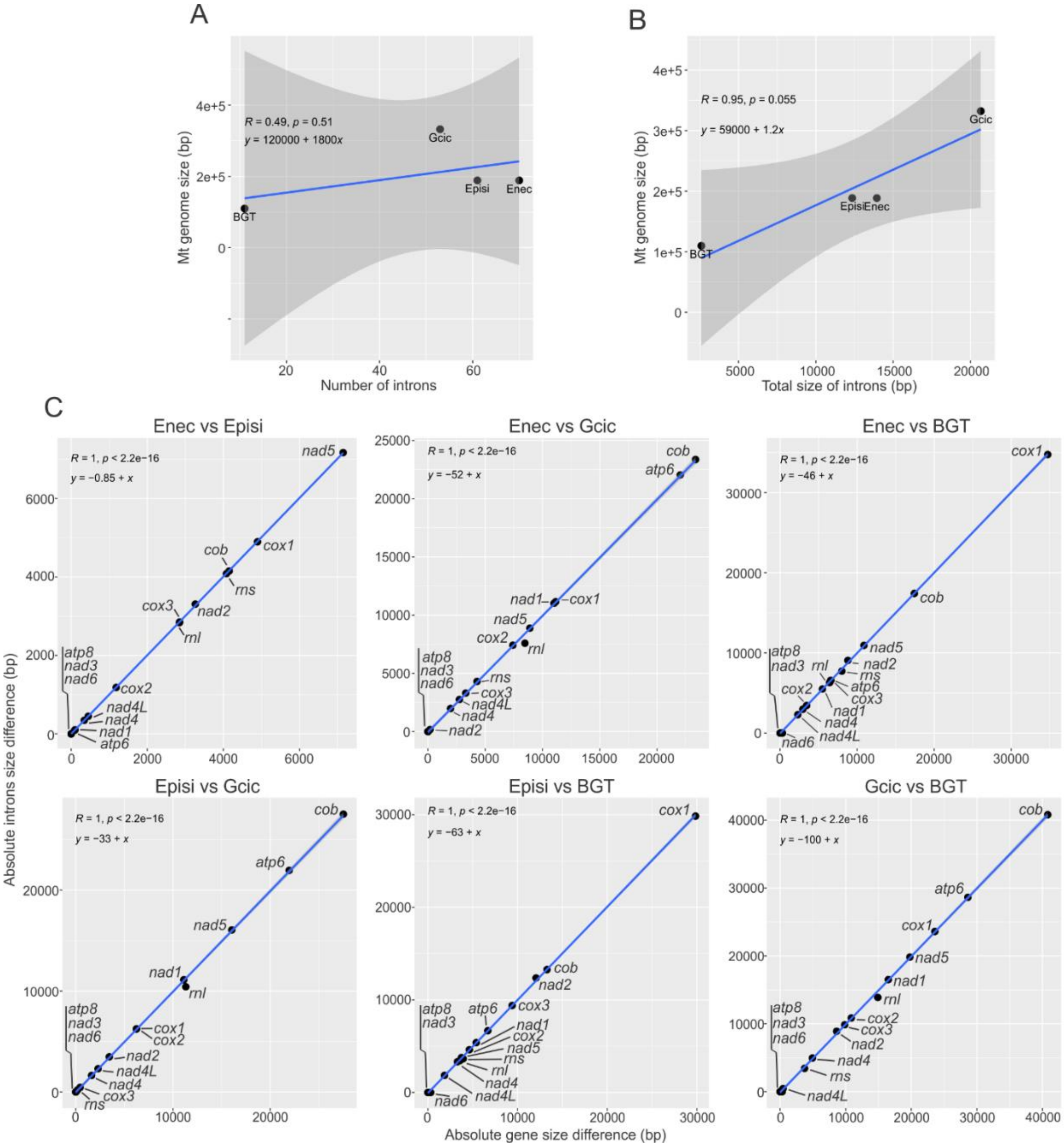


Figure 7.S5: Size of mitochondrial (mt) genomes and size of mt genes correlate with the size of intronic sequences among powdery mildew pathogens. (A) Scatter plot showing the weak correlation between the size of the mt genomes of *Blumeria graminis* f. sp. *tritici* (Bgt), *Erysiphe necator* (En), *E. pisi* (Ep), and *Golovinomyces cichoracearum* (Gc), and the number of introns present in them. (B) Scatter plot showing the strong correlation between the size of the mt genomes of Bgt, En, Ep and Gc and the total length of introns present in them. (C) Scatter plots showing the strong correlation between differences in gene and intron length for all core mt genes among pairwise comparisons of Bgt, En, Epi and Gc. The plots indicate that the differences in gene length among the four powdery mildew pathogens is explained by their differences in intron length. For all scatter plots in (A), (B) and (C), regression lines are shown in blue and were determined with the *geom_smooth* function from the R package *ggplot2*, utilizing the *lm* method. Dark areas represent confidence intervals (95%). Correlation coefficient, *p*-value and the equation of the regression line are shown at the top left corner of each plot. Data for the mt genome of *E. necator* were obtained from (Zaccaron et al. 2021).

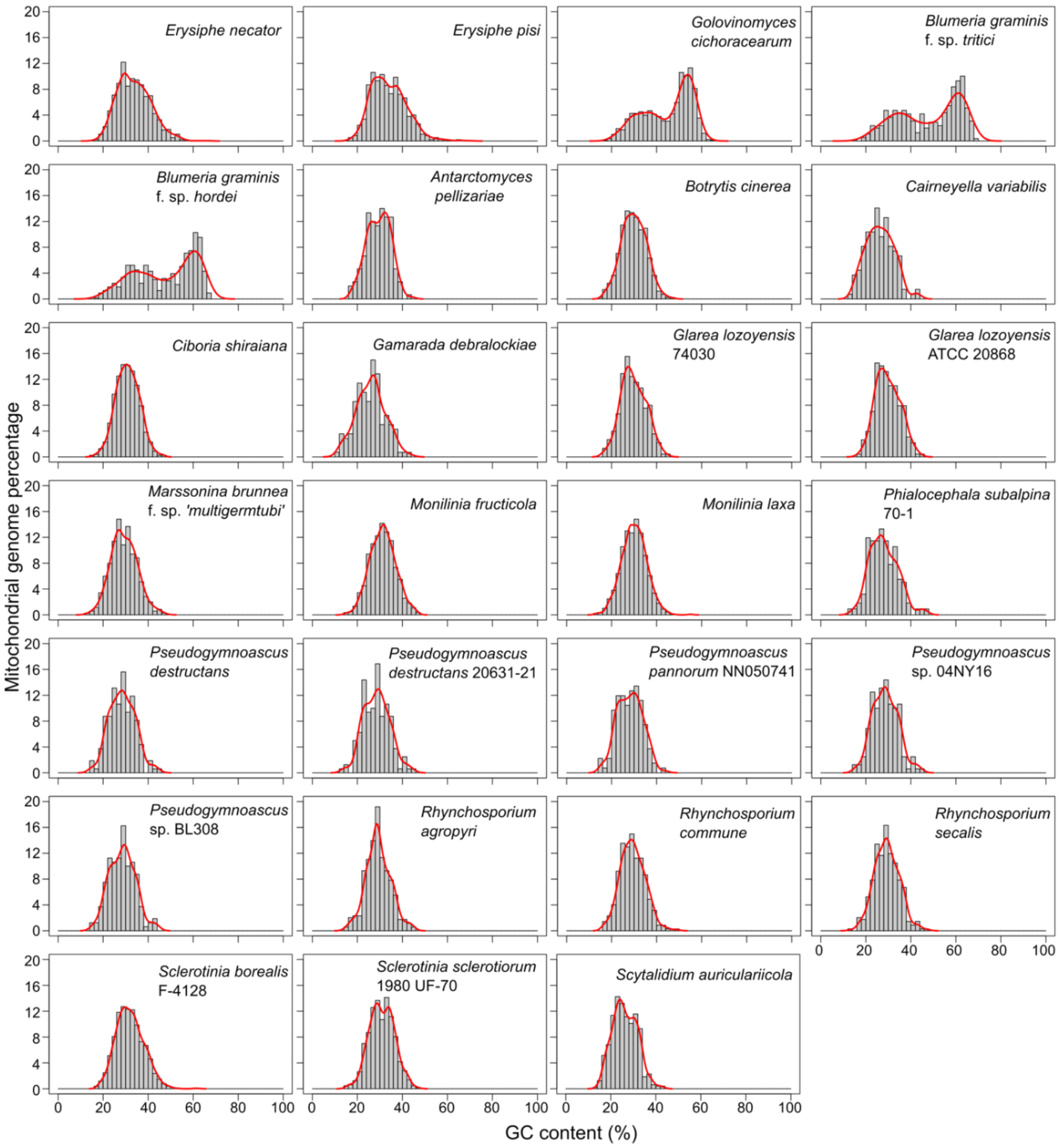


Figure 7.S6: GC content distribution in the mitochondrial (mt) genomes of members of Leotiomycetes. Histograms were generated by calculating the GC content within a non-overlapping sliding window of 200 bp across each respective mt genome. Solid lines represent the distributions estimated based on kernel density of the histograms. The graphs show that the mt genomes of the powdery mildew species *Golovinomyces cichoracearum*, *Blumeria graminis* f. sp. *tritici*, and *B. graminis* f. sp. *hordei* exhibit a bimodal distribution of GC content with two peaks at 36% and 53%, 35 and 61%, and 35% and 61%, respectively. In contrast, the mt genomes of *E. necator, E. pisi* and the other Leotiomycetes have a unimodal distribution of GC content. Accession numbers of the mt genomes are shown in Table 7.S1. Data for the mitochondrial genome of *E. necator* were obtained from (Zaccaron et al. 2021). Percentage values for the mt genome of *B. graminis* f. sp. *hordei* RACE1 were obtained after trimming one of the 32,179 bp overlapping ends of the original assembled contig (139,274 bp).

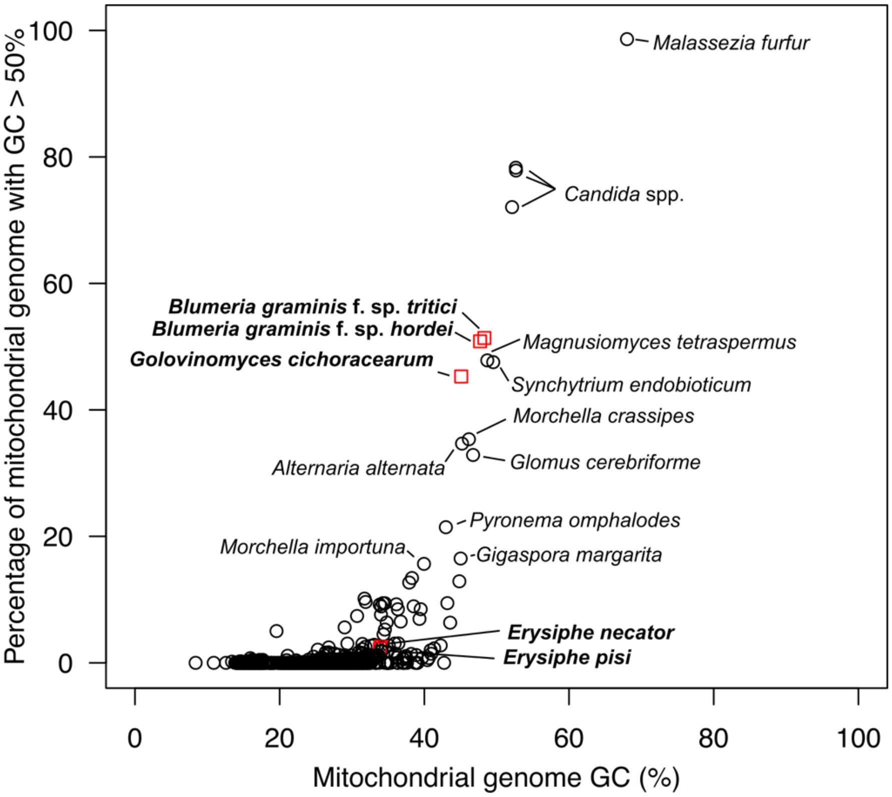


Figure 7.S7: GC content comparison of among 949 fungal mitochondrial (mt) genomes. Each circle in the scatterplot represents a mt genome, plotted on the X-axis based on its overall GC content and on the Y-axis based on what percentage of it has a GC content >50%. The mt genomes of the four powdery mildew species are shown as red squares. The scatterplot shows that the mt genomes of *Blumeria graminis* f. sp. *tritici* (GC=48%; $GC_{>50\%}$: 52%), *Blumeria graminis* f. sp. *hordei* (GC=47.7%; $GC_{>50\%}$: 50%), and *Golovinomyces cichoracearum* (GC=45%; $GC_{>50\%}$: 46%) have from the highest percentages of GC content among all other fungal mt genomes. In contrast, *Erysiphe necator* (GC=55%; $GC_{>50\%}$: 3%), and *E. pisi* (GC=33%; $GC_{>50\%}$: 4%) have mt genomes with an average GC content. The fungal species represented in this graph are listed in Table 7.S1. Percentages of the mt genomes with GC > 50% were determined using a nonoverlapping sliding window of 200 bp. Percentage values for the mt genome of *B. graminis* f. sp. *hordei* RACE1 were obtained after trimming one of the 32,179 bp overlapping ends of the original assembled contig (139,274 bp).

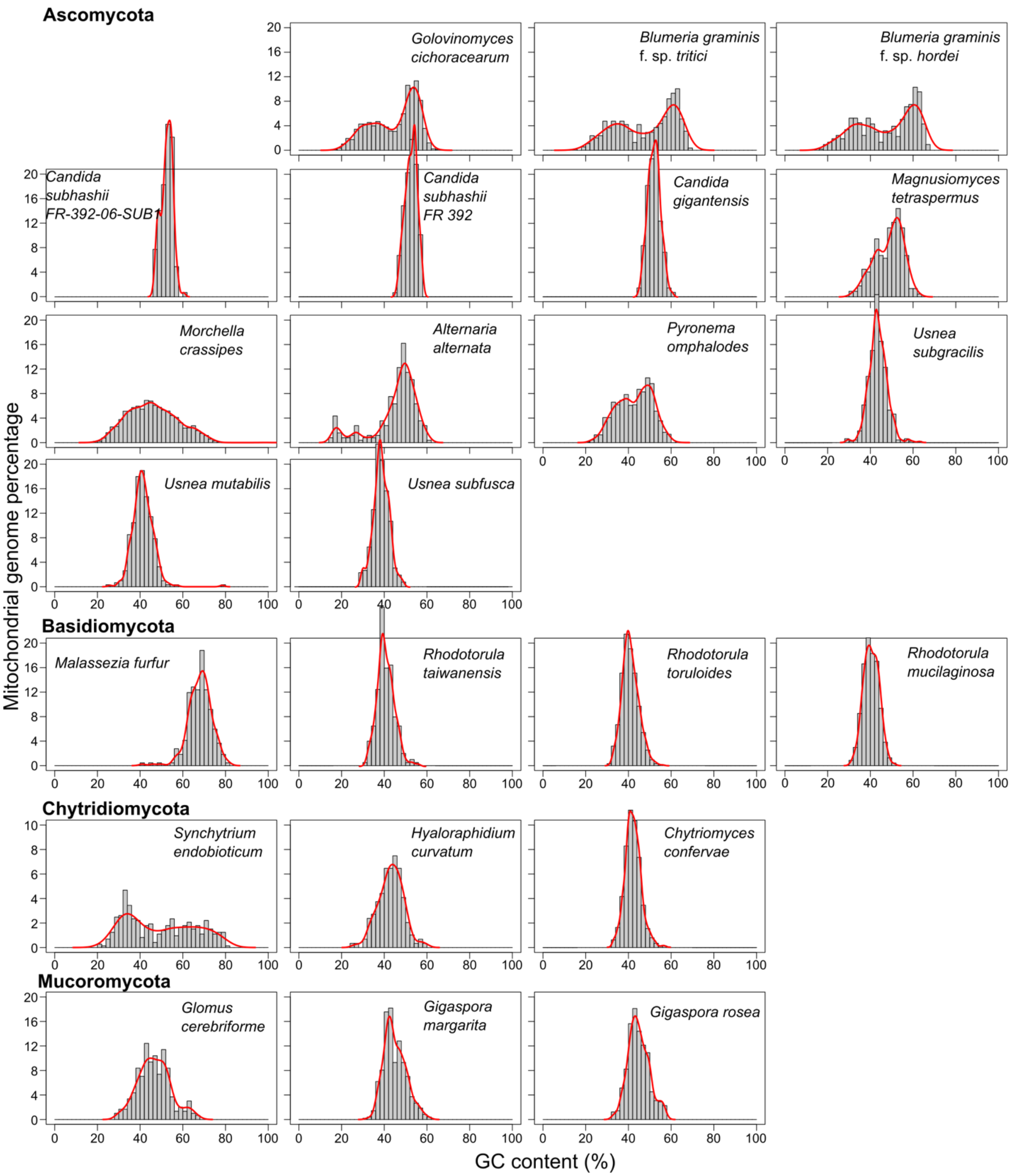


Figure 7.S8: GC content distribution in fungal mitochondrial (mt) genomes with an overall GC content higher than 40%. Histograms were generated by calculating the GC content within a non-overlapping sliding window of 200 bp across each respective mt genome. Solid lines represent the distributions estimated based on the kernel density of the histograms. The graphs show that the mt genomes of the powdery mildew species *Golovinomyces cichoracearum*, *Blumeria graminis* f. sp. *tritici*, and *B. graminis* f.

sp. *hordei* (shown at the top) exhibit a bimodal distribution of GC content with two peaks at 36 and 53%, 35 and 61%, and 35 and 61%, respectively. A similar pattern of GC content distribution although less pronounced is seen in the mt genomes of *Magnusiomyces tetraspermus, Morchella crassipes, Alternaria alternata, Pyronema ompalodes, Synchytrium endobioticum,* and *Glomus cerebriforme*. All other mt genomes with a GC content of higher than 40% have a unimodal distribution of GC content. Accession numbers of the mt genomes are shown in Table 7.S1. Percentage values for the mt genome of *B. graminis* f. sp. *hordei* RACE1 were obtained after trimming one of the 32,179 bp overlapping ends of the original assembled contig (139,274 bp).

7.7.2 Supplementary tables

Table 7.S1: Statistics of fungal mitochondrial (mt) genomes found in the NCBI Nucleotide and Organelle databases as of May 6, 2021. Mt genomes with no release date were not present in the NCBI Organelle database, but were obtained from the NCBI Nucleotide database. Genomes were sorted by length. Powdery mildew pathogens are highlighted. Columns 11 to 15 show the respective organism's taxonomy classification. The last column shows the percentage of the respective mt genome that are GC-rich (GC>50%) and calculated using a non-overlapping sliding window of 200 bp. Data for the mt genome of *Erysiphe necator* were obtained from Zaccaron et al (2021). This table is available at https://zenodo.org/records/11211529.

Table 7.S2: Short tandem repeats within the mitochondrial (mt) genomes of the powdery mildews *Blumeria graminis* f. sp. *tritici*, *Erysiphe pisi*, *Golovinomyces cichoracearum*, and *Erysiphe necator*. Tandem repeats were identified with the Tandem Repeat Finder program v4.09. Data for the mt genome of *E. necator* were obtained from Zaccaron et al (2021). This table is available at https://zenodo.org/records/11211529.

Table 7.S3: Classification of introns in the mitochondrial (mt) genomes of the powdery mildew pathogens *Blumeria graminis* f. sp. *tritici*, *Erysiphe pisi*, *Golovinomyces cichoracearum*, and *Erysiphe necator*. Information of intronic ORFs encoding homing endonucleases or reverse transcriptase is shown. Start and end coordinates are relative to the respective mt genome. Intron IDs are composed of the gene name that they are from followed by the intron number from 5' to 3' end. Data for the mt genome of *E. necator* were obtained from Zaccaron et al (2021). This table is available at https://zenodo.org/records/11211529.

Table 7.S4: Core mitochondrial genes of powdery mildew species vary greatly in size and intron content. Shown are the total size (from start to stop codons) of each gene, the size of the mature transcripts (introns spliced out), the number of introns, and the total size of introns. Data for the mitochondrial genome of *Erysiphe necator* were obtained from Zaccaron et al (2021). This table is available at https://zenodo.org/records/11211529.

Table 7.S5: Mitochondrial (mt) introns are overall poorly conserved at the nucleotide level among the powdery mildew pathogens *Blumeria graminis* f. sp. *tritici* (Bg), *Erysiphe necator* (En), *Erysiphe pisi* (Ep), and *Golovinomyces cichoracearum* (Gc). The table shows pairwise identity values (%) of introns inserted into the same site between two species. If an intron is absent in one or both species, the corresponding identity value is denoted as NA. Introns were aligned globally using the Needleman -Wunsch algorithm implemented in the R package Biostrings. Identity values were calculated with the formula 100 * (identical positions)/(aligned positions + internal gap positions). Data for the mt genome of *E. necator* were obtained from Zaccaron et al (2021). This table is shown in the next page.

Intron insertion site	En vs. Ep	En vs. Gc	En vs. Bg	Ep vs. Gc	Ep vs. Bg	Gc vs. Bg
cox1-153	NA	61.1	NA	NA	NA	NA
cox1-244	NA	NA	NA	NA	59.5	NA
cox1-257	89.1	54.8	NA	54.8	NA	NA
cox1-326	88.7	48.3	NA	47.2	NA	NA
cox1-417	NA	46.7	NA	NA	NA	NA
cox1-438	91.5	NA	NA	NA	NA	NA
cox1-538	83.1	NA	NA	NA	NA	NA
cox1-636	97.2	NA	NA	NA	NA	NA
cox1-660	71.7	NA	NA	NA	NA	NA
cox1-765	76.7	89.8	NA	78.2	NA	NA
cox1-776	66.8	NA	NA	NA	NA	NA
cox1-852	94.8	NA	NA	NA	NA	NA
cox1-866	88.3	NA	NA	NA	NA	NA
cox1-913	96.4	NA	71.2	NA	72.7	NA
cox1-939	99.1	NA	NA	NA	NA	NA
cox1-1016	77.9	NA	NA	NA	NA	NA
cox1-1102	95.8	72.6	NA	72.2	NA	NA
cox1-1153	94	NA	NA	NA	NA	NA
cox1-1170	63.2	56.2	57.4	47	55.1	56.7
cox1-1326	NA	NA	45	NA	NA	NA
cox2-87	92.5	97.1	NA	91	NA	NA
cox2-307	54.5	NA	NA	NA	NA	NA
cox2-363	75.3	48.6	NA	45.5	NA	NA
cox2-558	74.2	52.8	46.5	51.6	44.9	45.1
cox3-225	97	54.3	NA	52.8	NA	NA
cox3-339	97.5	96.9	NA	98.5	NA	NA
cox3-646	93.6	95.1	NA	94.5	NA	NA
cox3-733	92.3	NA	NA	NA	NA	NA
cox3-784	NA	NA	NA	46.9	NA	NA
nad1-291	71.1	62.1	NA	51.1	NA	NA
nad1-388	NA	NA	49.2	NA	NA	NA
nad1-636	56.4	46.7	NA	46.3	NA	NA
nad2-481	64.6	64.3	NA	71.6	NA	NA
nad2-1020	69	NA	NA	NA	NA	NA
nad2-1218	91.6	NA	NA	NA	NA	NA
nad2-1308	79.4	52.5	NA	53.2	NA	NA
nad4-484	89	53.1	NA	51	NA	NA
nad4-714	80.9	NA	NA	NA	NA	NA
nad4L-240	73.8	43	44.9	45	43.8	46.1
nad5-324	87.5	49.4	NA	47.6	NA	NA
nad5-417	NA	NA	NA	NA	NA	51
nad5-570	NA	50	NA	NA	NA	NA
nad5-717	50.5	NA	NA	NA	NA	NA
nad5-924	55.5	52.3	NA	58.1	NA	NA
atp6-347	72.4	48.3	NA	48.2	NA	NA
atp6-524	NA	NA	NA	48.5	NA	NA
atp6-575	76.4	54.3	NA	65.6	NA	NA
cob-99	84.5	65.2	53.8	61.8	52.8	60.1
cob-159	89.8	67.4	NA	66.4	NA	NA
cob-201	90.3	54.8	NA	52.7	NA	NA
cob-358	96.6	NA	NA	NA	NA	NA
cob-379	71.4	60	NA	55.5	NA	NA
cob-439	92.5	48.7	NA	49.9	NA	NA
cob-506	99.6	61.5	NA	61.7	NA	NA
cob-619	66.1	61.8	NA	66.2	NA	NA
cob-824	69.1	51.8	NA	52.1	NA	NA
rns-1414	94.3	76.9	NA	76.3	NA	NA
rns-1525	89.5	87.6	NA	90.8	NA	NA
rnl-1184	97.5	52.2	NA	52.2	NA	NA
rnl-2155	87.9	72.9	NA	72.3	NA	NA
rnl-2923	76.6	31.6	52	30.4	55.4	27.4

Table 7.S6: GC-rich regions in the mitochondrial (mt) genomes of *Blumeria graminis* f. sp. *tritici* and *Golovinomyces cichoracearum* are poorly conserved in other species. Queries correspond to GC-rich (GC%>50) segments within the mt genomes of *B. graminis* f. sp. *tritici* (bgt) and *G. cichoracearum* (gcic). Homology searches were performed with BLASTn against the nucleotide nr database (NCBI) with an e-value < 1e-5. From the 69 GC-rich regions of *B. graminis* f. sp. *tritici*, 58 matched only to a sequence (accession LR026995.1) that contains the mt genome of *B. graminis* f. sp. *tritici*, and was assembled along with its nuclear genome. From the 230 GC-rich regions of *G. cichoracearum*, 212 matched scaffolds from the genome assembly of the Oomycete *Albugo laibachii*. This table is available at https://zenodo.org/records/11211529.

Table 7.S7: Features and statistics of the cytochrome b (*cob*) gene from four powdery mildew pathogens. Data for the mitochondrial genome of *Erysiphe necator* were obtained from Zaccaron et al (2021).

Cytochrome b feature	***B. graminis* f. sp. *tritici***	***E. necator***	***E. pisi***	***G. cichoracearum***
Length (bp)	4534	21964	17812	45335
CDS length (bp)	1161	1170	1170	1170
Protein length (aa)	386	389	389	389
Introns length (bp)	3373	20794	16642	44165
Number of introns	1	10	9	13
Group I introns	0	6	6	6
Group II introns	1	3	3	4
Unclassified introns	0	1	0	3
LAGLIDADG-coding intronic ORFs	0	5	4	1
GIY-YIG-coding intronic ORFs	0	0	0	1
RT-coding intronic ORFs	1	3	3	4
CDS identity with *B. graminis* f. sp. *tritici* (%)	100	89.4	89.7	92.1
CDS identity with *E. necator* (%)	89.4	100	97.7	92.5
CDS identity with *E. pisi* (%)	89.7	97.7	100	93.2
CDS identity with *G. cichoracearum* (%)	92.1	92.5	93.2	100
Protein identity with *B. graminis* f. sp. *tritici* (%)	100	94.9	94.9	95.6
Protein identity with *E. necator* (%)	94.9	100	99	97.9
Protein identity with *E. pisi* (%)	94.9	99	100	97.9
Protein identity with *G. cichoracearum* (%)	95.6	97.9	97.9	100

Table 7.S8: Summary of statistics, features and accession numbers of cytochrome b (*cob*) genes from different fungal species. Members of the Leotiomycetes are shown at the top and powdery mildew pathogens are highlighted. For each species, NCBI accession numbers for the mt genome and cob gene are shown, followed by the size of the *cob* coding sequence, coordinate of the exons, number of introns, *cob* start and end coordinates, total size of *cob*, average number of introns per kb of *cob* coding sequence, taxID of the species, description of the mt genome shown at NCBI, mt genome length, taxonomic lineage and genetic code number. This table is available at https://zenodo.org/records/11211529.

Table 7.S9: Fungal mitochondrial (mt) genomes containing cytochrome b (*cob*) genes harboring reverse transcriptase (RT)-encoding ORFs within its introns. RT-encoding ORFs are embedded within introns inserted into nine sites of *cob* coding sequence, i.e., 99 (cob-99), 159 (cob-159), 247 (cob-212), 277 (cob-277), 311 (cob-311), 358 (cob-358), 541 (cob-541), and 687 (cob-687), using *Erysiphe necator cob* as reference. For each fungal mt genome in the table, presence of an RT-encoding ORF within the respective intron insertion site is indicated with the coordinates of the respective ORF (5' to 3'). Absence of a RT-encoding ORF is indicated with black cells. Coordinates of the ORFs are in reference to the respective mt genome, for which the GenBank accession number is given in the third column. This table is available at https://zenodo.org/records/11211529.

Chapter 8

Concluding remarks and future work

Alex Z. Zaccaron

Summary

The sequencing of genomes is often driven by questions related to the understanding of genetic traits and molecular mechanisms. When it comes to fungal pathogens, genome sequencing allows the investigation of major questions such as which and how molecular mechanisms mediate pathogenicity, antifungal tolerance, and adaptation to adverse conditions. In this respect, chromosome-level genome assemblies have been essential to enriching our understanding of the genome organization and the landscape of structural variations associated with evolutionary adaptation in fungal pathogens. In this dissertation, I have applied whole genome sequencing using PacBio long-read technology combined with chromatin conformation capture (Hi-C) to obtain, for the first time, chromosome-level genome assemblies for the tomato pathogen *Cladosporium fulvum* and the grapevine pathogen *Erysiphe necator*. I used advanced bioinformatics and comparative genomic techniques to reveal the contrasting genome organizations of these two fungal pathogens. The genome of *C. fulvum* has a discernible 'checkerboard' pattern composed of gene-dense, repeat-poor regions interspersed with gene-sparse, repeat-rich regions, and the presence of 13 core chromosomes, two of which are dispensable. In contrast, the genome of *E. necator* does not exhibit evident compartmentalization of gene-dense and gene-poor regions, and no dispensable chromosomes were detected. Furthermore, the genome of *E. necator* was distinguished for its long pericentromeric regions, a high percentage of duplicated genes, and a reduced gene complement that underlies its obligate lifestyle. I anticipate that the results presented herein will serve as a solid basis for future targeted functional studies and comparative genomic analyses with other fungal pathogens. Next, I discuss the major findings presented in this dissertation in the context of fungal genetics and plant pathology.

8.1 Transposable elements as catalysts of fungal genome organization

The amount and estimated age of transposable elements (TEs) are remarkable features that contrast the genomes of *C. fulvum* and *E. necator*. As shown in **Chapter 2**, the genome of *C. fulvum* appears to have undergone a single burst of TEs relatively recently in its evolutionary history. Interestingly, the genome of the pine tree pathogen *Dothistroma septosporum*, a close relative of *C. fulvum*, is almost void of TEs (De Wit et al. 2012). This indicates that the surge of TEs in the genome of *C. fulvum* occurred after divergence from *D. septosporum*, which occurred approximately 20 Mya (Ohm et al. 2012). In contrast, as shown in **Chapter 5**, the genome of *E. necator* has a high abundance of ancient TEs and exhibits signatures of at least two TE bursts in its evolutionary history, as seen in other powdery mildew fungi as well (Frantzeskakis et al. 2018). The extent to which the genomes of powdery mildew fungi share the same TE families is currently unknown. However, it is reasonable to assume that different species of powdery mildew fungi still carry in their genomes TEs from an ancient burst that took place in the genome of their common ancestor species. Combining these observations with the contrasting organization of the genomes of *C. fulvum*, which shows clear compartmentalization, and *E. necator*, which shows weak compartmentalization, one speculation is that the extent to which the genomes of fungal pathogens are compartmentalized depends on the points in their evolutionary history that they experienced bursts of TE proliferation. Because the distribution of TEs drastically change the organization of fungal genomes and further impacts the frequency of mutations, e.g., by promoting structural variations (SVs), TE bursts in fungal genomes may have profound impacts on fungal genome evolution and the ability of these organisms to adapt to new hosts and environments. Further genomic analyses can investigate this speculation by searching for a possible correlation between the amount of ancient TEs and the degree of compartmentalization in fungal genomes from different lineages. Such investigation could lead to the conclusion that not only the amount of TEs, but also their age are factors contributing to contrasting genome organizations among fungal species.

8.2 Compartmentalization as a possible transitional stage of repeat-rich genomes

Compartmentalization of genomes into repeat-rich and repeat-poor regions arises from a biased distribution of TEs that are typically found clustered within a genome. It is widely accepted that clustering of TEs partially stems from selection against the insertion of TEs into genomic regions containing essential genes, thereby minimizing the chances of disrupting their coding or regulatory regions that could h ave a negative impact on fitness of the organism. At the same time, insertion of TEs into repeat -rich regions is under more relaxed selection because important genes are less likely to be affected. Many fungal pathogens, including *C. fulvum*, exhibit pronounced compartmentalization in their genomes. Interestingly, genome compartmentalization is not unique to fungi, because it is also observed in the genomes of other eukaryotes that span more than one billion years of evolution (Gozashti et al. 2023). Thus, one reasonable speculation is that compartmentalization is a convergent feature of repeat -rich genomes, that emerged independently in different lineages. One advantage of compartmentalized genomes is that repeat -rich compartments are often affected by point mutations and structural variations, such as duplications and deletions, which can serve as a cradle of genetic diversity leading to adaptation. Considering the long-term benefits of genome compartmentalization, one question that arises is why it is not ubiquitous in fungi and other organisms. Perhaps the answer is that genome compartmentalization in fungi and other eukaryotes is a transitional stage that naturally arises after a first burst of TE proliferation, and which subsequently starts to disappear as TEs continue to proliferate over the period of millions of years, thereby ‘diluting’ the gene-rich regions. Genomic studies comparing age of TEs and genome compartmentalization in fungi and other eukaryotes could support such a speculation.

8.3 Pericentromeres in powdery mildews, large graveyards of young transposable elements

As described in **Chapter 5**, the pericentromeres in the genome of *E. necator* could be clearly distinguished from the rest of the genome because they were long regions containing much higher amounts of TEs. The low levels of mRNA transcription observed in these regions align well with the heterochromatic state expected for pericentromeric regions (Hall et al. 2012), and indicates that the TEs in these regions are mostly inactive. In eukaryotes, there is evidence of a positive correlation between total genome size and centromere size (Zhang and Dawe 2012). In plants, which have much larger genomes compared to fungi, centromeres can be up to 4 Mb in size. However, Mb-sized centromeres are unusual in fungi, which have an average genome size of 40.8 Mb. In powdery mildew fungi, Mb-sized centromeres appear to be common (Müller et al. 2019, 2021), indicating that this is a distinctive feature of Erysiphales. Interestingly, when comparing the chromosomes of *E. necator* with the chromosomes of the wheat powdery mildew *Blumeria graminis* f. sp. *tritici*, the pericentromeric regions were poorly conserved compared to the remaining regions of the chromosomes. This observation is in accord with the phenomenon described as the 'centromere paradox', which refers to the rapid change of these loci, even among closely related species, despite having a well-conserved and essential function (Henikoff et al. 2001). However, the overall low sequence diversity of TEs in the pericentromeric regions of the *E. necator* chromosomes is particularly intriguing and indicates that the long pericentromeric regions originated from recent bursts of TEs, likely after divergence from *B. graminis* f. sp. *tritici*. This would explain the overall poor sequence conservation in these genomic regions between the two pathogens. Based on these observations, I hypothesize that the long pericentromeric regions in the genomes of powdery mildew fungi are almost entirely composed of inactive TEs that emerged from recent TE bursts after speciation. Future research can shed light on the evolution of centromeres in powdery mildew fungi. For example, chromatin immunoprecipitation using antibodies against centromere protein A (CENP-A), which determines the position of the active centromere and kinetochore (Westhorpe and Straight 2015), can reveal the exact location and the repetitive sequence of centromeres. Moreover, the acquisition of chromosome-level genome assemblies from other powdery mildew species will allow more detailed discoveries related to the sequence conservation, differences in TE content, and possible differences in the structure of pericentromeric regions within Erysiphales. These steps would be key to generating hypotheses about differences in the evolution of centromeres in fungi, which can largely contribute to the genome organization as observed in powdery mildew fungi.

8.4 The genome of *C. fulvum* is a fertile ground for repeat-induced structural variations

The presence of TEs in genomes often promotes non-allelic homologous recombination (NAHR), which occurs during mispairing of similar DNA sequences and results in structural variations (SVs) that can be deletions, segmental duplications, or inversions. As reported in **Chapter 3**, nearly all SVs in the genome of *C. fulvum* consist of TE insertions or deletions of a few thousand base-pairs. One possibility is that the occurrence of these SVs is the result of retrotransposition of active TEs. However, another possibility is that most of the SVs in the genome of *C. fulvum* are the result of NAHR, mediated by the presence of the TEs. In either case, TEs have undoubtedly played a key role in the increase of genetic diversity in the asexually reproducing pathogen *C. fulvum*. Nearly all SVs in the genome of *C. fulvum* do not directly affect predicted protein-coding genes, and therefore they probably do not have a major impact in the fitness of the pathogen. However, a few SVs seem to have played a key role in enabling *C. fulvum* to overcome resistance of its host tomato. This was the case for the three SVs that resulted in the deletion of the effector genes *Avr9*, *Avr5,* and *Avr4E*. Loss of these effector genes prevents the recognition of their encoded proteins by their matching resistance genes *Cf-9*, *Cf-5*, and *Cf-4E* in tomato, thus avoiding effector-triggered immunity and resulting in compatible interactions. The borders of the SVs that deleted *Avr9*, *Avr5,* and *Avr4E* co-localized with highly similar copies of TEs, suggesting deletion events mediated by NAHR. While *Avr5* and *Avr4E* were likely deleted via unequal crossing over, deletion of *Avr9* appears to have occurred after a double strand break at its locus. This subsequently triggered the break-induced replication pathway and allowed recombination with homologous TEs located at ectopic positions, thereby resulting in a non-reciprocal translocation. While these molecular mechanisms described to delete *Avr9*, *Avr5,* and *Avr4E* are speculative, it is possible to carry out future experiments to support deletion of *Avr9* through the DNA repair pathway, by inducing a double strand break at the *Avr9* locus, using for example the CRISPR/Cas9 system, in isolates with wild-type *Avr9*. Observation of an ectopic recombination after inducing a double-strand break would support the hypothesis of the mechanism behind the deletion of *Avr9*.

It is intriguing to imagine a different outcome in a hypothetical scenario where *Avr9*, *Avr5*, and *Avr4E* are not located in TE-rich regions. Would these genes still be deleted without TEs to promote NAHR? Would *C. fulvum* take longer to overcome the matching resistance genes *Cf-9*, *Cf-4E*, and *Cf-5* in tomato? If so, the location of effector genes, or rather the TE content surrounding them, in the genome of pathogens is relevant information in crop breeding to increase durability of resistance genes.

8.5 Dispensable chromosomes of *C. fulvum* as a possible source of genetic diversity

As described in **Chapter 3**, we discovered two dispensable chromosomes in the genome of *C. fulvum*, i.e. Chr14 and Chr15, whose existence was previously unknown in this pathogen. Despite the importance of this discovery, genes present in Chr14 and Chr15 exhibit low transcriptional activity during interaction with tomato and encode hypothetical proteins with limited similarity to proteins from other fungal species. Thus, no role in pathogenicity or other putative functions could be assigned for the disposable chromosomes of *C. fulvum*. Functional studies could potentially provide clues about the importance of Chr14 and Ch15 for *C. fulvum*, but these will be very challenging in this pathogen. For example, an experiment to induce chromosome loss would be difficult considering that *C. fulvum* reproduces almost always asexually. One notable feature observed in Chr14 though, is an abundance of Repeat-Induced Point (RIP) mutations that was much higher than in all the other 13 core chromosomes. This raises the possibility that the dispensable chromosomes of *C. fulvum* can rapidly accumulate RIP mutations during rounds of sexual reproduction, and thus increase the genetic diversity in this fungus. However, an experiment similar to the one performed in the wheat pathogen *Zymoseptoria tritici* (Komluski et al. 2023), aiming on verifying the accumulation of RIP mutations in the genome of *C. fulvum,* would be impractical considering the near absence of sexual reproduction in this pathogen. Recently, a dispensable chromosome has been hypothesized to increase genetic diversity of clonal populations of the rice blast fungus *Magnaporthe oryzae*. In this pathogen, a dispensable chromosome was recurrently transmitted horizontally from a lineage that causes disease in Indian goosegrass to a lineage that causes disease in rice (Barragan et al. 2024). Even though no evidence supporting the horizontal transfer of a dispensable chromosome was found in *C. fulvum*, possibly due to the small number of isolates analyzed, it may still be the case among lineages of this pathogen. Therefore, one hypothesis is that dispensable chromosomes and SVs induced by TEs, collectively promote genetic diversity in asexual pathogens such as *C. fulvum*. A large-scale population genomic analysis that includes hundreds of isolates collected world-wide could be carried out to investigate this hypothesis, and whether Chr14 and Chr15 can be horizontally transferred among isolates of the fun gus.

8.6 Alternative splicing, a possible major source of diversity in fungal pathogens

Alternative splicing (AS) in fungal pathogens has been largely unexplored. One reason for this is the difficulty in obtaining full-length transcripts from short reads because of the short intergenic regions in fungal genomes, with frequent overlap of untranslated regions (UTRs). While analyzing the transcriptome of *C. fulvum*, I adopted the strategy to assemble transcripts gene-by-gene, which almost solved the problem of chimeric transcripts assembled from physically close genes. The results presented in **Chapter 4** show that AS in *C. fulvum* genes during the course of tomato infections, is frequently observed in genes encoding major facilitator superfamily (MFS) and sugar-like transporters, transcription factors, and cytochrome P450 enzymes. This implies that AS in these genes could be promoting adaptation to the host environment by regulating nutrition, metabolism, and cellular detoxification processes. Future functional analyses could explore this possibility and determine the extent to which AS in pathogen genes during host infections is used as a means to finetune virulence on the host. Another interesting finding is that AS genes in the genome of *C. fulvum* have significantly longer 5' intergenic regions with higher repetitive DNA content compared to non-AS genes. This observation led to the hypothesis that a causal relation may exist between the presence of TEs and AS, meaning that next to SVs, repeat-rich compartments of the genome are subjected to higher levels of AS as well. In *C. fulvum*, candidate effector genes were much less likely to undergo AS compared to other functional gene categories. This observation is somewhat surprising considering that, as compared to accumulation of mutations, AS could provide a faster way to modulate effector sequences during host infections in order to avoid recognition by matching resistance genes. Further experiments in *C. fulvum* and other pathogens can elucidate the benefits of AS, extending from different functions of isoforms to possible benefits in regulation, stability, and translation efficiency of mRNA with varying UTRs. Notably, experiments using liquid chromatography mass spectrometry (LC-MS) can confirm the translation of different isoforms by identifying peptides unique to each isoform. Also, ribosome profiling (Ribo-seq) combined with RNA-seq could be used to identify differences in translation efficiency among isoforms encoding the same protein, but having different UTRs (Ingolia et al. 2019).

8.7 Change in sterol composition through an evolutionary gene loss

One peculiar feature about the biology of *E. necator* and other powdery mildew fungi is that the main sterol in their cell membranes is not ergosterol as in other Ascomycete fungi, but instead 24-methylenecholesterol (ergosta-5,24(24^1)-dienol) (Debieu et al. 1995; Loeffler et al. 1992). As demonstrated in **Chapter 5**, the absence of *ERG4* and *ERG5* genes, which are crucial in the final steps of ergosterol biosynthesis, in *E. necator* suggests that the lack of ergosterol in powdery mildew fungi may be attributed to the loss of *ERG4* and *ERG5*. In *Saccharomyces cerevisiae*, ERG5 desaturates ergosta-5,7,24(28)-trienol into ergosta-5,7,22,24(28)-tetraenol, which is subsequently converted to ergosterol by ERG4 (Lees et al.

1995). It is reasonable to speculate that, in powdery mildew fungi, the absence of *ERG4* and *ERG5* would lead to the accumulation of ergosta-5,7,24(28)-trienol, necessitating conversion to 24-methylenecholesterol. The mechanism underlying this conversion in powdery mildew fungi remains undiscovered. However, insights can be gleaned from a study that engineered a strain of *S. cerevisiae* capable of producing 24-methylenecholesterol (Yang et al. 2021). This study replaced the *ERG4* and *ERG5* genes with *DHCR7*, encoding a 7-dehydrocholesterol reductase involved in the cholesterol biosynthesis pathway in animals. While *S. cerevisiae* lacks an ortholog of *DHCR7*, *E. necator* and other powdery mildew fungi possess such an ortholog. Hence, I hypothesize that the *DHCR7* ortholog in powdery mildew fungi is required for the biosynthesis of their primary membrane sterol, 24-methylenecholesterol. Further experiments, integrating gene silencing with gas chromatography-mass spectrometry to assess sterol composition, can corroborate the essential role of DHCR7 in the biosynthesis of 24-methylenecholesterol. Moreover, phenotypic characterization may elucidate the extent to which *DHCR7* silencing impedes the growth of powdery mildew fungi. Speculating on potential benefits of the loss of ergosterol biosynthesis in powdery mildews, it is conceivable that its absence aids in evading sterol-triggered immunity in the host. Furthermore, the loss of ergosterol may confer advantages to the obligate biotrophic lifestyle of powdery mildew fungi, considering the pivotal roles sterols play in cell membrane fluidity and permeability. Notably, a study demonstrated that altering sterol composition in *S. cerevisiae* from ergosterol to fecosterol provided a phenotype with increased thermotolerance (Caspeta et al. 2014). Therefore, it is plausible that the substitution of ergosterol with 24-methylenecholesterol, through the loss of *ERG4* and *ERG5*, represents an adaptive evolutionary response within the Erysiphales lineage.

8.8 A new family of fungal carboxylesterases

As described in **Chapter 5**, the identification of the *E. necator* gene *HI914_00624* that encodes a putative carboxylesterase (CE) is one of the highlights of this dissertation, given that it is a member of putative powdery mildew-specific family of CEs. The extensive variation in copy number of *HI914_00624* among isolates of *E. necator* implies an important but so far unknown role of this CE for the fungus. Supporting this hypothesis is the observed effect of extensive copy number variations in the *Cyp51* gene that enhance resistance to azole fungicides in *E. necator* (Jones et al. 2014). While catalytic competence is considered the ancestral state of CEs (Oakeshott et al. 1999), lineages of catalytically inactive CEs have evolved in ancestors of vertebrate, serving functions unrelated to hydrolysis. This is the case of neuroligins, gliotactins, and neurotactins that are involved in neurodevelopmental signaling and cell adhesion (Dean and Dresbach 2006; Sun et al. 2011; Speicher et al. 1998; Oakeshott et al. 2005). One notable observation about the gene *HI914_00624* is that its homologs in other powdery mildew fungi lack the conserved catalytic triad, while its closest homologs in fungi outside Erysiphales retained the catalytic triad. Based on the observations described above, I propose that the gene *HI914_00624* belongs to a novel family of CEs that evolved from a catalytic competence state to serve distinct functions within the Erysiphales lineage. Future research will shed light on the function and significance of *HI914_00624* or its homologs in powdery mildew fungi for growth and virulence on the host. As a starting point, employing site-directed mutagenesis aiming to replace the residue Gly251 with Ser251, a nucleophile capable to attack the carbonyl carbon of the ester substrate (Oakeshott et al. 2005), could potentially restore the enzymatic activity of HI914_00624. Subsequently, CE activity could be assayed by measuring p-nitrophenyl acetate hydrolysis (Hosokawa and Satoh 2001), thus confirming the loss of catalytic activity in the wild-type HI914_00624. Additionally, RNA interference-mediated silencing could unveil the importance of *HI914_00624* for growth and virulence, although phenotyping *E. necator* growing on grape leaves remains challenging. Finally, detecting copy-number variations in orthologs of *HI914_00624* in other powdery mildew fungi would bolster the hypothesis that the dosage effect resulting from gene duplication confers advantages in adverse conditions.

8.9 Duplication of effector genes, the more the better?

As described in **Chapter 2**, among all candidate effector genes in the genome of *C. fulvum*, only *Ecp11-1* was found to be recently duplicated. In contrast, as shown in **Chapter 5**, nearly half of the candidate effector genes in the genome of *E. necator* were predicted to be the result of duplication events. Similar high rates of candidate effector gene duplication are also observed in other powdery mildew fungi (Menardo et al. 2017; Müller et al. 2019), indicating that this is a common feature within Erysiphales. However, the benefits of high rates of effector gene duplication in powdery mildew fungi and other fungal pathogen remain elusive. One possibility is that the dosage effect caused by an increase in copy number benefits the pathogen, for example in overcoming inhibitory proteins in the host. On the other hand, recognition in the host of duplicated effectors could be disadvantageous for the pathogen, because many copies of the effector genes would need to be mutated to avoid effector-triggered immunity. Recent studies have been revealing important clues about the evolution of fungal effector genes. Notably, the ongoing debate about the origin of species-specific effector genes has been shifting away from the hypothesis of *de novo* origin, towards the hypothesis of gene duplication followed by heavy accumulation of mutations leading to great loss of sequence similarity (Seong and Krasileva 2023). Therefore, one possibility is that duplication of effector genes in powdery mildew fungi provides the means to increase beneficial functional diversity of effectors. Notably, as reported in **Chapter 5**, there are two major clades of RNase-like candidate effectors in *E. necator* that are likely the result of gene duplications followed by fast divergence. The two lineages could be clearly distinguished by protein sequence length, similarity, and presence/absence of a conserved microbial RNase-like domain. These observations lead to the assumption that these two clades also differ in molecular function, which remains to be discovered. In future studies, valuable insights about the evolution of effector genes within Erysiphales can be obtained by performing large-scale comparative evolutionary analyses of candidate effectors from different powdery mildew species. Moreover, important information would be obtained by analyzing the transcriptome of *E. necator* during interaction with grape to estimate transcriptional activity of copies of candidate effector genes, and to have a narrower set of target genes important for pathogenicity, e.g., those highly induced during infection, before starting with functional studies to provide clues about their function.

8.10 Extensive diversity of powdery mildew mitochondrial genomes

Next to describing the nuclear genome of *E. necator*, my studies as part of this dissertation extended to assembling and comparing the mitochondrial (mt) genome of this pathogen as well as of another three powdery mildew species. As discussed in **Chapter 6**, the bicistronic expression of the gene *atp6* and *nad3* in *E. necator* is a notable feature of its mt genome, that is potentially unique to powdery mildew fungi. Another intriguing observation is the precited high copy number of mtDNA per cell in *E. necator*, varying between 124 and 322. In comparison, the mtDNA copy number per haploid cell was predicted to vary between 18 and 80 in *Saccharomyces cerevisiae* (De Chiara et al. 2020). In the reptile fungal pathogen *Nannizziopsis barbatae*, less than 10 copies of the mtDNA per haploid cell were estimated (Powell et al. 2023). These observations indicate that *E. necator* has an unusually high number of mt genomes per cell, which could be related to its obligate biotrophic lifestyle. Future studies to investigate the mtDNA copy number variation between isolates of *E. necator* or other powdery mildew fungi using whole-genome sequencing or real-time PCR could reveal an association between mtDNA copy number and a particular phenotype. Notably, increase in mtDNA copy number could be associated with increase in tolerance to fungicides that target mitochondrial respiration.

As shown in **Chapter 7**, the mt genomes of powdery mildew fungi have a conserved arrangement and orientation of the 13 core mt protein-coding genes commonly found in other fungal mt genomes. This observation indicates that no rearrangement of the mt genomes occurred throughout the evolutionary history of Erysiphales. In contrast, comparative analyses revealed remarkable diversity on the number and size of mt introns, which contributed significantly to the notable difference in size observed among the mt genomes of the four powdery mildew species. Overall, the insertion sites of the introns were well conserved, in accordance to the common assumption that mt introns are acquired through activity of homing endonucleases that recognize and insert themselves in specific sites in the mt DNA (Goddard and Burt 1999; Gogarten and Hilario 2006). Thus, the origin of introns in fungal mt genomes is typically attributed to horizontal transfer of homing endonucleases, which can also explain the large variation of mt intron content among powdery mildew species. Another unusual aspect of the mt genomes of some powdery mildew fungi is the presence of GC-rich islands in non-coding regions. However, the driver and potential benefits of these islands, such as better DNA stability, of high GC content in the mt genomes are unknown. Future studies can help elucidate the origin, evolution, and impact of introns and GC-rich islands in the mt genomes of powdery mildew fungi.

8.11 General conclusion

Taken together, the results presented in this dissertation reveal new evolutionary aspects of the *C. fulvum* genome and provide new insights on the importance of genomic structural variations in overcoming host resistance in fungal plant pathogens. Furthermore, the work herein illuminates higher-order genomic architectural features of the nuclear and mitochondrial genomes of *E. necator* and unearthed previously unknown genomic features associated with its biology. Finally, the high-quality genomes of these two pathogens are a valuable resource for future work, including population genomics, functional studies, and epigenetic modifications in *C. fulvum* and *E. necator*, that will further illuminate the evolutionary processes underlying adaptation of fungal pathogens.

8.12 References

Barragan, A. C., Latorre, S. M., Malmgren, A., Harant, A., Win, J., Sugihara, Y., Burbano, H. A., Kamoun, S., and Langner, T. 2024. Multiple horizontal mini-chromosome transfers drive genome evolution of clonal blast fungus lineages. bioRxiv. :2024.02.13.580079

Caspeta, L., Chen, Y., Ghiaci, P., Feizi, A., Buskov, S., Hallström, B. M., Petranovic, D., and Nielsen, J. 2014. Altered sterol composition renders yeast thermotolerant. Science. 346:75–78

De Chiara, M., Friedrich, A., Barré, B., Breitenbach, M., Schacherer, J., and Liti, G. 2020. Discordant evolution of mitochondrial and nuclear yeast genomes at population level. BMC Biol. 18:49

De Wit, P. J., Van Der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., Bahkali, A. H., Beenen, H. G., Chettri, P., Cox, M. P., and others. 2012. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. PLoS Genet. 8:e1003088

Dean, C., and Dresbach, T. 2006. Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. Trends Neurosci. 29:21–29

Debieu, D., Corio-Costet, M.-F., Steva, H., Malosse, C., and Leroux, P. 1995. Sterol composition of the vine powdery mildew fungus, *Uncinula necator*: comparison of triadimenol-sensitive and resistant strains. Phytochemistry. 39:293–300

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., and 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

Goddard, M. R., and Burt, A. 1999. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci U A. 96:13880–13885

Gogarten, J. P., and Hilario, E. 2006. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. BMC Evol. Biol. 6:94

Gozashti, L., Hartl, D. L., and Corbett-Detig, R. 2023. Universal signatures of transposable element compartmentalization across eukaryotic genomes. bioRxiv. :2023–10

Hall, L. E., Mitchell, S. E., and O'Neill, R. J. 2012. Pericentric and centromeric transcription: a perfect balance required. Chromosome Res. 20:535–546

Henikoff, S., Ahmad, K., and Malik, H. S. 2001. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. Science. 293:1098–1102

Hosokawa, M., and Satoh, T. 2001. Measurement of Carboxylesterase (CES) Activities. Curr. Protoc. Toxicol. 10:4.7.1 - 4.7.14

Ingolia, N. T., Hussmann, J. A., and Weissman, J. S. 2019. Ribosome profiling: global views of translation. Cold Spring Harb. Perspect. Biol. 11:a032698

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K. C. H., McGuire, B., Gubler, W. D., Walker, M. A., and Cantu, D. 2014. Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. BMC Genomics. 15:1081

Komluski, J., Habig, M., and Stukenbrock, E. H. 2023. Repeat-Induced Point Mutation and Gene Conversion Coinciding with Heterochromatin Shape the Genome of a Plant-Pathogenic Fungus. Mbio. :e03290-22

Lees, N., Skaggs, B., Kirsch, D., and Bard, M. 1995. Cloning of the late genes in the ergosterol biosynthetic pathway of *Saccharomyces cerevisiae*-A review. Lipids. 30:221–226

Loeffler, R. T., Butters, J. A., and Hollomon, D. W. 1992. The sterol composition of powdery mildews. Phytochemistry. 31:1561–1563

Menardo, F., Praz, C. R., Wicker, T., and Keller, B. 2017. Rapid turnover of effectors in grass powdery mildew (*Blumeria graminis*). BMC Evol. Biol. 17:1–14

Müller, M. C., Kunz, L., Graf, J., Schudel, S., and Keller, B. 2021. Host adaptation through hybridization: genome analysis of triticale powdery mildew reveals unique combination of lineage-specific effectors. Mol. Plant. Microbe Interact. 34:1350–1357

Müller, M. C., Praz, C. R., Sotiropoulos, A. G., Menardo, F., Kunz, L., Schudel, S., Oberhänsli, S., Poretti, M., Wehrli, A., Bourras, S., Keller, B., and Wicker, T. 2019. A chromosome-scale genome assembly reveals a highly dynamic effector repertoire of wheat powdery mildew. New Phytol. 221:2176–2189

Oakeshott, J., Claudianos, C., Campbell, P., Newcomb, R., and Russell, R. 2005. Biochemical genetics and genomics of insect esterases. Pages 309–381 in: Biochemical genetics and genomics of insect esterases, Elsevier, Oxford, Amsterdam.

Oakeshott, J., Claudianos, C., Russell, R., and Robin, G. 1999. Carboxyl/cholinesterases: a case study of the evolution of a successful multigene family. Bioessays. 21:1031–1042

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., Condon, B. J., Copeland, A. C., Dhillon, B., Glaser, F., and others. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 8:e1003037

Powell, D., Schwessinger, B., and Frère, C. H. 2023. Whole-mitochondrial genomes of *Nannizziopsis* provide insights in evolution and detection. Ecol. Evol. 13:e9955

Seong, K., and Krasileva, K. V. 2023. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. Nat. Microbiol. 8:174–187

Speicher, S., García-Alonso, L., Carmena, A., Martín-Bermudo, M. D., de la Escalera, S., and Jimenez, F. 1998. Neurotactin functions in concert with other identified CAMs in growth cone guidance in Drosophila. Neuron. 20:221–233

Sun, M., Xing, G., Yuan, L., Gan, G., Knight, D., He, C., Han, J., Zeng, X., Fang, M., Boulianne, G. L., and others. 2011. Neuroligin 2 is required for synapse development and function at the *Drosophila* neuromuscular junction. J. Neurosci. 31:687–699

Westhorpe, F. G., and Straight, A. F. 2015. The centromere: epigenetic control of chromosome segregation during mitosis. Cold Spring Harb. Perspect. Biol. 7:a015818

Yang, J., Li, C., and Zhang, Y. 2021. Engineering of *Saccharomyces cerevisiae* for 24-Methylene-Cholesterol Production. Biomolecules. 11:1710

Zhang, H., and Dawe, R. K. 2012. Total centromere size and genome size are strongly correlated in ten grass species. Chromosome Res. 20:403–412

List of publications

Chapter 1 has been submitted for publication in the Journal of Genetics and Genomics as:

Zaccaron AZ and Stergiopoulos I. The dynamics of fungal genome organization and its impact on host adaptation and antifungal resistance.

Chapter 2 has been published as:

Zaccaron AZ, Chen LH, Samaras A, Stergiopoulos I. (2022). A chromosome-scale genome assembly of the tomato pathogen *Cladosporium fulvum* reveals a compartmentalized genome architecture and the presence of a dispensable chromosome. *Microbial Genomics*, *8*(4), 000819.

Chapter 3 has been published as:

Zaccaron AZ, Stergiopoulos I. (2024). Analysis of five near-complete genome assemblies of the tomato pathogen *Cladosporium fulvum* uncovers additional accessory chromosomes and structural variations induced by transposable elements effecting the loss of avirulence genes. *BMC Biology*, *22*(1), 25.

Chapter 4 to be submitted for publication as:

Zaccaron AZ, Sanchez JN, Chen LH, Stergiopoulos I. Transcriptome analysis of two isolates of the tomato pathogen *Cladosporium fulvum* uncovers genome-wide patterns of alternative splicing events during a host infection cycle.

Chapter 5 has been published as:

Zaccaron AZ, Neill T, Corcoran J, Mahaffee WF, Stergiopoulos I. (2023). A chromosome-scale genome assembly of the grape powdery mildew pathogen *Erysiphe necator* reveals its genomic architecture and previously unknown features of its biology. *MBio*, *14*(4), e00645-23.

Chapter 6 has been published as:

Zaccaron AZ, Stergiopoulos I. (2021). Characterization of the mitochondrial genomes of three powdery mildew pathogens reveals remarkable variation in size and nucleotide composition. *Microbial Genomics*, *7*(12), 000720.

Chapter 7 has been published as:

Zaccaron AZ, De Souza JT, Stergiopoulos I. (2021). The mitochondrial genome of the grape powdery mildew pathogen *Erysiphe necator* is intron rich and exhibits a distinct gene organization. *Scientific Reports*, *11*(1), 13924.

Other publications not included in this dissertation are shown below. A complete list of my publications is shown in my Google Scholar profile (online link).

Zaccaron AZ, Stergiopoulos I. (2020). First draft genome resource for the tomato black leaf mold pathogen *Pseudocercospora fuligena*. *Molecular Plant-Microbe Interactions*, *33*(12), 1441-1445.

Mesarich CH, Barnes I, Bradley EL, de la Rosa, S, de Wit PJ, … Zaccaron AZ, Bradshaw RE. (2023). Beyond the genomes of *Fulvia fulva* (syn. *Cladosporium fulvum*) and *Dothistroma septosporum*: New insights into how these fungal pathogens interact with their host plants. *Molecular Plant Pathology*, *24*(5), 474-494.