Evolutionary and molecular basis of host adaptation in fungal plant pathogens: insights from *Zymoseptoria* pathosystems

Dissertation

in fulfilment of the requirements for the degree of "Dr. rer. nat." of the Faculty of Mathematics and Natural Sciences at the Christian Albrechts University of Kiel

submitted by

Wagner Calegari Fagundes

Kiel, December 2022

First examiner: Prof. Dr. Eva Holtgrewe Stukenbrock

Second examiner: Prof. Dr. Hinrich Schulenburg

Date of oral examination: February 14th, 2023

Contents

Summary

Host adaptation is considered a strong driver of pathogen evolution. With the ongoing development of modern agriculture, it is clear the importance in applying evolutionary analyses to access the risk of potential future pathogens, particularly in pathosystems that transition between wild and agricultural environments. Implementing population and functional genomics, transcriptome and infection assay approaches, we aimed to identify evolutionary and phenotypic patterns of host adaptation in fungal plant pathogens using the fungal pathogen *Zymoseptoria tritici* as a model of study. Unique collections of *Z. tritici* were isolated from wild (*Aegilops* spp.) and domesticated (*Triticum aestivum*) grass hosts, and different genomic and virulence-related aspects have been studied to identify key genomic traits playing a role in host adaptation of *Z. tritici* specimens in these different hosts. In **Chapter 1**, I describe methods to handle *Z. tritici* in laboratory and greenhouse conditions that are used throughout different experiments in the scope of this thesis. In **Chapter 2**, I focus on population genomics and virulence analyses of these host-diverging *Z. tritici* populations and reveal particular features of the *Aegilops*-infecting population that may have shaped its host adaptation and evolutionary history. In **Chapter 3**, I describe a new chromosome present in *Aegilops*-infecting *Z. tritici* isolates and propose its possible origin from introgression events with the closely related fungal species *Zymoseptoria ardabiliae*. Finally, in **Chapter 4**, I outline functional genetics analyses of candidate genes for host adaptation in the *Aegilops*-infecting *Z. tritici* population. As a foreseen impact, the results hereby generated are of relevance for translational evolutionary biology and plant pathology, highlighting the potential interaction between pathogens of wild and cultivated plants and a possible route of new pathogen lineages emergence.

Zusammenfassung

Die wichtigste Triebfeder der Evolution von Pathogenen ist die Anpassung an ihren Wirt. Im Hinblick auf die Weiterentwicklung der modernen Landwirtschaft wird deutlich, wie wichtig es ist, evolutionäre Analysen zur Bewertung des Risikos möglicher zukünftiger Krankheitserreger heranzuziehen. Das gilt insbesondere für Gruppen nahverwandter Pathogene, die sowohl Wirtspflanzen in natürlichen als auch in landwirtschaftlichen Ökosystemen infizieren. Mithilfe von Populations- und funktioneller Genomik, Transkriptom- und Infektions-Analysen haben wir evolutionäre und phänotypische Wirtsanpassungen bei pilzlichen Pflanzenpathogenen identifiziert. Das Pilzpathogen *Zymoseptoria tritici* war dabei unser Modellsystem. Wir haben eine einzigartige Sammlung von *Z. tritici* Isolaten generiert, in dem wir den Pilz aus den Blättern von wilden (*Aegilops* spp.) sowie domestizierten (*Triticum aestivum*) Gräsern isolieren konnten. Die Virulenz und die Genome dieser Isolate wurden untersucht mit dem Ziel essentielle Schlüsselmerkmale zu identifizieren, die an der Anpassung an die verschiedenen Wirte beteiligt sind. In **Kapitel 1** beschreibe ich die Methoden zur Handhabung von *Z. tritici* unter Labor- und Gewächshausbedingungen, die in verschiedenen Experimenten im Rahmen dieser Arbeit verwendet wurden. In **Kapitel 2** konzentriere ich mich auf Populationsgenomik und Virulenzanalysen der wirtsspezifischen *Z. tritici*-Populationen. Ich hebe besondere Merkmale der *Aegilops*-infizierenden Population hervor, die ihre Wirtsanpassung und Evolutionsgeschichte geprägt haben könnten. In **Kapitel 3** beschreibe ich ein neuartiges Chromosom, das in *Aegilops*-infizierenden *Z. tritici*-Isolaten vorkommt, und stelle die Hypothese auf, dass es durch Introgression mit der eng verwandten Pilzart *Zymoseptoria ardabiliae* entstanden ist. Im finalen **Kapitel 4** stelle ich funktionelle Analysen von Kandidatengenen vor. Diese Gene haben vermutlich eine wichtige Rolle für die Wirtsanpassung der *Aegilops*-infizierenden *Z. tritici*-Population. Die hier erzielten Ergebnisse stellen eine wichtige Verbindung zwischen Evolutionsbiologie und Phytopathologie her, da sie die mögliche Interaktion zwischen Pathogenen von Wild- und Kulturpflanzen demonstrieren und damit einen möglichen Weg für die Entstehung neuer Pflanzenpathogene aufzeigen.

General

Introduction

General Introduction

During 1960s at the Boyce Thompson Institute for Plant Research, George McNew formalized the disease triangle concept to understand the relationship of different factors that contribute for a plant disease epidemy: host, pathogen and environment (McNew, 1960). In his work, these factors were broadly defined as "the inherent susceptibility of the host, the inoculum potential of the parasite, and the impact of the environment on parasitism and pathogenesis" (McNew, 1960; Scholthof, 2007). Over the years, the disease triangle has been widely employed in fields on plant pathology and host-pathogen interactions in order to better elucidate plant diseases emergence and spread (Agrios, 2005; Schumann, 1991). Considering the components that influence a plant disease epidemic (host, pathogen and environment), selection of hosts, particularly crops, have been the one most manipulated by humanity over the past 10,000 years of agricultural practices (Balter, 2007; McNew, 1960; Scholthof, 2007).

The impact of host domestication goes beyond plant health and yield production as it is strongly correlated with the adaptative evolution of associated pathogens. Host-driven selection is considered the strongest driver of pathogen's evolution, since a successful infection, colonization and completion of life cycle within the host is crucial for dispersal of plant pathogens and, therefore, their own fitness (Möller & Stukenbrock, 2017). The tight association between host and pathogen evolution can be described by the "trench-warfare" and "arms-race" evolution (Brown & Tellier, 2011; Möller & Stukenbrock, 2017; Stukenbrock & McDonald, 2009; Tellier, Moreno-Gámez, & Stephan, 2014). In the trench-warfare scenario, multiple alleles at a co-evolving locus are maintained in intermediate frequencies in a population by balancing selection (Möller & Stukenbrock, 2017; Tellier et al., 2014). Under the arms-race scenario, there is a continuous replacement and fixation of advantageous alleles at a co-evolving locus in a population mediated by positive selection, generating recurrent selective sweeps (Möller & Stukenbrock, 2017; Tellier et al., 2014). In these evolutionary dynamics, hosts evolve different defense mechanisms to target specific pathogen genotypes while pathogens evolve virulence strategies that can evade host recognition leading to a constant coevolutionary fight (Jones & Dangl, 2006). Genes playing a role in virulence may diversity at a high rate in repeat-rich genome compartments and thereby evolve new virulence specificities faster (Dong, Raffaele, & Kamoun, 2015; Feurtey et al., 2020). Virulence-related genes in pathogen genomes generally encode small, secreted proteins with often no sequence homology known as effectors (Lo Presti et al., 2015). These small molecules can modulate host cell structure and plant metabolism and therefore evade host defense responses

through several and diverse functions, sometimes in a highly specific molecular dialogue (Raffaele & Kamoun, 2012; Stergiopoulos & de Wit, 2009). Besides specific effectors, different classes of proteins (i.e. plant cell wall–degrading enzymes - PCWDEs and carbohydrate degradation enzymes – CAZymes) and metabolites (i.e. toxins) play also an important role in such host-pathogen interactions and therefore can be also under selective pressure imposed by hosts (Lo Presti et al., 2015). Pathogens that specialize to specific hosts or host genotypes may become genetically differentiated from their source population even when they occur in sympatry (i.e. at the same location) with an overlapping range of dispersal, potentially leading to a so-called sympatric speciation (Stukenbrock & McDonald, 2008). Therefore, different host species can be considered divergent ecological niches for plant pathogens, in which host specialization sets the stage for adaptive evolution and divergent selection on pathogen populations and species (Gladieux et al., 2011; McCoy, 2003; Silva et al., 2012).

Understanding how species and populations evolve is one of the main goals in many host-pathogen genomic studies. The neutral theory of molecular evolution (Kimura, 1968) states that most of the variation within and between species is neutral or slightly deleterious (i.e. it does not affect the fitness of the organisms) and that new mutations that arise within populations (or species) change their frequency due to genetic drift (Kimura, 1968; King & Jukes, 1969; Nielsen, 2005; Ohta, 1973). In this case, the levels of polymorphism within species and the rate of divergence between species are governed by the effective population size (Ne) and the neutral mutation rate (μo), respectively (Biswas & Akey, 2006). Controversially, the Darwinian selection theory states that most of the genomic variation observed will affect the fitness of an organism to survive, and therefore will be prone to natural selection (Gillespie, 1994). With the growing technological advances in both computational tools and high-throughput sequencing, population and comparative genomics studies have focused in the identification of genomic regions shaped by positive selection due to its association with novel functions and adaptive evolution (Nielsen, 2005). In host-pathogen interactions, comparative and population genomics approaches can help to elucidate the genetic basis of variation found between different host and pathogen populations in terms of genomic and phenotypic structure as well as the genes underlying adaptation to divergent niches, besides recapturing the evolutionary history of a species (Grünwald, McDonald, & Milgroom, 2016; Stukenbrock, 2013). For plant pathogens, the identification of genomic regions under selection can further pinpoint candidate genes involved in pathogenicity (e.g. effectors), fungicide/antibiotic resistance and host specialization, all of which are major goals in the study and development of more sustainable control strategies (Grünwald et al., 2016; Stukenbrock & Bataillon, 2012; Stukenbrock & McDonald, 2008). Therefore, population and comparative

genomic approaches are powerful tools to understand the origin and population history of species besides supporting the assessment of future pathogens emergence and spread (de Vries, Stukenbrock & Rose, 2020; Stukenbrock & McDonald, 2008).

Among plant pathogens, fungal specimens have demonstrated a large array of genomic variation even within species, in which the composition and organization of the genome architecture can be very distinct (Feurtey & Stukenbrock, 2018; Möller & Stukenbrock, 2017; Stajich, 2017). The adaptive genome evolution of fungal pathogens is underlined by different genomic mechanisms as duplication, deletions, chromosomal rearrangements, new mutations as well as hybridizations and horizontal gene transfers (HGT) – all of which may contribute to an enhanced adaptive potential upon colonization of a novel host (Feurtey & Stukenbrock, 2018; Friesen et al., 2006; Gladieux et al., 2011; Plissonneau et al., 2017). Among these mechanisms, hybridization and introgression/introgressive hybridization (recurrent backcrossing between hybrids and parental species) have been shown to play an important role in many clades of fungal plant pathogens, in which under appropriate conditions may generate new adaptive traits in existing species or give rise to new pathogen lineages (Möller & Stukenbrock, 2017). Some intriguing examples of how divergent genomic mechanisms may contribute to host adaptation and speciation in fungal pathogens are already described in the literature (Couch et al., 2005; Depotter, Seidl, van den Berg, Thomma, & Wood, 2017; Gladieux et al., 2018; Islam et al., 2016; Leroy et al., 2016; Menardo et al., 2016; Yoshida et al., 2016; Zhong et al., 2016). A population genomics study on the cereal mildew pathogen *Blumeria graminis* revealed that the genome of the forma specialis species *B. graminis* f. sp. triticale (lineage that infects triticale) is a hybrid of two other species, *B. graminis* f. sp. tritici and *B. graminis* f. sp. secalis (Menardo et al., 2016). The triticale lineage is characterized by few genomic regions with polymorphisms that are only present in this species, and therefore consistent with a recent hybridization event (Menardo et al., 2016; Möller & Stukenbrock, 2017). The authors suggested that the emergence and spread of *B. graminis* f. sp. triticale occurred only around 30 years ago, coinciding with the domestication and use of triticale (Menardo et al., 2016). Other comparative genomics studies on the prominent blast pathogen *Magnaporthe oryzae* revealed that gain and losses of genes associated with transposable elements and excess of non-synonymous substitutions may have shaped the host specialization of this fungal species in different host grasses-subgroups (Oryza-Setaria vs. Avena-Triticum) (Yoshida et al., 2016). Besides that, it seems that host shifting from wild plants to rice and speciation of *M. oryzae* have been associated with rice domestication (Couch et al., 2005). These examples indicate how strong selections imposed by (new) hosts may facilitate host shifts and, in special cases, ecological speciation in fungal pathogens (Giraud, Gladieux, & Gavrilets, 2010; Gladieux et al., 2011).

Apart from plant health and economical importance, *Zymoseptoria tritici* and closely related species have served as powerful model systems for evolutionary and molecular genetic studies of crop-infecting pathogens. *Zymoseptoria tritici* is one of the most prominent pathogens in wheat throughout the world, causing the Septoria tritici Blotch (STB) disease and leading to large yield losses annually (Fones & Gurr, 2015; Torriani et al., 2015). Studies of the evolutionary history of *Z. tritici* indicate it has originated from a wild grass-associated ancestor in the Fertile Crescent region in the Middle East around 11,000 years ago, coinciding with the domestication of its host wheat (Banke, Peschon, & McDonald, 2004; Stukenbrock, Banke, Javan-Nikkhah & McDonald, 2007). Since then, *Z. tritici* is considered to be a highly specialized and genetically diverse wheat pathogen while the closest known relatives to *Z. tritici*, *Z. pseudotritici* and *Z. ardabiliae* are found associated to different wild grasses in the Middle East (Stukenbrock et al., 2012; Stukenbrock et al., 2011). Population genomic studies have confirmed high levels of genetic variation in populations of *Z. tritici*, frequent sexual recombination and extensive gene flow at regional as well as continental scales (Linde, Zhan, & McDonald, 2002). Evidence also suggests that gene flow and interspecific hybridization occur between different species of *Zymoseptoria* and can contribute to the formation of new genetic diversity and maybe even be advantageous during host jumps (Feurtey, Stevens, Stephan, & Stukenbrock, 2019; Stukenbrock, Christiansen, Hansen, Dutheil, & Schierup, 2012). Besides evolutionary aspects, *Z. tritici* has one of the best assembled fungal genomes, composed by 13 core chromosomes and 8 accessory chromosomes fully sequenced on the reference isolate IPO323 (Goodwin et al., 2011). Additionally, *Z. tritici* is amenable to culturing methods *in vitro*, and infection assays can be replicated in greenhouse or phytochamber conditions in a relative short time (Fagundes, Haueisen, & Stukenbrock, 2020). These characteristics make *Z. tritici* and closely-related *Zymoseptoria* species powerful resources to understand the impact of host-driven selective pressures on fungal evolution transitioning between wild and agricultural ecosystems.

Scope of this Thesis

Host adaptation is a crucial step towards compatible reactions in new, emergent pathosystems (De Vienne, Hood, & Giraud, 2009; Giraud et al., 2010). Fungal diseases, besides emerging at an increasing rate on several different host plants, serve as great examples of speciation due to host-driven selective pressures (Jones et al., 2008). During the development of agriculture and crop domestication, changes and selection of plant genotypes from wild progenitors led simultaneously to the emergence of new pathogens and variation in pre-existing pathogen populations changing dramatically their evolutionary potential (Stukenbrock & McDonald, 2008). Therefore, the analysis of evolutionary processes involving host adaptation is essential to access the risk of potential future pathogens.

The overall aim of this thesis has been to gain insights into the evolutionary, molecular and phenotypic patterns of host adaptation using host-divergent *Z. tritici* isolates as models of study. Unique collections of *Z. tritici* isolates have been collected from uncultivated wild grasses belonging to the *Aegilops* genus as well as from the cultivated common wheat (*Triticum aestivum*) in Iran. The analyses described here focus on these two collections - *Aegilops* and wheat-infecting *Z. tritici* - given the (i) resources of *Z. tritici* as a model organism for genetic and evolutionary studies; (ii) the known *Z. tritici* host adaptation and speciation following wheat domestication (Stukenbrock et al., 2007) and; (iii) the contribution of *Aegilops* spp. to domestication of the hexaploid bread wheat (Glémin et al., 2019; Salamini, Özkan, Brandolini, Schäfer-Pregl, & Martin, 2002). The four chapters comprising this thesis address the following research questions:

- How to handle *Z. tritici* under laboratory conditions and make experiments more comparable?

Chapter I is a methodological chapter which outlines important details for practical handling of *Z. tritici* under laboratory and greenhouse/phytochamber conditions. *Zymoseptoria tritici* can show extensive genetic variability and morphology plasticity that can make experimental studies and comparability of results obtained in different laboratories challenging. With the aim of standardizing experimental methods, here we describe steps ranging from isolating, handling and properly storing *Z. tritici* specimens to the inoculation and analyses of *Z. tritici* virulence *in planta*. We also discuss important points that should be addressed in order to reduce variability between experiments and to avoid the occurrence of aneuploidy and mutation accumulation in laboratory strains.

- What are the population genomic signatures of host adaptation between *Z. tritici* isolates infecting wild (*Aegilops* spp.) and domesticated (wheat) hosts?

In **Chapter II**, I have addressed this question by analyzing the genomes of host-diverging *Z. tritici* isolates with the aim of identifying and characterizing host-specific populations. I applied population genomics approaches to understand how the *Aegilops*- and wheat-infecting *Z. tritici* collections genetically differ and to access signatures of divergence that may have shaped their evolutionary history. In combination with the genomics analyzes, I also performed experimental infection assays and microscopy analyses to assess and characterize the host specificity of the sequenced *Z. tritici* isolates. I observed a clear separation between the host-diverging populations and could show that *Z. tritici* isolates collected from *Aegilops* spp. could only infect their respective host species and not *T. aestivum*. Based on genome screenings, I could also identify genomic regions under selection that harbor potential candidate genes involved in virulence and host specialization. Together with other aspects, my findings suggest that the *Aegilops*-infecting *Z. tritici* population represents a unique host-specific lineage that sympatrically diverged from the wheat one and further hypothesize that this lineage may be at the first stages of incipient speciation.

- Do host-diverging *Z. tritici* isolates differ in genome architecture and structure?

In **Chapter III**, I used comparative genomics analyses to illustrate how the genome architecture of host-diverging *Z. tritici* isolates is shaped. My aim in this chapter was to observe if host-diverging *Z. tritici* isolates have similar genomic structure in terms of organization (e.g., genome compartmentalization and gene synteny) and epigenetic landscape, and if these similarities can also be found in other closely related *Zymoseptoria* species. Using long-read sequencing technology (PacBio), transcriptome sequencing *in vitro* and *in planta,* and chromatin immunoprecipitation followed by sequencing (ChIP-seq), I observed interesting findings in the analyzed *Aegilops*-infecting *Z. tritici* isolate including the occurrence of a new accessory chromosome. Intriguingly, this new accessory chromosome does not occur in any wheat-infecting *Z. tritici* isolate analyzed, particularly in the reference wheat-infecting *Z. tritici* isolate IPO323, but show to be syntenic with a chromosome in the closely related species *Z. ardabiliae*. Analyses on the syntenic chromosome in *Z. ardabiliae* also demonstrated that this chromosome shows hallmarks of an accessory chromosome, but lacks the characteristic enrichment of heterochromatin histone methylation marks (e.g. H3K27me3). I further observed a higher activation of Transposable Elements (TEs) and a lower RIP (Repeat-Induced Point mutations) index in the syntenic *Z. ardabiliae* chromosome that may suggest a distinct chromosome formation. Along with other results, my findings lead us

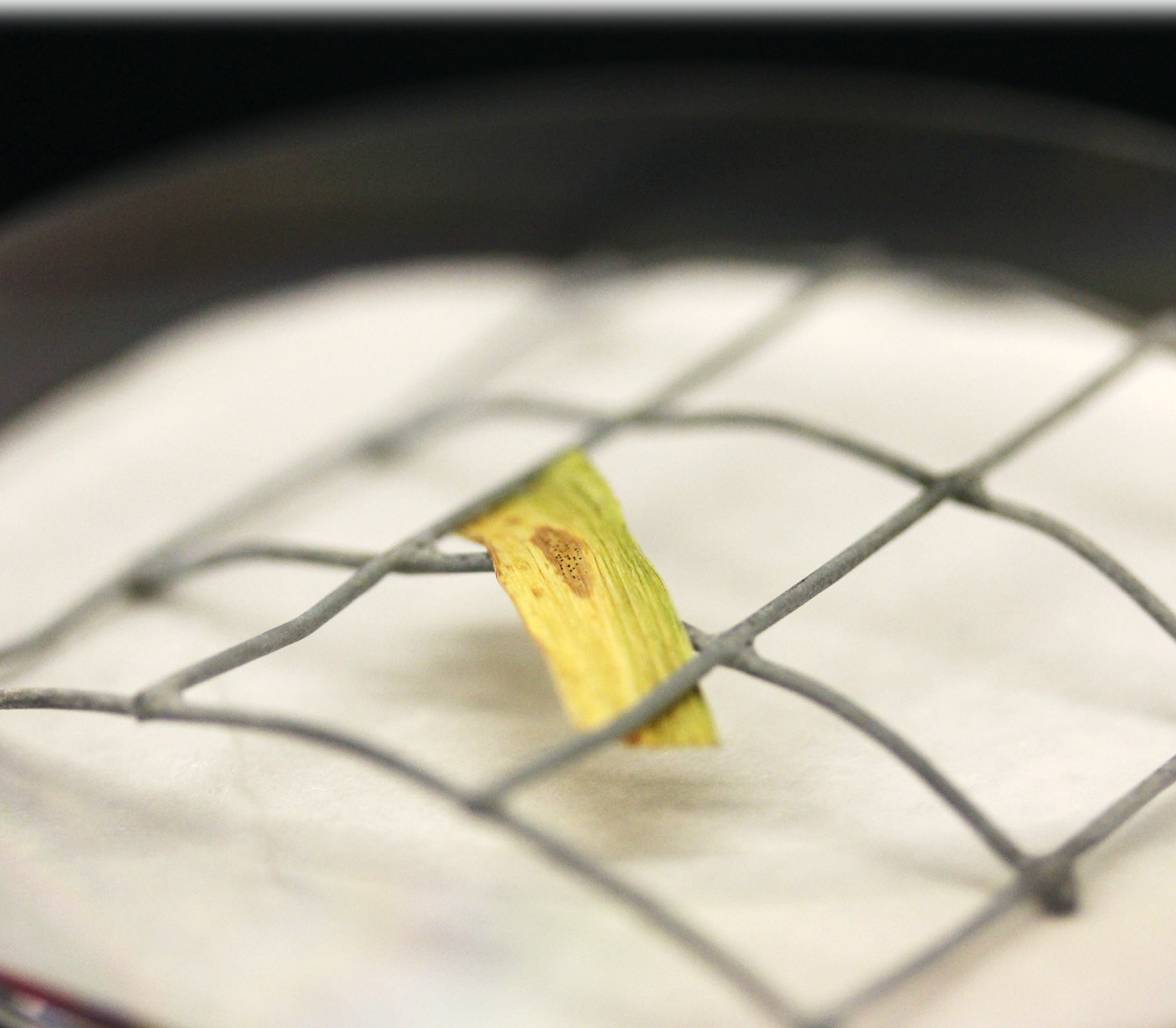
to the hypothesis that introgression events happened between *Aegilops*-infecting *Z. tritici* isolates and *Z. ardabiliae* with the transfer of an accessory chromosome between the two species. I also discuss possible introgression directions and the findings that may support each one of these direction scenarios.

- Are candidate effector genes potentially under selection involved in host adaptation and virulence specificity in *Aegilops*-infecting *Z. tritici* isolates?

In **Chapter IV**, I aim to answer this question by the analyses of *Aegilops*-infecting *Z. tritici* mutants lacking a genomic region potentially under positive selection. Here my goal was to generate mutants of a *Aegilops*-infecting *Z. tritici* strain (Zt495) lacking different portions of an identified genomic region that have been potentially under a selective sweep (i.e. recent positive selection) and test these mutants *in planta* and *in vitro*. This region, characterized in a previous study (Chapter II of this thesis), has a size of around 30 kb, it is located on chromosome 7 of the *Z. tritici* IPO323 reference genome (Goodwin et al., 2011) and harbors 13 predicted genes and 3 Transposable Elements (TEs). Among the predicted genes, 3 of them are candidate effector genes and 1 gene encodes a carbohydrate-degrading enzyme (CAZyme). Considering the importance of these gene categories in host-pathogen interactions, we hypothesize that this genomic region under selection, and particularly the 3 effector genes, play a role in the pathogenicity and host specificity of Zt495 on its host *Aegilops cylindrica*. We further compared this genomic region between different *Z. tritici* isolates and observed interesting findings as rearrangements and inversions. Together with other aspects, I expect that the results hereby generated will give further insights into to host adaptation mechanisms of *Z. tritici* and the possible virulence determinants in *Aegilops* hosts.

Chapter I

Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow



Chapter I

Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow

Published manuscript

Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. (2020). Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow. Current Protocols in Microbiology, 59(1), e128.

Abstract

The fungus *Zymoseptoria tritici* is one of the most devastating pathogens of wheat. Aside from its importance as a disease-causing agent, this species has emerged as a powerful model system for evolutionary genetic studies of crop-infecting fungal pathogens. *Z. tritici* exhibits exceptionally high levels of genetic and phenotypic diversity as well as morphological plasticity, which can make experimental studies and comparability of results obtained in different laboratories, e.g., from infection assays, challenging. Therefore, standardized experimental methods are crucial for research on *Z. tritici* biology and the interaction of this fungus with its wheat host. Here, we describe a suite of well-tested and optimized protocols ranging from isolation of *Z. tritici* field specimens to analyses of virulence assays under controlled conditions. Several biological and technical aspects of working with *Z. tritici* under laboratory conditions are considered and carefully described in each protocol.

Basic Protocol 1: Purification of *Z. tritici* field isolates from leaf material

Basic Protocol 2: Molecular identification of *Z. tritici* isolates

Support Protocol 1: Rapid extraction of *Z. tritici* genomic DNA

Support Protocol 2: Extraction of high-quality *Z. tritici* genomic DNA

Basic Protocol 3: *In vitro* culture and long-term storage of *Z. tritici* isolates

Basic Protocol 4: Analysis of *Z. tritici* virulence in wheat

Support Protocol 3: Preparation of *Z. tritici* inoculum

Introduction

The ascomycete fungus *Zymoseptoria tritici* (synonym: *Mycosphaerella graminicola*) is the causal agent of the foliar disease Septoria tritici blotch (STB), one of the most devastating diseases of wheat. This plant pathogenic fungus infects wheat worldwide and can cause up to 50% yield losses per year in the major European Union wheat-growing countries (Fones & Gurr, 2015). *Z. tritici* is a member of the Dothideomycetes, a class of fungi comprising several thousand species, including many plant pathogens (Goodwin, Waalwijk, & Kema, 2004, 2011). Members of this order are generally heterothallic, with sexual (teleomorph) and asexual (anamorph) stages, and plant-infecting species are found on a wide range of monocot and dicot hosts (Goodwin et al., 2004). *Z. tritici* has a hemibiotrophic lifestyle, with a prolonged biotrophic phase followed by a necrotic phase where plant cell death occurs and necrotic lesions develop (Ponomarenko, Goodwin, & Kema, 2011).

Population genetic and genomic studies based on different types of genetic and genomic markers have documented high levels of genetic variation in populations of *Z. tritici*, even within single fields and individual lesions (Banke, Peschon, & McDonald, 2004; Hartmann, McDonald, & Croll, 2018; Jürgens, Linde, & McDonald, 2006; Linde, Zhan, & McDonald, 2002; Zhan, Pettway, & McDonald, 2003). Moreover, the species history has been well documented, and it was shown that speciation of the wheat pathogen coincided with domestication of the wheat host (Stukenbrock, Banke, Javan-Nikkhah, &Mc- Donald, 2007). The high-quality genome sequence of the reference isolate IPO323 comprises 13 core and eight accessory chromosomes that are fully assembled from telomere to telomere (Goodwin et al., 2011). This reference genome has provided a unique resource for several comparative genome studies as well as transcriptomic and epigenomic analyses (Badet, Oggenfuss, Abraham, McDonald, & Croll, 2020; Haueisen et al., 2019; Möller et al., 2019).

Given the economic importance of *Z. tritici*, there is an urgent need to better understand the biology of this fungus. Here, we present some considerations for field collections and experimental work with *Z. tritici*, and we discuss the need for standardized protocols for research on this important pathogen. First of all, experimental work with *Z. tritici* has to account for the excessive amounts of genetic and phenotypic diversity. The high genetic diversity found among and within *Z. tritici* populations is reflected in phenotypic diversity. Virulence of *Z. tritici* is quantitative and determined by the extent of host damage and the density of asexual fruiting structures, or so-called pycnidia. Necrotic leaf area and pycnidia density on susceptible wheat cultivars can vary significantly between isolates including progenies originating from the same sexual cross (Abrinbana, 2017; Habig, Quade, & Stukenbrock, 2017; Stewart &McDonald, 2014; Zhan et al.,

2005). Moreover, variation among replicate leaves can be substantial. Comparative analyses of virulence phenotypes between *Z. tritici* isolates or mutant strains therefore rely on robust methods to quantify virulence traits. Automated leaf image–based analyses have been established as a more precise measure for virulence quantification that ideally should be applied for *in planta* studies of *Z. tritici* (Karisto et al., 2018; Stewart & McDonald, 2014; Stewart et al., 2016).

Typical STB symptoms can appear on leaves of seedling to adult wheat plants and are usually easy to spot. However, other wheat diseases with similar symptoms can co-occur in the same field, including the second major *Septoria* disease, *Septoria nodorum* blotch, caused by the Dothideomycete fungus *Parastagonospora nodorum* (synonyms: *S. nodorum*, *Phaeosphaeria nodorum*) (Fig. 1A). Co-occurring pycnidia of the two species have been reported, and *P. nodorum* may occasionally be confounded with *Z. tritici* (Fig. 1B) in the field (Eyal, Scharen, Prescott, & van Ginkel, 1987). To this end, collections of *Z. tritici* in the field should ideally be accompanied by species confirmation of the retrieved isolates.

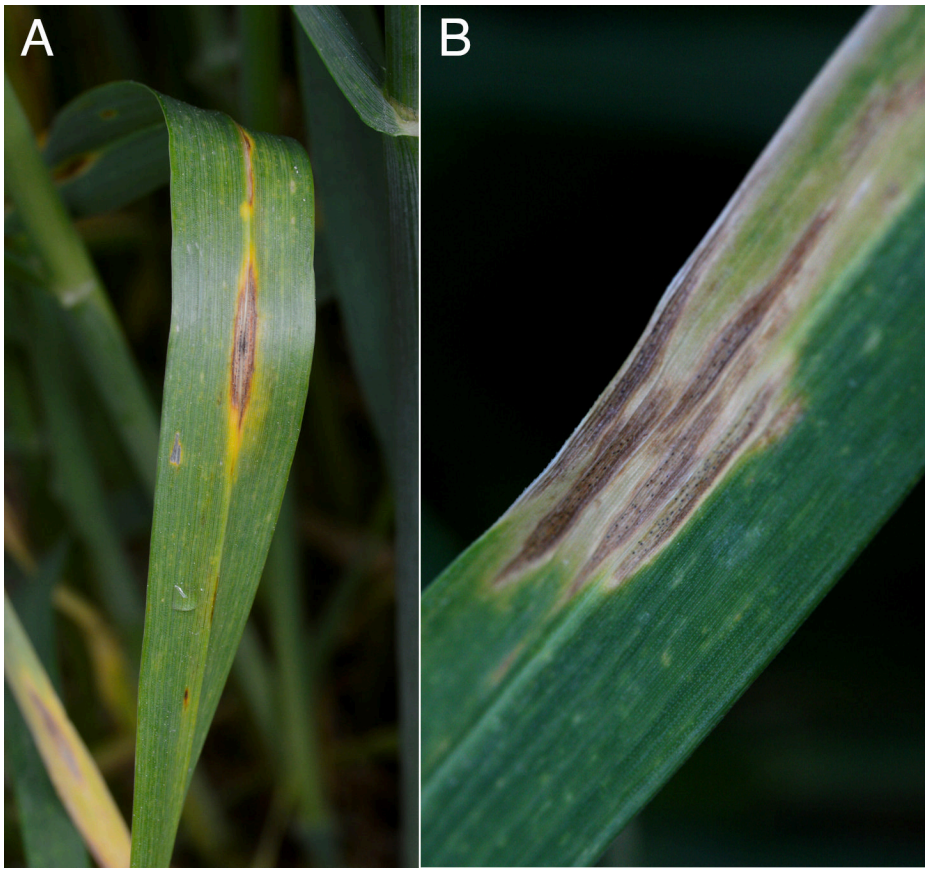


Figure 1. *Septoria* leaf blotch diseases caused by *P. nodorum* and *Z. tritici*. Depending on the environmental conditions, wheat cultivar, and disease stage, lesions due to *P. nodorum* **(A)** and *Z. tritici* **(B)** appear similar. Both pathogens cause "leaf blotch" diseases characterized by necrotic lesions with a brownish to tan color. Inside the lesions, small brown to black asexual pycnidia develop. Panel **(A)** courtesy of U. Adhikari, NC State University, USA.

Although *Z. tritici* is amenable to laboratory conditions, appropriate growth conditions and maintenance of fungal cultures are extremely important for reproducible experiments. *Z. tritici* is a pleomorphic fungus that can grow as yeast-like blastospores, filamentous hyphae, or chlamydospores (Francisco, Ma, Zwyssig, McDonald, & Palma-Guerrero, 2019). The underlying molecular and genetic bases for this variation in growth morphology are so far not fully understood; however, specific culture conditions may induce the shift from one to another morphotype. As mentioned above, the genome of *Z. tritici* comprises a set of accessory chromosomes that can be readily lost during mitotic cell divisions *in vitro* as well as *in planta* (Möller, Habig, Freitag, & Stukenbrock, 2018). Given this chromosome instability, it is important to maintain *Z. tritici* isolates under the right conditions for long-term storage and thereby avoid continuous and unnecessary mitotic growth, which may result in a population of aneuploid cells (cells with aberrant chromosome sets).

Considering the above-mentioned aspects and the use of *Z. tritici* as a model organism, the protocols described here outline important details for practical handling of *Z. tritici* under laboratory and greenhouse/phytochamber conditions. These protocols can serve either as a baseline for researchers initiating work with *Z. tritici* or as a reference to make experiments more comparable across different research groups. The protocols included here present a workflow "from the field to the laboratory bench" and describe isolation and confirmation of *Z. tritici* specimens from wheat samples (Basic Protocols 1 and 2 as well as Support Protocols 1 and 2), culture and long-term storage of *Z. tritici* isolates (Basic Protocol 3), and inoculation and analysis of *Z. tritici* virulence on wheat plants (Basic Protocol 4 and Support Protocol 3).

Basic Protocol 1: Purification of *Z. tritici* field isolates from leaf material

Population genetic analyses of *Z. tritici* at different spatial scales have revealed high levels of genetic variation even within single lesions (Linde et al., 2002). Depending on the specific research objectives, this variation should already be considered when sampling *Z. tritici* in the field. For a representative sampling of genetic variation, a hierarchical sampling design can be applied, in which infected wheat leaves are collected from individual sub-plots along a transect (McDonald, 1997). Most importantly, samples should be properly labeled with a code that allows derivation of the origin of the isolate and that considers the different levels of sampling (field, plot, sub-plot, plant, leaf, lesion).

Fresh leaves can immediately be used for *Z. tritici* isolation, or leaves can be mounted on paper sheets and stored for later isolations. Such "herbarium" specimens should be well dried and stored in a dry and dark place at room temperature, where they can be kept up to several months until isolations are performed. Due to the destructive nature of isolation, we advise recording information about the leaf by taking a photograph before it is used for isolation, numbering each leaf as well as each lesion on each leaf separately (Fig. 2A). In this way, it will later be possible to trace back to the original virulence phenotype, e.g., as lesion size, pycnidia density, and overall appearance of the leaf.

NOTE: We recommend performing the following method of isolation using aseptic techniques under a laminar flow hood and/or close to a Bunsen burner.

Materials

Sterile water
1.2% (w/v) sodium hypochlorite in aqueous solution
Tween® 20 (e.g., Roth, cat. no. 9127.1)
70% (v/v) ethanol
Leaf samples with necrotic lesions containing pycnidia
YMS agar plates with and without 50 µg/ml kanamycin (see recipe)

Sterile tweezers
Sterile Whatman paper or filter paper (disc-shaped to fit in petri dish)
Sterile petri dishes
Sterile metal grid or equivalent 50-ml Falcon tubes
Sterile scissors
Parafilm
Microbiological incubator with temperature control, 18°C
Dissecting microscope
Sterilized fine dissecting needles or fine syringe needles (e.g., insulin needles, Gr. 14, Braun, cat. no. 4657640) (see step 9)
Sterile toothpicks Sterile 1000-µl pipet tips (optional)

1. Using sterile tweezers, put a sterile Whatman or filter paper inside bottom of a sterile petri dish for each sample. Place a sterile metal grid or equivalent on top of it and wet paper with 2 to 3 ml sterile water.

The metal grid or something similar should fit inside the petri dish and allow you to position the samples to have them fully exposed to the humidity without touching the bottom (see Fig. 2B and step 7).

2. For each sample, prepare the following for leaf surface sterilization in 50-ml Falcon tubes: 1.2% sodium hypochlorite in aqueous solution with a few drops of Tween® 20, 70% ethanol with a few drops Tween® 20, and two tubes of sterile water.

To avoid mixing up samples, each lesion (that is, one leaf section) should be treated individually throughout the whole isolation procedure. We recommend conducting the surface sterilization process for the individual leaf sections in separate Falcon tubes.

3. Excise 3- to 4-cm sections from leaf material (leaf samples with necrotic lesions containing pycnidia) using sterile scissors and select sections with lesions and pycnidia.

4. Submerge excised leaf sections in the 1.2% sodium hypochlorite solution and wash for 3 min by gently inverting each tube.

5. Transfer leaves with sterile tweezers into 70% ethanol and wash for 15 s by gently inverting each tube.

6. Transfer leaves with sterile tweezers into the first tube of sterile water and wash leaves for ≥30s by gently inverting each tube. Repeat wash step in the second tube of sterile water.

Wash the leaves until the chlorine smell is gone.

7. Place surface-sterilized leaves onto the metal grid inside a petri dish (see step 1) using sterile tweezers, as shown in Figure 2B. Prepare one petri dish per leaf section.

Avoid allowing the leaves to touch the filter paper.
Label the dishes with sample ID and lesion number.

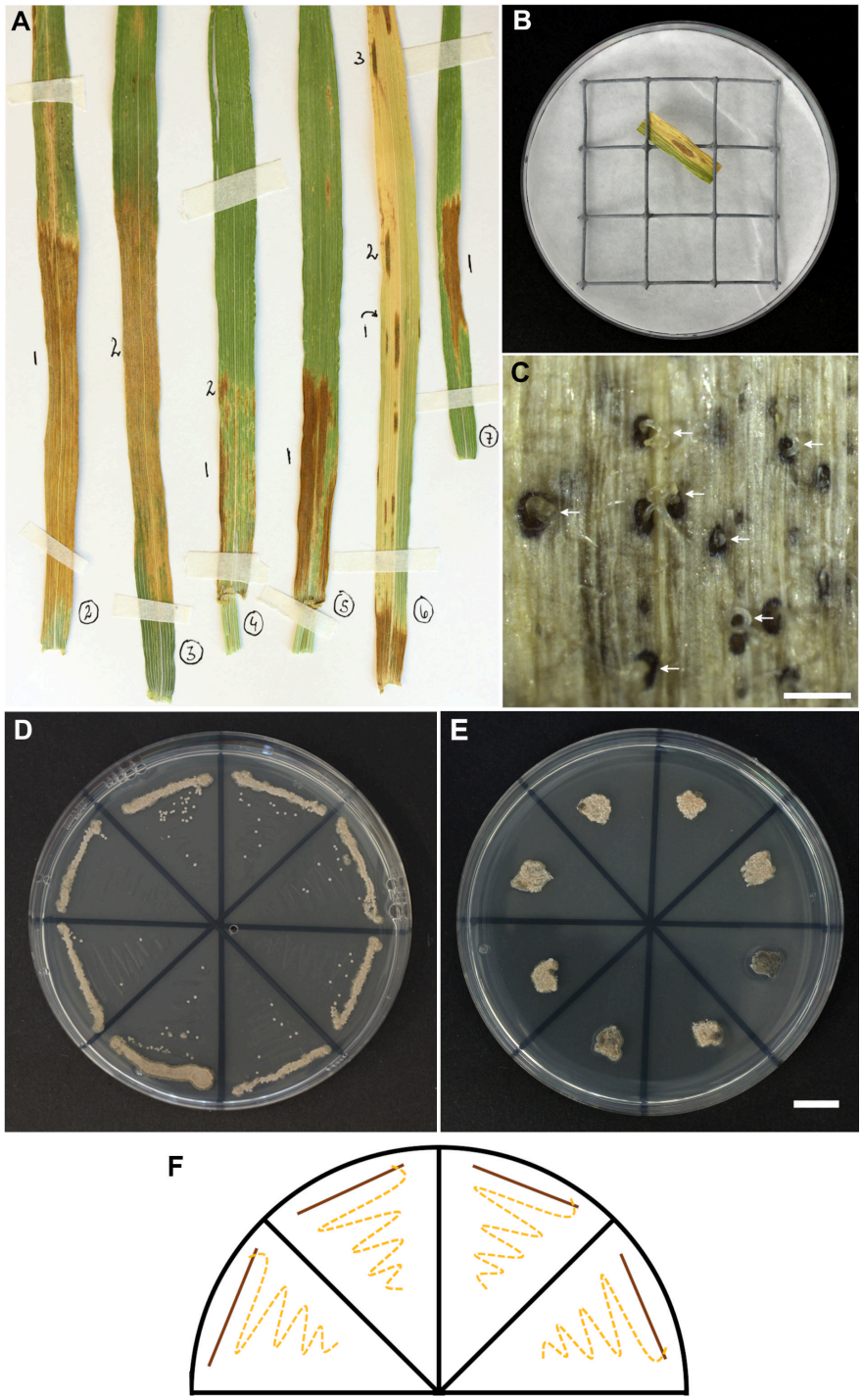


Figure 2. Workflow of *Z. tritici* isolation from wheat leaf lesions. **(A)** Symptomatic wheat leaves are detached from the plant, mounted to paper sheets, and labeled, with each leaf and each lesion numbered separately. **(B)** After surface sterilization, selected leaf sections with lesions are placed on metal grids inside petri dishes with wet Whatman paper. **(C)** After 5 days of incubation, pycnidiospores are released in cirrhi from pycnidia (white arrows). The scale bar represents 250 µm. Spores are then transferred to YMS agar plates with 50 µg/ml kanamycin, and colonies become visible after ~7days. **(D)** To obtain pure single-spore isolates, a few cells from the grown colonies are picked and streaked out onto a new YMS plate, following a "zig-zag" pattern. The shown plate was incubated at 18°C for 6 days. **(E)** Using a single colony from the last round of isolation, cells are spread on a small area (~4 mm2) of a fresh YMS plate. The scale bar represents 10 mm. **(F)** Schematic representation of the single-spore isolation pattern. Note that the first line (straight, solid brown line) and the second line ("zig-zag"-patterned, dashed yellow line) are done with separate sterile toothpicks.

8. Seal dishes with Parafilm and incubate at 20°C with a 16-hr light period until pycnidiospores are released in cirrhi (a mucilage that groups spores together; Fig. 2C), usually after 5 days.

Similar conditions for incubation can also work, e.g., simply keeping the plates on the laboratory bench as long as the temperature does not exceed 25°C.
Environmental samples can be highly variable and can contain a large diversity of endophytic, putatively unwanted microorganisms. Hence, it is important to monitor the leaf sections daily for Z. tritici cirrhi emergence and growth of other endophytes that might contaminate the isolation. Incubation periods should be kept as short as possible to avoid such contaminations.

9. Under a dissecting microscope, pick pycnidiospores (emerging from one pycnidium in one cirrhus) with a sterilized fine dissecting needle or fine syringe needle.

Sterilize the needle by dipping the tip in 70% ethanol and passing it through a Bunsen burner flame to burn off the ethanol. Ensure that the needle is cooled before touching pycnidiospores.
Avoid exposing the leaves to the light source for a prolonged time because it will dry out the cirrhi and make transfer difficult.
To facilitate pycnidiospore isolation, mount the leaf pieces on a sterile filter paper disc over a Styrofoam base using sterilized needles (e.g., simple pins). This fixates the leaves and helps in picking single cirrhi.

10. Transfer pycnidiospores to YMS agar plates with 50 µg/ml kanamycin by gently dipping the needle with the picked spores into the agar.

Pick pycnidiospores from each pycnidium just once and sterilize the needle between pycnidia. Divide YMS plates into eight compartments to reduce the total number of plates required (see Fig. 2D and 2E).

11. Seal plates with Parafilm and incubate at 18°C for 3 to 7 days, until small colonies are visible.

Colonies are often cream colored to pinkish, resembling yeast-like growth (see Fig. 2D). Monitor plates for growth of contaminating organisms frequently.

12. Pick cells from a colony with a sterile toothpick and streak out onto a new YMS agar plate with 50 µg/ml kanamycin.

Take only a few cells from the colony; the amount that sticks to the toothpick tip is sufficient. "Draw" a single straight line with this toothpick (see Fig. 2F).

13. On the same new YMS agar plate from step 12, use another sterile toothpick to streak out a second continuous line in "zig-zag" (Fig. 2D and 2F). Seal plates and incubate at 18°C for 5 to 7 days, until single colonies appear around "zig-zag" line.

14. Repeat steps 12 and 13 on YMS agar plates without kanamycin using cells from a well-separated single colony.

These two rounds of purification allow the researcher to obtain colonies grown from single fungal cells, which are mandatory to obtain pure, single cell–derived Z. tritici isolates.

15. Pick a single colony from this last round of isolation and spread it onto a small surface area (~4 mm2; Fig. 2E) on a fresh YMS agar plate without kanamycin using a sterile 1000-µl pipet tip or toothpick. Incubate plates for 5 days at 18°C. Use cells grown in this colony for any downstream procedures, such as species confirmation (see Basic Protocol 2) and inoculation of cultures for long-term storage (see Basic Protocol 3).

Basic Protocol 2: Molecular identification of *Z. tritici* isolates

Z. tritici species identification solely based on disease symptoms and *in vitro* phenotypes is challenging, so we recommend verifying the species identity of isolates obtained from field collections. To this end, polymerase chain reaction (PCR) followed by amplicon sequencing of a conserved nuclear locus (e.g., internal transcribed spacer rDNA, beta-tubulin, actin) represents a well-established procedure to discriminate *Z. tritici* specimens from other plant-associated fungi (Crous, Aptroot, Kang, Braun, &Wing- field, 2000; Fraaije, Lovell, Rohel, & Hollomon, 1999; Quaedvlieg et al., 2011). Based on the following protocol, isolates can be readily confirmed by PCR and amplicon sequencing using only small amounts of cells from the last single-colony isolation step in Basic Protocol 1 and a "rapid DNA extraction" method (Support Protocol 1). Once PCR and amplicon sequencing are done, online databases, alignment tools, and software such as NCBI (https://www.ncbi.nlm.nih.gov), BLAST (https://blast.ncbi.nlm.nih.gov),

Geneious (https://www.geneious.com), and Blast2GO (https://www.blast2go.com) can be used to align the sequences and support species identification by DNA sequence homology. The following PCR protocol uses crude template genomic DNA (gDNA) from a rapid extraction method (Support Protocol 1) and primer pairs for two loci, ITS and beta-tubulin (TUB2), that have been described previously (Glass & Donaldson, 1995; O'Donnell & Cigelnik, 1997; Weir, Johnston, & Damm, 2012; White, Bruns, Lee, & Taylor, 1990) (Table 1).

NOTE: Experiments involving PCR require extremely careful technique to prevent contamination.

Table 1. Primer Sequences for Amplification of ITS and Beta-Tubulin

Locus	Primer	Orientation	Sequence (5'-3')	Length (bp)	Reference
ITS	ITS5	Forward	GGAAGTAAAAGTCGTAACAAGG	22	White et al., 1990
ITS	ITS4	Reverse	TCCTCCGCTTATTGATATGC	20	White et al., 1990
TUB2	T1	Forward	AACATGCGTGAGATTGTAAGT	21	O'Donnell & Cigelnik, 1997
TUB2	Bt2b	Reverse	ACCCTCAGTGTAGTGACCCTTGGC	24	Glass & Donaldson, 1995

Materials

Template gDNA (see Support Protocol 1)

50 mM magnesium chloride (MgCl2; e.g., NEB, cat. no. B0510A)

5× Phusion High Fidelity (HF) Buffer (e.g., NEB, cat. no. B0518S)

100% dimethylsulfoxide (DMSO; e.g., NEB, cat. no. B0515)

5 M betaine (e.g., Sigma, cat. no. B0300-1VL)

10 mM deoxynucleotide (dNTP) mix

100 pmol/µl forward and reverse oligonucleotide primers

2 U/µl Phusion DNA polymerase (e.g., NEB, cat. no. M0530S)

Sterile water

2% (w/v) agarose gel with nucleic acid gel stain in 1× TAE buffer (from 50× TAE buffer; see recipe)

1× TAE buffer (from 50× TAE buffer; see recipe)

1-kb or 50-bp DNA ladder (e.g., GeneRuler 1 kb Plus or 50 bp, Thermo Fisher, cat. no. SM1331 or SM0371)

PCR cleanup kit (e.g., Wizard SV Gel and PCR Clean-Up System, Promega, cat. no. A9282; optional; alternatively, use cleanup services at commercial sequencing facility)

PCR tubes or plates

Thermocycler

Electrophoresis apparatus

Additional reagents and equipment for agarose gel electrophoresis (see Current Protocols article; Armstrong & Schulz, 2015) and Sanger sequencing at sequencing facility

1. Maintain all reagents and template gDNA on ice. Mix the following reagents in a PCR tube or plate well to a final reaction volume of 30 µl for each reaction:

15 µl template gDNA

6µl 5× Phusion HF Buffer (final concentration: 1×)

0.75 µl 50 mM MgCl2 (final concentration: 1.25 mM)

1 µl DMSO (final concentration: 3.3%)

6 µl 5 M betaine (final concentration: 1 M)

0.5 µl 10 mM dNTP mix (final concentration of each dNTP: 0.17 mM)

0.1 µl 100 pmol/µl forward oligonucleotide primer (final concentration: 0.3 pmol/µl)

0.1 µl 100 pmol/µl reverse oligonucleotide primer (final concentration: 0.3 pmol/µl)

0.5 µl 2 U/µl Phusion DNA polymerase (final concentration: 0.03 U/µl)

0.05 µl sterile water

Dilute the crude gDNA extract 1:10 using sterile water and use this diluted gDNA as a template for the PCR reaction.

For ITS PCR, use the primer pair ITS5 and ITS4 (White et al., 1990). For beta-tubulin PCR, primers T1 and Bt2b (Glass & Donaldson, 1995; O'Donnell & Cigelnik, 1997) should be used. See Table 1 for details.

It is recommended to prepare a master mix of the reaction components (minus the template gDNA) and to distribute the required volume into each tube/well prior to adding the template gDNA. Prepare the master mix with a slightly greater volume than required to avoid insufficient volume due to pipetting inaccuracy. Make sure to include negative- and positive-control reactions for each primer pair (e.g., substitute template gDNA with sterile water and use template gDNA from a previously confirmed Z. tritici isolate, respectively).

2. Preheat thermocycler to 98°C prior to entering the reaction tubes/plates and set lid temperature to 98°C to avoid condensation in lids. For the PCR reactions to amplify the target DNA, use the following program:

Initial step:	30 s	98°C (initial denaturation)
35 cycles:	8 s	98°C (denaturation)
	20 s	x°C (annealing)
	30 s	72°C (extension)
	8 min	72°C (final extension).

x°C = 50°C for ITS and 55°C for beta-tubulin.

PCR products can be stored at 4°C until they are analyzed via gel electrophoresis and submitted to sequencing. For longer storage, store at −20°C.

3. Separate PCR products on a 2% agarose gel with nucleic acid gel stain in 1× TAE buffer in an electrophoresis apparatus containing 1× TAE buffer. Set electrophoresis run to 120 V for 1 hr. Include controls and a 1-kb or 50-bp DNA ladder.

Supplement the gel with an appropriate volume of nucleic acid gel stain, e.g., add 5 µl Midori Green dye (Nippon Genetics, cat. no. MG04) per 100 ml gel.

Expected band sizes are around 550 to 600 bp for ITS and 750 bp for beta-tubulin. Only PCR fragments from samples that give single, clear bands should be used for sequencing. If necessary, excise bands with the expected fragment size from the gel.

4. Clean up PCR products or gel fragments prior to sequencing using a PCR cleanup kit following the manufacturer's recommendations or cleanup services at a commercial sequencing facility.

5. Submit PCR products for Sanger sequencing using the same primers as used for the PCR reactions.

Usually, one primer (sequencing in the forward or reverse direction) per locus is sufficient.
Follow the specific concentration requirements of the sequencing facility.

6. Use ITS and beta-tubulin sequences for alignments and similarity searches.

Highly similar sequences can provide information about the species.
Public tools and databases like BLAST (https://blast.ncbi.nlm.nih.gov) and NCBI (https://www.ncbi.nlm.nih.gov) or software like Geneious (https://www.geneious.com) and Blast2GO (https://www.blast2go.com) can be used for sequence analysis.

Support Protocol 1: Rapid extraction of *Z. tritici* genomic DNA

In order to facilitate the molecular identification of newly isolated fungal specimens (Basic Protocol 2), here we describe an easy method for PCR-quality gDNA extraction that is inexpensive and much faster compared to traditional protocols. It is a modified version of the previously published "HotSHOT" DNA extraction protocol (Truett et al., 2000). Besides being short, this protocol does not use potentially harmful organic compounds like chloroform and phenol and requires a low number of cells. These cells can easily be retrieved from the last step of single-colony isolation (Basic Protocol 1, step 15). It is important to note that the gDNA extract is crude and only suitable for amplification of products <1 kb. This protocol is not appropriate to isolate gDNA for whole-genome sequencing or for more delicate PCR-based analyses (e.g., amplification of genetic markers like inter-simple sequence repeats). For such applications, isolation of high-quality DNA as described in Support Protocol 2 is recommended.

NOTE: Considering the chemicals used in this protocol, all the steps can be performed on the laboratory bench using regular latex gloves.

Materials

Blastospores from axenic fungal cultures (e.g., from YMS agar plates with *Z. tritici* single colonies; see Basic Protocol 1, step 15)
25 mM sodium hydroxide (NaOH)

40 mM Tris·HCl, pH 5.5

Sterile toothpicks
PCR tubes or plates
98°C thermocycler or heat block

1. Pick a small number of cells (blastospores from axenic fungal cultures), approximately one-eighth of a single colony grown on YMS (see Fig. 2E), with a sterile toothpick and resuspend in 50 µl of 25 mM NaOH in a PCR tube or plate.

Swirl the toothpick in the solution to release the cells.

2. Incubate samples at 98°C for 10 min in a thermocycler or heat block.

3. Add 50 µl of 40 mM Tris·HCl (pH 5.5) to each sample.

This step ensures pH neutralization, which is important for DNA stability.

4. Store samples at −20°C until further use.

We recommend routinely freezing the samples after the 98°C incubation and neutralization, as we have observed that it increases DNA yield.

5. Thaw DNA extracts on ice prior to use in Basic Protocol 2.

Only the supernatant can be used as template gDNA for PCRs (e.g., amplification of ITS or beta-tubulin; see Basic Protocol 2). Pipet carefully and do not touch the cell debris on the bottom of the tube when transferring the supernatant. Alternatively, spin down cell debris using a microcentrifuge that can accommodate a PCR plate or tube. Dilute the supernatant 1:10 for use as template gDNA in PCR reactions (see Basic Protocol 2, step 1).

Support Protocol 2: Extraction of high-quality *Z. tritici* genomic DNA

The following protocol describes a cetyltrimethylammonium bromide (CTAB)-based method to isolate high-quality, high-molecular-weight gDNA from *Z. tritici* blastospores (Fig. 3A; Basic Protocol 3, step 2). This DNA can be used as input for genome sequencing libraries for short-read (e.g., Illumina) and long-read (e.g., SMRT sequencing by PacBio) technologies as well as for a broad range of molecular biology techniques, including PCR-based genotyping and cloning. We observe that DNA extracted using this protocol results in better overall sequencing quality compared to sequencing of DNA, e.g., extracted using a phenol/chloroform-based protocol. This protocol is a modified version of the CTAB method described by Allen, Flores-Vergara, Krasynanski, Kumar, & Thompson (2006).

CAUTION: Considering the handling of toxic organic compounds and liquid nitrogen, appropriate safety precautions, use of protective equipment (laboratory coat, goggles, and phenol-proof gloves), and working inside a chemical fume hood are extremely important.

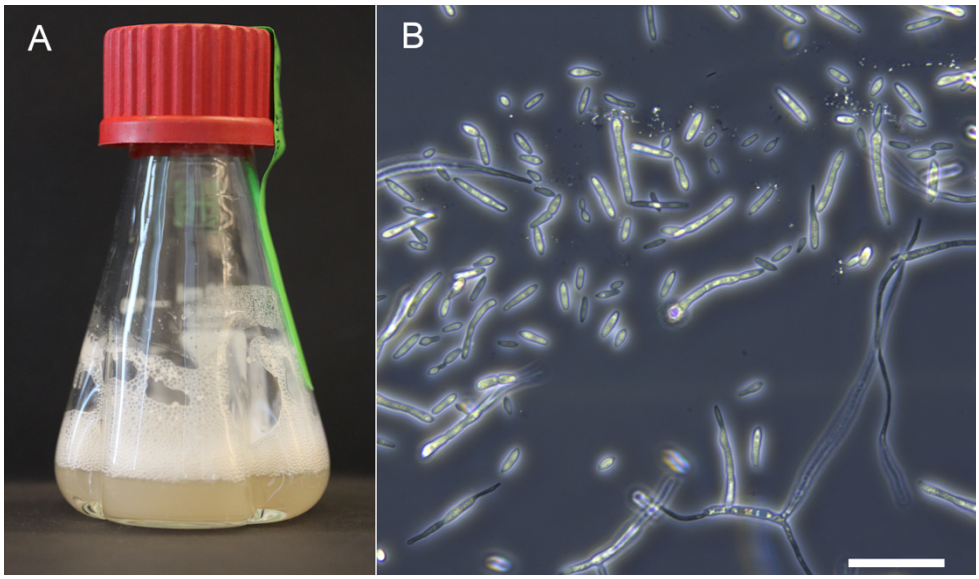


Figure 3. *Z. tritici* blastospore culture. **(A)** Example of a densely grown liquid culture of *Z. tritici* (OD600nm = 2.9). A 125-ml Erlenmeyer flask with baffles containing 20 ml liquid YMS was inoculated with a small amount of *Z. tritici* cells (approximately one-fourth of the single colony grown on YMS in Figure 2E) and incubated at 18°C and 200 rpm for 5 days. The screw cap was only loosely closed to allow gas exchange and was hence secured with tape. **(B)** *Z. tritici* blastospores grown in liquid YMS after 5 days of incubation. Formation of filaments is also observed. The scale bar represents 50 µm.

Materials

Densely grown liquid *Z. tritici* cultures (OD600nm = 2.8 to 3.2; see Basic Protocol 3, step 2)

Liquid nitrogen

1:1 (v/v) phenol/chloroform

Isopropanol, −20°C

1× TE buffer, pH 8 (see recipe)

10 mg/ml RNase A

3 M sodium acetate (NaOAc)

100% ethanol, −20°C

75% (v/v) ethanol

0.8% (w/v) agarose gel with nucleic acid gel stain in 1× TAE buffer (from 50× TAE buffer; see recipe)

1× TAE buffer (from 50× TAE buffer; see recipe)

1-kb DNA ladder (e.g., GeneRuler 1 kb Plus, Thermo Fisher, cat. no. SM1331)

Fluorometer DNA assay kit (broad range, measuring nucleic acid concentrations from 100 pg/µl to 1000 ng/µl; e.g., Qubit dsDNA BR Assay Kit, Thermo Fisher, cat. no. Q32853)

Water bath with temperature control

50-ml Falcon tubes

Thermo Scientific Heraeus Biofuge Stratos (cat. no. 10305002) with a fixed-angle rotor (Thermo Scientific HIGHConic, cat. no. 10203802)

Mortar and pestle (DNA- and DNase-free by sterilization at 180°C for 6 hr or chemical treatment)

Sterile 2-ml microcentrifuge tubes

Small spatulas (that can fit into 2-ml microcentrifuge tube)

Styrofoam box

Vortex (e.g., Vortex-Genie 2, Scientific Industries, cat. no. SI-0236)

Microcentrifuge

Nutator shaker platform (e.g., Grant-bio Sunflower Mini-Shaker, model PS-3D)

37°C thermomixer (for 2-ml microcentrifuge tubes; e.g., Eppendorf Thermomixer, Fisher Scientific, cat. no. 15346551)

Electrophoresis apparatus
Fluorometer (e.g., Qubit 4, Thermo Fisher, cat. no. Q33226)

Additional reagents and equipment for preparing DNA extraction buffer (see recipe) and for agarose gel electrophoresis (see Current Protocols article; Armstrong & Schulz, 2015)

1. Set temperature of the water bath to 65°C.

2. Prepare fresh DNA extraction buffer (see recipe) and pre-warm to 65°C in water bath.

3. Transfer densely grown liquid *Z. tritici* culture (~20 ml) to a 50-ml Falcon tube. Centrifuge 10 min at 4000 rpm in a Thermo Scientific Heraeus Biofuge Stratos with a fixed-angle rotor and remove supernatant.

If required, the cell pellet can be snap-frozen and stored at −80°C until further use. However, in order to prevent DNA degradation and maintain DNA integrity, prolonged storage should be avoided.

4. Grind cells in liquid nitrogen using a pre-cooled mortar and pestle. Ensure that cells are ground to a fine powder and that no frozen clumps remain.

Before transferring the cells into the mortar, cool the mortar and pestle, e.g., by adding a few drops of liquid nitrogen into the mortar or by placing both into liquid nitrogen. Ensure that the cells are also frozen while grinding.
The protocol can be paused after this step, and cell powder can be stored at −80°C until further use.

5. Transfer 100 to 200 mg cell powder (approximately until the 400-µl volume mark) into a pre-cooled, sterile 2-ml microcentrifuge tube using a pre-cooled small spatula. Use a different spatula for different samples. Close microcentrifuge tube and quickly place into a Styrofoam box containing liquid nitrogen.

Dip the spatula briefly into liquid nitrogen to cool it in order to avoid adhesion and thawing of cell powder.
To prevent DNA degradation, ground cells must not thaw before extraction buffer is added (see step 6). Hence, maintain samples in liquid nitrogen until proceeding with the next step.

6. Add 1.2 ml pre-warmed DNA extraction buffer (see step 2) to each sample. Vortex briefly.

7. Incubate sample at 65°C for 1 hr in the water bath (see step 1). Mix tube every 10 min by inversion.

8. Centrifuge sample for 10 min at 16,200 × g at room temperature.

9. Transfer supernatant (approximately 700 to 800 µl) to a new sterile 2-ml microcentrifuge tube containing 800 µl of 1:1 phenol/chloroform.

10. Place sample on a nutator shaker platform and gently shake for 20 min.

11. Centrifuge sample for 10 min at 16,200 × g at room temperature.

12. Transfer aqueous phase (top phase) to a new sterile 2-ml microcentrifuge tube and repeat phenol/chloroform extraction (steps 9 to 11).

IMPORTANT NOTE: To obtain high-quality DNA, it is very important to avoid transferring any organic phase (bottom phase) to the subsequent steps as it contains proteins and cell debris. Avoid touching the bottom phase while pipetting the supernatant.

13. Transfer aqueous phase (top phase) to a new sterile 2-ml microcentrifuge tube containing 800 µl cold isopropanol (−20°C). Mix several times by inversion (no shaking).

14. Incubate sample for 15 min at −20°C followed by 10 min at room temperature for DNA precipitation.

15. Centrifuge sample for 10 min at 16,200 × g at room temperature.

16. Gently remove supernatant and resuspend pellet in 250 µl of 1× TE buffer (pH 8).

At this stage, the pellet is not completely stable. To avoid losing it, do not dislodge the pellet when you remove the supernatant.

17. Add 25 µl of 10 mg/ml RNase A to each sample and incubate at 37°C for 30 min.

18. Add 25 µl of 3 M NaOAc and mix briefly by inversion. Add 600 µl of cold 100% ethanol (–20°C), mix, and incubate sample for ≥ 1h at −20°C to precipitate DNA.

A prolonged precipitation step (e.g., overnight) is possible and preferred; it can help to increase the final DNA yield.

19. Centrifuge sample for 10 min at 16,200 × g at room temperature.

20. Gently remove supernatant and wash pellet two times with 500 µl of 75% ethanol.

Wash the DNA pellet entirely. If the pellet sticks to the tube wall, dislodge it carefully using a pipet tip to ensure that it is fully submerged.

21. Centrifuge sample for 5 min at 16,200 × g at room temperature.

22. Remove supernatant and air-dry pellet for ≥20 min at room temperature.

IMPORTANT NOTE: It is extremely important to remove any residual ethanol as it may interfere with any downstream DNA applications. It is also possible to remove residual ethanol using a SpeedVac for 10 to 15 min. Do not extend this incubation time or use a heating option; over-dried DNA pellets are difficult to dissolve.

23. Dissolve DNA pellet in 50 to 100 µl of 1×TE buffer (pH 8) by incubating the sample in the thermomixer at 37°C and 300 rpm for 45 min.

Large DNA pellets may take more time to fully dissolve, and therefore, the DNA concentration can change over time. Hence, always check the DNA quality and concentration immediately before proceeding with downstream applications that require accurate DNA concentrations (e.g., molecular cloning, preparation of sequencing libraries).

24. Check DNA purity by gel electrophoresis using a 0.8% agarose gel with nucleic acid gel stain in 1× TAE buffer in an electrophoresis apparatus containing 1× TAE buffer. Run gel at 100 V for 45 min. Include a 1-kb DNA ladder.

Supplement the gel with an appropriate volume of nucleic acid gel stain, e.g., add 5 µl Midori Green dye (Nippon Genetics, cat. no. MG04) per 100 ml gel.
A single, high-molecular-weight band with little-to-no smear should be observed. In the case that smears or multiple bands are visible, in particular at the bottom of the gel, additional cleanup steps, e.g., with phenol/chloroform (1:1),

are recommended, as smears or multiple bands indicate either DNA fragmentation or contamination with proteins. Alternatively, spin columns for DNA purification can be used. However, every additional cleanup step will reduce DNA yield.

The Joint Genome Institute (JGI) provides a full protocol for gDNA quality control prior to sequencing. Check their Genomic DNA Sample Protocol for more information (https: //jgi.doe.gov/user-programs/pmo-overview/project-materials-submission-overview/).

25. Quantify DNA concentration using a fluorometer and fluorometer DNA assay kit following the manufacturer's recommendations.

Fluorometer measurements are more accurate compared to visual or spectrophotometer methods.

26. Keep DNA sample at 4°C for short-term storage or at −20°C for long-term storage prior to use in any analysis that requires high-quality DNA (e.g., cloning, preparation of genome sequencing libraries).

Avoid repeated thawing-freezing cycles, as they may accelerate DNA degradation.

Basic Protocol 3: *In vitro* culture and long-term storage of *Z. tritici* isolates

Proper storage and maintenance of new *Z. tritici* isolates (Basic Protocol 1) are important to keep isolates viable and to ensure their genomic integrity. Cells of isolates that are stored under cryogenic conditions can be preserved for decades and therefore serve as valuable resources to answer different research questions throughout history (Hartmann, Sánchez-Vallet, McDonald, & Croll, 2017; Oggenfuss et al., 2020). The Dutch *Z. tritici* isolate IPO323 was originally isolated in 1981 and has been used as a reference in several studies worldwide since then (Kema & Van Silfhout, 1997). For most experiments, including preparation of long-term stocks, blastospore cultures ("yeast-like" morphology) are used as input. For many isolates, these cells can be obtained from cultures grown in nutrient-rich medium (e.g., yeast malt sucrose, or YMS, medium) at 18°C (Fig. 3).

This protocol describes cultivation and long-term cryopreservation of *Z. tritici* blastospores and can be initiated using a small number of cells from single-spore colonies immediately after isolation (Basic Protocol 1). Considering the importance of long-term storage stocks, we recommend

preparing the cultures using aseptic techniques in a laminar flow hood. Additionally, we propose preparing two distinct sets of cryo-stocks: a safe "backup" stock that will only be used in cases of great necessity and an "in use" stock, which will be used routinely to cultivate cells for experiments. Digital records of the stored isolates can be combined with different types of metadata (e.g., location and date of isolation, original wheat cultivar).

Materials

YMS agar plate with *Z. tritici* single colony (see Basic Protocol 1, step 15)
Liquid YMS medium (see recipe)
Sterile silica (e.g., Sigma-Aldrich, cat. no. 60741; sterilized at 180°C for 6 hr)
Liquid YMS-glycerol medium (see recipe)
Liquid nitrogen

Sterile toothpicks
Sterile culture tubes or sterile Erlenmeyer flasks with baffles
Orbital shaker with adapter for culture tubes or flasks and temperature control, 18°C
Spectrophotometer
Sterile 2-ml cryovials
Small spatula (that can fit inside 2-ml cryovials)
Shaker with adapter for cryovials (e.g., IKA VXR Basic Vibrax, cat. no. 0002819000 & 0020018016)

1. Pick a small number of cells (from a YMS agar plate with *Z. tritici* single colony), e.g., approximately one-fourth of a single colony grown on YMS (Fig. 2E), using a sterile toothpick and inoculate ≥5 ml liquid YMS medium in a sterile culture tube or a sterile Erlenmeyer flask with baffles. Label tube with the isolate ID.

The volume of medium is dependent on the culture tube or flask used. A maximum of one- third of the culture tube or flask volume should be filled with YMS. To allow CO2 circulation, do not close the culture vessel to be airtight; secure the cap using tape (Fig. 3A).

2. Incubate cultures using an orbital shaker with an adapter for culture tubes or flasks at 18°C and 200 rpm for 5 days or until the cultures are densely grown (OD600nm = 2.8 to 3.2 via a spectrophotometer) (Fig. 3A).

Due to the phenotypic diversity of Z. tritici, some isolates may take more or less time to reach this culture density.

3. Prepare two or more sterile 2-ml cryovials for each isolate, with vials for both "backup" and "in use" stocks, as follows:

a. *Silica stocks*: Scoop sterile silica into cryovials using a small spatula until the vials are filled to the 300-µl mark.
b. *Glycerol stocks*: Add 900 µl liquid YMS-glycerol medium to the cryovials.
Glycerol stocks are easier to handle and are recommended for "in use" stocks.

4. Add 50 and 900 µl of densely grown culture from step 2 to the cryovials containing silica and YMS-glycerol, respectively. Close cryovial caps tightly.

Label the cryovials with the isolate ID and other relevant information (e.g., date of preparation). It is important to use freezer-ready labels or pens, as lettering may fade over time or labels may come off; regular paper labels easily detach if not covered with clear tape.

5. Mix liquid culture and YMS-glycerol for the glycerol stocks thoroughly by inversion. Likewise, mix culture and silica for the silica stocks for ≥5 min on a shaker with an adapter for cryovials.

The shaker does not need to be at high speed (500 rpm is sufficient). This step is required to dry the blastospores. Check if there are still clumps of "wet" silica after 5 min by holding the tubes horizontally. If this is the case, add more silica to the cryovials and leave them in the shaker until the silica-culture mix is dry and homogeneous.

6. Snap-freeze stock cultures using liquid nitrogen and store tubes at −80°C.

7. Keep cryo-stocks on ice or in a cooled rack during use.

It is good practice to not let samples thaw too long on ice (e.g., for ≥10 min). Ideally, the YMS-glycerol culture should still be soft, not entirely thawed. Use and return the stocks to the −80°C freezer as quickly as possible.

8. Remake "in use" stocks before these stocks run too low to avoid samples loss, preferentially retrieving material for culture inoculation from the safe "back up" stocks.

Basic Protocol 4: Analysis of *Z. tritici* virulence in wheat

Ranging from virulence quantification to mutant phenotyping, *in planta* infection assays are an important component of many *Z. tritici* studies. The infection outcome, namely wheat susceptibility or qualitative or quantitative resistance, can provide information about the presence or absence of avirulence factors in *Z. tritici* isolates as well as about the relevance of candidate gene products during host infection (Brown, Chartrain, Lasserre- Zuber, & Saintenac, 2015). For example, the wheat cultivar Chinese Spring harbors the resistance gene *Stb6*, which confers resistance against isolates carrying the avirulence factor *AvrStb6*, like the *Z. tritici* reference isolate IPO323 (Kema et al., 2018; Saintenac et al., 2018; Zhong et al., 2017). Hence, Chinese Spring can be used to study the molecular and physiological mechanisms of wheat resistance against *Z. tritici AvrStb6* genotypes. Infection of other highly susceptible wheat cultivars can be used to compare the virulence of different *Z. tritici* genotypes. The cultivars Obelisk, Riband, and Drifter are highly susceptible to IPO323 and to a large number of other isolates and are therefore commonly used in pathogenicity assays to quantify and compare virulence (Chartrain, Brading, & Brown, 2005; Habig et al., 2017; Haueisen et al., 2019; Karisto, Dora, & Mikaberidze, 2019; Stewart &McDonald, 2014; Zhong et al., 2017).

Previous studies revealed that quantitative virulence of *Z. tritici* can be highly variable within and between experiments, generally indicated by a broad data distribution for quantitative measures like pycnidia density (Habig et al., 2017; Stewart & McDonald, 2014). Thus, technical and experimental replicates are extremely important to be able to discern treatment effects from background variation. We recommend using a minimum of two biological replicates (e.g., two independent deletion mutants) and two technical replicates (e.g., two independent experiments) to obtain at least 30 leaves per treatment and replicate (*Z. tritici* isolates or mock). Randomization within experiments (randomization of pot placement and randomization of treatment order) is required to minimize confounding effects of environmental factors and experimental procedures. Furthermore, cross-contaminations or other abiotic and biotic biological aspects (e.g., senescence) that may influence the final results must be monitored by the inclusion of a negative control (mock

treatment). In addition, observer's bias can be avoided by making the experimenter unaware of the treatment type throughout the whole experiment.

The following protocol describes a method to quantitatively evaluate *Z. tritici* virulence in wheat hosts under controlled greenhouse or growth chamber conditions. This protocol can be used whenever phenotyping *in planta* is required. The complete workflow is summarized in Figure 4.

NOTE: Use of latex gloves and goggles is recommended during the inoculation procedures.

Materials

Sterile water

Wheat seeds

Peat potting soil (e.g., Fruhstorfer Erde Typ T, Hawita, cat. no. 01020)

70% (v/v) ethanol

Z. tritici inoculum (see Support Protocol 3)

0.1% (v/v) Tween® 20 (e.g., Roth, cat. no. 9127.1) in sterile water

Whatman paper

Plastic box (for seed germination; e.g., size 11 × 7cm)

Growth chamber and/or greenhouse

Plastic tags

Pots and trays (for plant growth; e.g., 9 × 9–cm pots; 60 × 40–cm trays)

Small dibble (optional)

Permanent black pen with soft tip (e.g., Lumocolor permanent, Staedtler, cat. no. 317-9)

Spray gun (from airbrush spray set; 0.5-mm needle, 0.5-mm nozzle, and air hose and compressor; e.g., Timbertech set, cat. no. ABPST06)

Tape or plastic locking clips

Large plastic bags (e.g., size 60 × 80 cm)

White paper sheets (A4 or letter format, depending on scanner)

White regular tape

Scissors

Flatbed scanner (e.g., HP Photosmart C4580; HP for A4 format)

ImageJ software

1. Initiate germination of wheat seeds: Place Whatman paper at bottom of a plastic box and add sterile water to moisten it (~12.5 ml is appropriate for an 11 × 7–cm box). Place wheat seeds inside box and close it. Germinate seeds for 4 days under plant growth conditions (20°C and a 16-hr light period) in a growth chamber (Fig. 4A).

Ensure that the Whatman paper is covering the bottom of the box and that the seeds are well dispersed and are touching the filter paper. Include seeds for at least 30 plants per treatment (Z. tritici isolate replicates and mock). This step ensures that enough wheat seedlings will be viable and at a similar developmental stage when they are planted and inoculated. It is recommended to germinate a number of additional seeds. In our experience, ~4.5 g of Obelisk seeds is sufficient to produce ~100 seedlings.

2. Fill pots with peat potting soil and organize them on trays. Add a plastic tag with information about wheat cultivar, treatment (e.g., anonymized isolate and replicate ID), and date to each pot.

Twenty-four 9 × 9–cm pots can be placed onto one 60 × 40–cm tray.

3. To plant wheat seedlings, first wet soil in the pots. Make four holes in soil (~1cm in depth), e.g., by using a small dibble or a finger. Transplant 4-day old seedlings (see step 1) into holes and cover roots with soil.

4. Grow seedlings for 7 days in soil prior to inoculation. Water plants regularly by pouring water in the trays instead of overhead watering.

Seedlings of different wheat cultivars require different times to develop. Plants can be inoculated as soon as the second leaf is expanded, or at stage 1 on the Feekes scale (Large, 1954).

5. One day before inoculation, mark a distinct section (e.g., 5 to 10 cm) on adaxial side of the second leaf of each plant using a permanent black pen with a soft tip (Fig. 4B).

This leaf section will be used as a reference to compare Z. tritici virulence between plants and treatments. To simplify this step, mark the length of the area to be inoculated on your glove, preferably on the pointer (index) finger, and bend the leaves over this finger (Fig. 4B).

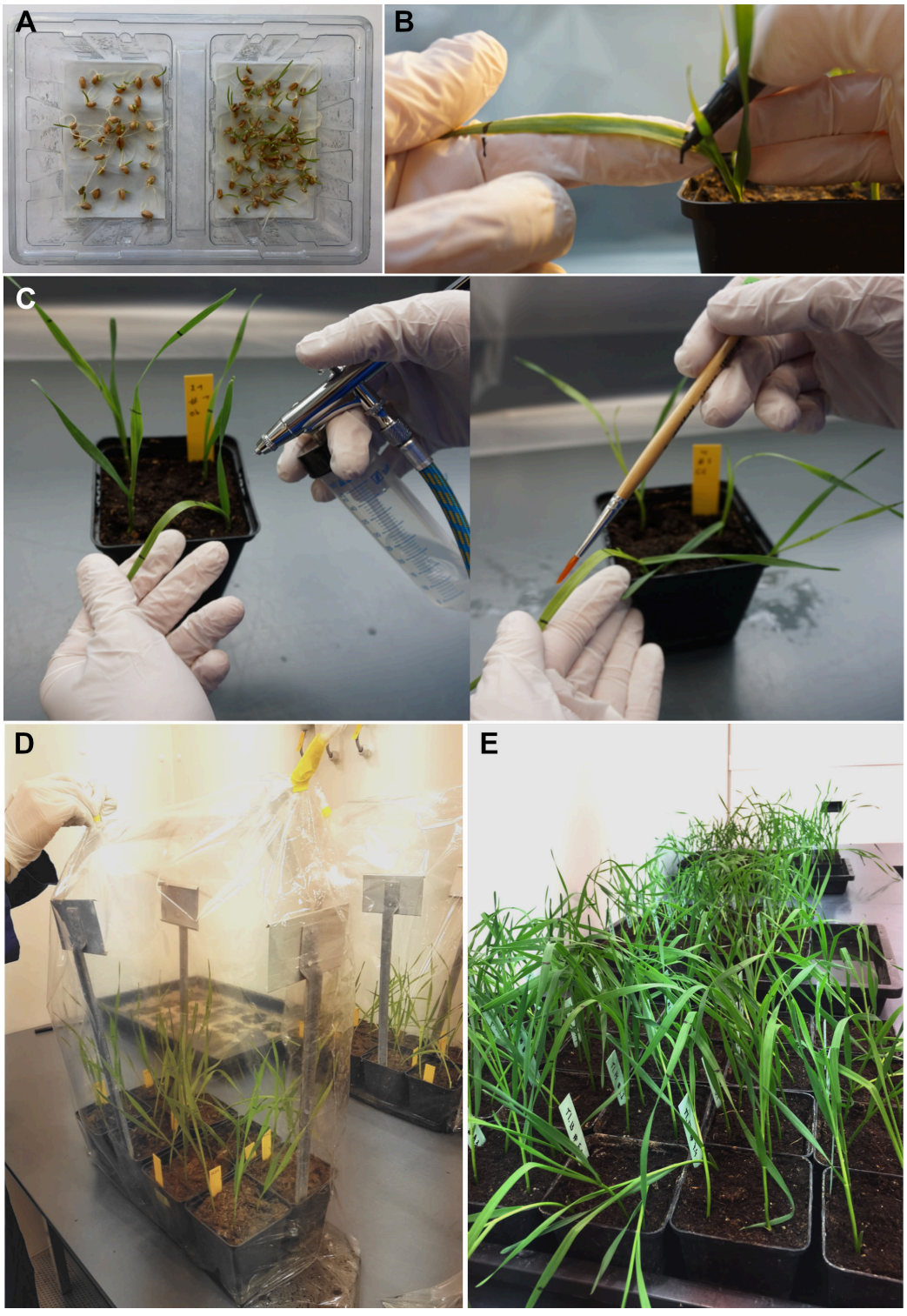


Figure 4. Inoculation of wheat leaves with *Z. tritici*. **(A)** Wheat seeds are germinated in plastic boxes containing wet Whatman paper for 4 days at 20°C and a 16-hr light period. Seedlings are subsequently transferred to pots filled with organic soil. **(B)** Seedlings are grown for 7 days in soil, and distinct leaf sections are marked on the adaxial side, e.g., of the second leaf. **(C)** Marked leaf areas are spray-inoculated (left) or brush-inoculated (right) with *Z. tritici* or mock suspension until runoff. **(D)** After inoculation, pots (separated by treatment) are placed inside large plastic bags with water and sealed to generate maximal humidity. **(E)** After 48hr, the plants are removed from the plastic bags and placed back onto trays. The plants are grown at 70% to 90% relative humidity, ~20°C (day)/~12°C (night), and a 16-hr day/8-hr night cycle until 21 days post-inoculation (dpi).

6. On the inoculation day, organize pots for each treatment and ensure that all pots are properly labeled.

7. To clean the spray gun, spray 70% ethanol 20 to 30 times for 2 to 3 s each followed by spraying sterile water 20 to 30 times for 2 to 3 s each using 2.0 bar of pressure.

IMPORTANT NOTE: This cleaning procedure is required between all treatments. It is extremely important to prevent cross-contamination and must not be omitted.

8. To inoculate the plants for each treatment, use cleaned spray gun and a pressure of 2.0 bar to inoculate the marked leaf sections of each plant with the *Z. tritici* inoculum by evenly spraying until runoff (Fig. 4C). Clean spray gun after each treatment as described in step 7. Follow same procedure for the mock treatment by spraying 0.1% Tween® 20 in sterile water onto the marked leaf sections.

Alternatively, sterile paint brushes can be used to apply Z. tritici inoculum (Fig. 4C). At least one paint brush should be used per treatment. Brush inoculum gently onto the marked leafarea, without applying pressure, until the leaf surface is completely covered. Follow the same procedure for the mock treatment.

9. Let inoculum dry on leaf surface (~15 min) and place pots (sorted by treatment) in large plastic bags containing water (~1L). Seal bags using tape or plastic locking clips to generate an environment with maximal relative humidity (Fig. 4D). Incubate plants in the bags at ~20°C (day)/~12°C (night) and a 16-hr day/8-hr night cycle under controlled greenhouse conditions. Alternatively, in growth chambers, use constant conditions of 20°C and 16 hr of light (~200 µmol/m2/s).

High humidity induces opening of stomata and facilitates even Z. tritici infections.

10. After 48hr, remove pots from the bags and place plants back onto the trays (Fig. 4E). Ensure randomized placement. Water regularly and maintain a relative humidity of 70% to 90%. Grow wheat for 21 days post-inoculation (dpi) at the same temperature and light conditions as described in step 9.

For cultivar Obelisk and the reference IPO323, 21 dpi is adequate to detect STB lesions and mature pycnidia. However, for other wheat cultivar–Z. tritici isolate combinations, this period can be shorter or longer.

Another important phenotypic trait is temporal development of disease. Timing of the transition from biotrophic to necrotrophic growth varies between Z. tritici isolates and might be impacted by the presence/absence of putative effector candidates. We suggest screening inoculated leaves of individual plants for the onset of necrosis and first visible pycnidia daily or every other day. Keep a record of the disease development (like the average day of the first observed symptom) for each isolate and mutant, e.g., for future experiments and as an additional virulence measurement.

11. At 21 dpi, harvest inoculated leaf sections by cutting ~1cm above and below the pen marks using scissors and mount leaves to white paper sheets using white regular tape.

Cover the black pen marks on the leaves with the white tape. Black marks will interfere with the leaf image analysis. Attach the leaves to the paper sheets with a distance of ≥5 cm from each other.

12. Scan paper sheets using a flatbed scanner, scanning each leaf individually. Use an image resolution of 2400 dots per inch.

Keep each scanned leaf in a separate image file, naming each file with the treatment followed by a consecutive number (e.g., Mock_1, Mock_2, …; IPO323_1, IPO323_2, …). Alternatively, QR codes can be used to automatically name image files; a related workflow was developed by Karisto et al. (2018).

13. Analyze scanned images using ImageJ software (Schneider, Rasband, & Eliceiri, 2012) and the published macros for *Z. tritici* virulence quantification (Karisto et al., 2018; Stewart & McDonald, 2014; Stewart et al., 2016).

The macros automatically output a spreadsheet summarizing the total leaf area, lesion area, percent of leaf area covered by lesions (PLACL), total pycnidia count, and pycnidia per cm2 of lesion leaf area, among other parameters, for each analyzed leaf.

Macro adjustments may be necessary depending on the hardware used and the leaf conditions (e.g., color intensity). Customize and test the macro settings to obtain more accurate results.

14. Compare virulence readouts (e.g., PLACL, normalized pycnidia count) within and between treatments using statistical tests in the R environment (R Core Team, 2013) or Open Office Calc (https://www.openoffice.org/product/calc.html).

Statistical methods should be performed to validate that the observed differences are significant, i.e., could only occur at a low probability assuming that the null hypothesis is true (which, in most cases, states that there is no difference between treatments). Different statistical tests have different assumptions that should be met for a correct statistical analysis. Among other premises, parametric tests (e.g., Student's t-test, analysis of variance/ANOVA) rely on the assumption that the data follow a normal distribution, with similar variances between treatments (Altman & Bland, 1995). If these assumptions are not met by the data, a statistical test will produce incorrect results. Hence, it is essential to verify that these assumptions are met. A visual inspection using a normal Q-Q plot and a Shapiro-Wilk test of normality (Shapiro & Wilk, 1965) (check the function "shapiro.test" of the "stats" R package) can be used to test the normal distribution of the data. Be aware that a significant result in the Shapiro-Wilk test indicates that the data are not distributed normally. Moreover, Levene's test for homogeneity of variances (also called homoscedasticity) (Levene, 1961) is recommended, in particular for the analysis of variance between treatments that have small or different sample sizes. The function "leveneTest" of the R package "car" can be used to perform this test.

Although it is not always the case, Z. tritici in planta phenotyping data generally fail one or more assumptions for parametric tests. Therefore, nonparametric tests such as the Mann-Whitney U test (or Wilcoxon rank sum test) and Kruskal-Wallis test can be used as good alternatives for the t-test and ANOVA, respectively. These tests do not assume a certain distribution of the data and have been largely used in studies analyzing Z. tritici virulence data (Haueisen et al., 2019; Karisto et al., 2018; Möller et al., 2018; Poppe, Dorsheimer, Happel, & Stukenbrock, 2015; Stewart et al., 2016). Check the functions "wilcox.text" and "kruskal.test" ofthe R package "stats" to perform such analyses.

Considering that high variation between experiments (i.e., technical replicates) could obscure real differences between treatments, a method where the variation between experiments can be removed and the true differences between treatments can be obtained is desirable. In general, multifactorial statistical analyses are extremely powerful in detecting effects and interactions. For nonparametric tests, the inclusion of multiple factors is difficult. ANOVA, however, can include and account for multiple factors. A promising way to apply ANOVA to non-normal data is the rank transformation, which has been suggested as a link between parametric and nonparametric statistical tests (Conover&Iman, 1981). Once rank-transformed, the data often follow a normal distribution, which allows the use of ANOVA (Habig et al., 2017; Habig, Bahena-Garrido, Barkmann, Haueisen, & Stukenbrock, 2019). Inclusion of a model (e.g., pycnidia density ~ treatment experiment) can allow dissection of the effect of treatment without confounding influence of the experiments (technical replicates). Applying the post hoc test Tukey's honestly significant difference (HSD) to the results of the ANOVA then allows for testing of individual treatment effects (Tukey, 1949).*

Support Protocol 3: Preparation of *Z. tritici* inoculum

In addition to their use in long-term cryo-stocks (Basic Protocol 3), *Z. tritici* blastospore cultures ("yeast-like" form) are routinely required to prepare inoculum for virulence assays (Basic Protocol 4). The following protocol describes a method to initiate blastospore culture, collect the cells, and adjust their density for plant infections. This protocol should be started 5 days prior to plant inoculation (Basic Protocol 4).

Materials

Z. tritici glycerol stock of *Z. tritici* isolate to be inoculated (see Basic Protocol 3)
YMS agar plate without kanamycin (see recipe)
Sterile water
70% (v/v) ethanol
0.1% Tween® 20 (e.g., Roth, cat. no. 9127.1) in sterile water

Sterile pipet tips and/or inoculation loops
Microbiological incubator with temperature control, 18°C
Sterile 2-ml microcentrifuge tubes
Vortex (e.g., Vortex-Genie 2, Scientific Industries, cat. no. SI-0236)
Hemocytometer and coverslip (e.g., Neubauer improved chamber, 0.100 mm, Marienfeld Superior, cat. no. 0640010)
Manual counter (i.e., tally counter)
Compound microscope
Sterile 50-ml Falcon tube

1. Retrieve *Z. tritici* glycerol stock of *Z. tritici* isolates to be inoculated from the −80°C freezer.

2. Dip a sterile pipet tip in soft (not entirely thawed) *Z. tritici* glycerol culture and transfer cells to a YMS agar plate without kanamycin. Using the pipet tip or a sterile inoculation loop, spread glycerol culture onto a small surface area (~4 mm2).

To avoid contamination of the YMS plate and the cryo-stocks, perform this step using aseptic techniques under a laminar flow hood.

Blastospores from one YMS plate are usually sufficient to inoculate up to 96 wheat plants. Inoculate extra YMS plates if additional inoculum is required.

3. Incubate plate for 5 days at 18°C and constant darkness.

After this period, a pinkish colony patch of blastospores should form. Some Z. tritici isolates may show different colony phenotypes. The color can range from pale pink to orange to black (melanized) and the texture from very smooth to hard and flaky, as displayed in Figure 5.

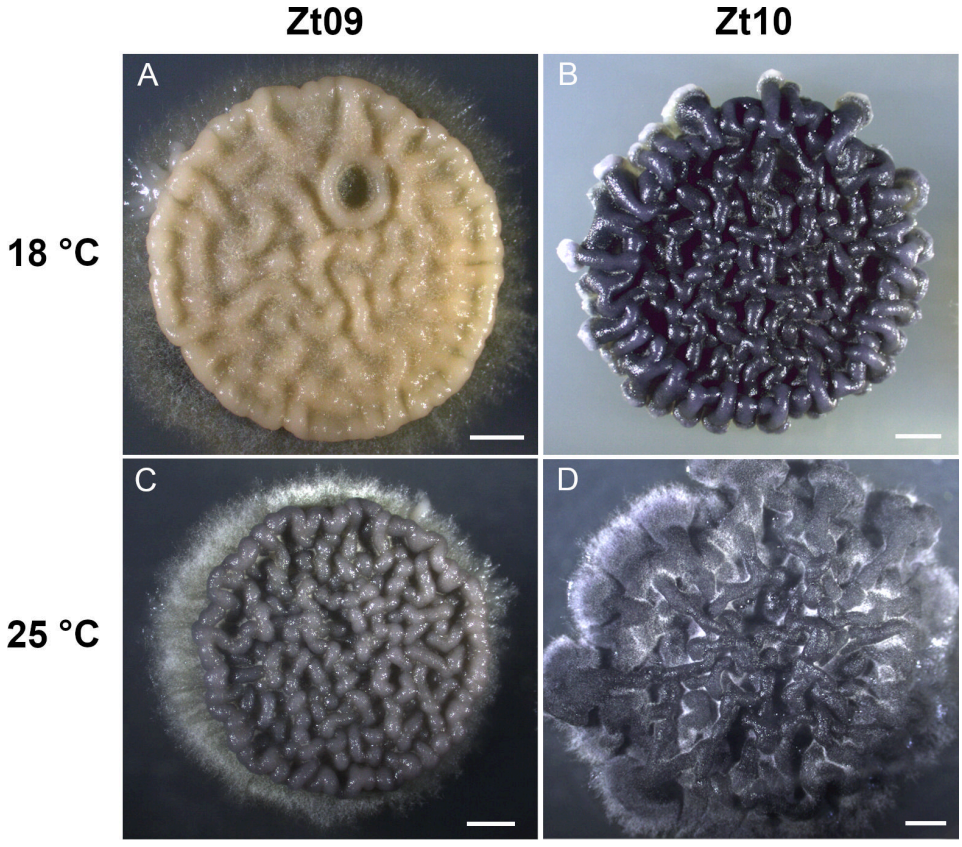


Figure 5. *Z. tritici* colony phenotypes are variable. Phenotypic plasticity is observed for two *Z. tritici* isolates grown under two different temperature conditions: isolates were grown at 18°C (**A** and **B**) and 25°C (**C** and **D**) for 5 days on YMS plates in constant darkness. For each isolate, the same inoculum source was used. The scale bars represent 1000 μm.

4. Scrape cells from the colony surface using a sterile pipet tip and resuspend them in a sterile 2-ml microcentrifuge tube containing 1 ml sterile water. Swirl pipet tip to release the cells and ensure that the suspension is homogeneous (no cell clumps).

Colonies of Z. tritici isolates may be formed by blastospores and hyphae. In the case of strong filamentous growth, filter the cell suspension through sterile miracloth to retain hyphal clumps.

5. Make at least two serial dilutions (1:10 and 1:100) of blastospore suspension in additional sterile 2-ml microcentrifuge tubes. Vortex tubes briefly between dilutions.

Two serial dilutions are generally enough to facilitate cell counting. Additional dilutions may be necessary if a larger number of cells has been resuspended.

6. Carefully clean hemocytometer and coverslip with 70% ethanol before use and position coverslip correctly (indicated by the appearance of Newton's rings). Select appropriate dilution, briefly vortex blastospore suspension, and pipet 10 µl into each compartment of hemocytometer.

Check the specifications of the counting chamber to select the appropriate cell dilution. Verify if the cell suspension is well homogenized and avoid injecting bubbles into the counting chamber.

7. Observe cells using a compound microscope (e.g., with a 10× objective lens).

8. Count number of cells using a manual counter. Record cell number.

9. Calculate cell density considering the average number of counted cells, the specifications of the hemocytometer in use, and the dilution of the cell suspension that was counted.

Verify the specific volume of the used counting chamber to calculate a correct estimate of the cell density.

10. Adjust inoculum to the required cell density (e.g., 10^6 or 10^7 cells/ml) in 0.1% Tween® 20 in sterile water. Keep final inoculum in a sterile 50-ml Falcon tube and use within a few hours.

A total volume of 15 ml is sufficient to inoculate ~96 wheat leaves. However, prepare as much as possible to avoid exhausting the inoculum during inoculation.

Reagents and Solutions

DNA extraction buffer

For 10 ml buffer, mix 1 ml 1 M Tris·HCl (Roth, cat. no. 9090.3), pH 8.0; 2.8 ml 5 M NaCl (Roth, cat. no. 3957.2); and 400 µl 0.5 M EDTA (Roth, cat. no. 8043.2), pH 8.0, using a stir bar. Completely dissolve 0.2 g cetyltrimethylammonium bromide (CTAB; Roth, cat. no. 9161.1) in 3 ml deionized water. Add CTAB solution to the buffer, mix well, and bring up volume to 10 ml with deionized water. Add 50 µl β-mercaptoethanol (Sigma-Aldrich, cat. no. M3148) before use. Prepare fresh immediately before use.

A larger volume of DNA extraction buffer (without β-mercaptoethanol) can be prepared and stored at room temperature for several months. Before DNA extraction, aliquot the required volume of extraction buffer and add the appropriate volume of β-mercaptoethanol [final concentration: 0.5% (v/v)].

Liquid YMS-glycerol medium

For 1 L medium, dissolve 4 g yeast extract (BD, cat. no. 212720), 4 g malt extract (Roth, cat. no. AE68.1), and 4 g sucrose (Roth, cat. no. 4621.2) in ~200 ml deionized water. Add 894.85 g of 98% (v/v) glycerol (Roth, cat. no. 7530.4) and stir well. Bring up volume to 1000 ml using deionized water. Aliquot in 100-ml autoclavable glass bottles and autoclave for 5 min at 121°C. Store ≤3 months at room temperature.

Glycerol is a very viscous substance that is easier to weigh instead of using a measuring cylinder. If glycerol with a different percentage (e.g., 99% or 78%) is used, the amount of glycerol added to the medium needs to be adjusted to achieve a final concentration of 69.6% (v/v).

Liquid YMS medium

For 1 L liquid YMS medium, solve 4 g yeast extract (BD, cat. no. 212720), 4 g malt extract (Roth, cat. no. AE68.1), and 4 g sucrose (Roth, cat. no. 4621.2) in ~200 ml deionized water. Stir well and bring up volume to 1000 ml using deionized water. Aliquot in autoclavable glass bottles and autoclave for 5 min at 121°C. Store ≤3 months at room temperature.

TAE buffer (Tris-acetate-EDTA), 50×

For 1 L buffer, add 242 g Tris (Roth, cat. no. 4855.3), 57.1 ml pure acetic acid (Roth, cat. no. 3738.4), and 100 ml 0.5 M EDTA (Roth, cat. no. 8043.2), pH 8.0, to 500 ml deionized water. Adjust pH to 8.3 with NaOH. Mix well using a stir bar and bring up volume to 1000 ml using deionized water.

Dilute to a working concentration of 1×, e.g., by diluting 40 ml 50× TAE buffer in 1960 ml deionized water. Store ≤1 year at room temperature.

TE buffer (Tris-EDTA) (pH 8), 1×

For 100 ml buffer, mix 200 µl 0.5 M EDTA (Roth, cat. no. 8043.2), pH 8.0, and 1 ml 1 M Tris·HCl (Roth, cat. no. 9090.3), pH 8.0, in 50 ml deionized water. Adjust pH to 8.0 with HCl if necessary and bring up volume to 100 ml with deionized water. Autoclave buffer for 5 min at 121°C. Store ≤1 year at room temperature.

YMS agar plates with and without 50 µg/ml kanamycin

For YMS solid medium, add 16 g agar (BD, cat. no. 214030) to an autoclavable 1-L glass bottle and add liquid YMS medium (see recipe) to 800 ml total volume. Add a stir bar and autoclave for 5 min at 121°C. Let cool down while stirring and pour plates under sterile conditions inside a laminar flow hood. Store ≤3 months at 4°C.

For YMS solid medium with 50 µg/ml kanamycin, cool down medium to ~50°C and add appropriate volume of kanamycin stock. Stir well and pour plates immediately. Store ≤1 month at 4°C.

Commentary

Background Information

Isolates of *Z. tritici* can switch their growth morphology when environmental conditions change. The fungus grows as blastospores (a "yeast-like" form), as filamentous hyphae, or as chlamydospores, depending on nutrient availability and temperature (Francisco et al., 2019). The molecular pathways that induce these morphological transitions are still largely unknown. Optimal culture conditions, i.e., nutrient-rich medium and an incubation temperature of 18°C, generally promote the formation of blastospores. Stress conditions such as nutrient-limited medium and temperatures between 25°C and 28°C induce filamentous hyphal growth in the *Z. tritici* reference strain IPO323 (Francisco et al., 2019; Motteram et al., 2011). It is possible that different morphotypes of the same isolate or mutant show different phenotypes in experimental assays. Hence, it is necessary to use standardized growth conditions to obtain consistent phenotypic results. Most laboratory approaches and infection experiments require yeast-like *Z. tritici* cells. Above, we have outlined protocols that produce such blastospore cultures for long-term storage (Basic Protocol 3) and as inoculum (Support Protocol 3) for *in planta* virulence assays (Basic Protocol 4).

In vitro growth and storage conditions influence the phenotype but can also impact the genome architecture of *Z. tritici* isolates. Exceptionally high rates of chromosome losses have been reported for *Z. tritici* after *in vitro* growth for only 3 weeks, and chromosome instability increased even more when the temperature was elevated to 28°C (Möller et al., 2018). These observations are biologically fascinating but clearly underline the importance of appropriate growth and storage conditions wherein karyotypes are maintained intact. Not only can chromosomes be lost, but spontaneous mutations may also occur. We observed this in a derived lineage of IPO323: the *Z. tritici* isolate Zt09, shown in Figure 5, has lost chromosome 18 as well as having partially lost the white collar complex (WCC) transcription factor gene homolog *wc-1* during *in vitro* propagation (Habig et al., 2019; Kellner et al., 2014; Möller et al., 2018).

The phenotypic plasticity that we observe in *Z. tritici* (Fig. 5) can explain the high level of variability between replicates in experiments and can bias comparative analyses of mutant and wild-type isolates. Phenotypic changes may be caused by small fluctuations in the local environment of the fungus. Abiotic factors like foliar temperature and moisture have been shown to affect infection outcomes and overall STB disease dynamics (Bernard, Sache, Suffert, & Chelle, 2013; Boixel,

Gélisse, Marcel, & Suffert, 2019). Biotic factors like host senescence, or the natural yellowing of leaves triggered by stress or aging-related factors, can also interfere with the quantification of *Z. tritici* virulence (Suffert, Sache, & Lannou, 2013). Hence, virulence phenotyping based on pycnidia densities has become more meaningful than leaf color–based readouts like PLACL and provides an indirect estimate of *Z. tritici* fitness. Experimental factors like wheat cultivar, inoculum cell density, and inoculum suspension time can further influence infection success (Fones, Steinberg, & Gurr, 2015; Kay, Fones, & Gurr, 2019; Suffert et al., 2013). Although many environmental and technical aspects can influence *Z. tritici* quantitative virulence, the method of inoculation (spray inoculation or paint-brush inoculation; Basic Protocol 4) seems to have little or no impact. We find that both methods can be used equivalently. However, we recommend that the two methods are not used in combination within one experiment. When establishing *Z. tritici* experimental infections, we recommend performing test infection experiments to select and adapt the inoculation method based on the *Z. tritici* isolates and the plant growth conditions that are available.

Although widely used as a model organism, many aspects of the biology of *Z. tritici* and its interaction with wheat are still largely unknown. Hence, the use of standardized experimental protocols is crucial for *Z. tritici* research. We hope that the suite of protocols, from field to laboratory, presented here can serve to improve comparability of experimental results between research groups and may provide a valuable guide for researchers new to *Z. tritici* research.

Critical Parameters and Troubleshooting

As emphasized throughout this article, a range of biological and technical aspects can have significant impacts on *Z. tritici* data and their interpretation. Genetic and phenotypic diversity, morphological plasticity, and large variation in virulence responses are common characteristics of *Z. tritici* and can bias the interpretation of results if not accounted for.

Chromosome losses and mutations are possible during prolonged periods of mitotic cell divisions in *Z. tritici* (Habig et al., 2019; Möller et al., 2018). Thus, strict temporal limitation of *in vitro* cultivation and proper long-term storage of *Z. tritici*, as shown in Basic Protocol 3, are essential to avoid the occurrence of aneuploidy and mutation accumulation in laboratory strains. Together with these measures, karyotype analyses using pulsed- field gel electrophoresis (PFGE) (Habig et al., 2017; Haueisen et al., 2019) and qualitative PCR assays using specific primer pairs for each accessory chromosome (Habig et al., 2017) can be applied to assess changes in chromosome composition.

In virulence assays of *Z. tritici*, quantitative disease measures like pycnidia density and lesion size can vary greatly between isolates (Habig et al., 2017; Stewart & McDonald, 2014). As discussed in Basic Protocol 4, experimental replicates and randomization of treatments are important to reduce both variability within *Z. tritici* treatments and "background noise" caused by abiotic and biotic factors in the experiment. Additionally, the use of experimental controls is extremely important. Negative-control treatments, i.e., "mock" controls, compensate for possible artifacts due to abiotic or technical influences and may help to spot possible cross-contaminations between treatments, mainly when using spray inoculation methods. A control for reduced virulence together with a positive control for disease can also be helpful, for example the *Z. tritici Mg3LysM* deletion mutant, which is impaired in leaf colonization and pycnidia production compared to a wild-type strain (Lee, Rudd, Hammond-Kosack, & Kanyuka, 2014; Marshall et al., 2011).

Apart from technical and experimental aspects, robust virulence quantification methods are essential for *Z. tritici* pathogenicity assays (Basic Protocol 4). High-throughput analysis of scanned leaf images provides a robust and reliable method (Karisto et al., 2018; Stewart & McDonald, 2014; Stewart et al., 2016). This yields readouts of disease symptoms (based on PLACL, for example) and an indirect measure of the potential degree of *Z. tritici* reproduction in host tissues (i.e., based on pycnidia density per leaf area) that are more accurate and precise than manual visual methods. Image analysis has two main advantages over visual estimates: (i) objectivity and reproducibility independent from the observer/experimenter and (ii) quantitative readouts that allow for more powerful statistical analysis methods. As for any other automated analyses, few adjustments (e.g., healthy leaf and lesion color thresholds) are necessary depending on the conditions of the scanned samples. Additionally, the results should always be double-checked for any spurious outcomes. Soil particles in scanned leaves, in particular, can introduce bias, as they can be counted as pycnidia and give an overestimation of the normalized pycnidia density values.

Understanding Results

Please see the introductions of the individual protocols and the Critical Parameters and Troubleshooting section for this information.

Time Considerations

Please see the time considerations in the individual protocols, including potential stopping points.

Acknowledgments

The authors would like to thank Michael Habig and Mareike Möller for helpful comments on the previous versions of the protocols and Marcello Zara for support in optimizing the CTAB DNA extraction protocol. This work was supported by the Canadian Institute for Advanced Research (CIFAR) and a Max Planck fellowship to EHS. Open access funding enabled and organized by Projekt DEAL.

Author Contributions

Wagner C. Fagundes: Conceptualization; investigation; methodology; validation; visualization; writing-original draft; writing-review & editing. **Janine Haueisen**: Conceptualization; formal analysis; investigation; methodology; supervision; writing-review & editing. **Eva H. Stukenbrock**: Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision; writing-review & editing.

Literature Cited

Abrinbana, M. (2017). Variation in aggressiveness components of *Zymoseptoria tritici* populations in Iran. *Journal of Phytopathology*, 166(1), 10– 17. doi: 10.1111/jph.12654.

Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., & Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols*, 1(5), 2320– 2325. doi: 10.1038/nprot.2006.384.

Altman, D., & Bland, J. (1995). Statistics notes: The normal distribution. *BMJ*, 310, 298. doi: 10.1136/bmj.310.6975.298.

Armstrong, J. A., and Schulz, J. R. (2015). Agarose gel electrophoresis. *Current Protocols Essential Laboratory Techniques*, 10, 7.2.1–7.2.22. doi: 10.1002/9780470089941.et0702s10.

Badet, T., Oggenfuss, U., Abraham, L., Mc- Donald, B. A., & Croll, D. (2020). A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici. BMC Biology*, 18(1), 1–18. doi: 10.1186/ s12915-020-0744-3.

Banke, S., Peschon, A., & McDonald, B. A. (2004). Phylogenetic analysis of globally distributed *Mycosphaerella graminicola* populations based on three DNA sequence loci. *Fungal Genetics and Biology*, 41(2), 226–238. doi: 10.1016/j. fgb.2003.09.006.

Bernard, F., Sache, I., Suffert, F., & Chelle, M. (2013). The development of a foliar fungal pathogen does react to leaf temperature! *New Phytologist*, 198(1), 232–240. doi: 10.1111/nph. 12134.

Boixel, A., Gélisse, S., Marcel, T. C., & Suffert, F. (2019). Differential tolerance to changes in moisture regime during early infection stages in the fungal pathogen *Zymoseptoria tritici*. *bio*R*xiv*. doi: 10.1101/867572.

Brown, J. K. M., Chartrain, L., Lasserre-Zuber, P., & Saintenac, C. (2015). Genetics of resistance to *Zymoseptoria tritici* and applications to wheat breeding. *Fungal Genetics and Biology*, 79, 33– 41. doi: 10.1016/j.fgb.2015.04.017.

Chartrain, L., Brading, P. A., & Brown, J. K. M. (2005). Presence of the *Stb6* gene for resistance to *Septoria tritici* blotch (*Mycosphaerella graminicola*) in cultivars used in wheat-breeding programmes worldwide. *Plant Pathology*, 54(2), 134–143. doi: 10.1111/j. 1365-3059.2005.01164.x.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3), 124–129. doi: 10.1080/00031305. 1981.10479327.

Crous, P. W., Aptroot, A., Kang, J. C., Braun, U., & Wingfield, M. J. (2000). The genus *Mycosphaerella* and its anamorphs. *Studies in Mycology*, 2000(45), 107–121.

Eyal, Z., Scharen, A. L., Prescott, J. M., & van Ginkel, M. (1987). The *Septoria* diseases of wheat: *Concepts and methods of disease management*. Mexico: CIMMYT.

Fones, H., & Gurr, S. (2015). The impact of *Septoria tritici* Blotch disease on wheat: An EU perspective. *Fungal Genetics and Biology*, 79, 3–7. doi: 10.1016/j.fgb.2015.04.004.

Fones, H. N., Steinberg, G., & Gurr, S. J. (2015). Measurement of virulence in *Zymoseptoria tritici* through low inoculum-density assays. *Fungal Genetics and Biology*, 79, 89–93. doi: 10. 1016/j.fgb.2015.03.020.

Fraaije, B. A., Lovell, D. J., Rohel, E. A., & Hollomon, D. W. (1999). Rapid detection and diagnosis of Septoria tritici epidemics in wheat using a polymerase chain reaction/PicoGreen assay. *Journal of Applied Microbiology*, 86(4), 701–708. doi: 10.1046/j. 1365-2672.1999.00716.x.

Francisco, C. S., Ma, X., Zwyssig, M. M., McDonald, B. A., & Palma-Guerrero, J. (2019). Morphological changes in response to environmental stresses in the fungal plant pathogen *Zymoseptoria tritici*. *Scientific Reports*, 9(1), 1–18. doi: 10.1038/s41598-019-45994-3.

Glass, N. L., & Donaldson, G. C. (1995). Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Applied and Environmental Microbiology*, 61(4), 1323–1330. doi: 10.1128/ aem.61.4.1323-1330.1995.

Goodwin, S. B., M'Barek, S. B., Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., … Kema, G. H. J. (2011). Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLOS Genetics*, 7(6), e1002070. doi: 10.1371/ journal.pgen.1002070.

Goodwin, S. B., Waalwijk, C., & Kema, G. H. J. (2004). Genetics and genomics of *Mycosphaerella graminicola*, a model for the Dothideales. In D. K. Arora, &G. G. Khachatourians (Eds.), Applied Mycology and Biotechnology. Volume 4: Fungal Genomics (pp. 315–330). Elsevier. doi: 10.1016/S1874-5334(04)8001 6-9.

Habig, M., Bahena-Garrido, S. M., Barkmann, F., Haueisen, J., & Stukenbrock, E. H. (2019). The transcription factor Zt107320 affects the dimorphic switch, growth and virulence of the fungal wheat pathogen *Zymoseptoria tritici*. *Molecular Plant Pathology*, 21(1), 124–138. doi: 10.1111/ mpp.12886.

Habig, M., Quade, J., & Stukenbrock, E. H. (2017). Forward genetics approach reveals host genotype-dependent importance of accessory chromosomes in the fungal wheat pathogen *Zymoseptoria tritici*. *mBio*, 8(6), 1–16. doi: 10. 1128/mBio.01919-17.

Hartmann, F. E., McDonald, B. A., & Croll, D. (2018). Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Molecular Ecology*, 27(12), 2725–2741. doi: 10.1111/mec.14711.

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME Journal*, 11(5), 1189–1204. doi: 10.1038/ismej.2016.196.

Haueisen, J., Möller, M., Eschenbrenner, C. J., Grandaubert, J., Seybold, H., Adamiak, H., & Stukenbrock, E. H. (2019). Highly flexible infection programs in a specialized wheat pathogen. *Ecology and Evolution*, 9(1), 275– 294. doi: 10.1002/ece3.4724.

Jürgens, T., Linde, C. C., & McDonald, B. A. (2006). Genetic structure of *Mycosphaerella graminicola* populations from Iran, Argentina and Australia. *European Journal of Plant Pathology*, 115(2), 223–233. doi: 10.1007/ s10658-006-9000-0.

Karisto, P., Dora, S., & Mikaberidze, A. (2019). Measurement of infection efficiency of a major wheat pathogen using time-resolved imaging of disease progress. *Plant Pathology*, 68(1), 163–172. doi: 10.1111/ppa.12932.

Karisto, P., Hund, A., Yu, K., Anderegg, J., Walter, A., Mascher, F., …Mikaberidze, A. (2018). Ranking quantitative resistance to *Septoria tritici* blotch in elite wheat cultivars using automated image analysis. *Phytopathology*, 108(5), 568–581. doi: 10.1094/PHYTO-04-17-0163-R.

Kay, W. T., Fones, H. N., & Gurr, S. J. (2019). Rapid loss of virulence during submergence of *Z. tritici* asexual spores. *Fungal Genetics and Biology*, 128, 14–19. doi: 10.1016/j.fgb.2019. 03.004.

Kellner, R., Bhattacharyya, A., Poppe, S., Hsu, T. Y., Brem, R. B., & Stukenbrock, E. H. (2014). Expression profiling of the wheat pathogen *Zymoseptoria tritici* reveals genomic patterns of transcription and host-specific regulatory programs. *Genome Biology and Evolution*, 6(6), 1353–1365. doi: 10.1093/gbe/evu101.

Kema, G. H. J., Mirzadi Gohari, A., Aouini, L., Gibriel, H. A. Y., Ware, S. B., van Den Bosch, F., … Seidl, M. F. (2018). Stress and sexual reproduction affect the dynamics of the wheat pathogen effector *AvrStb6* and strobilurin resistance. *Nature Genetics*, 50, 1–6. doi: 10.1038/s41588-018-0052-9

Kema, Gert H. J., & van Silfhout, C. H. (1997). Genetic variation for virulence and resistance in the wheat-*Mycosphaerella graminicola* pathosystem. III. Comparative seedling and adult plant experiments. *Phytopathology*, 87(3), 266–272. doi: 10.1094/PHYTO.1997.87.3.2 66.

Large, E. C. (1954). Growth stages in cereals illustration of the Feekes scale. *Plant Pathology*, 3(4), 128–129. doi: 10.1111/j.1365-3059.1954. tb00716.x.

Lee,W. S., Rudd, J. J., Hammond-Kosack, K. E., & Kanyuka, K. (2014). *Mycosphaerella graminicola LysM* effector-mediated stealth pathogenesis subverts recognition through both CERK1 and CEBiP homologues in wheat. *Molecular Plant-Microbe Interactions*, 27(3), 236–243. doi: 10. 1094/MPMI-07-13-0201-R.

Levene, H. (1961). Robust tests for equality of variances. In I. Olkin et al. (Eds.), Contributions to probability and statistics: *Essays in honor of Harold Hotelling* (pp. 279–292). Palo Alto, CA: Stanford University Press.

Linde, C. C., Zhan, J., & McDonald, B. A. (2002). Population structure of *Mycosphaerella graminicola*: From lesions to continents. *Phytopathology*, 92(9), 946–955. doi: 10.1094/ phyto.2002.92.9.946.

Marshall, R., Kombrink, A., Motteram, J., Loza- Reyes, E., Lucas, J., Hammond-Kosack, K. E., … Rudd, J. J. (2011). Analysis of two in planta expressed *LysM* effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiology*, 156(2), 756–769. doi: 10.1104/pp.111. 176347.

McDonald, B. A. (1997). The population genetics of fungi: Tools and techniques. *Phytopathology*, 87, 448–453. doi: 10.1094/PHYTO.1997.87.4. 448.

Möller, M., Habig, M., Freitag, M., & Stukenbrock, E. H. (2018). Extraordinary genome instability and widespread chromosome rearrangements during vegetative growth. *Genetics*, 210(2), 517–529. doi: 10.1534/genetics. 118.301050.

Möller, M., Schotanus, K., Soyer, J. L., Haueisen, J., Happ, K., Stralucke, M., … Stukenbrock, E. H. (2019). Destabilization of chromosome structure by histone H3 lysine 27 methylation. *PLOS Genetics*, 15(4), e1008093. doi: 10.1371/ journal.pgen.1008093.

Motteram, J., Lovegrove, A., Pirie, E., Marsh, J., Devonshire, J., van de Meene, A., … Rudd, J. J. (2011). Aberrant protein N-glycosylation impacts upon infection-related growth transitions of the haploid plant-pathogenic fungus *Mycosphaerella graminicola. Molecular Microbiology*, 81(2), 415–433. doi: 10.1111/j. 1365-2958.2011.07701.x.

O'Donnell, K., & Cigelnik, E. (1997). Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Molecular Phylogenetics and Evolution*, 7(1), 103–116. doi: 10.1006/mpev.1996.0376.

Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F. E., Singh, N. K., Abraham, N., … Biology, M. (2020). A population-level invasion by transposable elements in a fungal pathogen. *bioRxiv*, doi: 10.1101/2020.02.11.944652

Ponomarenko, A., Goodwin, S. B., & Kema, G. H. J. (2011). *Septoria tritici* blotch (STB) of wheat. *Plant Health Instructor*, 1–7. doi: 10. 1094/PHI-I-2011-0407-01.

Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. (2015). Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLOS Pathogens*, 11(7), 1–21. doi: 10.1371/journal. ppat.1005055.

Quaedvlieg, W., Kema, G. H. J., Groenewald, J. Z., Verkley, G. J. M., Seifbarghi, S., Razavi, M., … Crous, P. W. (2011). *Zymoseptoria* gen. nov.: A new genus to accommodate *Septoria*-like species occurring on graminicolous hosts. *Persoonia: Molecular Phylogeny and Evolution of Fungi*, 26, 57–69. doi: 10. 3767/003158511X571841.

R Core Team (2013). R: A language and environment for statistical computing.

Saintenac, C., Lee, W. S., Cambon, F., Rudd, J. J., King, R. C., Marande, W., … Kanyuka, K. (2018). Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to fungal pathogen *Zymoseptoria tritici*. *Nature Genetics*, 50, 1–7. doi: 10.1038/s41588-018-0051-x.

Schneider, C.A., Rasband, W. S.,&Eliceiri,K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671–675. doi: 10.1038/nmeth.2089.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591. doi: 10.2307/ 2333709.

Stewart, E. L., Hagerty, C. H., Mikaberidze, A., Mundt, C. C., Zhong, Z., & McDonald, B. A. (2016). An improved method for measuring quantitative resistance to the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. *Phytopathology*, 106(7), 782–788. doi: 10.1094/ PHYTO-01-16-0018-R.

Stewart, E. L., &McDonald, B. A. (2014). Measuring quantitative virulence in the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. *Phytopathology*, 104(9), 985–992. doi: 10.1094/PHYTO-11-13- 0328-R.

Stukenbrock, E. H., Banke, S., Javan-Nikkhah, M., & McDonald, B. A. (2007). Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, 24(2), 398–411. doi: 10.1093/molbev/msl1 69.

Suffert, F., Sache, I., & Lannou, C. (2013). Assessment of quantitative traits of aggressiveness in *Mycosphaerella graminicola* on adult wheat plants. *Plant Pathology*, 62(6), 1330–1341. doi: 10.1111/ppa.12050.

Truett, G. E., Heeger, P., Mynatt, R. L., Truett, A. A., Walker, J. A., & Warman, M. L. (2000). Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (Hot-SHOT). *BioTechniques*, 29(1), 52–54. doi: 10. 2144/00291bm09.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99. doi: 10.2307/3001913.

Weir, B. S., Johnston, P. R., & Damm, U. (2012). The *Colletotrichum gloeosporioides* species complex. *Studies in Mycology*, 73, 115–180. doi: 10.3114/sim0011.

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications*, 18(1), 315–322.

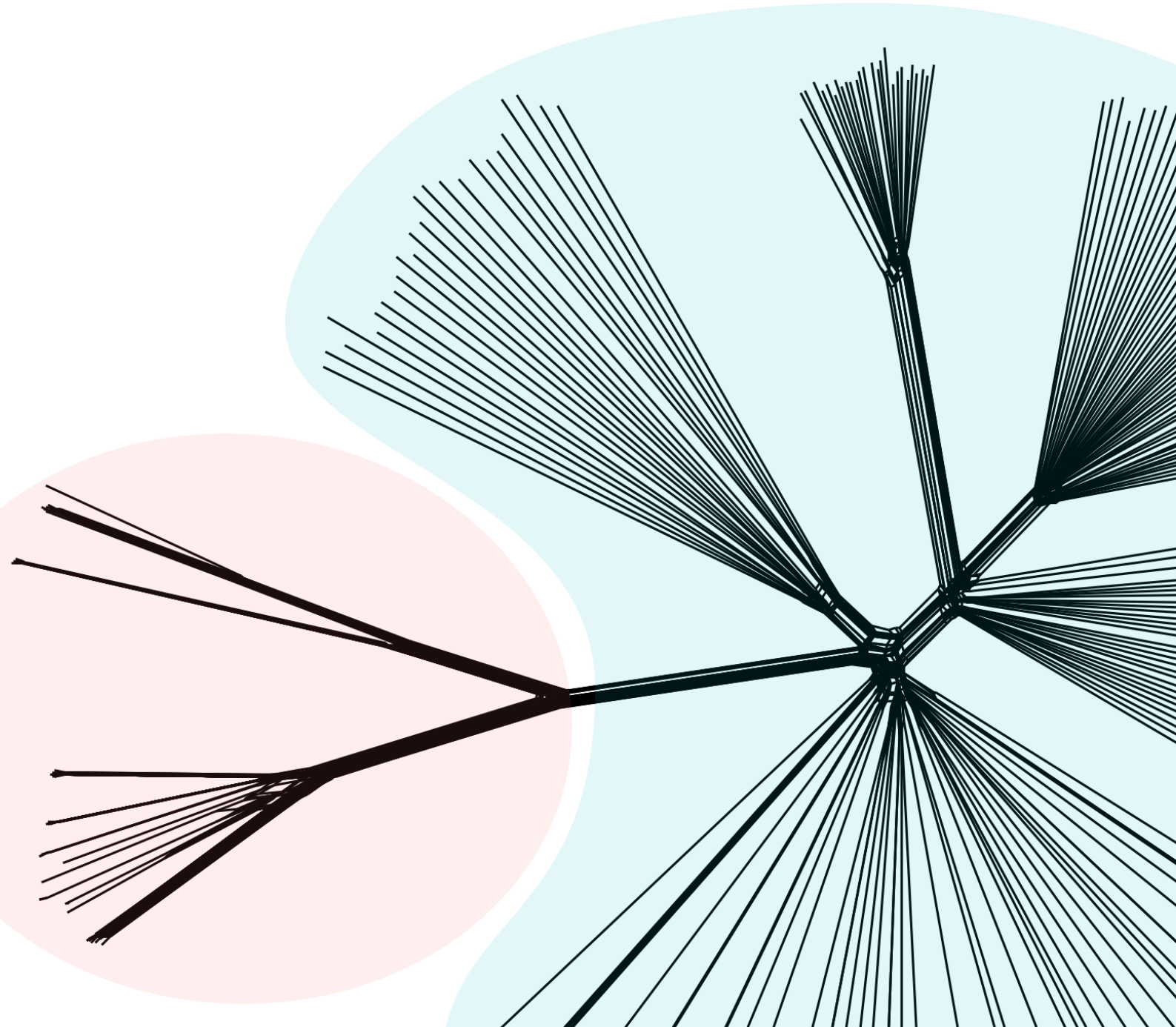
Zhan, J., Linde, C. C., Jürgens, T., Merz, U., Steine- brunner, F., & McDonald, B. A. (2005). Variation for neutral markers is correlated with variation for quantitative traits in the plant pathogenic fungus *Mycosphaerella graminicola*. *Molecular Ecology*, 14(9), 2683–2693. doi: 10.1111/j. 1365-294X.2005.02638.x.

Zhan, J., Pettway, R. E., & McDonald, B. A. (2003). The global genetic structure of the wheat pathogen *Mycosphaerella graminicola* is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. *Fungal Genetics and Biology*, 38(3), 286–297. doi: 10.1016/S1087-1845(02) 00538-8.

Zhong, Z., Marcel, T. C., Hartmann, F. E., Ma, X., Plissonneau, C., Zala,M.,…Palma-Guerrero, J. (2017). A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the *Stb6* resistance gene. *New Phytologist*, 214(2), 619–631. doi: 10.1111/nph. 14434.

Chapter II

Host specificity defines a new fungal plant pathogen population



Chapter II

Host specificity defines a new fungal plant pathogen population

Research paper manuscript

Authors: **Wagner C. Fagundes**, Idalia C. Rojas Barrera, Rune Hansen, Alice Feurtey, Fatemeh Salimi, Janine Haueisen, Alireza Alizadeh and Eva H. Stukenbrock

Abstract

Host-driven mutations can be considered strong drivers of pathogen evolution. To successfully infect, colonize and complete the life cycle, plant pathogens are under constant selective pressures imposed by hosts, leading to adaptative genomic evolution and possibly pathogen species or lineages radiation. Population and comparative genomics approaches are powerful tools to identify signatures of selection associated with host adaptation in pathogens genomes and further allow to recapture their evolutionary history. Implementing such approaches, we identified evolutionary and genomic patterns of host specialisation in fungal plant pathogens using the fungal pathogen *Zymoseptoria tritici* as a model of study. Unique collections of *Z. tritici* were isolated from wild (*Aegilops* spp.) and domesticated (*Triticum aestivum*) host grasses in the Middle East and whole-genome sequencing was performed in a subset of isolates from each collection. We observed distinct population structure between the two host-diverging collections and particular genomic features in the *Aegilops*-infecting isolates that may have shaped their evolutionary history. Phylogenetic analyses indicated that the *Aegilops*-infecting population forms a separate *Z. tritici* cluster when compared to worldwide collections of *Z. tritici* and to closely-related species. Population genomic analyses and demographic inference allowed us to characterize the distinct populations and to detect signatures of recent selection. Moreover we find evidence that divergence of the *Z. tritici* populations on wild and domesticated hosts likely coincided with wheat domestication. Using infection experiments, we show that *Z. tritici* isolates collected from *Aegilops* spp. only infect their respective host species and not *T. aestivum*. Together with other aspects, our findings highlight the interplay between agricultural and wild hosts on the evolution of fungal plant pathogens and illustrate a possible route of new pathogen lineages emergence.

Introduction

Host specialization is considered a strong driver of pathogen evolution (de Vienne, Hood, & Giraud, 2009; de Vries, Stukenbrock, & Rose, 2020; Tatiana Giraud, Gladieux, & Gavrilets, 2010; Gladieux et al., 2011). Host range is determined by a variety of extrinsic and intrinsic factors that govern the pathways for a successful host specialization and pathogen colonization. Extrinsic factors as geographical or ecological barriers can prevent host-pathogen contact while intrinsic physiological, molecular and biochemical properties of a specific host (i.e., resistance genes, antimicrobial metabolites, etc.) can determine the extent and duration of compatible interactions (de Vienne et al., 2009; Jaenike, 1985). Adaptation to a novel host can occur due to spill-over (cross-species disease transmission without establishment of pathogen population in the new host); host range expansion (colonization of a new host species while still colonizing the ancestral host); or host shift (colonization of a new host coupled with the loss of ability to infect the ancestral host) (Gladieux et al., 2011). Depending on the intrinsic and extrinsic factors affecting the pathogen population on a specific host, reproductive isolation following host shift may be considered a driver of pathogen speciation and regarded as a special case of ecological speciation (Giraud et al., 2010; Rundle & Nosil, 2005; Schluter, 2000). Not only specialist pathogens, but also generalist pathogens must adapt to overcome host defenses (Stukenbrock, 2013). However, selection for host specialization occurs at every generation in a specialist pathogen, but infrequently in generalists infecting distinct host species, which increases the chances of a specialist pathogen to persist in a certain host (de Vienne et al., 2013). Therefore, for plant pathogens, different host species can be considered divergent ecological niches (Gladieux et al., 2011; McCoy, 2003; Silva et al., 2012), in which host specialization sets the stage for adaptive evolution and divergent selection on pathogen populations.

Plant pathogens deploy several mechanisms to adapt to their hosts. The adaptative evolution of host-pathogen interactions can be described by the "trench-warfare" evolution scenario, which depicts evolution through balancing selection, and the "arms-race" evolution scenario, which refers to evolution by directional selection (Brown & Tellier, 2011; Möller & Stukenbrock, 2017; Stukenbrock & McDonald, 2009; Tellier, Moreno-Gámez, & Stephan, 2014). In these evolutionary models, hosts evolve different defense mechanisms to target specific pathogen genotypes while pathogens evolve virulence strategies that can evade host recognition leading to a constant coevolutionary fight (Jones & Dangl, 2006). Virulence-related genes in pathogen genomes generally encode small, secreted proteins with often no sequence homology known as effectors (Lo Presti et al., 2015). These small molecules can modulate host cell structure and plant

metabolism and therefore evade host defense responses through several and diverse functions, sometimes in a highly specific molecular dialogue (Raffaele & Kamoun, 2012; Stergiopoulos & de Wit, 2009). All hemibiotroph and biotroph pathogens studied so far have shown extensive effector reservoirs and have evolved different mechanisms to rapidly variate this gene repertoire, particularly in regards of host adaptation (Bertazzoni et al., 2018; Dong, Raffaele, & Kamoun, 2015; Thines, 2019). Besides effectors, different classes of proteins (e.g. plant cell wall–degrading enzymes - PCWDEs and carbohydrate-degradading enzymes – CAZymes) and secondary metabolites (e.g. toxins) play also an important role in such host-pathogen interactions and therefore can be also under selective pressures imposed by hosts (Lo Presti et al., 2015). Footprints of strong positive selection have already been found in effector genes (Aguileta, Refregier, Yockteng, Fournier, & Giraud, 2009) as well as in toxins and genes encoding PCWDEs (Brunner, Torriani, Croll, Stukenbrock, & McDonald, 2013; McDonald, Oliver, Friesen, Brunner, & McDonald, 2013) in pathogen populations, suggesting the importance of such gene categories for pathogen fitness and adaptation to hosts.

Distinct methods can help to detect signatures of host specialization in pathogen populations. Population genomics approaches in particular have the power to identify signatures of selection associated with speciation and host adaptation events besides recapturing the evolutionary history of species (Stukenbrock, 2013). Methods that assess population divergence such as the Fixation index (e.g, Wright's Fst) and nucleotide diversity as Tajima's D and π are applicable and provide insights into the evolutionary history of plant pathogens (Aguileta et al., 2009; Grünwald, McDonald, & Milgroom, 2016; Meirmans & Hedrick, 2011; Plissonneau et al., 2017). Adaptive changes and recent positive selection (e.g. selective sweeps), for example, can be identified through genomic scans as regions with skews in allele frequency distribution and distinct levels of linkage disequilibrium (Braverman, Hudson, Kaplan, Langley, & Stephan, 1995; Smith & Haigh, 1974; Ronald & Akey, 2005). Selective sweep scans have been previously applied to fungal populations and identified several outlier regions containing genes potentially involved in host adaptation (Badouin et al., 2017) and adaptation to a range of environmental factors (Branco et al., 2017, 2015; Ellison et al., 2011; Hartmann, McDonald, & Croll, 2018; Mohd-Assaad, McDonald, & Croll, 2018). Application of these tools to understand evolutionary processes involved in host adaptation and emergence of new pathogen species is crucial to access the risks of potential future pathogens, particularly in agricultural ecosystems (Stukenbrock & McDonald, 2008).

Zymoseptoria tritici and closely related species have served as reference organisms for evolutionary genetic studies of crop-infecting fungal pathogens. *Zymoseptoria tritici* is a widely distributed wheat

pathogen and the most devasting pathogen for wheat production in Europe (Fones & Gurr, 2015). The fungus causes the disease *Septoria tritici* blotch in wheat leaves, a disease that is characterized by a latent phase of biotrophy followed by a necrotic phase, in which host cell death and lesions become apparent. *Zymoseptoria tritici* reproduces both asexually and sexually and can undergo several cycles of reproduction in a year (Eyal, 1999). The levels of genetic variation and gene flow found in populations of *Z. tritici* around the world are very high and levels of genetic variation within wheat fields are comparable to the level of variation found at global scales, indicating large effective population sizes and a high adaptive potential (Linde, Zhan, & McDonald, 2002; Zhan, Pettway, & McDonald, 2003). A study reconstructing the evolutionary history of *Z. tritici* indicated that it originated from a wild grasses-associated ancestor in the fertile crescent region in the Middle East around 11,000 years ago, coinciding with the domestication of wheat and hence providing a great example of species emergence due to host specialization (Banke & McDonald, 2005; Stukenbrock, Banke, Javan-Nikkhah, & McDonald, 2007). Endemic relatives to *Z. tritici* are still found in the Middle East, however on different wild grass hosts (Stukenbrock et al., 2012). As example of such species, *Z. pseudotritici* and *Z. ardabiliae* are described as the closest known relatives of *Z. tritici* originally isolated from wild grass hosts in Iran (Stukenbrock et al., 2012; Stukenbrock et al., 2011). Comparative genomics analysis between "wild" *Z. pseudotritici* and the "domesticated" *Z. tritici* strains revealed patterns of hybridization in the wild species probably followed by a strong population bottleneck, leading to the emergence of this new species (Stukenbrock, Christiansen, Hansen, Dutheil, & Schierup, 2012). Moreover, recent studies suggest that recurrent introgression events between closely related species of *Zymoseptoria* occur at the center of origin in the Middle East and can leave footprints of nonmonophyletic high variable regions in the genome (Feurtey, Stevens, Stephan, & Stukenbrock, 2019; Moller et al., 2021). However, the evolutionary mechanisms underlying such host-pathogen interactions in wild and agricultural ecosystems are still poorly understood in the *Zymoseptoria* genus.

In this study, we aimed to identify evolutionary and molecular signatures of host adaptation using host-divergent *Z. tritici* isolates as models of study. Unique collections of *Z. tritici* strains have been isolated from uncultivated wild grasses belonging to the *Aegilops* genus as well as from the cultivated common wheat (*Triticum aestivum*) in the Middle East. Our hypotheses were that the different host species were determinants of the overall population and genomic structure of these host-diverging isolates and that distinct molecular and evolutionary mechanisms were underpinning their host adaptation. Implementing population genomics analyses, we observed distinct population structure between the sympatric host-diverging collections and particular genomic features in the *Aegilops*-infecting *Z. tritici* isolates that may have shaped their evolutionary

history. Genomic scans for selective sweeps identified diverse genomic regions under recent positive selection containing candidate genes potentially involved in host adaptation. Moreover, demographic inference suggests that divergence of the *Z. tritici* populations on wild and domesticated hosts likely coincided with wheat domestication. Using comparative infection experiments, we also show that *Z. tritici* isolates collected from *Aegilops* spp. could only infect *Aegilops* species and not *T. aestivum*, indicating a high degree of host specialization. Taken together, our results represent a unique analysis on host-diverging *Z. tritici* populations and suggest the divergence of a host-specific *Z. tritici* lineage.

Material and Methods

Fungal isolates and growth conditions

Aegilops spp. and wheat (*Triticum aestivum* L.) samples presenting *Zymoseptoria* symptoms were collected in the fertile crescent region of the Middle East at different locations in Iran (Supplementary Table S1). Fungal specimens were isolated, stored in cryo-stocks at -80˚C and the species confirmed using *ITS* and *Beta-tubulin* amplicon sequencing following protocols and primers described previously (Fagundes, Haueisen, & Stukenbrock, 2020). A total of 310 isolates were collected (135 from *Aegilops* spp. and 175 from wheat). All fungal isolates were grown at 18˚C in YMS medium (4 g yeast extract, 4 g malt, 4 g sucrose per 1 L, 20 g agar per L for plates) (Fagundes et al., 2020). For DNA extraction, fungal cells were retrieved directly from -80˚C glycerol stocks, grown in liquid YMS for 5 days at 18˚C and 200 rpm, and harvested by centrifugation (3,500 rpm for 10 min).

DNA extraction and whole-genome sequencing

High-quality genomic DNA extractions were performed for all 310 *Z. tritici* isolates following a CTAB DNA extraction protocol described previously (Fagundes et al., 2020). Prior to whole-genome sequencing, isolates were clone-corrected by PCR using ISSR (Inter Simple-Sequence Repeat) markers (Czembor & Arseniuk, 1999; Ratanacherdchai et al., 2010; Nirmaladevi et al. 2016). The list of primers used is described in Supplementary Table S2. Isolates originated from the same leaf lesion presenting the same ISSR pattern were considered clones and therefore disregarded for genome sequencing. A subset of 123 clone-corrected isolates (48 isolated from *Aegilops* spp. and 75 isolated from wheat; Supplementary Table S1) were then submitted for whole-genome sequencing using paired-end reads of 150bp on an Illumina HiSeq 3000 platform. Library

preparations and sequencing were performed at the Max Planck Genome Center, Cologne, Germany (http://mpgc.mpipz. mpg.de).

Read mapping and variant calling

Raw sequencing reads were first trimmed for adapters and sequencing quality. Raw reads were trimmed using Trimmomatic v 0.39 (Bolger, Lohse, & Usadel, 2014) using the following parameters: LEADING:20 SLIDING- WINDOW:4:30 AVGQUAL:30 MINLEN:50. Trimmed reads of all isolates were mapped to the *Z. tritici* reference genome IPO323 (Goodwin et al., 2011) using the short-read aligner bwa-mem v. 0.7.17 (Li & Durbin, 2010). Conversion, sorting, merging, and indexing of alignment files were performed using SAMtools v. 1. 13 (Danecek et al., 2021). PCR duplicates and read groups were determined using Picard tools v. 2.26.2 (https://broadinstitute.github.io/picard/). Single nucleotide polymophism (SNP) calling and variant filtration were carried out using the Genome Analysis Toolkit (GATK) v. 4.1.4.1 (McKenna et al., 2010). The GATK HaplotypeCaller was first used on each isolate individually with the commands -ERC GVCF and -ploidy 1. Then, gVCFs were merged using the GATK CombineGVCFs program and joint variant calling was performed on the merged gvcf variant file using GATK GenotypeGVFs. Only SNPs were maintained on the joint variant call file (VCF) using the GATK SelectVariants tool with the command --select-type-to-include SNP. We performed hard-filtering on SNPs based on quality cut-offs and following the GATK Best Practices recommendations (Auwera et al., 2013) using the GATK VariantFiltration and SelectVariants tools. First, we filtered low depth genotypes using the options --genotype-filter-expression "DP < 3", --set-filtered-genotype-to-no-call and --genotype-filter-name "low_depth". Then, the following filter conditions were applied: DP > 5000.0; QD < 20.0; MQ < 50.0; FS > 20.0; ReadPosRankSum, MQRankSum, and BaseQRankSum between -2 and 2. SNPs that fail to pass these criteria were subsequently removed from the VCF file using GATK SelectVariants with the options --exclude-non-variants, --remove-unused-alternates and --exclude-filtered. Genotyping accuracy of GATK was previously analyzed in *Z. tritici* populations and shown to be highly congruent with other SNP callers (Hartmann, McDonald, & Croll, 2018). We further excluded SNPs located on accessory chromosomes (i.e., chromosomes not present in all isolates) and retained only biallelic SNPs with a genotyping rate of ≥ 90% using VCFtools v 0.1.13 (Danecek et al., 2011). Based on the 1,357,300 remaining SNPs, we computed the pairwise genomic relatedness among the 123 isolates by constructing an Identity-By-State (IBS) similarity matrix using PLINK v. 1.07 (Purcell et al., 2007). Individuals that showed an IBS similarity > 0.9999 to any other were excluded for further analysis as they represent highly similar isolates and

potentially clones. This filtered dataset of 1,355,994 biallelic SNPs across 85 individuals was used for the subsequent population genomics analyses.

Linkage disequilibrium analysis

In order to access the extent of linkage disequilibrium (LD) for the *Z. tritici* collections (*Aegilops*- and wheat-infecting), we estimated the LD decay for SNPs located on chromosome 1. To this end, we retained only SNPs with a minor allele frequency (MAF) of ≥ 5% in each individual collection using VCFtools v 0.1.13 (Danecek et al., 2011). We calculated the LD coefficient of correlation (r^2) between all markers pairs up to a distance of 20 kb or up to 100 kb using PLINK v. 1.07 (Purcell et al., 2007). The analysis was performed in each collection individually to prevent possible population structure and demography interference. The pairwise r^2 value was then plotted over genomic distance using the package 'ggplot2' v. 3.3.2 (Wickham, 2016) in R software v. 3.6.3 (Team, 2013). We also determined the approximate genomic distance required to reach 50% LD decay for each collection. R script to plot LD results is available at Supplementary Text S1.

Population structure and population genomics analyses

Different methods to estimate the population structure between the wheat- and *Aegilops*-infecting *Z. tritici* isolates were used. For these analyses, we retained SNPs with a MAF of ≥ 5% and further pruned the filtered SNP dataset to select genome-wide SNPs at equidistant intervals of 12 kb along the chromosomes using VCFtools v 0.1.13 (Danecek et al., 2011). This pruning step ensured the removal of SNPs in strong linkage disequilibrium based on our analysis and on previous reports of linkage disequilibrium decay in *Z. tritici* populations (Croll, Lendenmann, Stewart, & McDonald, 2015; Hartmann, Sánchez-Vallet, McDonald, & Croll, 2017). Based on the 2507 remaining SNPs, we performed a Principal Component Analysis (PCA) using the R Bioconductor package 'SNPRelate' v. 1.20.1 (Zheng et al., 2012). Population structure was further inferred using the individual ancestry coefficients calculated by the sNMF v. 2.0 algorithm implemented in the R package 'LEA' (Frichot & François, 2015; Frichot, Mathieu, Trouillon, Bouchard, & François, 2014). The *K* parameter was tested ranging from 1 to 10 with 100 repetitions for each tested *K* value. The best run was chosen and plotted based on the lowest cross-entropy value across the 100 repetitions for each *K*. Finally, we calculated nucleotide diversity per site (π) (Nei & Li, 1979) in 100 kb sliding windows using the LD-unpruned SNP dataset and a MAF > 5% for each *Z. tritici* collection individually. Genotypic diversity calculations were obtained using VCFtools v 0.1.13 (Danecek et al., 2011) with the haploid mode fork provided by Julien Y. Dutheil

(https://github.com/vcftools/vcftools/pull/69). R script for population structure and population genomics analyses is available at Supplementary Text S1.

Whole genome de novo assemblies, mating types and phylogenomics analyses

To understand the evolutionary relationship between the sequenced isolates, we analyzed draft genome assemblies of different *Zymoseptoria* species. In addition to our dataset, we used previously published population genome Illumina sequencing data of *Z. tritici* isolates collected in Oregon (US), Switzerland, Australia and Israel (Hartmann et al., 2017), and from the closely related *Zymoseptoria* species *Z. ardabiliae*, *Z. pseudotritici* and *Z. brevis* (Grandaubert, Bhattacharyya, & Stukenbrock, 2015; Stukenbrock et al., 2012; Stukenbrock et al., 2011; Stukenbrock & Dutheil, 2018) (Supplementary Table S1). Illumina paired-end reads of ≥ 100bp were trimmed for adapter and sequencing quality using Trimmomatic v 0.39 (Bolger, Lohse, & Usadel, 2014) with the following parameters: LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:50. To account for older Illumina sequencing libraries (paired-end reads of 75bp), Trimmomatic v 0.39 settings were modified as follow: LEADING:10 TRAILING:10 SLIDINGWINDOW:5:10 MINLEN:50. We generated *de novo* genome assemblies for each isolate with SPAdes v. 3.14.1 (Bankevich et al., 2012) using the following parameters: --careful and a kmer range (-k) of '21,33,55,77'.

In order to infer evolutionary relationships between *Zymoseptoria* isolates, two phylogenetic methods were used. A Jukes-Cantor distance matrix was constructed using the Illumina draft genomes and the Andi v. 0.12 software (Haubold, Klötzl, & Pfaffelhuber, 2015). Then, a Neighbor-net network was produced using the generated distance matrix and the SplitsTree v.4 software (Huson, 1998; Huson & Bryant, 2005) to detect evidence of phylogenetic reticulation. We also generated a maximum likelihood (ML) tree based on filtered genome-wide SNPs using the substitution model TVM+F+P+N9+G4 and 100 bootstraps in PoMo software implemented in IQ-TREE 1.6.12 (De Maio, Schrempf, & Kosiol, 2015; Nguyen, Schmidt, Von Haeseler, & Minh, 2015; Schrempf, Minh, De Maio, von Haeseler, & Kosiol, 2016; Schrempf, Minh, Von Haeseler, & Kosiol, 2019). PoMo estimates population genetic parameters (mutation rates and fixation biases) using genetic variation between and within species and populations and accounts for phylogenetic tree discordances as ILS (incomplete lineage sorting) (De Maio et al., 2015; Schrempf et al., 2016). The software analyzes the provided genetic variation and chooses automatically the best substitution model to the data using ModelFinder (-m MFP option) (Kalyaanamoorthy, Minh, Wong, Von Haeseler, & Jermiin, 2017). The ML tree was visualized

using the online tool iTOL v.6 (Letunic & Bork, 2021). The VCF files for this step were generated using GATK v. 4.1.4.1 (McKenna et al., 2010) and *Z. tritici* reference genome IPO323 (Goodwin et al., 2011) as described in the "Read mapping and variant calling" section. After low depth genotype filtering, SNPs were quality-filtered using the GATK VariantFiltration and SelectVariants tools based on the following filter conditions: DP > 8000.0; QD < 20.0; MQ < 50.0; FS > 20.0; ReadPosRankSum, MQRankSum, and BaseQRankSum between -2 and 2. Only biallelic SNPs present in core chromosomes and with no missing genotypes were retained using VCFtools v 0.1.13 (Danecek et al., 2011). Using this generated VCF, we also estimated population differentiation by calculating all pairwise Weir and Cockerham's Fst fixation indices among the *Z. tritici* populations (Weir & Cockerham, 1984). For this calculation and to avoid SNPs in high linkage disequilibrium, we further pruned the filtered SNP dataset to select genome-wide SNPs at equidistant intervals of 12 kb along the chromosomes using VCFtools v 0.1. 13 (Danecek et al., 2011) as previously performed. Fst calculations were obtained using VCFtools v 0.1.13 (Danecek et al., 2011) with the haploid mode fork provided by Julien Y. Dutheil (https://github.com/vcftools/vcftools/pull/69).

To analyze the distribution of MAT1-1 and MAT1-2 idiomorphs in the sequenced *Aegilops*- and wheat-infecting populations, we performed nucleotide BLAST (blastn) (Altschul, Gish, Miller, Myers, & Lipman, 1990) searchers (e-value 1e−6, identity ≥ 80%) in the draft genome assemblies using the *Z. tritici* MAT1-1 and MAT1-2 loci as query sequences (Waalwijk, Mendes, Verstappen, de Waard, & Kema, 2002). The number of idiomorphs were counted for each *Z. tritici* population individually and chi-square tests were performed in R v. 3.6.3 (Team, 2013).

Selective sweep analyses

Selective sweep scans were based on SNPs that could have ancestral states assigned. In order to identify ancestral SNP alleles, we analyzed whole-genome population sequencing data of the two sister species of *Z. tritici*, *Z. ardabiliae* and *Z. pseudotritici*. The same *Z.pseudotritici* and *Z. ardabiliae* isolates used for the phylogenomics analyses were used in these steps (Supplementary Table S1). Illumina paired-end reads of 150bp were trimmed for adapter and sequencing quality using Trimmomatic v 0.39 (Bolger et al., 2014) with the following parameters: LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:50. To account for older Illumina sequencing libraries (paired-end reads of 75bp), Trimmomatic v 0.39 settings were modified as follow: LEADING:10 TRAILING:10 SLIDINGWINDOW:5:10 MINLEN:50. Read mapping and GATK SNP calling procedures were the same as described for *Z. tritici* populations above.

We merged all gvcf variant files and performed a final joint variant calling to include all *Z. tritici*, *Z. pseudotritici* and *Z. ardabiliae* isolates. We filtered SNPs based on quality cut-offs as described above with the following filter conditions: DP > 5000.0; QD < 20.0; MQ < 50.0; FS > 20.0; ReadPosRankSum, MQRankSum, and BaseQRankSum between -2 and 2. For each of the *Z. tritici* sister species, we retained only biallelic SNPs present in core chromosomes with a genotyping rate >50% and with no interspecific polymorphism using VCFtools v 0.1.13 (Danecek et al., 2011). To determine the ancestral state of variable positions in *Z. tritici* genomes, we used *Z. ardabiliae* and *Z. pseudotritici* as outgroups. A *Z. tritici* SNP allele was assigned as ancestral if the allele was fixed in both *Z. ardabiliae* and *Z. pseudotritici* species.

Signals of selection in selective sweep analyses can be erroneously detected due to confounding factors as demography (Nielsen et al., 2005; Pavlidis & Alachiotis, 2017). To decrease the number of false positives, we used two methods to detect selective sweeps that are robust to demography effects: the Composite Likelihood Ratio (CLR) method, implemented in SweeD (Pavlidis, Živković, Stamatakis, & Alachiotis, 2013); and the μ statistic, implemented in the RAiSD software (Alachiotis & Pavlidis, 2018). RAiSD is a composite evaluation test that simultaneously evaluates hallmarks of selective sweep regions by accounting changes in the Site Frequency Spectrum (SFS), the levels of LD, and the amount of genetic diversity along chromosomes (Alachiotis & Pavlidis, 2018). Both methods were run for each core chromosome and each *Z. tritici* population individually at grid points of 1kb using only the SNPs that could have the ancestral state assigned. To avoid erroneous CLR and μ scores due to low SNP density, we calculated genome-wide SNP density in 50kb non-overlapping windows for each population and retained only grid point scores contained in windows of at least 100 SNPs. The 99.5th percentile score distribution for each test and for each population was used as a threshold to identify outlier grid points. Selective sweep regions were then determined based on previous linkage disequilibrium estimations. For the *Aegilops* population 1, outlier grid points were merged if their distance was less than 12kb, whereas for the wheat-infecting population outlier grid points were merged if they were within a 5kb distance apart. To account for potential high linkage disequilibrium blocks around these identified selective sweep regions, we further extended these regions by adding 10kb to each end of *Aegilops* population 1 regions and 3kb for the wheat population ones. At last, we added 15kb to each end of the identified selective sweep regions in both populations to visualize the surrounding genomic locations.

Analysis of genomic content in selective sweep regions

Gene and Transposable Element (TE) content in each selective sweep region was analysed using the gene and TE models and functional annotations of the *Z. tritici* IPO323 reference genome published previously (Feurtey et al., 2020; Grandaubert et al., 2015; Lorrain et al., 2021). Population genomics statistics (Dxy, Fst, Tajima's D and nucleotide diversity, π) were calculated using VCFtools v 0.1.13 (Danecek et al., 2011) with the haploid mode fork provided by Julien Y. Dutheil (https://github.com/vcftools/vcftools/pull/69) and with the scripts provided by Simon Martin (https://github.com/simonhmartin/genomics_general). Calculations were performed in non-overlapping widows of 1kb (Dxy, Fst and π) or 5kb (Tajima's D). Linkage disequilibrium across the selective sweep region was calculated using VCFtools v 0.1.13 (Danecek et al., 2011) and PLINK v. 1.07 (Purcell et al., 2007). To this end, we used SNPs with a MAF of ≥ 1% for each population individually. Heatmap plots were created using the R package "LDheatmap" (Shin, Blay, McNeney, & Graham, 2006) and gene/TE models were drawn using the R package "gggenes" (https://cran.r-project.org/web/packages/gggenes/index.html). For the Gene Ontology (GO) enrichment analyses, we used the R package "topGO" (Alexa, Rahnenführer, & Lengauer, 2006) with the GO terms for each gene model published previously (Grandaubert et al., 2015). *p* values were calculated using Fischer's exact test applying the topGO algorithm "weight01" considering GO term hierarchy. We reported significant GO categories for the "Biological Process" ontology when $p \leq 0.05$. PFAM domain enrichment analyses were performed using a custom python script (Haueisen et al., 2019) with *p* values being calculated using chi-square tests. Lastly, we performed the McDonald & Kreitman (MK) test (McDonald & Kreitman, 1991) to test if non-neutral evolution (e.g. positive selection) is acting on any of the genes in the selective sweep region. For this, we used the script "vcf2fasta.py" (https://github.com/santiagosnchez/ vcf2fasta) to obtain gene sequences with polymorphisms for each host-diverging *Z. tritici* population individually. The orthologs genes in the *Z. ardabiliae* Za17 isolate (Feurtey et al., 2020) served as outgroup for these tests. For some of the analysed genes, no orthologs could be identified in *Z. ardabiliae* Za17 isolate, and therefore no MK tests could be performed on them (Zt09_chr_7_00296, Zt09_chr_7_00298, Zt09_chr_7_00303, Zt09_chr_7_00308, Zt469_000007F_arrow_0296). The MK tests were calculated with the web server MKT (http://mkt.uab.es/mkt/MKT.asp; Egea, Casillas, & Barbadilla, 2008) using the settings "Standard MKT" and "divergence corrected by Jukes & Cantor".

Identification of herbarium species

We performed amplicon sequencing of plant barcode loci to determine the species present in our herbarium material. To this end, a combined PTB-phenol protocol was used to extract genomic DNA from herbarium leaves (Fagundes, Haueisen, & Stukenbrock, 2020; Latorre, Lang, Burbano, & Gutaker, 2020; Shapiro & Hofreiter, 2009). Briefly, 10mg of leaves from each herbarium sample were place in 2mL sterile cryotubes containing three sterile 2,8 mm ceramic beads (biolab products GmbH, Bebensee, Germany) and snap-frozen in liquid nitrogen. Samples were then moved to a Precellys tissue homogenizer (Precellys Evolution, Bertin instruments GmbH, Frankfurt, Germany) and the plant tissue was ground for two cycles of 30 seconds with 5500 RPM and at –15°C. After grinding, 1.2 ml of PTB-Buffer (PTB-stock solution, 0,4mg/ml proteinase K, 50mM DTT, 2,5 mM PTB) was added to the powdery plant material and was stored overnight at 37°C under constant rotation (Latorre et al., 2020; Shapiro & Hofreiter, 2009). Samples were then centrifuged for 10 minutes at 16.000 x g and the supernatant was added to 800µL of 1:1 phenol/chloroform solution. The next steps were followed as described at the CTAB DNA extraction protocol of Fagundes et al. 2020 (step 10 onwards) (Fagundes et al., 2020). PCR amplification of the plant nuclear region *ITS2* was performed using the primers described in Supplementary Table S2 and submitted for Sanger sequencing at Eurofins Genomics (Ebersberg, Germany). Homology-based searches were performed for each sequenced amplicon using nucleotide BLAST (blastn) (Altschul et al., 1990) optimized for discontiguous megablast against the National Center for Biotechnology Information (NCBI) nucleotide database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). To observe phylogenetic relationships, multiple sequence alignments were performed using ClustalW (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Thompson, Higgins, & Gibson, 1994) and genetic distances were calculated using MEGA 5.2.2 (Tamura et al., 2011) according to the Kimura 2-Parameter (K2P) model. Positions with less than 95% site coverage were eliminated from the alignments. Neighbor-Joining (NJ) consensus dendograms were then constructed with 1000 bootstrap replicates also using MEGA 5.2.2 (Tamura et al., 2011). Additional barcode loci sequences from *Aegilops* and *Triticum* species as well as from barley (*Hordeum vulgare*) and rye (*Secale cereale*) were obtained from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) (Supplementary Table S3).

Demography analyses

To access the evolutionary history of the sequenced *Z. tritici* isolates, we surveyed the demography of the *Aegilops*- and wheat-infecting *Z. tritici* populations using MSMC2 (Schiffels & Durbin, 2014; Schiffels & Wang, 2020). For these analyses, we used a genomic dataset of 45 wheat-infecting isolates and 28 *Aegilops*-infecting isolates from the *Aegilops* population 1. Isolates in high IBS and the ones from the *Aegilops* population 2 were not included considering their high clonality. First, we built pseudodiploid bam files aligned to the 13 core chromosomes of the *Z. tritici* IPO323 reference genome (Goodwin et al., 2011), randomly combining two haploid bam files from genotypes coming from the same *Z. tritici* population. This step was performed with the Samtools v. 1.9 "merge" function (Danecek et al., 2021) and the "AddOrReplaceReadGroups" function in Picard tools v. 2.25.0 (https://broadinstitute.github.io/picard/). Subsequently, SNP calling was performed with bcftools v. 1.3.1 (Danecek et al., 2021) according to the pipeline suggested in Schiffels & Wang, 2020 considering we merged haploid genotypes and the genotype phasing step was omitted (Schiffels & Wang, 2020). Gene coding regions were included due to the high gene density of the *Z. tritici* genome. We computed the effective population size (Ne) for each population individually using 10 different combinations of 8 haploid genotypes for the wheat-infecting population and 10 different combinations of 8 haploid genotypes for the *Aegilops*-infecting population. For the relative cross coalescent rate (RCC) calculations, we used 10 different combinations that consist of 16 haploid genomes, 8 from wheat-infecting and 8 from *Aegilops*-infecting isolates. We assumed a mutation rate of 3.3e-8 per cell cycle (Lynch et al., 2008) and 1 generation of sexual reproduction per year as in Stukenbrock et al. (2011) (Stukenbrock et al., 2011). R script to plot demography results is available at Supplementary Text S1.

Virulence assays

We accessed the host specificity of the sequenced isolates collected from wheat and *Aegilops* spp. using comparative infection assays. Independent qualitative and quantitative experiments were performed under greenhouse conditions [~20°C (day)/~12°C (night) and ~70% humidity with a cycle of 16hr day/8hr night] using different wheat accessions and wild grass relatives as hosts (Supplementary Table S4). Seeds of wild wheat and *Aegilops* accessions were obtained from Dreschflegel Bio-Saatgut (Witzenhausen, Germany) and from NordGen (Lomma, Sweden). Wheat cultivar Riband seeds were kindly provided by Jason Rudd (Rothamsted Research, Harpenden, United Kingdom). Seeds from the wheat cultivar Chinese Spring were kindly provided by Bruce McDonald (ETH Zurich, Switzerland) and seeds from the cultivar Titlis was obtained

from DSP AG (Delley, Switzerland). For all experiments, blastospore cultures of isolates were used for inoculum preparation as described previously (Fagundes et al., 2020). To ensure all host species were in similar sizes at the time of inoculation, we used 14-days old seedlings of the wheat cultivars Chinese Spring, Riband and Titlis, and 20 days-old seedlings of all the other wild wheat and *Aegilops* accessions. Seed sowing and plant preparations were performed as described previously (Fagundes et al., 2020).

Qualitative disease scoring was performed manually 21 or 28 days post-inoculation (dpi) and inoculation procedures were followed as previously described (Fagundes et al., 2020). In order to avoid other biotic factors that can be confounded with disease symptoms as e.g. senescence, isolates were only considered to be virulent on a specific host if at least one of the inoculated plants had visual presence of pycnidia at 21 or 28 dpi. In the first experiment, isolates were inoculated in "batches" of populations containing 12 isolates each, namely "*Aegilops* pop 1" batch, representing the *Aegilops* population 1, "*Aegilops* pop 2" batch, representing the *Aegilops* population 2 and the "Wheat pop" batch, representing the wheat population. Isolates were selected to represent all herbarium samples and lesions of origin and to have an equal number of isolates per batch. In addition, we selected isolates based on their IBS (identity-by-state) scores. Isolates that showed the lowest IBS score per herbarium sample and per lesion were chosen. We also used mock [sterile water with 0.1% Tween 20 (Roth, Karlsruhe, Germany)] and the reference *Z. tritici* isolate IPO323 (strain Zt244) as control treatments. Blastospore suspensions for each isolate were individually adjusted to an OD600nm = 0.2 (approximately $1x10^6$ cells/mL) in 0.1% (v/v) Tween20 (Roth, Karlsruhe, Germany) when merging into the batches. Each batch was inoculated on three plants of each of the 31 plant hosts described in Supplementary Table S4. For a second qualitative experiment, we confirmed the observed virulent phenotypes of the *Aegilops* population 1 by individually inoculating the *Z. tritici* isolates on three *Aegilops cylindrica* and three wheat cultivar Riband plants. We selected two to three sequenced isolates per herbarium sample and per lesion based on the lowest IBS (identity-by-state) scores and adjusted the inoculum concentrations as experiment one (OD600nm = 0.2). Mock and the reference *Z. tritici* isolate IPO323 (strain Zt244) were used as control treatments.

At last, to get insights of the infection progress and quantitative virulence in the newly described *A. cylindrica-Z. tritici* pathosystem, we performed a disease progression and quantitative virulence analyses in a subset of isolates from *Aegilops* population 1. For the disease progression analyses, one sequenced isolate per herbarium sample was selected based on the symptoms observed in the qualitative virulence assays. Inoculum concentrations were adjusted to $1x10^5$ cells/mL in 0.1%

(v/v) Tween20 (Roth, Karlsruhe, Germany) and each *Aegilops*-infecting isolate was inoculated on 40 *Aegilops cylindrica* plants following the same inoculation procedures described previously (Fagundes et al., 2020). To compare the disease progression to a wheat-infecting isolate, we also inoculated 40 wheat cultivar Riband plants with an adjusted blastospore suspension ($1x10^5$ cells/mL) of the *Z. tritici* reference isolate IPO323 (strain Zt244). Each inoculated leaf was manually inspected between 7 and 21 dpi and the occurrence of first visible symptoms (necrosis and pycnidia) recorded every two days. Mock treatments were used as a negative control. For the quantitative virulence assay, we used the same *Z. tritici* isolates and inoculum concentrations used for the disease progression analyses. *Aegilops*-infecting *Z. tritici* isolates were inoculated on 50 *A. cylindrica* plants and the *Z. tritici* reference isolate IPO323 (strain Zt244) was inoculated on 50 wheat cultivar Riband plants. Mock treatments were inoculated on 37 plants of each species as negative controls. At 21 dpi, individual inoculated leaves were scanned and analyzed using automated image analysis as previously described (Fagundes et al., 2020). Pycnidia density (number of pycnidia per square centimeter of leaf surface) was used as readout of virulence. Statistical analyses were conducted in R version 4.0.5 (Team, 2013).

Confocal microscopy analyses

In order to further characterize the infection progress of *Aegilops cylindrica*-infecting *Z. tritici* isolates, we analyzed the morphological development of the *Z. tritici* isolate Zt469 on *A. cylindrica* plants using confocal laser scanning microscopy (CLSM). The *Z. tritici* isolate Zt469 was included on the *Aegilops* population 1 dataset for both population genomics and virulence assays and it was randomly selected to represent this host-specific population. *Aegilops cylindrica* plants were inoculated with an adjusted Zt469 blastospore suspension of $1x10^5$ cells/mL in 0.1% (v/v) Tween20 (Roth, Karlsruhe, Germany) as described previously (Fagundes et al., 2020). Mock treatment (sterile water with 0.1% Tween 20) was used as a negative control. Plants preparation and inoculation procedures were followed as described in the "Virulence assays" section above. We harvested inoculated *A. cylindrica* leaves at 7, 10, 12, 15, 18 and 21 days post-inoculation (dpi) to investigate the compatible interaction between Zt469 and the wild grass host. These time points were chosen to cover the expected four infection stages of *Z. tritici* ranging from early biotrophy to late necrotrophy as previously described (Haueisen et al., 2019). Mock-treated plants were also harvested as negative controls. Leaf material was prepared and stained with wheat germ agglutinin conjugated to fluorescein isothiocyanate (WGA-FITC) in combination with propidium iodide (PI) as performed previously (Haueisen et al., 2019). In total, we examined 36 *A. cylindrica* leaves for Zt469, 8 leaves for mock treatment and analyzed up to 43 infection events per leaf sample by

CLSM, creating a total of 44 confocal image z-stacks. Microscopy was conducted using a Zeiss LSM880 microscope (Carl Zeiss Microscopy, Germany). FITC was excited at 488 nm (argon laser) and detected between 500 and 540 nm. PI was excited at 561 nm (diode-pumped solid-state laser) and detected between 600 and 670 nm. We used ZEN black and ZEN blue (Carl Zeiss Microscopy, Germany) software for analyses, visualization, and processing of image z-stacks. Animations of image z-stacks in ".avi" format can be played in VLC media player (available at http://www.videolan.org/vlc/) and are deposited on the supplementary USB key.

The infection stages were determined by examining central leaf sections (1-2cm) of the harvested leaves by CSLM as performed previously (Haueisen et al., 2019). Considering the heterogeneity of infection events happening in a single leaf (Fantozzi, Kilaru, Gurr, & Steinberg, 2021), we determined the infection stage of each leaf sample by observing the most progressed infection events (ranging from stage "A" to "D" as previously described by Haueisen et al. 2019) along a transect of six 0.25 mm2 squares of leaf area (Supplementary Figure S1). That is, if in a leaf transect we observed most of the infection events representing specific infection stages as e.g. stage "B" and stage "C", we assigned the infection stage "C" to the respective leaf sample. The determination and assignment of infection stages was performed in a single-blinded procedure in which the examiner did not know the time point post-inoculation of the leaf sample *a priori*.

Results

Host-diverging *Z. tritici* isolates differ in genomic diversity and linkage disequilibrium

Prior accessing the population structure of the sequenced isolates, we aimed to compare the overall genomic similarity between the isolates and the levels of nucleotide diversity and linkage disequilibrium (LD) within each host-diverging collection. These analyses allowed us to have a first glimpse in the genomic structure of the host-diverging collections and to detect potential clonal isolates.

In order to access the genomic relatedness between the isolates and detect highly similar individuals (e.g. clones), we calculated genome-wide rates of Identity-By-State (IBS) among all 123 sequenced *Z. tritici* isolates. To this end, we used a quality-filtered dataset of 1,357,300 SNPs distributed among the thirteen core chromosomes of *Z. tritici*. We observed higher rates of IBS similarity within the *Aegilops*-infecting isolates (IBS similarity average: 0.86) when compared to wheat-infecting ones (IBS similarity average: 0.79) (Figure 1A and Supplementary Figure S2). This result suggests a difference in the overall genotypic composition between the two host-diverging collections, in which the *Aegilops*-infecting collection shows a higher genetic homogeneity when compared to the wheat-infecting one. Then, based on a pairwise threshold of IBS > 0.9999, we excluded highly similar individuals from the dataset as they can represent potential clonal isolates. This filtering step resulted in a removal of 8 (out of 48) isolates and a reduction of 0.04% in the total number of SNPs for the *Aegilops*-infecting collection. For the wheat-infecting collection, we removed 30 (out of 75) isolates and observed a decrease of 0.12% in the total number of SNPs for this collection.

Next, using the isolate-filtered SNP datasets, we analyzed the levels of nucleotide diversity and accessed the linkage disequilibrium (LD) in each host-diverging collection. For these analyses we only used SNPs with a MAF of ≥ 5% within each collection. In line with the IBS results, we observed a lower genome-wide nucleotide diversity (π) for the *Aegilops*-infecting isolates when compared to the wheat-infecting collection (Wilcoxon rank test, P- value = 2e-14, Figure 1B). We also observed a high differential LD decay between the two collections, in which *Aegilops*-infecting isolates had LD decayed to 50% of its maximum value at approximately 1580 bp and the wheat-infecting collection to approximately 230 bp (Figure 1C). Moreover, the *Aegilops*-infecting collection holds a higher overall degree of LD between SNPs (average LD r2 value: 0.30) than the wheat-infecting collection (average LD r2 value: 0.10). Altogether, these results indicate that both

Z. tritici collections are genetically distinct and have possibly been under different demographic scenarios throughout history (Hohenlohe, Phillips, & Cresko, 2011). The differential high degree of LD and low nucleotide diversity in the *Aegilops*-infecting collection, in particular, also suggests that asexual reproduction is occurring more frequently in these isolates (Grünwald, McDonald, & Milgroom, 2016).

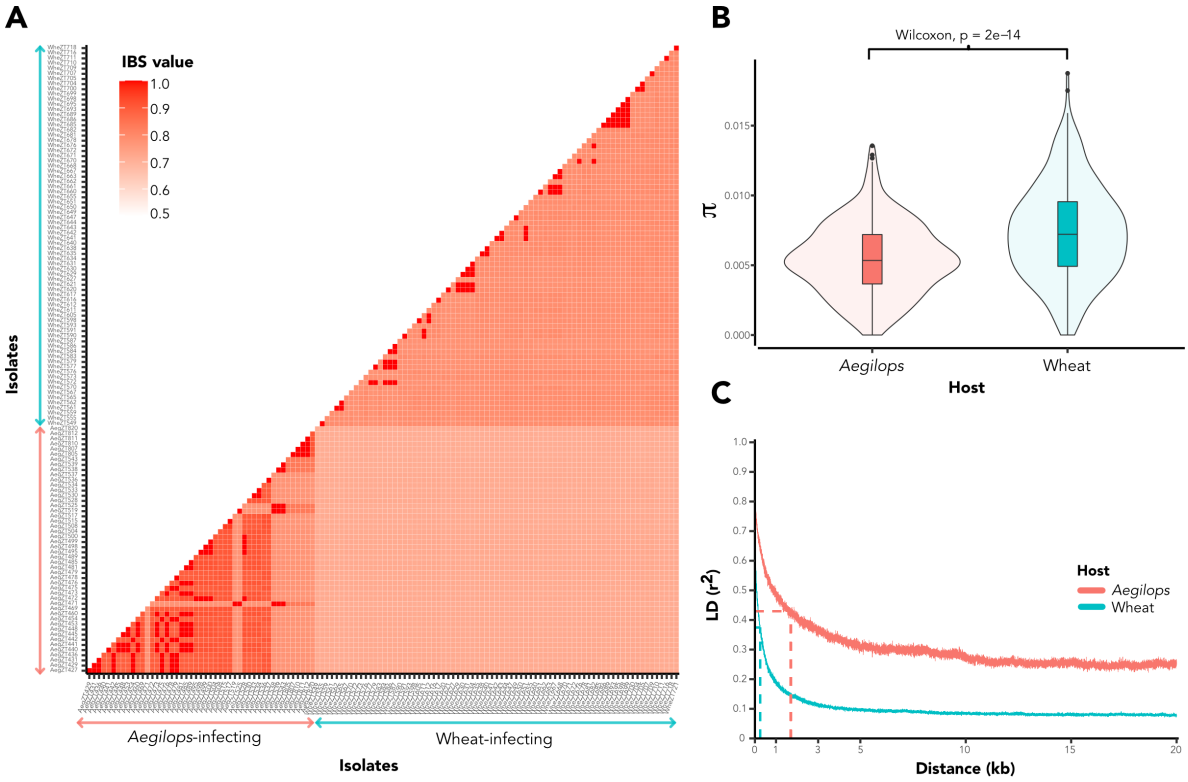


Figure 1. Host-diverging *Z. tritici* collections are genetically distinct. (A) Pairwise matrix representing genome-wide IBS values between *Aegilops* (red arrows) and wheat-infecting (blue arrows) *Z. tritici* isolates. Darker red colors represent higher IBS values and therefore higher similarity. **(B)** Genome-wide nucleotide diversity (π) calculated for each host collection individually. **(C)** Linkage disequilibrium decay across 20 kb of chromosome 1. Top, red line represents the LD decay for *Aegilops*-infecting isolates while the bottom, blue line represents the LD decay for the wheat-infecting collection. Dashed perpendicular lines represent distance to reach 50% maximum LD value in each collection.

Population structure is correlated with host species

We accessed the population structure among host-diverging *Z. tritici* isolates using different methods. To this end, we restricted the analyses to 2507 genome-wide independent (i.e. un-linked) SNPs distributed among the thirteen core chromosomes of *Z. tritici*. Principal component analysis (PCA) revealed an evident separation between the isolates collected from *Aegilops* spp. and from

wheat (Figure 2A and Supplementary Figure S3). Intriguingly, the *Aegilops*-infecting cluster showed a further separation, in which a smaller subcluster was composed of twelve isolates obtained from *Aegilops* spp. samples collected at a unique geographical location in two different years (Figure 2A). Corroborating these results, clustering analysis based on ancestry coefficients implemented in the sNMF software (Frichot & François, 2015; Frichot et al., 2014) also showed a clear separation between the *Aegilops*- and wheat-infecting isolates and the further substructure within the *Aegilops*-infecting collection (K=2 and K=3; Figure 2B). No large clustering overlaps were observed between the host-diverging collections at higher clustering models (K=4 and K=5; Supplementary Figure S4). Taking these results together, we could clearly define three populations based on hosts including a population substructure within the *Aegilops*-infecting population. Hereafter, we refer to these populations as "*Aegilops* population 1", representing the larger *Aegilops*-infecting *Z. tritici* population; "*Aegilops* population 2", representing the smaller population of twelve *Aegilops*-infecting *Z. tritici* isolates and "wheat population", representing the wheat-infecting *Z. tritici* isolates.

Considering the population segregation observed for the *Aegilops* population 2, we hypothesized that host specificity to different *Aegilops* species or genotypes was playing a role in the population structure of *Z. tritici* isolates. To explore this hypothesis, we identified the species of the original hosts from all analyzed *Z. tritici* isolates using amplicon sequencing of the *ITS2* barcode locus from the herbarium samples. Interestingly, *ITS2* sequences of leaf samples from the *Aegilops*-infecting population 1 revealed that the *Z. tritici* strains were potentially isolated from *Aegilops cylindrica* or a closely-related *Aegilops* species (e.g. *Ae. triuncialis or Ae. tauschii*) while *Z. tritici* isolates from the *Aegilops* population 2 were obtained from *A. tauschii* or other closely-related *Aegilops* species (e.g. *Ae. geniculata*; Supplementary Figure S5 and Supplementary Table S5). For the wheat-infecting *Z. tritici* isolates, amplicon analyses suggested that their hosts were indeed from the common wheat species (*Triticum aestivum* L.) or from closely-related polyploid wheat species (e.g. *T. durum*) (Supplementary Figure S5 and Supplementary Table S5). Taken together, these results support the hypothesis that distinct host species are shaping the structure of the *Aegilops*-infecting *Z. tritici* populations.

Aegilops-infecting populations form distinct *Z. tritici* clades

To obtain an evolutionary perspective at the species level, we estimated the phylogenetic relationship between the host-diverging *Z. tritici* populations by analyzing draft genome assemblies. Neighbor-net network analysis based on whole genome distances revealed a clear separation

between the *Aegilops*-infecting *Z. tritici* isolates from the ones isolated from wheat at different geographical locations (Figure 2C). We also observed that the *Aegilops* population 2 was segregating from *Aegilops* population 1 (Figure 2C), suggesting a strong signal of divergence between these isolates and corroborating the results observed for the population structure analyses (Figure 2A, B). Similar results were obtained using a maximum-likelihood phylogenetic approach based on intraspecific genomic variation (PoMo; De Maio et al., 2015; Schrempf et al., 2016), in which we also observed a clear separation between the *Aegilops*-infecting *Z. tritici* isolates from the wheat-infecting ones (Supplementary Figure S6). Genome-wide divergence values based on the Weir & Cockerham's Fst fixation indices (Weir & Cockerham, 1984) also indicated that the *Aegilops*-infecting populations were highly divergent from other *Z. tritici* populations and between themselves (Supplementary Figure S7), which implies that host specificity rather than geographical origin is determining *Z. tritici* population structure and population divergence.

Next, considering the divergence between the *Z. tritici* isolates, we hypothesized that the *Aegilops*-infecting populations were more closely related to other *Zymoseptoria* species than to *Z. tritici* itself. To test this hypothesis, we used the same Neighbor-net network approach based on genomic distances including draft genome assemblies from the closely related *Zymoseptoria* species *Z. passerini*, *Z. ardabiliae* and *Z. brevis*. We observed an evident separation of the *Z. tritici* cluster from the closely related species, although some edges (internal branches connecting taxa nodes) were observed connecting the *Aegilops*-infecting isolates to the other *Zymoseptoria* species (Supplementary Figure S8). Phylogenetic network analyses can illustrate possible evolutionary relationships as recombination, horizontal gene transfer, or hybridization between taxa, in which the edges represent such reticulate evolutionary events (Huson & Bryant, 2006). Based on these results, we suggest that the *Aegilops*-infecting isolates represent *Z. tritici* lineages distinct from the wheat-infecting one and further hypothesize that these lineages may have experienced gene flow events with other *Zymoseptoria* species throughout evolution.

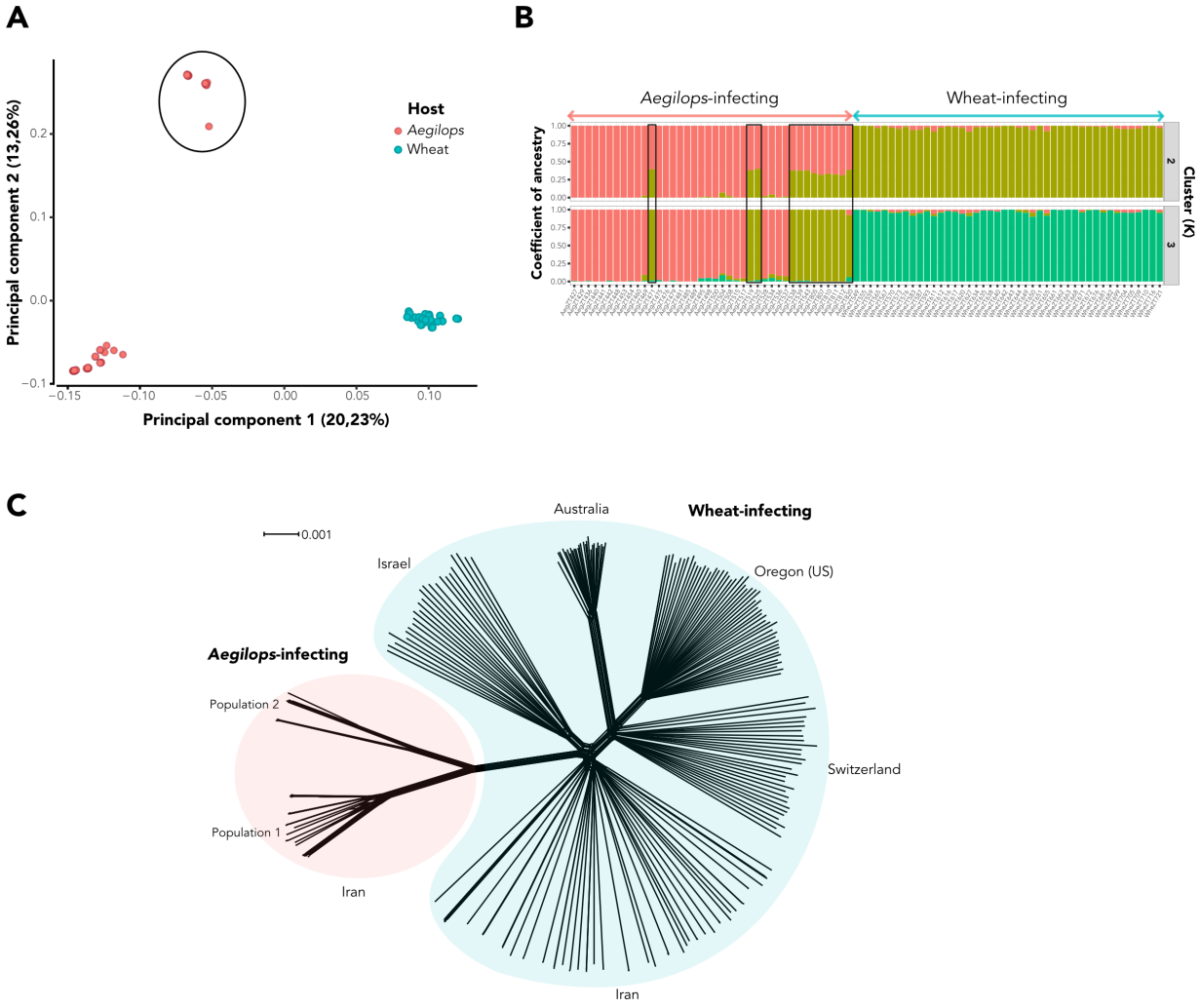


Figure 2. Population structure of *Z. tritici* isolates is correlated with host species. (A) PCA analyses based on genome-wide independent SNPs. The first two principal components (PCs) are shown with the variance explained by each component in parenthesis. Red dots represent *Z. tritici* isolates collected from *Aegilops* species while blue dots represent the ones isolated from wheat. *Aegilops* population 2 is indicated with a black ellipse. **(B)** Ancestry coefficients and clustering assignments using the sNMF software. Each vertical bar represents an isolate from the *Aegilops* (red arrow) and wheat-infecting (blue arrow) *Z. tritici* collection with each color indicating one genetic cluster. The color height in each vertical bar represents the probabilities of cluster assignment based on genome-wide independent SNPs. Isolates belonging to the *Aegilops* population 2 are highlighted in black. **(C)** Neighbor-net network based on whole genome distances. Red ellipse represents the two *Aegilops*-infecting *Z. tritici* populations while the blue elliptical shape represents worldwide wheat-infecting *Z. tritici* populations. Geographical locations of origin are also indicated. Scale bar indicates branch distances.

Mating type distribution differs between host-diverging populations

The high levels of linkage disequilibrium and IBS and low genome-wide nucleotide diversity observed in the *Aegilops*-infecting collection (Figure 1) led us to the hypothesis that these *Z. tritici* isolates reproduce mostly asexually (e.g. by clonal propagation). One of the strongest evidences for clonal reproduction in ascomycete fungi is the uneven distribution of both mating types in a population or the complete lack of one of the idiomorphs. Several genes are involved in mating, and two major idiomorphs (MAT1-1 and MAT1-2) are described in *Z. tritici* (Waalwijk et al., 2002). To analyze the distribution of such idiomorphs, BLAST searches using nucleotide sequences of both mating type loci against whole-genome assemblies were performed in all sequenced individuals, and the number of idiomorphs counted for each *Z. tritici* population individually (wheat population, *Aegilops* population 1 and the *Aegilops* population 2). Both wheat and *Aegilops* population 1 did not deviate from the even (1:1) distribution of mating types ($\chi 2$ P=0.204, wheat population and $\chi 2$ P=0.3173, *Aegilops* population 1), suggesting that sexual reproduction is occurring in such populations (Supplementary Table S6). In contrast, the *Aegilops* population 2 had no MAT1-1 idiomorphs present, and therefore may represent a population of asexually-propagated individuals (Supplementary Table S6). We also observed a lack of LD decay to reach 50% of its maximum value over 100 kb of chromosome 1 and a low genome-wide nucleotide diversity in the *Aegilops* population 2 (Supplementary Figure S9), which also suggests the potential asexual reproduction and high clonality of the isolates belonging to this population.

***Aegilops*-infecting *Z. tritici* lineage diverged around 10,000 years ago**

The phylogenetic and population divergence observed between the *Z. tritici* populations infecting *Aegilops* spp. from the ones infecting wheat prompted us to further investigate the demography history of these isolates. To this end, we used MSMC2 (Schiffels & Durbin, 2014; Schiffels & Wang, 2020) to estimate the coalescence rates and divergence times between haplotypes from the wheat population and the *Aegilops* population 1. We could not perform the demography analyses on the *Aegilops* population 2 due to the potential asexual propagation and clonality of this population. Based on the mutation rate estimated for *S. cerevisiae* (3.3e-8 per cell cycle) (Lynch et al., 2008) and assumming 1 generation of sexual reproduction per year, we observed that the wheat-infecting population experienced a population expansion followed by a bottleneck between 30,000 and 10,000 years ago (Supplementary Figure S10A). A similar pattern was identified in the *Aegilops* population 1, although a large variation in effective population size (Ne) was observed for recent times (Supplementary Figure S10B). Next, to access the timing of divergence between the

two populations, we conducted a relative cross-coalescence rate (RCCR) analysis based on 16 haploid genomes (8 from each population). The results suggested that the two host-diverging *Z. tritici* populations started to diverge around 10,000 years ago, period in which a decrease of the RCCR is apparent (Supplementary Figure S11). We observed similar results when we permuted the samples between the two populations, indicating that this divergence signal is strong and well supported by the data (Supplementary Figure S11). Altogether, these findings suggest that the two host-diverging *Z. tritici* populations have diverged and expanded at approximately 10,000 years ago and corroborate our hypothesis that these populations represent distinct *Z. tritici* lineages. Moreover, the time frame of divergence detected here correlates with previous coalescence analyzes on *Z. tritici*, which indicates that *Z. tritici* has split from its closest sister species *Z. pseudotritici* approximately 10,500 years ago coinciding with the domestication of its host wheat (Stukenbrock et al., 2007). Regarding the isolates from the *Aegilops* population 2, additional sampling and demographic analyses are necessary to explore the potential divergence time of this population.

Aegilops-infecting *Z. tritici* isolates are host-specific

Considering the evolutionary divergence observed between the wheat and *Aegilops*-infecting populations, we investigated if the *Z. tritici* isolates from these populations also had distinct host ranges. To this end, we performed independent qualitative and quantitative *in planta* virulence assays under greenhouse conditions. In a first experiment, we inoculated batches of sequenced isolates based on the identified populations, namely "*Aegilops* pop 1" batch, representing the *Aegilops* population 1, "*Aegilops* pop 2" batch, representing the *Aegilops* population 2 and the "Wheat pop" batch, representing the wheat population. These batches of isolates were inoculated individually along mock and the reference *Z. tritici* isolate IPO323 as control treatments in different wheat and wild grasses accessions (Supplementary Table S4), and the disease symptoms were assessed at 21 and 28 days post-inoculation (dpi). Out of the 31 different wheat and wild-grass relatives tested, only the host species *Aegilops cylindrica* could be infected by the isolates present in the "*Aegilops* pop 1" batch. No symptoms were observed for any *Triticum* species inoculated with this batch (Supplementary Figure S12 and Supplementary Table S7). The isolates contained in the "*Aegilops* pop 2" batch also could not infect any *Triticum* species analyzed but could be virulent on *Aegilops cylindrica* and *Aegilops tauschii* plants (Supplementary Figure S12 and Supplementary Table S7). Isolates belonging to the "Wheat pop" batch could infect *Triticum aestivum* and other closely-related *Triticum* species in addition to *Aegilops tauschii* (Supplementary Figure S12 and Supplementary Table S7). Then, given the clear virulence dichotomy on *A. cylindrica* observed

between isolates from the *Aegilops* population 1 and the wheat population, we performed a second qualitative infection assay focusing just on these two populations. In this experiment, we aimed to confirm the host specificity among the collection of sequenced isolates from the *Aegilops* population 1 and wheat population by selecting two to three isolates per herbarium sample and inoculated them individually on *Aegilops cylindrica* and *Triticum aestivum* plants. Consistent with the previous inoculation assay, *Aegilops*-infecting *Z. tritici* isolates could only infect *A. cylindrica* plants, while the wheat-infecting isolates could only cause symptoms on *T. aestivum* plants (Figure 3 and Supplementary Table S8). Taken together, these results fully support the hypothesis that host specificity to different *Aegilops* species is shaping the population structure of these sympatric yet divergent *Z. tritici* populations and reinforce the assumption of different host-specific *Z. tritici* lineages. In addition, we could confirm the virulence of individual *Z. tritici* isolates in *A. cylindrica* plants and establish a new, powerful wild pathosystem for *Z. tritici*.

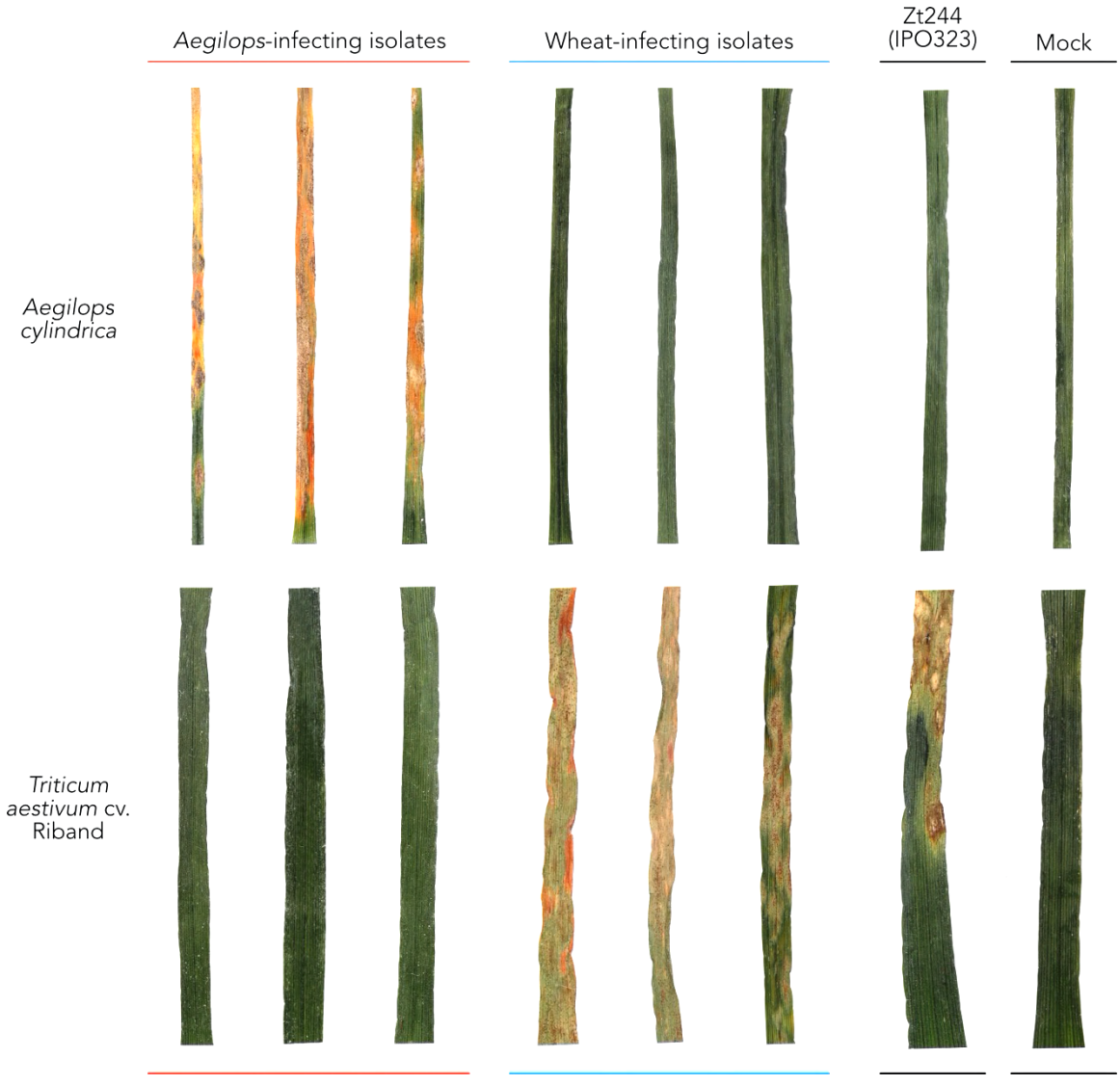


Figure 3. Host specificity among *Z. tritici* isolates. Summary of qualitative virulence assay performed by isolates under greenhouse conditions. *Aegilops cylindrica* (top panels) and wheat (*Triticum aestivum* cv. Riband; bottom panels) were inoculated with adjusted blastospore suspensions of each individual *Z. tritici* isolate. Each grey column represents leaves inoculated with a different isolate. Mock treatments received 0.1% Tween® 20 in sterile water. Pictures were taken at 21 dpi (days-post inoculation). Complete description of qualitative virulence results can be found at Supplementary Table S8.

To better characterize the *Z. tritici* infection program in the newly established *A. cylindrica*-*Z. tritici* pathosystem, we performed a disease progression analysis and quantitative measures of disease symptoms. In these assays, we used a subset of *Aegilops*-infecting *Z. tritici* isolates from *Aegilops* population 1 along with the *Z. tritici* IPO323 isolate (strain Zt244) and mock treatments as a wheat-infecting reference and negative control, respectively. All isolates/treatments were inoculated individually at similar inoculum concentrations on its correspondent hosts: *Aegilops cylindrica*, for the *Aegilops*-infecting *Z. tritici* isolates and *Triticum aestivum* cv. Riband, for the *Z. tritici* IPO323 isolate. Disease progression analyses revealed that leaves inoculated with the *Aegilops*-infecting *Z. tritici* isolates showed, on average, the first signs of necrosis and pycnidia at 9-10 dpi and 12-13 dpi, respectively (Supplementary Figure S13). This onset of visible symptoms was earlier than those observed for the wheat-infecting *Z. tritici* isolate Zt244 (IPO323), in which visible symptoms became apparent between 15 dpi to 18 dpi on *T. aestivum* plants (Supplementary Figure S13). Similar results for other wheat-infecting isolates were observed previously, with medians for onset of necrosis and pycnidia happening between 12-19 dpi and 17-23 dpi, respectively (Haueisen et al., 2019). Next, we performed a quantitative measure of disease symptoms. To this end, we inoculated *Aegilops cylindrica* and wheat plants with the same isolates used for the disease progression analyses and collected inoculated leaves at 21 days post-inoculation (dpi) to evaluate infections using pycnidia density as a virulence readout. We observed statistical difference in pycnidia levels among all treatments tested (Kruskal-Wallis test, chi-squared = 218.08, df = 8, p-value < 2.2e-16; Supplementary Figure S14). The isolates Zt495 and Zt536 showed statistically higher pycnidia densities compared to the other isolates, although no statistical difference was observed between them two (Wilcoxon Rank Sum test, p-value = 0.65). Altogether, the differences in the temporal disease development and in the levels of virulence observed among the *Aegilops*- and wheat-infecting isolates suggest that the infection process on the the wild pathosystem is, besides host-specific, a highly variable trait and happens at a different pace than in *Z. tritici* on wheat.

Infection of *Z. tritici* in *A. cylindrica* is also characterized by four developmental stages

Next, we addressed if the infection development of *Z. tritici* isolates on *Aegilops cylindrica* plants was similar to the one previously observed for *Z. tritici* on wheat (Haueisen et al., 2019). To this end, we morphologically characterize Zt469 host colonization using detailed confocal scanning laser microscopy (CSLM) analyses in inoculated *A. cylindrica* leaves collected between 7 and 21 dpi (see Methods and Supplementary Figures S1). The isolate Zt469 was randomly selected to represent the *Aegilops* population 1 and it was included in all population genomics and virulence assays datasets analyzed in this study. Leaf samples were characterized based on the previously described infection development stages of *Z. tritici* on wheat (Haueisen et al., 2019) in a single-blinded procedure (see Methods). Briefly, we characterized infection events belonging to the following four developmental stages: stage "A", or infection establishment stage, in which fungal hyphae penetrate leaf tissue via stomata; stage "B", characterized by biotrophic, symptomless colonization of leaf mesophyll by fungal hyphae; stage "C", which comprises the transition from biotrophic to necrotrophic growth when first disease symptoms are apparent; and the stage "D", characterized by necrotrophic colonization and asexual reproduction. In this stage, mesophyll tissue is heavily colonized by hyphae and substomatal cavities are occupied by mature pycnidia harboring asexual pycnidiospores (Haueisen et al., 2019). Based on these characteristics, we identified leaf samples that could most represent these four infection stages, namely 7dpi (stage "A"), 10dpi (stage "B"), 15dpi (stage "C") and 21dpi (stage "D"; Supplementary Figure S15). These results indicate that, although infecting a distinct host and belonging to a divergent lineage, *Aegilops*-infecting *Z. tritici* isolates have a similar infection development to wheat-infecting isolates.

Host-diverging *Z. tritici* populations exhibit diverse selective sweep footprints

Based on the findings that the analyzed sympatric *Z. tritici* populations were diverging by hosts, we performed genome scans to detect signatures of selective sweep in each individual population that could have been driven by host adaptation. For these analyses, we focused on the *Aegilops* population 1 and the wheat population only, given that the *Aegilops* population 2 was consisted of only one mating type and therefore potentially a clonal population. Our analyses were also restricted to SNPs with known ancestral states, which were determined by comparing genome-wide SNP data of our *Z. tritici* populations to the ones from the closest *Z. tritici* species, *Z. ardabiliae* and *Z. pseudotritici*. Ancestral SNP alleles were designated if the same SNP alleles was fixed in both sister species. Using this approach, our selective sweep scans were based on 149,716 SNPs with known ancestral states dispersed over the thirteen core chromosomes of *Z. tritici*.

Signals of selection can be erroneously detected in genome-wide scans. Demographic events as population bottlenecks or migrations, in particular, can introduce confounding factors to the analyses and result in false positives (Nielsen, 2005; Nielsen et al., 2005; Pavlidis & Alachiotis, 2017). To decrease the number of false positives, we used two methods and stringent outlier thresholds to detect selective sweeps: the Composite Likelihood Ratio (CLR) method, implemented in SweeD (Pavlidis et al., 2013); and the μ statistic, implemented in the RAiSD software (Alachiotis & Pavlidis, 2018), which detects different signatures of selective sweeps simultaneously across chromosomes. These two methods slightly differ in terms of detection of hard selective sweeps (i.e. selection and spread of a single advantageous allele in a population; Hermisson & Pennings, 2005; Messer & Petrov, 2013) while being robust to demographic scenarios (Alachiotis & Pavlidis, 2018; Nielsen et al., 2005; Pavlidis & Alachiotis, 2017). After clustering sweep regions based on high 99.5% outlier thresholds, several and widespread selective sweep regions were detected across all chromosomes among the host-diverging populations (Figure 4 and Supplementary Tables S9 and S10). Selective sweep regions were on average 25066 bp and 8100 bp in length in the *Aegilops* and wheat-infecting *Z. tritici* populations, respectively (Supplementary Table S10). Two region overlaps were detected between the two different methods in the *Aegilops* population 1 while five regions were in common between the two methods in the wheat population (Supplementary Tables S9). Besides widespread, selective sweep regions were also highly divergent between the two analyzed populations with only one region overlapping between the *Aegilops* and the wheat-infecting *Z. tritici* populations (Supplementary Table S9). These findings indicate that both populations have experienced selection affecting different traits and different selective sweeps have shaped their genomes.

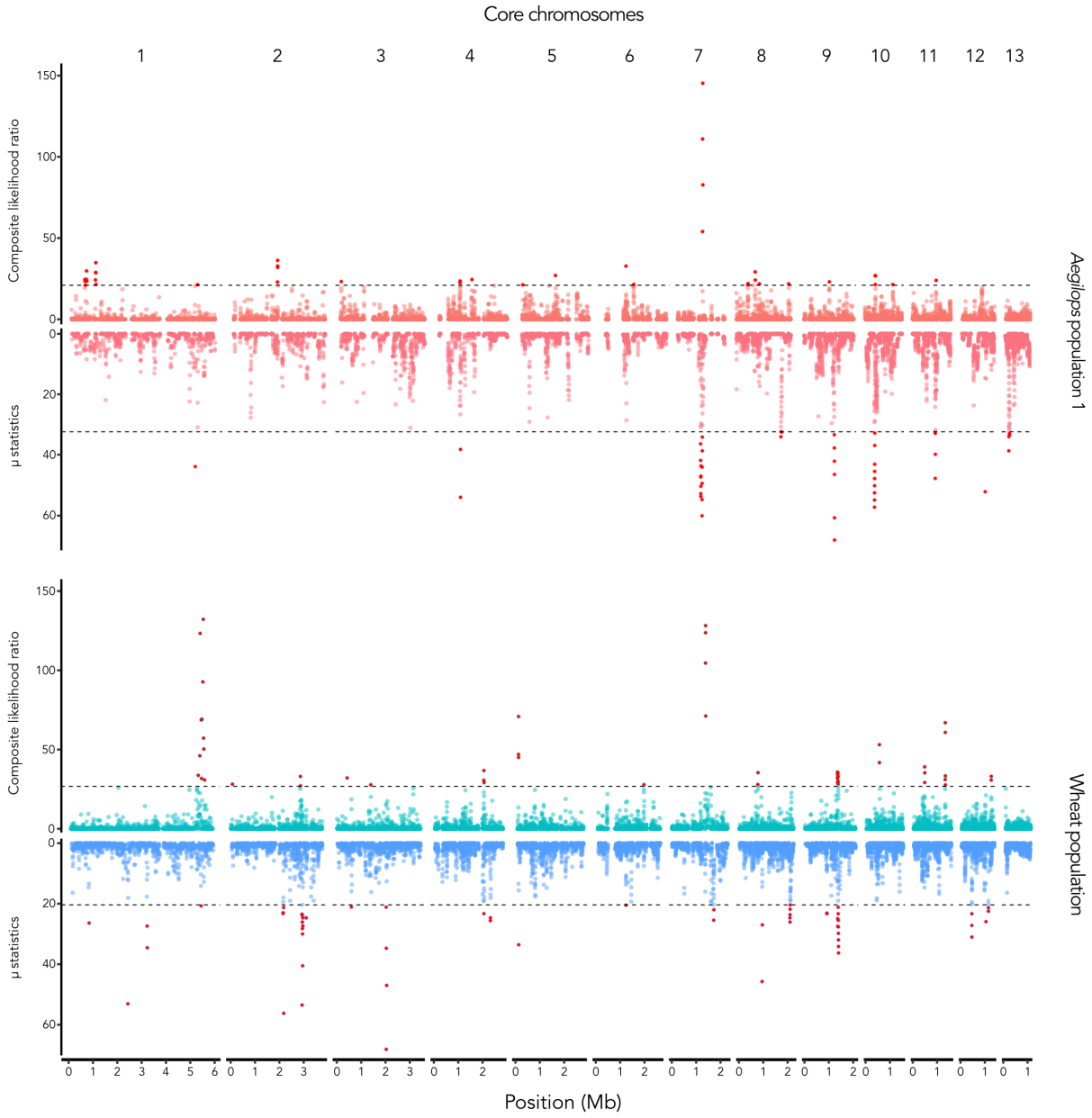


Figure 4. Diverse and widespread selective sweep signatures in host-diverging *Z. tritici* populations. Genomes scans for selective sweeps were performed across the thirteen core chromosomes of *Z. tritici*. Two methods were used for each host-diverging *Z. tritici* population. Top panels show the CLR and μ statistics scores for the *Aegilops* population 1 while the bottom panels show the same statistics for the wheat population. Horizontal dashed lines represent the 99.5th quantile threshold in each method and individual population. Above the thresholds, outlier regions potentially under selective sweep are highlighted in red.

Selective sweep regions harbor candidate genes for host adaptation

To investigate genes that are putatively under recent selection and potentially involved in host adaptation, we characterized the genome landscape of selective sweep regions in each population

individually. We used the gene and Transposable Element (TE) models and functional annotations in the reference *Z. tritici* isolate IPO323 published previously (Feurtey et al., 2020; Grandaubert et al., 2015; Lorrain et al., 2021) to identify the features present in the regions under selection. In total, selective sweep regions harbored 301 genes and 184 genes in the *Aegilops* and wheat-infecting *Z. tritici* populations, respectively (Supplementary Table S11). Regarding TEs, selective sweep regions contained a total of 43 TEs for the *Aegilops* population 1 and 46 TEs in the wheat population (Supplementary Table S11). We also performed Gene Ontology (GO) analyses and tested the enrichment of GO terms and PFAM domains in the selective sweep regions compared to whole genome. For the *Aegilops* population 1, we observed that GO terms for biological processes related to RNA splicing and DNA replication and transcription were overrepresented among the selective sweep regions (Fischer's exact test, p-values = 0.0089 - 0.045; Supplementary Table S12). In the selective sweep regions of the wheat population, we found that GO terms related to biosynthetic processes and transport were significantly enriched (Fischer's exact test, p-values = 0.029 - 0.049; Supplementary Table S12). PFAM domain analyses revealed that in the wheat population, selective sweep regions are enriched in genes encoding cytochrome P450-like and polyketide synthase-like proteins, chitin-binding lysin motif (LysM), among others ($\chi 2$ test, $p < 0.05$; Supplementary Table S13). In the *Aegilops* population 1, we observed a significant enrichment of genes encoding glycosyl hydrolases, peptidases, antibiotic biosynthesis monooxygenase along with other diverse domains ($\chi 2$ test, $p < 0.05$; Supplementary Table S13).

In host-pathogen interactions, different pathogen virulence factors, including genes encoding effectors and carbohydrate-degrading enzymes (CAZymes), are important components of the molecular arms race with host plants (Lo Presti et al., 2015; van der Does & Rep, 2017). Focusing on these gene categories, we observed that five candidate effector genes were present in the selective sweep regions of the *Aegilops* population 1 while two were identified in the wheat population (Supplementary Table S11). One of the candidate effector genes in the *Aegilops* population 1 was present in the intersection region detected between the two selective sweep methods (Zt09_chr_7_00299). For genes encoding CAZymes, while the selective sweep regions in *Aegilops* population 1 comprised thirteen CAZymes-encoding genes, in the wheat population nine CAZymes-encoding genes were identified, with one of them present in the intersection region observed between the two selective sweep methods (Zt09_chr_9_00422) (Supplementary Table S11). No effector or CAZyme-encoding genes present in the selective sweep regions were in common between the two host-diverging *Z. tritici* populations. Interestingly, among the different selective sweep regions in the *Aegilops* population 1, one region stood out. We found that from all the five candidate effector genes detected in the selective sweep regions of the *Aegilops* population

1, three candidate effectors (Zt09_chr_7_00299, Zt09_chr_7_00305 and Zt09_chr_7_00308) were localized in proximity at chromosome 7, spanning a unique selective sweep region identified from position 1,000,062 to 1,032,926bp using the CLR test. Moreover, adjacent to this selective sweep region, RAiSD also identified a region potentially under selection from position 980,772 to 1,013,262bp containing one of the thirteen CAZyme-encoding genes (Zt09_chr_7_00296) detected in the selective sweep regions of the *Aegilops* population 1. Part of this genomic region (from 1,000,062 to 1,013,262 bp) was the one observed to be in common between the two selective sweep metrices used for the *Aegilops*-infecting population (Supplementary Table S9).

Next, considering the detection overlap and the importance of effector and CAZyme gene categories in host-pathogen interactions, we focused on the further characterization of the selective sweep region identified at chromosome 7 in the *Aegilops* population 1. We used the genomic coordinates detected by the CLR test as they comprised the three candidate effector genes within the same region. We also analyzed part of the selective sweep region identified by RAiSD and surrounding genomic locations by adding 15 kb to the up and downstream flanks of the region potentially under selection as defined by the CLR test. Moreover, to compare the local genomic landscape between the two host-diverging *Z. tritici* populations, we analyzed the region using SNPs from the wheat population and *Aegilops* population 1 individually. We observed peaks of high absolute (Dxy) and relative (Fst) divergences between the *Aegilops*- and wheat-infecting populations across the region under selection, particularly in the loci encoding the three candidate effector genes (Figure 5A, B). We observed a lower nucleotide diversity (π) for the *Z. tritici* isolates infecting *Aegilops* spp. throughout the whole selective sweep region as well as in the surrounding flanks (Figure 5A). A decrease in nucleotide diversity was also observed in the wheat-infecting population for the regions encoding the three candidate effector genes (Figure 5A). We also identified negative Tajima's D values in the *Aegilops* population 1, especially in the region comprising the CAZyme gene (Zt09_chr7_00296) and two candidate effector genes (Zt09_chr7_00299 and Zt09_chr7_00305), while the isolates infecting wheat showed a positive Tajima's D value in the detected selective sweep locus (Figure 5A, B). A high linkage disequilibrium between SNP pairs was observed across the region detected in the selective sweep scan for the *Aegilops*-infecting *Z. tritici* isolates (Figure 5C), particularly in the region spanning the genes Zt09_chr7_00296 and Zt09_chr7_00308. In contrast, the wheat-infecting population showed a low linkage disequilibrium throughout the whole selective sweep region and surrounding flanks (Figure 5C). To complement these results, we performed the McDonald & Kreitman (MK) test for each gene within the selective sweep region using the polymorphisms of each host-diverging population individually and *Z. ardabiliae* orthologs as outgroup. No significant differences in the

ratio for synonymous versus non-synonymous sites within and between species was detected in any gene analyzed in the *Aegilops* population 1. In the wheat-infecting population, evidence for non-neutral evolution was found for only one gene, namely Zt09_chr_7_00300 (p-value < 0.001). Taken together, these results show clear patterns of selection in the selective sweep region of chromosome 7, even though the MK tests did not detect significant ratios for positive selection. The negative Tajima's D as well as the high linkage disequilibrium values observed for the *Aegilops* population 1, in particular, support the signatures of a recent selective sweep at this locus (Nielsen, 2005; Nielsen et al., 2005; Stephan, 2016). More importantly, the high peaks of absolute (Dxy) and relative (Fst) divergence detected between the two host-diverging population on candidate effectors further pinpoint candidate genes that may contribute for host adaptation on these distinct populations.

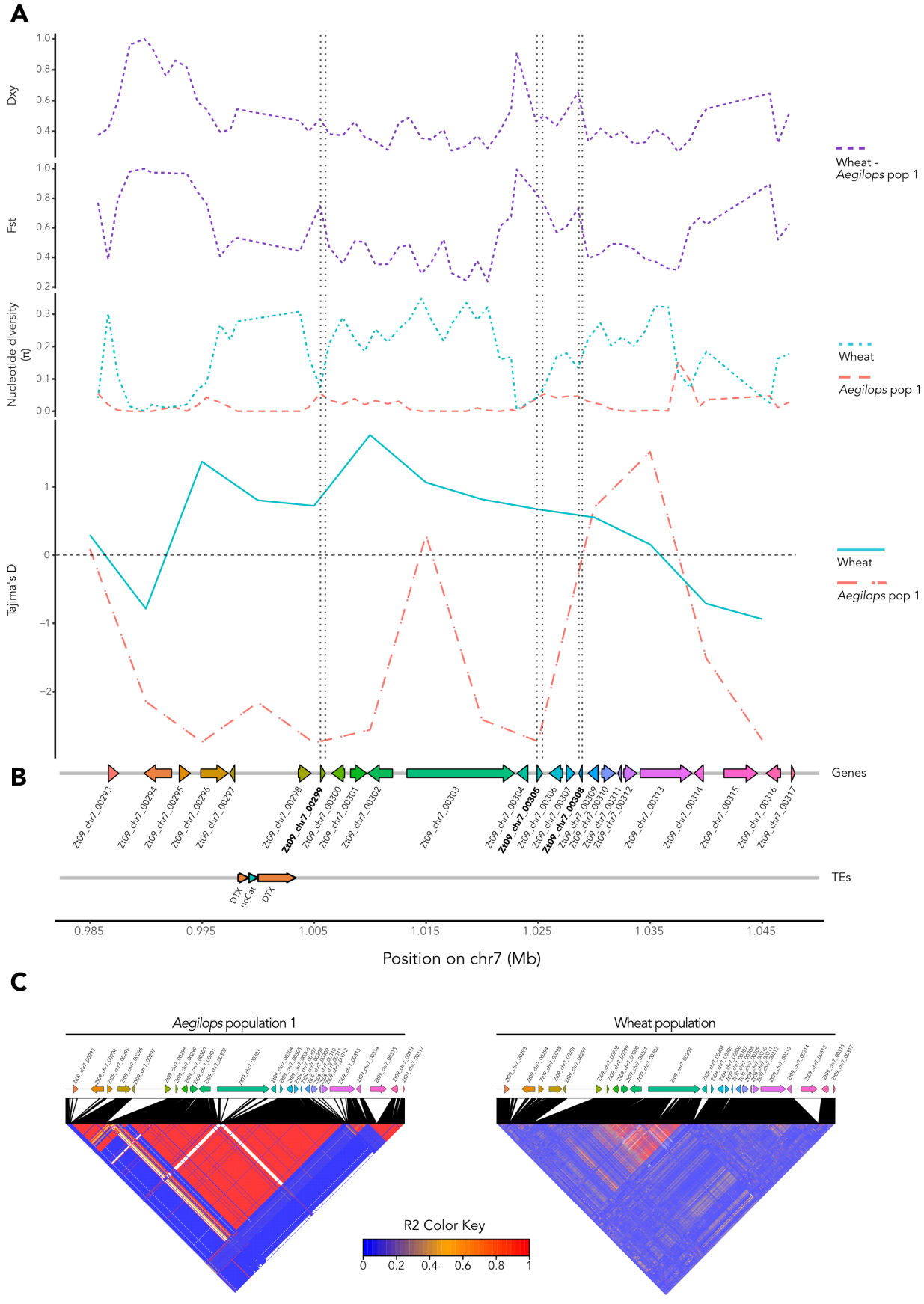

A
Dxy
Fst
Nucleotide diversity (π)
Tajima's D
Wheat - Aegilops pop 1
Wheat
Aegilops pop 1
Wheat
Aegilops pop 1
B
Genes
TEs
DTX
noCat
DTX
0.985
0.995
1.005
1.015
1.025
1.035
1.045
Position on chr7 (Mb)
C
Aegilops population 1
Wheat population
R2 Color Key
0
0.2
0.4
0.6
0.8
1

Figure 5. *Aegilops* population 1 shows selective sweep signatures at chromosome 7. Genomic summary statistics were calculated across the selective sweep region with 15kb up- and downstream flanks for the *Aegilops* population 1 and wheat population individually. **(A)** Absolute (Dxy) and relative (Fst) divergences between the two host-diverging *Z. tritici* populations (top panels) and nucleotide diversity (π) for each population (middle panel) were calculated in 1kb windows. Tajima's D (lower panel) was calculated in 5kb windows. Vertical dashed lines delimit the genomic locations of the three candidate effector genes. **(B)** Genes and TE models annotated across the selective sweep and surrounding genomic regions (DTX = Class II transposons, TIR; noCat = no category). Candidate effector genes are in bold. **(C)** Linkage disequilibrium heatmap based on coefficient of correlation (r2) between SNPs in the *Aegilops* population 1 (left) and wheat population (right). Correspondence between SNPs location in the genomics regions (e.g. genes) and on the heatmap are depicted by black segments.

Discussion

Host adaptation is a crucial step towards compatible plant-pathogen interactions and has a direct impact on the evolution of pathogen populations and species (Giraud et al., 2010; Gladieux et al., 2011; Stukenbrock & McDonald, 2008). In this study, we present comprehensive analyses on unique host-diverging *Z. tritici* populations and further demonstrate the impact of host adaptation on fungal evolution and population divergence. Using population genomics analyses and comparative infection assays, we show that host specificity is a strong determinant of *Z. tritici* population structure and identify candidate genes potentially involved in host adaptation of this important fungal pathogen.

Previous studies have already shown that *Z. tritici* populations are highly diverse within wheat fields and show a strong continental population structure consistent with the geographic origin of the isolates (Hartmann et al., 2018, 2017; Linde, Zhan, & McDonald, 2002; Zhan, Pettway, & McDonald, 2003). Host cultivar specificity has also been shown to be an important factor in the interaction between wheat and *Z. tritici* populations (Ahmed, Mundt, & Coakley, 1995; Arraiano & Brown, 2006; Brown et al., 2001; Eyal, Amiri, & Wahl, 1973; Gurung, Goodwin, Kabbage, Bockus, & Adhikari, 2011; Kema, Annone, Sayoud, & Van Silfhout, 1996; Kema, Sayoud, Annone, & Van Silfhout, 1996; Zhan, Mundt, Hoffer, & McDonald, 2002). Other studies however have not found evidence for population subdivision according to wheat genotype in *Z. tritici* isolates collected from different cultivars in the same field (Hartmann et al., 2018; Linde et al., 2002). Besides in the common bread wheat host (*T. aestivum*), mechanisms of *Z. tritici* host specialization have also been studied in durum wheat (*T. durum*) (Eyal et al., 1973; Kema, Annone, et al., 1996; Kema, Sayoud, et al., 1996; Torriani, Brunner, & McDonald, 2011; Zhan, Kema, & McDonald, 2004) and in the noncompatible grass host *Brachypodium distachyon* (Kellner et al., 2014). However,

while these studies have shown important factors underlying the interaction between *Z. tritici* and specific hosts, none has actually focused on the analyses of *Z. tritici* population structure and genome evolution in these distinct host species and genotypes. Here we show that sympatric *Z. tritici* isolates are organized in three distinct populations based on hosts with little admixture and high divergence between them. Corroborating the genomics analyses, we show by barcode-amplicon sequencing of herbarium specimens that different host species are indeed the determinants of the population structure in the presented Iranian *Z. tritici* isolates and further confirm their host specificity through infection assays.

We observed distinct patterns of linkage disequilibrium (LD) and differences in the overall genetic diversity in our host-diverging *Z. tritici* populations. These results suggest that besides genetically distinct, these populations also have differential sexual recombination frequencies. In line with this, we also observed that these populations differ in the number of mating types idiomorphs detected, with the *Aegilops* population 2 representing a potentially clonal population with only one mating type and the other two populations with a balanced distribution of both mating types idiomorphs, which indicates sexual reproduction within these populations. During the process of adaptation to a new host habitat, fungi can undergo frequent asexual propagation with rare events of sexual reproduction (Giraud et al., 2010; McDonald & Linde, 2002). The lack of sexual recombination can still enhance the accumulation of genomic variants by genetic drift, which in turn can act as barriers of sexual recombination and gene flow with nonself individuals from other host habitats and therefore increase the divergence between these populations (Stukenbrock, 2013). Within generations, the adaptation to a new host can drive the evolution of reproductive isolation between different populations and potentially set the stage for the emergence of new pathogen species or lineages (Giraud et al., 2010; Gladieux et al., 2011; Hendry, Nosil, & Rieseberg, 2007). The association between low sexual recombination and emergence of new pathogen species or lineages has been well documented in the literature, as for example in the potato late blight pathogen *Phythophthora infestans* (Goodwin, Cohen, & Fry, 1994), the yellow rust pathogen *Puccinia striiformis* (Bahri, Leconte, Ouffroukh, De Vallavieille-Pope, & Enjalbert, 2009), and the blast pathogen *Magnaporthe oryzae* in rice (Couch et al., 2005; Gladieux, Ravel, et al., 2018) and wheat (Islam et al., 2016; Latorre et al., 2022). The emergence of these species and lineages was associated to the spread of few clonal lineages that specialized in distinct hosts, leading to a strong divergence between the different clades. Thus, while our observations may be a reflection of the sampling we used, they already suggest that specialization onto a different host impacts the frequency of sexual reproduction in *Z. tritici* and can further underlie the divergence we observe between the different host-infecting *Z. tritici* populations.

Intriguingly, however, were the genetic diversity results we identified for each host-diverging population. Previous comparisons between pathogens occurring on domesticated versus wild host species have shown a substantial loss of genetic diversity in the pathogen populations occurring in the domesticated hosts, as it is the case for the fungal plant pathogen *Rhynchosporium secalis* (Zaffarano, McDonald, & Linde, 2008) and the apple scab fungal pathogen *Venturia inaequalis* (Gladieux, Zhang, et al., 2010). Yet, our nucleotide diversity results show actually the opposite: the wild *Aegilops*-infecting *Z. tritici* populations exhibit a lower genetic diversity compared to the domesticated wheat-infecting population. We hypothesized that this is due to two factors: (i) large effective population sizes in the domesticated pathosystem and (ii) a fluctuating population dynamic in the wild pathosystem. In one hand, the dense and uniform host populations found in agricultural ecosystems allow specialized pathogen populations to rapidly propagate, and in the case of sexually reproductive species, increase the genetic diversity of the local population by recurring cycles of reproduction throughout the season (Stukenbrock & McDonald, 2008). This increase in population size directly impacts the maintenance of genetic diversity given that populations with a larger population size are less prone to genetic drift (Charlesworth, 2009). In the case of *Z. tritici* on wheat, several population genomics studies have shown that the mixed reproductive system found in the *Septoria tritici* blotch (STB) epidemics (characterized by both asexual and sexual reproduction) allows this fungus to maintain high levels of both gene and genotype diversity in field populations around the world (Chen & McDonald, 1996; Jürgens, Linde, & McDonald, 2006; Linde et al., 2002; McDonald & Martinez, 1990; Morais et al., 2019; Singh, Karisto, & Croll, 2021; Zhan, Mundt, & McDonald, 1998; Zhan et al., 2003). In fact, a recent study estimated that wheat fields showing typical STB infections can carry up to 14.0 million *Z. tritici* genotypes per hectare, indicating a large evolutionary potential of the local *Z. tritici* populations and maintenance of high genomic diversity (McDonald, Suffert, Bernasconi, & Mikaberidze, 2022; McDonald & Linde, 2002). On the other hand, host populations in wild ecosystems generally consists of genetically diverse individuals with a heterogeneous distribution in space and time (Möller & Stukenbrock, 2017). This sparse distribution of host populations leads to fluctuations in the local pathogen population sizes and contribute to the metapopulation dynamics of these pathogens on wild plants (Hanski, 1998; Hastings & Harrison, 1994). The metapopulation dynamics of local pathogens in wild ecosystems is characterized by recurrent cycles of population extinction and re-colonization, leading to frequent population bottleneck events and a severe loss of genetic variation at local scales (Laine, Burdon, Nemri, & Thrall, 2014; Möller & Stukenbrock, 2017; Thrall et al., 2012). Considering the annual and variable distribution of *Aegilops* species in different ecosystems and climates throughout Iran (Hosseini, Mehrabian, & Mostafavi, 2022), the

metapopulation dynamics can be a strong factor contributing to the low genetic diversity found in the *Aegilops*-infecting *Z. tritici* populations.

A previous phylogenetic study in *Zymoseptoria* species have pointed out the presence of a host-diverging phylogenetic cluster within the *Z. tritici* clade collected from *Aegilops tauschii* plants in Iran (Quaedvlieg et al., 2011). The small clade composed of three isolates showed a strong bootstrap support value of 94%, supporting the idea of a possible cryptic or ancestral lineage so far considered as *Z. tritici* (Quaedvlieg et al., 2011). In agreement with Quaedvlieg and collaborators (2011), we also demonstrate using different phylogenomic methods that *Aegilops*-infecting *Z. tritici* isolates form distinct clades when compared to worldwide populations of wheat-infecting *Z. tritici* isolates and to closely-related *Zymoseptoria* species. More interestingly, we observed network edges connecting the *Aegilops*-infecting *Z. tritici* clades to the closely related *Zymoseptoria* species (Supplementray Figure S8). Phylogenetic networks have the power to represent conflicting signals or alternative phylogenetic histories that cannot be represented by simple branching trees as e.g. recombination, hybridizations and parallel evolution (Bryant & Moulton, 2004; Fitch, 1997; Huson & Bryant, 2006). Among the phylogenetic network methods, the Neighbor-net can construct networks based on genomic distances and provide a valuable tool to access such evolutionary events (Bryant & Moulton, 2004). Thus, based on the phylogenetic network results observed in this study, we suggest that the *Aegilops*-infecting isolates represent not only divergent *Z. tritici* populations but distinct lineages from the wheat-infecting one. Moreover, the network edges observed between the *Aegilops*-infecting isolates and closely related *Zymoseptoria* species further suggests that these lineages may have experienced gene flow events or hybridization with other *Zymoseptoria* species throughout evolution. Considering the sympatric occurrence of *Z. tritici* and closely related *Zymoseptoria* species in the Middle East, such gene flow events are thought to occur in a considerable frequency. In fact, a study by Feurtey and collaborators (2019) has suggested that regions of high variability in the *Z. tritici* genome are potentially products of recurrent hybridization between different *Zymoseptoria* species, and that these frequent hybridization events may even be a mechanism for the exchange of adaptations and host range expansion between species (Feurtey, Stevens, Stephan, & Stukenbrock, 2019). Additional analyses on gene flow and introgression between the *Aegilops*-infecting isolates and closely related *Zymoseptoria* populations need to be performed to access the occurrence and extent of such evolutionary events.

Adaptation to distinct hosts has impacted not only the population structure but has also left different footprints of recent positive selection on the analyzed *Z. tritici* genomes. We performed genome scans using the CLR and μ statistics methods to detect selective sweeps in each host-

diverging *Z. tritici* population that could have been driven by host adaptation. Regions under recent positive selection were widespread and diverse among the populations analyzed with only two regions overlapping between the *Aegilops* and the wheat-infecting *Z. tritici* populations. These observed differences in the number and location of selective sweeps indicate that these host-diverging populations have experienced selection in different portions of their genomes. A previous study analyzing selective sweeps across the *Z. tritici* genome has also found differences in the number and position of selected regions among four allopatric wheat field populations (Hartmann et al., 2018). The authors argued that differences in fungicide usage, annual mean temperatures and deployed host cultivars in the fields are likely the factors contributing to such heterogeneous selection observed between the analyzed populations (Hartmann & Croll, 2017; Hartmann et al., 2018; Zhan et al., 2005; Zhan, Stefanato, & McDonald, 2006; Zhan & McDonald, 2011). Regarding wild ecosystems, other studies have pointed out that divergent selection can also act upon fungal populations isolated across distinct ecological niches and leave signatures of positive selection in different loci (Branco et al., 2015, 2017; Ellison et al., 2011). Thus, we hypothesize that the distinct selective sweep regions detected between the *Aegilops*- and wheat-infecting *Z. tritici* populations can be explained by the divergent evolutionary history of these populations and the concomitant local adaptation of these isolates to their host environment, as exemplified by our demography and population structure analyses.

Besides highly divergent, selective sweep regions harbored several genes potentially involved in host adaptation. We characterized the selective sweep regions based on previously published gene models and functional annotations on the *Z. tritici* IPO323 reference genome (Feurtey et al., 2020; Grandaubert et al., 2015). In the wheat population, we observed that selective sweep regions were enriched in genes encoding polyketide synthase-like proteins and chitin-binding lysin motifs (LysM; Supplementary Table S13). Polyketide synthase genes have shown to be upregulated in *Z. tritici* during wheat infection and suggested to play a role during disease progression by producing yet unknown secondary metabolites targeting the host (e.g. toxins) (Palma-Guerrero et al., 2017; Rudd et al., 2015). Many fungal pathogens have evolved effector proteins containing LysM motifs that can either protect their cell walls against plant chitinases or prevent the recognition and elicitation of chitin-triggered host immunity, including the fungal pathogens *Cladosporium fulvum*, *Magnaporthe oryzae*, *Colletotrichum higginsianum*, *Verticillium dahlia,* and *Z. tritici* (de Jonge et al., 2010; Kombrink et al., 2017; Marshall et al., 2011; Mentlak et al., 2012; Ökmen et al., 2018; Rovenich, Zuccaro, & Thomma, 2016; Sánchez-Vallet, Mesters, & Thomma, 2015; Takahara et al., 2016). In *Z. tritici*, the LysM effectors Mg1LysM and Mg3LysM have been identified and functionally characterized, and more recently the crystal structure of the effector Mg1LysM has also been

elucidated (Marshall et al., 2011; Sánchez-Vallet et al., 2020). These analyses have revealed that Mg1LysM and Mg3LysM have the ability to protect fungal hyphae against host chitinase hydrolysis and contribute for fungal virulence (Marshall et al., 2011).

In the *Aegilops* population 1, we observed that selective sweep regions were enriched with genes encoding peptidases, diverse glycosyl hydrolases (GHs), and antibiotic biosynthesis monooxygenases (Abm; Supplementary Table S13). In fungal pathogens, secreted peptidases can act as virulence factors and contribute for pathogenesis, as already demonstrated in the fungal pathogens *Botrytis cinerea* (Urbanek & Kaczmarek, 1985), *Fusarium culmorum* (Urbanek & Yirdaw, 1984), *Glomerella cingulata* (Clark, Templeton, & Sullivan, 1997) *and Sclerotinia sclerotiorum* (Poussereau, Creton, Billon-Grand, Rascle, & Fevre, 2001). Glycosyl hydrolases (GHs), which represent the largest class of carbohydrate-active enzymes (CAZymes) and many times act as plant cell wall degrading enzymes (PCWDEs), have also shown to be an important group of secreted proteins contributing to fungal virulence and endophytic lifestyle (Henrissat, 1991; Henrissat & Davies, 1997; Knapp et al., 2018; Kubicek, Starr, & Glass, 2014; Mesny et al., 2021; Zhao, Liu, Wang, & Xu, 2014). However, analyses of the reference *Z. tritici* IPO323 genome have revealed a small complement of PCWDEs compared with other fungi (Goodwin et al., 2011; Ohm et al., 2012) and a low expression of these genes during the biotrophic phase of wheat infection (Brunner, Torriani, Croll, Stukenbrock, & McDonald, 2013), suggesting a biphasic mechanism of nutrient-stealth pathogenesis in the host apoplast as proposed by Goodwin and collaborators (2011) (Goodwin et al., 2011). The presence of two genes encoding antibiotic biosynthesis monooxygenases (Abm; PFAM PF03992) within the selective sweep regions of the *Aegilops*-infecting population, including the selective sweep region on chromosome 7, was also an interesting finding (Supplementary Table S11). Some years ago, a study led by Patkar and collaborators (2015) have shown that the rice blast fungus *Magnaporthe oryzae* secretes an analog of the phytohormone Jasmonic Acid (JA) to modulate host immunity (Patkar et al., 2015). *Magnaporthe oryzae* uses the antibiotic biosynthesis monooxygenase (Abm) to convert intrinsically produced as well as host-derived JA into 12-hydroxyjasmonic acid (12OH-JA) during infection, which in turn suppresses JA-mediated signaling defense responses in the rice host and thus facilitates host colonization (Patkar et al., 2015; Patkar & Naqvi, 2017). The authors also showed that mutant *M. oryzae* strains lacking Abm (*abm*Δ) accumulate methyl JA (MeJA), strongly inducing host defense responses and blocking further host invasion and infection (Patkar et al., 2015; Patkar & Naqvi, 2017). This plethora of genes and gene functions detected between the two host-diverging *Z. tritici* populations suggests that adaptation to distinct hosts drives selection on a wide range of traits targeting several host defense mechanisms.

Effectors often are determinant components of plant-pathogen interactions (de Jonge, Bolton, & Thomma, 2011; Lo Presti et al., 2015; Rep, 2005; van der Does & Rep, 2017). Considering their importance for pathogenesis and host adaptation, we particularly focused on the distribution of genes encoding candidate effector proteins throughout the selective sweep regions. Interestingly, we found a specific selective sweep region on chromosome 7 in the *Aegilops*-infecting population harboring three candidate effector genes in close proximity (Figure 5). Adjacent to these genes, we also found a CAZyme-encoding gene (Zt09_chr_7_00296) and one of the genes with the antibiotic biosynthesis monooxygenase (Abm) domain (Zt09_chr_7_00297). Part of this selective sweep region was detected by both SweeD and RAiSD methods used (Supplementary Table S9), which suggests a strong selective sweep signature. In fact, our analyses on the genomic landscape in this selective sweep region revealed remarkable footprints of recent positive selection and high divergence between the two host-diverging *Z. tritici* populations. As expected in regions under selective sweep (Biswas & Akey, 2006; Nielsen, 2005; Smith & Haigh, 1974), we observed a low nucleotide diversity, a negative Tajima's D and a high LD throughout the genomic region in the *Aegilops*-infecting population (Figure 5). We also observed a similar pattern in specific loci in this region for the wheat-infecting population, particularly at the candidate effector genes Zt09_chr_7_00299 and Zt09_chr_7_00305, although positive Tajima's D values were observed. A previous study analyzing isolate- and temporal-specific infection programs of *Z. tritici* in wheat has shown the expression of particular candidate effector genes and pointed out their potential contribution for virulence in different infection stages (Haueisen et al., 2019). In this study, Haueisen and colleagues (2019) demonstrated that in the *Z. tritici* isolate IPO323 (strain Zt09), the candidate effector gene Zt09_chr_7_00299 was upregulated during wheat infection, and this expression had a constant increase from 4 days-post inoculation (dpi) to 20 dpi (Haueisen et al., 2019). The two other candidate effector genes Zt09_chr_7_00305 and Zt09_chr_7_00308 also showed a peak of expression at 4dpi but reducing the expression afterwards (Haueisen et al., 2019). The high expression *in planta* of these genes supports their role as effectors and further suggests a stage-specific expression during wheat infection, particularly for the genes Zt09_chr_7_00305 and Zt09_chr_7_00308. Hence, we suggest that this region under selection harbors genes potentially involved in *Z. tritici* virulence and host adaptation. However, if these genes play a particular role during *Z. tritici* infection in *Aegilops* species remains a question for further investigation. Experimental validation of this selective sweep region through reverse genetic approaches, especially of the candidate effector genes, is ultimately required to access the full contribution of these genes for the host-adaptive evolution of these unique *Aegilops*-infecting *Z. tritici* populations.

Although we detected several footprints of recent positive selection using stringent thresholds, the possibility of false positives cannot be discarded. Selective sweep signatures can be confounded by those left by the demography history of a population due to skews in the allele frequency distribution. Population expansions, for example, can lead to an excess of low frequency alleles in specific loci compared to the frequency expected under the standard neutral model, which directly impacts neutrality tests as Tajima's D and the inference of selection (Biswas & Akey, 2006; Nielsen, 2005; Ramos-Onsins & Rozas, 2002). Genome -wide selective scans are robust to demographic effects at some extent considering that selection affects specific loci whereas population demographic history and neutral processes are affecting all loci across the genome (Biswas & Akey, 2006; Vitti, Grossman, & Sabeti, 2013). Moreover, we restricted our analyses using only SNP alleles that could have the ancestral state assigned in well-defined populations supported by both demography and population structure analyses, and only reported regions with the strongest signals of selection. Therefore, even though false positive rates can be high due to confounding demography factors, we could detect regions as the one in chromosome 7 in the *Aegilops*-infecting population, which besides harboring footprints of selection, it had one the of the highest outlier scores and it was detected by both selective sweep methods, supporting a high probability of recent positive selection in this region.

Using MK tests, we could not detect a statistically significant shift on the rate of synonymous versus non-synonymous polymorphism on the genes of the selective sweep region in chromosome 7. We hypothesize that this due the differences of the time frame of selection that can be detected between the MK test and the selective sweep scans. Detection of positive selection based on substitutions and polymorphism rates, as the one performed by the MK test, expects that rate of synonymous versus non-synonymous substitutions between species is higher than the polymorphism within species, and therefore requires much longer time frames (McDonald & Kreitman, 1991; Moutinho, Bataillon, & Dutheil, 2020). Moreover, the MK test is unable to distinguish between past and present selection (McDonald & Kreitman, 1991; Nielsen, 2005). Thus, considering selective sweep scans detect recent signatures of positive selection (Nielsen, 2005; Pavlidis & Alachiotis, 2017), it is expected that changes in synonymous versus nonsynonymous polymorphism and divergence ratios will not statistically deviate from the genomic background, and no evidence of positive selection will be detected by the MK test.

Host specificity is a strong factor governing *Z. tritici* evolution. *Zymoseptoria tritici* is considered a specialized pathogen of wheat although several reports have demonstrated its ability to infect other grasses (Suffert, Sache, & Lannou, 2011). Previous studies aiming to understand the epidemiology

of STB (*Septoria tritici* blotch) in other grasses have reported that the species *Agrostis capillaris*, *Bromus hordeaceus* subsp. hordeaceus, *Anisantha sterilis*, *Festuca arundinacea*, *Poa annua* and *Poa pratensis* are susceptible to *Z. tritici* (Ao & Griffiths, 1976; Brokenshire, 1975; Haghdel & Banihashemi, 2005; Sprague, 1944; Weber, 1922; Williams & Jones, 1973). These species were reported in at least two independent studies and therefore can be considered as alternative hosts of *Z. tritici* (Suffert et al., 2011). The ability of *Z. tritici* to infect several species of the *Triticum-Aegilops* complex under controlled conditions has also been reported in other studies (Jlibene, Gustafson, & Amri, 1995; McKendry & Henke, 1994; Raman & Muthukathan, 2015; Seifbarghi, Razavi, Aminian, Zare, & Etebarian, 2009). However, previous coalescence analyses indicated that *Z. tritici* speciated and expanded following its wheat host in the Fertile Crescent, which suggests a so-called "host-tracking" coevolutionary scenario of the pathogen with its host and reinforces the idea that wheat is the main host of *Z. tritici* (Salamini, Özkan, Brandolini, Schäfer-Pregl, & Martin, 2002; Stukenbrock et al., 2007; Stukenbrock & McDonald, 2008; Suffert et al., 2011). Using qualitative and quantitative infection assays under greenhouse conditions, we could also demonstrate that the *Z. tritici* species is able to infect distinct grasses, but it is also highly host-specific. Isolates from the *Aegilops* population 1 could only infect *Aegilops cylindrica* plants, while the isolates belonging to the *Aegilops* population 2 could only infect *Aegilops cylindrica* and *Aegilops tauschii*. The isolates from the wheat-infecting population could only infect *Triticum aestivum* and other closely-related *Triticum* species (e.g. *Triticum turgidum*), although we observed lesions in one *Aegilops tauschii* plant inoculated with a batch of wheat-infecting *Z. tritici* isolates. Considering that the *Z. tritici* populations analyzed in this study come from the Fertile Crescent region, and therefore from the most probable center of origin of the *Z. tritici* species, it is not surprising to find such host range among the isolates. Our infections assays and amplicon sequencing of herbarium specimens have pointed out that *Z. tritici* isolates infect closely-related grass species from the *Triticum-Aegilops* complex that participated in the domestication of common bread wheat (*Triticum aestivum*) in the Fertile Crescent (Feldman & Levy, 2012; Glémin et al., 2019; Salamini et al., 2002). Moreover, our demography analyses show that the *Aegilops*- and wheat-infecting *Z. tritici* populations started to diverge around 10,000 years ago, which coincides with previous findings of *Z. tritici* speciation event (Stukenbrock et al., 2007). Thus, we suggest that these *Z. tritici* populations represent a snapshot of ancient *Z. tritici* lineages and may corroborate and reconstruct the evolutionary history of the *Z. tritici* species through host-tracking.

Host specialization can also lead to the emergence of new fungal species and lineages. Due to the intrinsic properties of different plant species (e.g. resistance genes; antimicrobial metabolites), different host species can be considered as divergent habitats and thus lead to a strong selection

on associated fungal pathogen populations (Giraud et al., 2010; Gladieux et al., 2011; McCoy, 2003). A strong divergent selection due to host shift or host range expansion, if persistent over generations and coupled with a reduction of gene flow between populations, can drive these populations to a process of ecological speciation (Giraud et al., 2010; Rundle & Nosil, 2005). Ecological speciation occurs in allopatry or sympatry and requires that ecologically-based divergent selection acts upon distinct populations and create barriers of gene flow between them, ultimately leading to reproductive isolation (Rundle & Nosil, 2005; Schluter, 2000, 2001). Reproductive isolation can then arise from pre- and post-zygotic reproductive barriers, such as temporal or habitat isolation; genetic incompatibilities; assortative mating and reduced viability of migrants between populations (Giraud, 2006; Giraud et al., 2010; Giraud, Refrégier, Le Gac, de Vienne, & Hood, 2008; Gladieux et al., 2011; Nosil, Vines, & Funk, 2005; Rundle & Nosil, 2005). In this matter, host specialization itself can act as a barrier of gene flow between fungal species or lineages, as already exemplified by studies in *Ascochyta* species (Peever, 2007); *Venturia inequalis* (Gladieux, Caffier, Devaux, & Le Cam, 2010; Le Cam, Parisi, & Arene, 2002); in the fungal pathogen *Colletotrichum kahawae* (Silva et al., 2012); and *Magnaporthe oryzae* lineages infecting distinct grass species (Gladieux, Condon, et al., 2018).

Regarding *Z. tritici*, it has been shown that divergence and reproductive isolation between *Zymoseptoria* species occurred with the presence of gene flow, which suggests that ecological divergence and continuous accumulation of minor incompatibilities may have contributed to the incipient speciation of this fungal species (Stukenbrock, 2013; Stukenbrock et al., 2007). This complex evolutionary history with gene flow in different directions at different time points involving multiple *Zymoseptoria* species makes the implementation of introgression tests as ABBA-BABA or other F-statistics between *Z. tritici* populations challenging (Durand, Patterson, Reich, & Slatkin, 2011; Feurtey et al., 2019; Martin, Davey, & Jiggins, 2015; Peter, 2016). However, our analyses show that divergence between the host-diverging populations driven by host specificity is strong enough to potentially serve as ecological barriers and reduce the gene flow between these host-specific lineages, as evidenced by the lack of reticulate branches observed connecting the *Aegilops*-infecting *Z. tritici* isolates to the worldwide collection of wheat-infecting *Z. tritici*. Moreover, considering the distribution of the hosts we sampled in the Middle East, differences in space and time, and therefore the extent of sympatry, may also contribute for a reduction of gene flow between these populations. Based on these factors, we suggest that these *Z. tritici* lineages may be at initial stages of an ongoing incipient speciation. Experimental crosses and further genealogical inferences using isolates from each host-specific lineage are required to access if

reproductive isolation, in pre- or post-zygotic stages, takes place and if the *Aegilops*-infecting *Z. tritici* can be considered a cryptic species.

Conclusion

Altogether, this study provides evidences of distinct host-specific *Z. tritici* lineages and further suggests that the divergence observed between these lineages is part of an ongoing incipient speciation event. Using population genomics analyses coupled with infection assays, our results underline that host specificity is a strong determinant of the population structure and divergence of the *Z. tritici* species and support a host-tracking coevolutionary scenario. Our work also presents a valuable wild pathosystem (*Aegilops-Z. tritici*) that can be used for functional studies and further evolutionary analyses on the impact of domestication of plants and associated pathogens. At last, our findings highlight the importance of epidemiological surveillance in non-cultivated plant species considering their contribution as reservoirs and "green bridges" of multiple pathogen lineages (Stukenbrock, 2013; Stukenbrock & McDonald, 2008), particularly in regions where wild and domesticated hosts are in close range and new diseases outbreaks are prone to emerge.

Acknowledgements

The authors would like to thank Danilo Pereira for support with the genomics analyses, Janine Müller for support with the infection assays, and the members of the Environmental Genomics group for helpful discussions.

Funding statement

This work was supported by intramural funding of the Max Planck Society and a personal grant from the State of Schleswig-Holstein to Eva H. Stukenbrock. The funders had no role in study design, data collection and analyses, decision to publish, or preparation of the manuscript.

Authors contributions

Conceptualization: WCF, JH, EHS; Samples collection: FS, AA; *Zymoseptoria tritici* isolation: WCF, JH; DNA extractions and clone correction: WCF; Genome sequencing and genomics analyses: WCF, AF; Demography analyses: ICRB; Greenhouse infection assays: WCF, RH; Herbarium

amplicon sequencing and analyses: WCF, RH; Confocal microscopy analyses: RH, JH; Preparation and writing of manuscript: WCF. Editing of original manuscript: WCF, EHS, JH and ICRB.

References

Aguileta, G., Refregier, G., Yockteng, R., Fournier, E., & Giraud, T. (2009). Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution*, 9(4), 656–670.

Ahmed, H. U., Mundt, C. C., & Coakley, S. M. (1995). Host-pathogen relationship of geographically diverse isolates of *Septoria tritici* and wheat cultivars. *Plant Pathology*, 44(5), 838–847.

Alachiotis, N., & Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications Biology*, 1(1). https://doi.org/10.1038/s42003-018-0085-8

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22(13), 1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, 19(10), 535–544. https://doi.org/https://doi.org/10.1016/j.tree.2004.07.021

Ao, H. C., & Griffiths, E. (1976). Change in virulence of *Septoria nodorum* and *S. tritici* after passage through alternative hosts. *Transactions of the British Mycological Society*, 66(2), 337–340.

Arraiano, L. S., & Brown, J. K. M. (2006). Identification of isolate-specific and partial resistance to *Septoria tritici* blotch in 238 European wheat cultivars and breeding lines. *Plant Pathology*, 55(6), 726–738.

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, 43(1), 483–492. https://doi.org/10.1002/0471250953.bi1110s43

Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguileta, G., Snirc, A., … Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology*, 26(7), 2041–2062. https://doi.org/10.1111/mec.13976

Bahri, B., Leconte, M., Ouffroukh, A., De Vallavieille-Pope, C., & Enjalbert, J. (2009). Geographic limits of a clonal population of wheat yellow rust in the Mediterranean region. *Molecular Ecology*, 18(20), 4165–4179. https://doi.org/10.1111/j.1365-294X.2009.04267.x

Banke, S., & McDonald, B. A. (2005). Migration patterns among global populations of the pathogenic fungus *Mycosphaerella graminicola*. *Molecular Ecology*, 14(7), 1881–1896.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. a. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

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(8), 779–788. https://doi.org/10.1094/mpmi-06-17-0135-fi

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, 22(8), 437–446. https://doi.org/10.1016/j.tig.2006.06.005

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Branco, S., Bi, K., Liao, H. L., Gladieux, P., Badouin, H., Ellison, C. E., … Bruns, T. D. (2017). Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*. *Molecular Ecology*, 26(7), 2063–2076. https://doi.org/10.1111/mec.13892

Branco, S., Gladieux, P., Ellison, C. E., Kuo, A., Labutti, K., Lipzen, A., … Bruns, T. D. (2015). Genetic isolation between two recently diverged populations of a symbiotic fungus. *Molecular Ecology*, 24(11), 2747–2758. https://doi.org/10.1111/mec.13132

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, 140(2), 783–796. https://doi.org/10.1093/genetics/140.2.783

Brokenshire, T. (1975). The Role of Graminaceous Species in the Epidemiology of *Septoria tritici* on Wheat. *Plant Pathology*, 24(1), 33–38.

Brown, J K M, Kema, G. H. J., Forrer, H., Verstappen, E. C. P., Arraiano, L. S., Brading, P. A., … Jenny, E. (2001). Resistance of wheat cultivars and breeding lines to Septoria tritici blotch caused by isolates of *Mycosphaerella graminicola* in field trials. *Plant Pathology*, 50(3), 325–338.

Brown, James K. M., & Tellier, A. (2011). Plant-Parasite Coevolution: Bridging the Gap between Genetics and Ecology. *Annual Review of Phytopathology*, 49(1), 345–367. https://doi.org/10.1146/annurev-phyto-072910-095301

Brunner, P. C., Torriani, S. F. F., Croll, D., Stukenbrock, E. H., & McDonald, B. A. (2013). Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. *Molecular Biology and Evolution*, 30(6), 1337–1347. https://doi.org/10.1093/molbev/mst041

Bryant, D., & Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, 21(2), 255–265. https://doi.org/10.1093/molbev/msh018

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), 195–205. https://doi.org/10.1038/nrg2526

Chen, R.-S., & McDonald, B. A. (1996). Sexual reproduction plays a major role in the genetic structure of populations of the fungus *Mycosphaerella graminicola*. *Genetics*, 142(4), 1119–1127. https://doi.org/10.1093/genetics/142.4.1119

Clark, S. J., Templeton, M. D., & Sullivan, P. A. (1997). A secreted aspartic proteinase from *Glomerella cingulata*: purification of the enzyme and molecular cloning of the cDNA. *Microbiology*, 143(4), 1395–1403.

Couch, B. C., Fudal, I., Lebrun, M. H., Tharreau, D., Valent, B., Van Kim, P., … Kohn, L. M. (2005). Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, 170(2), 613–630. https://doi.org/10.1534/genetics.105.041780

Croll, D., Lendenmann, M. H., Stewart, E., & McDonald, B. A. (2015). The impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics*, 201(3), 1213–1228. https://doi.org/10.1534/genetics.115.180968

Czembor, P. C., & Arseniuk, E. (1999). Study of genetic variability among monopycnidial and monopycnidiospore isolates derived from single pycnidia of *Stagonospor*a ssp. and *Septoria tritici* with the use of RAPD-PCR, MP-PCR and rep-PCR techniques. *Journal of Phytopathology*, 147(9), 539–546. https://doi.org/10.1046/j.1439-0434.1999.00434.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., … Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), 1–4. https://doi.org/10.1093/gigascience/giab008

De Jonge, R., Bolton, M. D., & Thomma, B. P. H. J. (2011). How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Current Opinion in Plant Biology*, 14(4), 400–406.

de Jonge, R., Peter van Esse, H., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., … Thomma, B. P. H. J. (2010). Conserved fungal Lysm effector Ecp6 prevents chitin-triggered immunity in plants. *Science*, 329(5994), 953–955. https://doi.org/10.1126/science.1190859

De Maio, N., Schrempf, D., & Kosiol, C. (2015). PoMo: An allele frequency-based approach for species tree estimation. *Systematic Biology*, 64(6), 1018–1031. https://doi.org/10.1093/sysbio/syv048

de Vienne, D. M., Hood, M. E., & Giraud, T. (2009). Phylogenetic determinants of potential host shifts in fungal pathogens. *Journal of Evolutionary Biology*, 22(12), 2532–2541. https://doi.org/10.1111/j.1420-9101.2009.01878.x

de Vienne, D. M., Refr, G., Tellier, A., Hood, M. E., Giraud, T., … López-Villavicencio, M. (2013). Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, 198(2), 347–385. https://doi.org/10.1111/nph.12150

de Vries, S., Stukenbrock, E. H., & Rose, L. E. (2020). Rapid evolution in plant–microbe interactions – an evolutionary genomics perspective. *New Phytologist*. https://doi.org/10.1111/nph.16458

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development*, 35, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Egea, R., Casillas, S., & Barbadilla, A. (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, 36, 157–162. https://doi.org/10.1093/nar/gkn337

Ellison, C. E., Hall, C., Kowbel, D., Welch, J., Brem, R. B., Glass, N. L., & Taylor, J. W. (2011). Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences*, 108(7), 2831–2836. https://doi.org/10.1073/pnas.1014971108

Eyal, Z, Amiri, Z., & Wahl, I. (1973). Physiologic specialization of *Septoria tritici*. *Phytopathology*.

Eyal, Zahir. (1999). The *Septoria tritici* and *Stagonospora nodorum* blotch diseases of wheat. *European Journal of Plant Pathology*, 105(7), 629–641. https://doi.org/10.1023/A:1008716812259

Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. (2020). Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow. *Current Protocols in Microbiology*, 59(1), 1–27. https://doi.org/10.1002/cpmc.128

Fantozzi, E., Kilaru, S., Gurr, S. J., & Steinberg, G. (2021). Asynchronous development of *Zymoseptoria tritici* infection in wheat. *Fungal Genetics and Biology*, *146*(December 2020), 103504. https://doi.org/10.1016/j.fgb.2020.103504

Feldman, M., & Levy, A. A. (2012). Genome evolution due to allopolyploidization in wheat. *Genetics*, 192(3), 763–774. https://doi.org/10.1534/genetics.112.146316

Feurtey, A., Lorrain, C., Croll, D., Eschenbrenner, C., Freitag, M., Habig, M., … Stukenbrock, E. (2020). Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*. *BMC Genomics*, 1–15. https://doi.org/10.1101/864561

Feurtey, A., Stevens, D. M., Stephan, W., & Stukenbrock, E. H. (2019). Interspecific gene exchange introduces high genetic variability in crop pathogen. *Genome Biology and Evolution*, 11(11), 3095–3105. https://doi.org/10.1093/gbe/evz224

Fitch, W. M. (1997). Networks and viral evolution. *Journal of Molecular Evolution*, 44(S1), S65–S75. https://doi.org/10.1007/PL00000059

Fones, H., & Gurr, S. (2015). The impact of *Septoria tritici* blotch disease on wheat: An EU perspective. *Fungal Genetics and Biology*, 79, 3–7. https://doi.org/10.1016/j.fgb.2015.04.004

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Ganopoulos, I., Kapazoglou, A., Bosmali, I., Xanthopoulou, A., Nianiou-Obeidat, I., Tsaftaris, A., & Madesis, P. (2017). Application of the ITS2 region for barcoding plants of the genus *Triticum* L. and *Aegilops* L. Cereal Research Communications, 45(3), 381-389. https://doi.org/10.1556/0806.45.2017.031

Giraud, T. (2006). Speciation: Selection against migrant pathogens: The immigrant inviability barrier in pathogens. *Heredity*, 97(5), 316–318. https://doi.org/10.1038/sj.hdy.6800890

Giraud, T., Gladieux, P., & Gavrilets, S. (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology and Evolution*, 25(7), 387–395. https://doi.org/10.1016/j.tree.2010.03.006

Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D. M., & Hood, M. E. (2008). Speciation in fungi. *Fungal Genetics and Biology*, 45(6), 791–802. https://doi.org/10.1016/j.fgb.2008.02.001

Gladieux, P., Caffier, V., Devaux, M., & Le Cam, B. (2010). Host-specific differentiation among populations of *Venturia inaequalis* causing scab on apple, pyracantha and loquat. *Fungal Genetics and Biology*, 47(6), 511–521.

Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J. L. N., Nhani, A., … Fournier, E. (2018). Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *MBio*, 9(1). https://doi.org/10.1128/mBio.01219-17

Gladieux, P., GuÉrin, F., Giraud, T., Caffier, V., Lemaire, C., Parisi, L., … Le Cam, B. (2011). Emergence of novel fungal pathogens by ecological speciation: Importance of the reduced viability of immigrants. *Molecular Ecology*, 20(21), 4521–4532. https://doi.org/10.1111/j.1365-294X.2011.05288.x

Gladieux, P., Ravel, S., Rieux, A., Cros-Arteil, S., Adreit, H., Milazzo, J., … Tharreau, D. (2018). Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. *MBio*, 9(2). https://doi.org/10.1128/mBio.01806-17

Gladieux, P., Zhang, X. G., Róldan-Ruiz, I., Caffier, V., Leroy, T., Devaux, M., … Le Cam, B. (2010). Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology*, 19(4), 658–674. https://doi.org/10.1111/j.1365-294X.2009.04498.x

Glémin, S., Scornavacca, C., Dainat, J., Burgarella, C., Viader, V., Ardisson, M., … Ranwez, V. (2019). Pervasive hybridizations in the history of wheat relatives. *Science Advances*, 5(5). https://doi.org/10.1126/sciadv.aav9188

Goodwin, S. B., Cohen, B. A., & Fry, W. E. (1994). Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), 11591–11595. https://doi.org/10.1073/pnas.91.24.11591

Goodwin, S. B., M'Barek, S. Ben, Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., … Kema, G. H. J. (2011). Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics*, 7(6). https://doi.org/10.1371/journal.pgen.1002070

Grandaubert, J., Bhattacharyya, A., & Stukenbrock, E. H. (2015). RNA-seq-based gene annotation and comparative genomics of four fungal grass pathogens in the genus *Zymoseptoria* identify novel orphan genes and species-specific invasions of transposable elements. *G3*, 5(7), 1323–1333. https://doi.org/10.1534/g3.115.017731

Grünwald, N. J., McDonald, B. A., & Milgroom, M. G. (2016). Population genomics of fungal and oomycete pathogens. *Annual Review of Phytopathology*, 54(1), 323–346. https://doi.org/10.1146/annurev-phyto-080614-115913

Gurung, S., Goodwin, S. B., Kabbage, M., Bockus, W. W., & Adhikari, T. B. (2011). Genetic differentiation at microsatellite loci among populations of *Mycosphaerella graminicola* from California, Indiana, Kansas, and North Dakota. *Phytopathology*, 101(10), 1251–1259. https://doi.org/10.1094/PHYTO-08-10-0212

Haghdel, M., & Banihashemi, Z. (2005). Survival and host range of *Mycosphaerella graminicola* the causal agent of *Septoria* leaf blotch of wheat. *Iranian Journal of Plant Pathology*, 41(4).

Hansen, E. M. (1987). Speciation in plant pathogenic fungi: the influence of agricultural practice. *Canadian Journal of Plant Pathology=Revue Canadienne de Phytopathologie*.

Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396(6706), 41–49.

Hartmann, F. E., & Croll, D. (2017). Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. *Molecular Biology and Evolution*, 34(11), 2808–2822. https://doi.org/10.1093/molbev/msx208

Hartmann, F. E., McDonald, B. A., & Croll, D. (2018). Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Molecular Ecology*, 27(12), 2725–2741. https://doi.org/10.1111/mec.14711

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME Journal*, 11(5), 1189–1204. https://doi.org/10.1038/ismej.2016.196

Hastings, A., & Harrison, S. (1994). Metapopulation dynamics and genetics. *Annual Review of Ecology and Systematics*, 167–188.

Haubold, B., Klötzl, F., & Pfaffelhuber, P. (2015). Andi: Fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics*, 31(8), 1169–1175. https://doi.org/10.1093/bioinformatics/btu815

Haueisen, J., Möller, M., Eschenbrenner, C. J., Grandaubert, J., Seybold, H., Adamiak, H., & Stukenbrock, E. H. (2019). Highly flexible infection programs in a specialized wheat pathogen. *Ecology and Evolution*, 9(1), 275–294. https://doi.org/10.1002/ece3.4724

Hendry, A. P., Nosil, P., & Rieseberg, L. H. (2007). The speed of ecological speciation. *Functional Ecology*, 21(3), 455–464. https://doi.org/10.1111/j.1365-2435.2006.01240.x

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, 280(2), 309–316.

Henrissat, B., & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology*, 7(5), 637–644.

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4), 2335–2352. https://doi.org/10.1534/genetics.104.036947

Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2011). Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Science*, 171(9), 1059–1071. https://doi.org/10.1086/656306.

Hosseini, N., Mehrabian, A., & Mostafavi, H. (2022). Modeling climate change effects on spatial distribution of wild *Aegilops* L. (Poaceae) toward food security management and biodiversity conservation in Iran. *Integrated Environmental Assessment and Management*, 18(3), 697–708. https://doi.org/10.1002/ieam.4531

Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* (Oxford, England), 14(1), 68–73.

Huson, D. H., & Bryant, D. (2005). Estimating phylogenetic trees and networks using SplitsTree 4. Manuscript in Preparation, Software Available from www.splitstree.org.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254–267. https://doi.org/10.1093/molbev/msj030

Islam, M. T., Croll, D., Gladieux, P., Soanes, D. M., Persoons, A., Bhattacharjee, P., … Kamoun, S. (2016). Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biology*, 14(1), 1–11. https://doi.org/10.1186/s12915-016-0309-7

Jaenike, J. (1985). Parasite pressure and the evolution of amanitin tolerance in *Drosophila*. *Evolution*, 39(6), 1295–1301.

Jlibene, M., Gustafson, J. P., & Amri, A. (1995). Evaluation of *Aegilops* for resistance to *Septoria tritici*. *Al Awamia*, 91, 83–91.

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323–329. https://doi.org/10.1038/nature05286

Jürgens, T., Linde, C. C., & McDonald, B. A. (2006). Genetic structure of *Mycosphaerella graminicola* populations from Iran, Argentina and Australia. *European Journal of Plant Pathology*, 115(2), 223–233. https://doi.org/10.1007/s10658-006-9000-0

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kellner, R., Bhattacharyya, A., Poppe, S., Hsu, T. Y., Brem, R. B., & Stukenbrock, E. H. (2014). Expression profiling of the wheat pathogen *Zymoseptoria tritici* reveals genomic patterns of transcription and host-specific regulatory programs. *Genome Biology and Evolution*, 6(6), 1353–1365. https://doi.org/10.1093/gbe/evu101

Kema, G. H. J., Annone, J. G., Sayoud, R., & Van Silfhout, C. H. (1996). Genetic variation for virulence and resistance in the wheat-*Mycosphaerella graminicola* pathosystem. I. Interactions between pathogen isolates and host cultivars. *Phytopathology*.

Kema, G. H. J., Sayoud, R., Annone, J. G., & Van Silfhout, C. H. (1996). Genetic variation for virulence and resistance in the wheat-*Mycosphaerella graminicola* pathosystem. II. Analysis of interactions between pathogen isolates and host cultivars. *Phytopathology*.

Knapp, D. G., Németh, J. B., Barry, K., Hainaut, M., Henrissat, B., Johnson, J., … Kovács, G. M. (2018). Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. *Scientific Reports*, 8(1), 1–13. https://doi.org/10.1038/s41598-018-24686-4

Kohn, L. M. (2005). Mechanisms of fungal speciation. *Annual Review of Phytopathology*, 43(1), 279–308. https://doi.org/10.1146/annurev.phyto.43.040204.135958

Kombrink, A., Rovenich, H., Shi-Kunne, X., Rojas-Padilla, E., van den Berg, G. C. M., Domazakis, E., … Thomma, B. P. H. J. (2017). *Verticillium dahliae* LysM effectors differentially contribute to virulence on plant hosts. *Molecular Plant Pathology*, 18(4), 596–608. https://doi.org/10.1111/mpp.12520

Kubicek, C. P., Starr, T. L., & Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology*, 52, 427–451. https://doi.org/10.1146/annurev-phyto-102313-045831

Laine, A. L., Burdon, J. J., Nemri, A., & Thrall, P. H. (2014). Host ecotype generates evolutionary and epidemiological divergence across a pathogen metapopulation. *Proceedings of the Royal Society B: Biological Sciences*, 281(1787). https://doi.org/10.1098/rspb.2014.0522

Latorre, S. M., Lang, P. L. M. M., Burbano, H. A., & Gutaker, R. M. (2020). Isolation, library preparation, and bioinformatic analysis of historical and ancient plant DNA. *Current Protocols in Plant Biology*, 5(4), 1–28. https://doi.org/10.1002/cppb.20121

Latorre, S. M., Were, V. M., Foster, A. J., Langner, T., Malmgren, A., Harant, A., … Kamoun, S. (2022). A pandemic clonal lineage of the wheat blast fungus. *BioRxiv*, 2022.06.06.494979. Retrieved from http://biorxiv.org/content/early/2022/06/07/2022.06.06.494979.abstract

Le Cam, B., Parisi, L., & Arene, L. (2002). Evidence of two formae speciales in *Venturia inaequalis*, responsible for apple and Pyracantha scab. *Phytopathology*, 92(3), 314–320. https://doi.org/10.1094/PHYTO.2002.92.3.314

Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Linde, C. C., Zhan, J., & McDonald, B. A. (2002). Population Structure of *Mycosphaerella graminicola*: From lesions to continents. *Phytopathology*, 92(9), 946–955. https://doi.org/10.1094/phyto.2002.92.9.946

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., … Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Lorrain, C., Feurtey, A., Möller, M., Haueisen, J., & Stukenbrock, E. (2021). Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. *G3 Genes|Genomes|Genetics*, 11(4). https://doi.org/10.1093/g3journal/jkab068

Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., … Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, 105(27), 9272-9277. https://doi.org/10.1073/pnas.0803466105

Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K. E., … Rudd, J. J. (2011). Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiology*, 156(2), 756–769. https://doi.org/10.1104/pp.111.176347

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, 32(1), 244–257. https://doi.org/10.1093/molbev/msu269

Mayr, E. (1982). The growth of biological thought. Cambridge, MA: Belknap.

McCoy, K. D. (2003). Sympatric speciation in parasites–what is sympatry? *Trends in Parasitology*, 19(9), 400–404.

McDonald, B. A., & Martinez, J. P. (1990). DNA restriction fragment length polymorphisms among *Mycosphaerella graminicola* (anamorph *Septoria tritici*) isolates collected from a single wheat field. *Phytopathology*, Vol. 80, p. 1368. https://doi.org/10.1094/phyto-80-1368

McDonald, Bruce A., & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, 40(1), 349–379. https://doi.org/10.1146/annurev.phyto.40.120501.101443

McDonald, Bruce A., Suffert, F., Bernasconi, A., & Mikaberidze, A. (2022). How large and diverse are field populations of fungal plant pathogens? The case of *Zymoseptoria tritici*. *Evolutionary Applications*, 1–14. https://doi.org/10.1111/eva.13434

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 351(6328), 652–654.

McDonald, M. C., Oliver, R. P., Friesen, T. L., Brunner, P. C., & McDonald, B. A. (2013). Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. *New Phytologist*, 199(1), 241–251. https://doi.org/10.1111/nph.12257

McKendry, A. L., & Henke, G. E. (1994). Evaluation of wheat wild relatives for resistance to *Septoria tritici* blotch. *Crop Science*, 34(4), 1080–1084.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, 11(1), 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Mentlak, T. A., Kombrink, A., Shinya, T., Ryder, L. S., Otomo, I., Saitoh, H., … Talbot, N. J. (2012). Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *The Plant Cell*, 24(1), 322–335. https://doi.org/10.1105/tpc.111.092957

Mesny, F., Miyauchi, S., Thiergart, T., Pickel, B., Atanasova, L., Karlsson, M., … Hacquard, S. (2021). Genetic determinants of endophytism in the *Arabidopsis* root mycobiome. *Nature Communications*, 12(1), 1–15. https://doi.org/10.1038/s41467-021-27479-y

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution*, 28(11), 659–669. https://doi.org/10.1016/j.tree.2013.08.003

Mohd-Assaad, N., McDonald, B. A., & Croll, D. (2018). Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. *Genome Biology and Evolution*, 10(5), 1315–1332. https://doi.org/10.1093/gbe/evy087

Moller, M., Habig, M., Lorrain, C., Feurtey, A., Haueisen, J., Fagundes, W. C., … Stukenbrock, E. H. (2021). Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. *PLoS Genetics*, 17(3), 1–27. https://doi.org/10.1371/journal.pgen.1009448

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, 15(12), 756–771. https://doi.org/10.1038/nrmicro.2017.76

Money, N. P. (2006). The triumph of the fungi: a rotten history. Oxford University Press.

Morais, D., Duplaix, C., Sache, I., Laval, V., Suffert, F., & Walker, A. S. (2019). Overall stability in the genetic structure of a *Zymoseptoria tritici* population from epidemic to interepidemic stages at a small spatial scale. *European Journal of Plant Pathology*, 154(2), 423–436. https://doi.org/10.1007/s10658-018-01666-y

Moutinho, A. F., Bataillon, T., & Dutheil, J. Y. (2020). Variation of the adaptive substitution rate between species and within genomes., 34(3), 315–338. https://doi.org/10.1007/s10682-019-10026-z

Nei, M., & Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10), 5269–5273.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, 15(11), 1566–1575. https://doi.org/10.1101/gr.4252305

Nirmaladevi, D., Venkataramana, M., Srivastava, R. K., Uppalapati, S. R., Gupta, V. K., Yli-Mattila, T., ... & Chandra, N. S. (2016). Molecular phylogeny, pathogenicity and toxigenicity of *Fusarium oxysporum* f. sp. lycopersici. *Scientific reports*, 6(1), 1-14. https://doi.org/10.1038/srep21367

Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, 59(4), 705–719. https://doi.org/10.1111/j.0014-3820.2005.tb01747.x

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., … Grigoriev, I. V. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen dothideomycetes fungi. *PLoS Pathogens*, 8(12). https://doi.org/10.1371/journal.ppat.1003037

Ökmen, B., Kemmerich, B., Hilbig, D., Wemhöner, R., Aschenbroich, J., Perrar, A., … Doehlemann, G. (2018). Dual function of a secreted fungalysin metalloprotease in *Ustilago maydis*. *New Phytologist*, 220(1), 249–261. https://doi.org/10.1111/nph.15265

Palma-Guerrero, J., Ma, X., Torriani, S. F. F., Zala, M., Francisco, C. S., Hartmann, F. E., … McDonald, B. A. (2017). Comparative transcriptome analyses in *Zymoseptoria tritici* reveal significant differences in gene expression among strains during plant infection. *Molecular Plant-Microbe Interactions*, 30(3), 231–244. https://doi.org/10.1094/MPMI-07-16-0146-R

Patkar, R. N., Benke, P. I., Qu, Z., Chen, Y. Y. C., Yang, F., Swarup, S., & Naqvi, N. I. (2015). A fungal monooxygenase-derived jasmonate attenuates host innate immunity. *Nature Chemical Biology*, 11(9), 733–740. https://doi.org/10.1038/nchembio.1885

Patkar, R. N., & Naqvi, N. I. (2017). Fungal manipulation of hormone-regulated plant defense. *PLoS Pathogens*, 13(6), e1006334. https://doi.org/10.1371/journal.ppat.1006334

Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research (Greece)*, 24(1), 1–17. https://doi.org/10.1186/s40709-017-0064-0

Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, 30(9), 2224–2234. https://doi.org/10.1093/molbev/mst112

Peever, T. L. (2007). Role of host specificity in the speciation of *Ascochyta* pathogens of cool season food legumes. In *Ascochyta blights of grain legumes* (pp. 119–126). Springer.

Peter, B. M. (2016). Admixture, population structure, and f-statistics. *Genetics*, 202(4), 1485–1501. https://doi.org/10.1534/genetics.115.183913

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., & Croll, D. (2017). Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. *Frontiers in Plant Science*, 8, 119. https://doi.org/10.3389/fpls.2017.00119

Poussereau, N., Creton, S., Billon-Grand, G., Rascle, C., & Fevre, M. (2001). Regulation of acp1, encoding a non-aspartyl acid protease expressed during pathogenesis of *Sclerotinia sclerotiorum*. *Microbiology*, 147(3), 717–726.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575.

Quaedvlieg, W., Kema, G. H. J., Groenewald, J. Z., Verkley, G. J. M., Seifbarghi, S., Razavi, M., … Crous, P. W. (2011). *Zymoseptoria* gen. nov.: A new genus to accommodate Septoria-like species occurring on graminicolous hosts. *Persoonia: Molecular Phylogeny and Evolution of Fungi*, 26, 57–69. https://doi.org/10.3767/003158511X571841

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10(6), 417–430. https://doi.org/10.1038/nrmicro2790

Raman, T., & Muthukathan, G. (2015). Host range of *Septoria* species on cereals and some wild grasses in Iran. *Phytopathologia Mediterranea*, 54(2), 241–252. https://doi.org/10.14601/Phytopathol

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19(12), 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034

Ratanacherdchai, K., Wang, H., Lin, F., Soytong, K (2010). ISSR for comparison of cross-inoculation potential of *Colletotrichum capsici* causing chilli anthracnose. *African Journal of Microbiology Research*, 4(2), 076-083.

Rep, M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiology Letters*, 253(1), 19–27. https://doi.org/10.1016/j.femsle.2005.09.014

Ronald, J., & Akey, J. M. (2005). Genome-wide scans for loci under selection in humans. *Human Genomics*, 2(2), 113–125. https://doi.org/10.1186/1479-7364-2-2-113

Rovenich, H., Zuccaro, A., & Thomma, B. P. H. J. (2016). Convergent evolution of filamentous microbes towards evasion of glycan-triggered immunity. *New Phytologist*, 212(4), 896–901. https://doi.org/10.1111/nph.14064

Rudd, J. J., Kanyuka, K., Hassani-Pak, K., Derbyshire, M., Andongabo, A., Devonshire, J., … Courbot, M. (2015). Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle. *Plant Physiology*, 167(3), 1158–1185. https://doi.org/10.1104/pp.114.255927

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., & Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*, 3(6), 429–441. https://doi.org/10.1038/nrg817

Sánchez-Vallet, A., Mesters, J. R., & Thomma, B. P. H. J. (2015). The battle for chitin recognition in plant-microbe interactions. *FEMS Microbiology Reviews*, 39(2), 171–183. https://doi.org/10.1093/femsre/fuu003

Sánchez-Vallet, A., Tian, H., Rodriguez-Moreno, L., Valkenburg, D. J., Saleem-Batcha, R., Wawra, S., … Thomma, B. P. H. J. (2020). A secreted LysM effector protects fungal hyphae through chitin-dependent homodimer polymerization. *PLoS Pathogens*, 16(6 June), 1–21. https://doi.org/10.1371/journal.ppat.1008652

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46(8), 919–925.

Schiffels, S., & Wang, K. (2020). MSMC and MSMC2: the multiple sequentially markovian coalescent. In *Statistical population genomics* (pp. 147-166). Humana, New York, NY.

Schluter, D. (2000). The ecology of adaptive radiation. OUP Oxford.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, 16(7), 372–380.

Schrempf, D., Minh, B. Q., De Maio, N., von Haeseler, A., & Kosiol, C. (2016). Reversible polymorphism-aware phylogenetic models and their application to tree inference. *Journal of Theoretical Biology*, 407, 362–370. https://doi.org/10.1016/j.jtbi.2016.07.042

Schrempf, D., Minh, B. Q., Von Haeseler, A., & Kosiol, C. (2019). Polymorphism-aware species trees with advanced mutation models, bootstrap, and rate heterogeneity. *Molecular Biology and Evolution*, 36(6), 1294–1301. https://doi.org/10.1093/molbev/msz043

Seifbarghi, S., Razavi, M., Aminian, H., Zare, R., & Etebarian, H. R. (2009). Studies on the host range of *Septoria* species on cereals and some wild grasses in Iran. *Phytopathologia Mediterranea*, 48(3), 422–429.

Shapiro, B., & Hofreiter, M. (2009). Methods in Molecular Biology - Ancient DNA. In *Springer Protocols* (Vol. 840).

Shin, J. H., Blay, S., McNeney, B., & Graham, J. (2006). LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of Statistical Software*, 16(3), 1–9. https://doi.org/10.18637/jss.v016.c03

Silva, D. N., Talhinhas, P., Cai, L., Manuel, L., Gichuru, E. K., Loureiro, A., … Batista, D. (2012). Host-jump drives rapid and recent ecological speciation of the emergent fungal pathogen *Colletotrichum kahawae*. *Molecular Ecology*, 21(11), 2655–2670. https://doi.org/10.1111/j.1365-294X.2012.05557.x

Singh, N. K., Karisto, P., & Croll, D. (2021). Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen *Zymoseptoria tritici*. Microbial *Genomics*, 7(10). https://doi.org/10.1099/MGEN.0.000678

Slippers, B., Stenlid, J., & Wingfield, M. J. (2005). Emerging pathogens: fungal host jumps following anthropogenic introduction. *Trends in Ecology & Evolution*, 20(8), 420–421.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23–35.

Sprague, R. (1944). *Septoria* Disease of Gramineae in Western United States. Corvallis, OR, USA: Oregon State College.

Stephan, W. (2016). Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, 25(1), 79–88. https://doi.org/10.1111/mec.13288

Stergiopoulos, I., & de Wit, P. J. G. M. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, 47(1), 233–263. https://doi.org/10.1146/annurev.phyto.112408.132637

Stukenbrock, E.H., Christiansen, F. B., Hansen, T. T., Dutheil, J. Y., & Schierup, M. H. (2012). Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences*, 109(27), 10954–10959. https://doi.org/10.1073/pnas.1201403109

Stukenbrock, E.H., & McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions*, 22(4), 371–380. https://doi.org/10.1094/MPMI-22-4-0371

Stukenbrock, E.H. (2013). Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist*, 199(4), 895–907. https://doi.org/10.1111/nph.12374

Stukenbrock, E.H., Banke, S., Javan-Nikkhah, M., & McDonald, B. A. (2007). Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, 24(2), 398–411. https://doi.org/10.1093/molbev/msl169

Stukenbrock, E.H., & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, 46(1), 75–100. https://doi.org/10.1146/annurev.phyto.010708.154114

Stukenbrock, E.H., Quaedvlieg, W., Javan-Nikhah, M., Zala, M., Crous, P. W., & McDonald, B. A. (2012). *Zymoseptoria ardabiliae* and *Z. pseudotritici*, two progenitor species of the *Septoria tritici* leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella graminicola*). *Mycologia*, 104(6), 1397–1407. https://doi.org/10.3852/11-374

Stukenbrock, E.H, Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M., … Schierup, M. H. (2011). The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research*, 21(12), 2157–2166. https://doi.org/10.1101/gr.118851.110

Stukenbrock, E.H, & Dutheil, J. Y. (2018). Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. *Genetics*, 208(3), 1209-1229. https://doi.org/10.1534/genetics.117.300502

Suffert, F., Sache, I., & Lannou, C. (2011). Early stages of *Septoria tritici* blotch epidemics of winter wheat: build-up, overseasoning, and release of primary inoculum. *Plant Pathology*, 60(2), 166–177. https://doi.org/10.1111/j.1365-3059.2010.02369.x

Takahara, H., Hacquard, S., Kombrink, A., Hughes, H. B., Halder, V., Robin, G. P., … O'Connell, R. J. (2016). *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. *New Phytologist*, 211(4), 1323–1337. https://doi.org/10.1111/nph.13994

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

Team, R. C. (2013). R: A language and environment for statistical computing.

Tellier, A., Moreno-Gámez, S., & Stephan, W. (2014). Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution*, 68(8), 2211–2224. https://doi.org/10.1111/evo.12427

Thines, M. (2019). An evolutionary framework for host shifts – jumping ships for survival. *New Phytologist*, 605–617. https://doi.org/10.1111/nph.16092

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Thrall, P. H., Laine, A. L., Ravensdale, M., Nemri, A., Dodds, P. N., Barrett, L. G., & Burdon, J. J. (2012). Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecology Letters*, 15(5), 425–435. https://doi.org/10.1111/j.1461-0248.2012.01749.x

Torriani, S. F. F., Brunner, P. C., & McDonald, B. A. (2011). Evolutionary history of the mitochondrial genome in *Mycosphaerella* populations infecting bread wheat, durum wheat and wild grasses. *Molecular Phylogenetics and Evolution*, 58(2), 192–197. https://doi.org/10.1016/j.ympev.2010.12.002

Urbanek, H, & Yirdaw, G. (1984). Hydrolytic ability of acid protease of *Fusarium culmorum* and its possible role in phytopathogenesis. *Acta Microbiologica Polonica*, 33(2), 131–136.

Urbanek, Henryk, & Kaczmarek, A. (1985). Extracellular proteinases of the isolate of *Botrytis cinerea* virulent to apple tissues. *PWN*.

van der Does, H. C., & Rep, M. (2017). Adaptation to the host environment by plant-pathogenic fungi. *Annual Review of Phytopathology*, 55(1), 427–450. https://doi.org/10.1146/annurev-phyto-080516-035551

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, 47, 97–120.

Waalwijk, C., Mendes, O., Verstappen, E. C. P., de Waard, M. A., & Kema, G. H. J. (2002). Isolation and characterization of the mating-type idiomorphs from the wheat *Septoria* leaf blotch fungus *Mycosphaerella graminicola*. *Fungal Genetics and Biology*, 35(3), 277–286.

Weber, G. F. (1922). Septoria diseases of cereals. II. *Septoria* diseases of wheat. Speckled leaf blotch of wheat. *Phytopathology*, 12, 558-585.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358. https://doi.org/10.2307/2408641

Wickham, H. (2016). Data analysis. In *ggplot2* (pp. 189-201). Springer, Cham.

Williams, J. R., & Jones, D. G. (1973). Infection of grasses by *Septoria nodorum* and *S. tritici*. *Transactions of the British Mycological Society*, 60(2), 355–358.

Zaffarano, P. L., McDonald, B. A., & Linde, C. C. (2008). Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution*, 62(6), 1418–1436. https://doi.org/10.1111/j.1558-5646.2008.00390.x

Zhan, J., Kema, G. H. J., & McDonald, B. A. (2004). Evidence for natural selection in the mitochondrial genome of *Mycosphaerella graminicola*. *Phytopathology*, 94(3), 261–267. https://doi.org/10.1094/PHYTO.2004.94.3.261

Zhan, J., Linde, C. C., Jürgens, T., Merz, U., Steinebrunner, F., & McDonald, B. A. (2005). Variation for neutral markers is correlated with variation for quantitative traits in the plant pathogenic fungus *Mycosphaerella graminicola*. *Molecular Ecology*, 14(9), 2683–2693. https://doi.org/10.1111/j.1365-294X.2005.02638.x

Zhan, J., Mundt, C. C., Hoffer, M. E., & McDonald, B. A. (2002). Local adaptation and effect of host genotype on the rate of pathogen evolution: an experimental test in a plant pathosystem. *Journal of Evolutionary Biology*, 15(4), 634–647. https://doi.org/10.1046/j.1420-9101.2002.00428.x

Zhan, J., Mundt, C. C., & McDonald, B. A. (1998). Measuring immigration and sexual reproduction in field populations of *Mycosphaerella graminicola*. *Phytopathology*, 88(12), 1330–1337. https://doi.org/10.1094/PHYTO.1998.88.12.1330

Zhan, J., Pettway, R. E., & McDonald, B. A. (2003). The global genetic structure of the wheat pathogen *Mycosphaerella graminicola* is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. *Fungal Genetics and Biology*, 38(3), 286–297. https://doi.org/10.1016/S1087-1845(02)00538-8

Zhan, J., Stefanato, F. L., & Mcdonald, B. A. (2006). Selection for increased cyproconazole tolerance in *Mycosphaerella graminicola* through local adaptation and in response to host resistance. *Molecular Plant Pathology*, 7(4), 259–268. https://doi.org/10.1111/j.1364-3703.2006.00336.x

Zhan, J., & McDonald, B. A. (2011). Thermal adaptation in the fungal pathogen *Mycosphaerella graminicola*. *Molecular Ecology*, 20(8), 1689–1701. https://doi.org/10.1111/j.1365-294X.2011.05023.x

Zhao, Z., Liu, H., Wang, C., & Xu, J. (2014). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, 15(1), 1-15. https://doi.org/10.1186/1471-2164-15-6

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Supplementary Material

Supplementary Tables

All supplementary tables in ".xlsx" format are deposited on the supplementary USB key.

Table S1. *Zymoseptoria* isolates used in this study.

List of *Zymoseptoria* isolates used in this study indicating their geographical origin, year of collection, host and references**.**

Table S2. List of primers used in this study.

Table S3. Specimens and *ITS2* GenBank accession numbers of the Triticeae species used for the herbarium phylogenetic analyses.

Table S4. List of host specimens used for the qualitative and quantitative infections assays under greenhouse conditions.

Table S5. Nucleotide BLAST results of *ITS2* sequences from the herbarium specimens.

Homology-based searches were performed using nucleotide BLAST (blastn) (Altschul et al., 1990) optimized for discontiguous megablast against the National Center for Biotechnology Information (NCBI) nucleotide database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Only the first three hits ordered by their percent identity values are listed for each sample.

Table S6. Mating type distribution between *Z. tritici* populations.

Mating type idiomorphs were detected in all analyzed *Z. trtici* isolates using BLAST searches of both mating type loci (MAT1-1 and MAT1-2; Waalwijk et al., 2002) sequences against whole-genome assemblies. Chi-square ($\chi 2$) tests were performed within each population based on the number of idiomorphs identified.

Table S7. Qualitative results of greenhouse infection assays by *Z. tritici* population batches.

Infection assays under greenhouse conditions were performed using 31 wild grass and wheat relatives (Supplementary Table S4) as hosts. In this experiment, *Z. tritici* isolates were inoculated by batches of 12 isolates each representing the three populations identified in this study: "*Aegilops* pop 1", representing the *Aegilops* population 1; "*Aegilops* pop 2", representing the *Aegilops* population 2; and the "Wheat pop", representing the wheat population. Mock (0.1% Tween20 in sterile water) and the *Z. tritici* isolate IPO323 (strain Zt244) were used as controls. Hosts were considered susceptible to a specific population if at least one inoculated plant showed presence of pycnidia. Numbers in brackets represent proportion of plants with presence of pycnidia at 21 or 28 dpi over total number of plants inoculated. "+" represents compatible interactions (visual presence of pycnidia) and "-" represents incompatible interactions. Compatible interactions are highlighted in bold.

Table S8. Qualitative results of greenhouse infection assays by individual *Z. tritici* isolates.

Infection assays were performed under greenhouse conditions. In this experiment, *Z. tritici* isolates were inoculated individually in *Aegilops cylindrica* and *Triticum aestivum* cultivar Riband plants (Supplementary Table S4). Mock (0.1% Tween20 in sterile water) and the *Z. tritici* isolate IPO323 (strain Zt244) were used as controls. Hosts were considered susceptible to a specific isolate if at least one inoculated plant showed presence of pycnidia. Numbers in brackets represent proportion of plants with presence of pycnidia over total number of plants inoculated. "+" represents compatible interactions (visual presence of pycnidia) and "-" represents incompatible interactions. Compatible interactions are highlighted in bold.

Table S9. Genomic coordinates of selective sweep regions identified in the two host-diverging *Z. tritici* populations *Aegilops* population 1 and wheat population using the CLR and μ statistics methods.

Grey-colored cells represent overlapping regions commonly detected by the two selective sweep methods within each population. Red-colored letters represent overlapping regions commonly detected between the two populations.

Table S10. Summary of selective sweep regions detected in the two host-diverging *Z. tritici* populations *Aegilops* population 1 and wheat population using the CLR and μ statistics methods.

Table S11. Summary of gene and TE models found in the selective sweep regions of *Aegilops* population 1 and wheat population using the CLR and μ statistics methods.

Gene/TE models and functional annotation of genes in the *Z. tritici* IPO323 reference genome were retrieved from Feurtey et al., 2020, Grandaubert et al., 2015 and Lorrain et al., 2021.

Table S12. Gene ontology (GO) terms linked to biological processes significantly enriched in the selective sweep regions of the *Aegilops* population 1 and wheat population.

Only GO terms with $p \leq 0.05$ are shown. p values were calculated using Fischer's exact test applying the topGO (Alexa, Rahnenführer, & Lengauer, 2006) algorithm "weight01" considering GO term hierarchy.

Table S13. PFAM domains significantly enriched in the selective sweep regions of the *Aegilops* population 1 and wheat population.

Only PFAM domains with $p \leq 0.05$ are shown. p values were calculated using a chi-square ($\chi 2$) test and a custom python script published previously (Haueisen et al., 2019).

Supplementary Figures

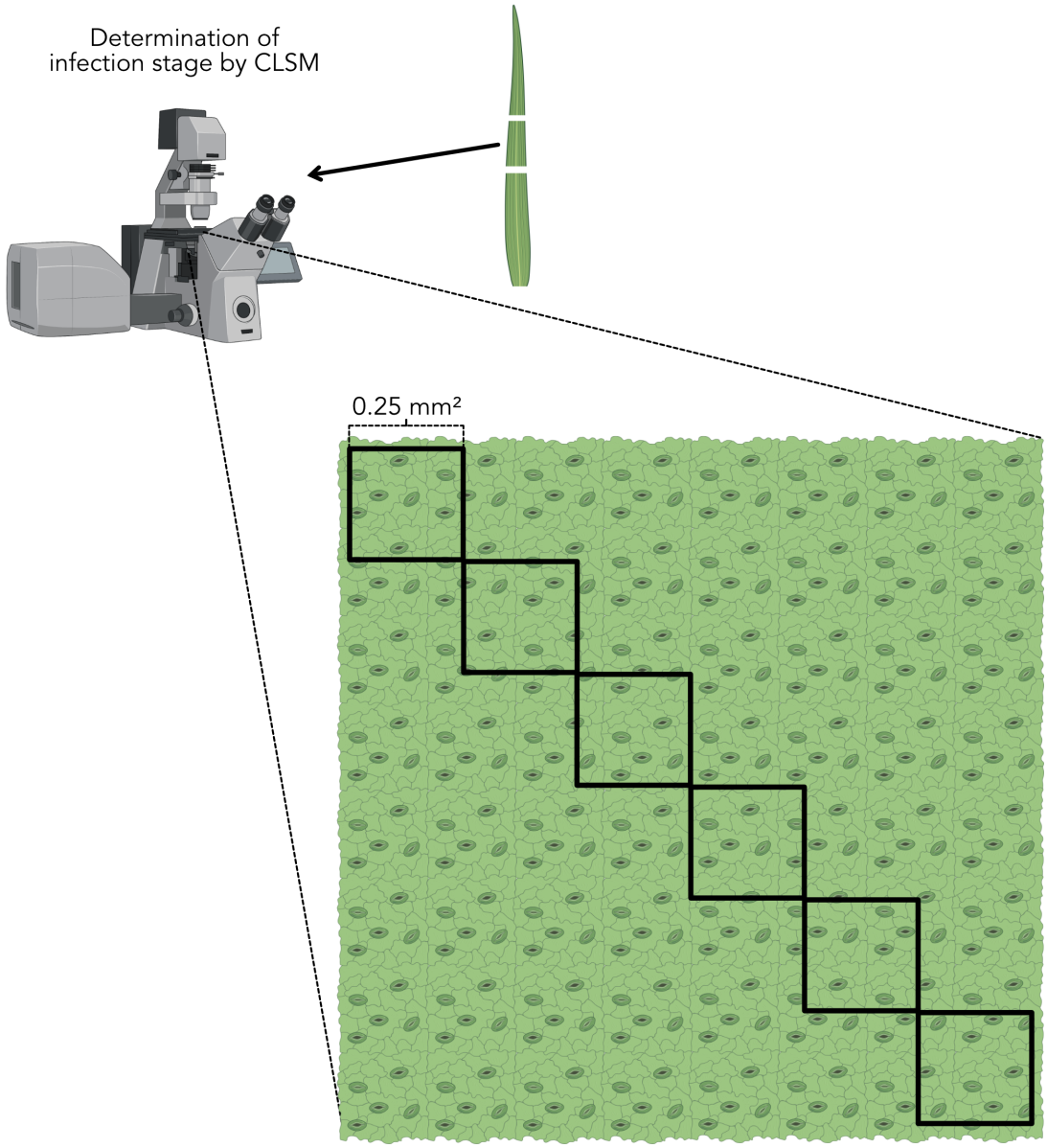


Figure S1. Schematic illustration representing the determination of infection stages by confocal microscopy analyses. Based on the morphological description of infection stages in *Z. tritici* published previously (Haueisen et al. 2019), we determined the infection stage of each inoculated leaf sample by observing the most progressed infection events (ranging from stage "A" to "D") along a transect of six 0.25 mm2 squares of leaf area. The determination and assignment of infection stages was performed in a single-blinded procedure (see Methods). Illustrations were designed on BioRender.com. Leaf area of a dicot plant is shown as example.

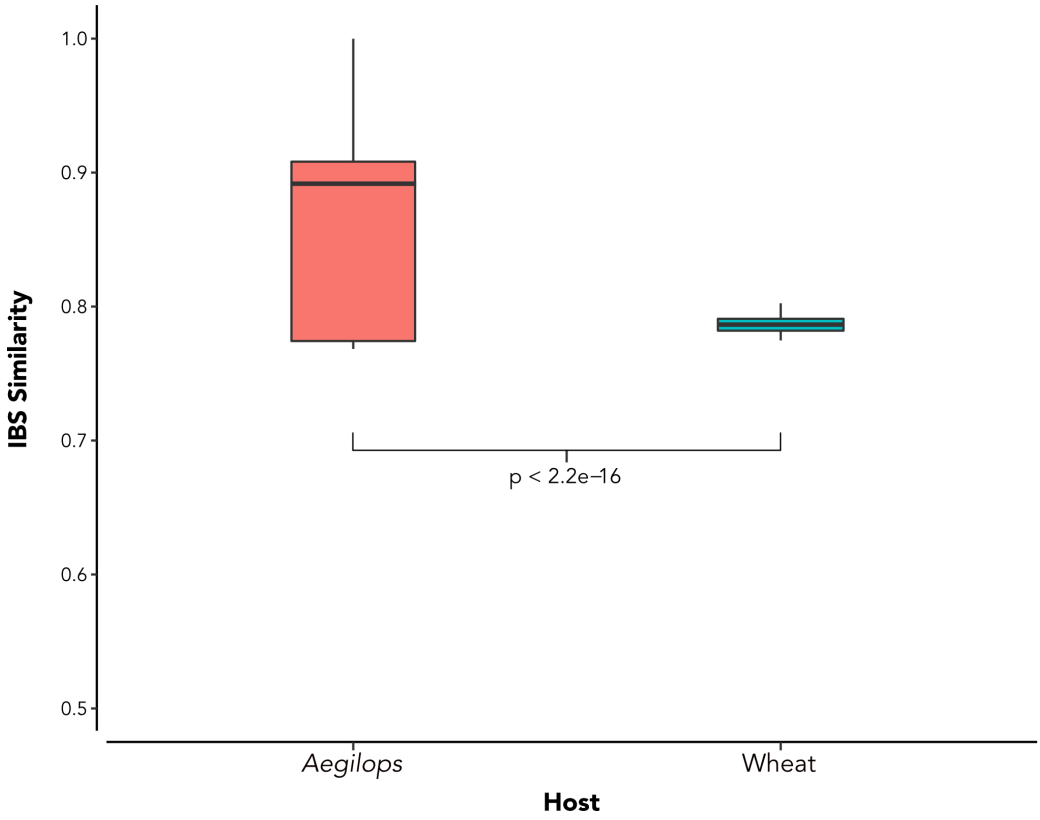


Figure S2. Genome-wide distribution of IBS values per host collection. Boxplot showing the distribution of IBS (Identity-By-State) values calculated genome-wide by pairs of *Z. tritici* isolates within each collection. P-value was calculated using Wilcoxon rank sum test.

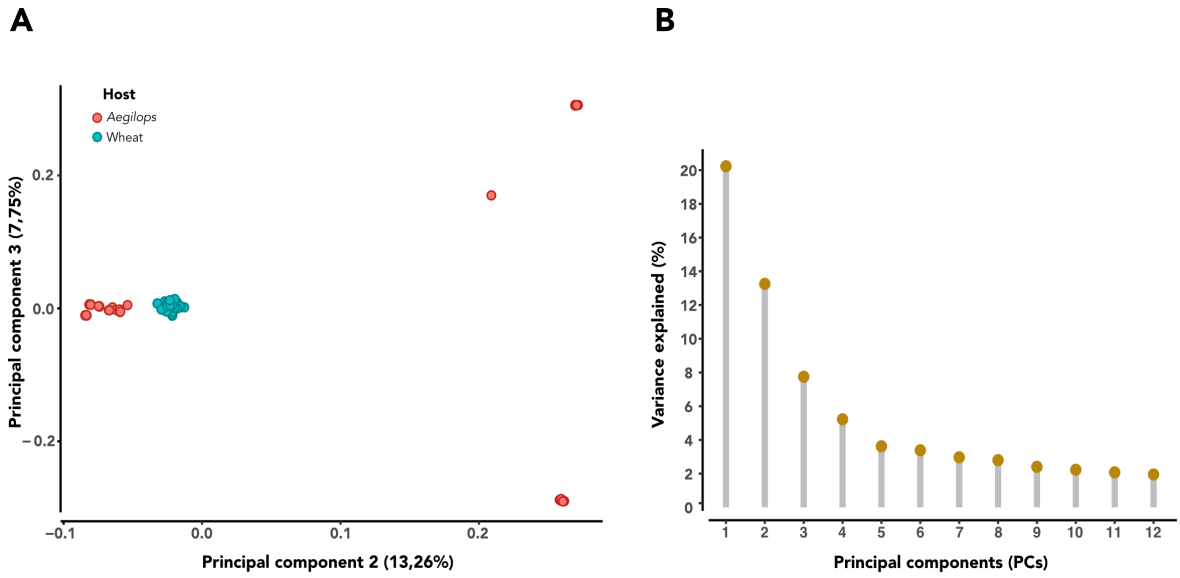


Figure S3. Principal component analysis on genome-wide independent SNPs among host-diverging *Zymoseptoria tritici* isolates. (A) PCA showing the second and third principal components. Percentage of variance explained by each component is shown in parentheses. **(B)** Lollipop plot showing the percentage of variance explained by the first 12 PCs.

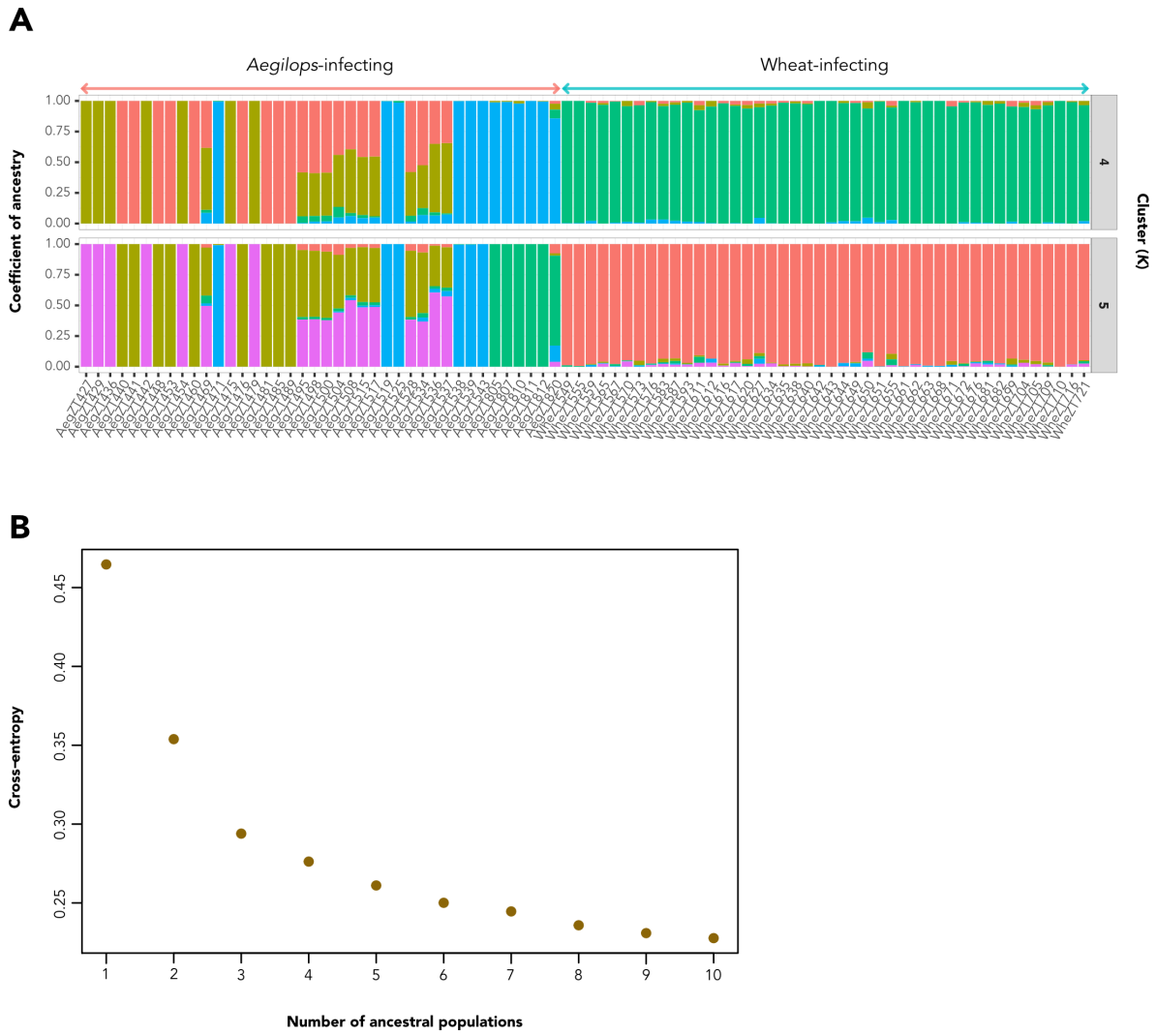


Figure S4. Admixture analyses of host-diverging *Z. tritici* isolates. (A) Ancestry coefficients and clustering assignments using the sNMF v2.0 software (Frichot & François, 2015; Frichot, Mathieu, Trouillon, Bouchard, & François, 2014). Each vertical bar represents an isolate from the *Aegilops* (red arrow) and wheat-infecting (blue arrow) *Z. tritici* collection with each color indicating one genetic cluster. The color height in each vertical bar represents the probabilities of cluster assignment based on genome-wide independent SNPs. **(B)** Dot plot showing the cross-entropy values calculated for each ancestral population tested (*K*=1 to *K*=10). Values represent the lowest score over 100 independent runs for each *K*.

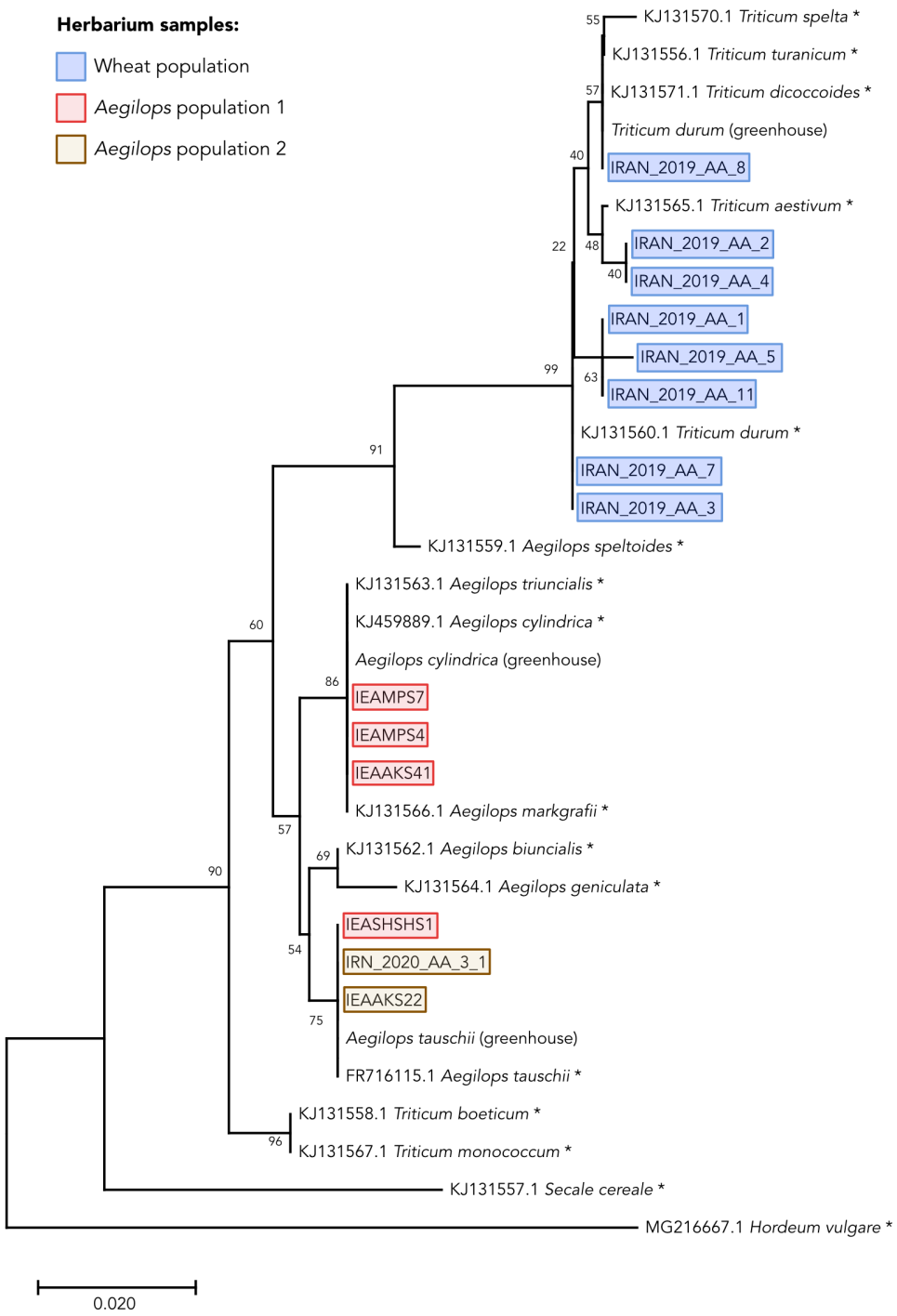


Figure S5. Herbarium species correlate with *Z. tritici* population structure. Neighbor-Joining (NJ) dendrogram performed with 1000 bootstrap replicates based on multiple sequence alignment of plant ITS2. The genetic distances were calculated according to the Kimura 2-Parameter model. Numbers at the branches show bootstrap values. Sequences amplified from greenhouse-grown plants (Supplementary Table S5) are indicated. Additional sequences were obtained from the NCBI database and are marked with asterisks and accession number. The *ITS2* sequence of *Hordeum vulgare* was used as an outgroup. List of accession numbers and references are described in Supplementary Table 3.

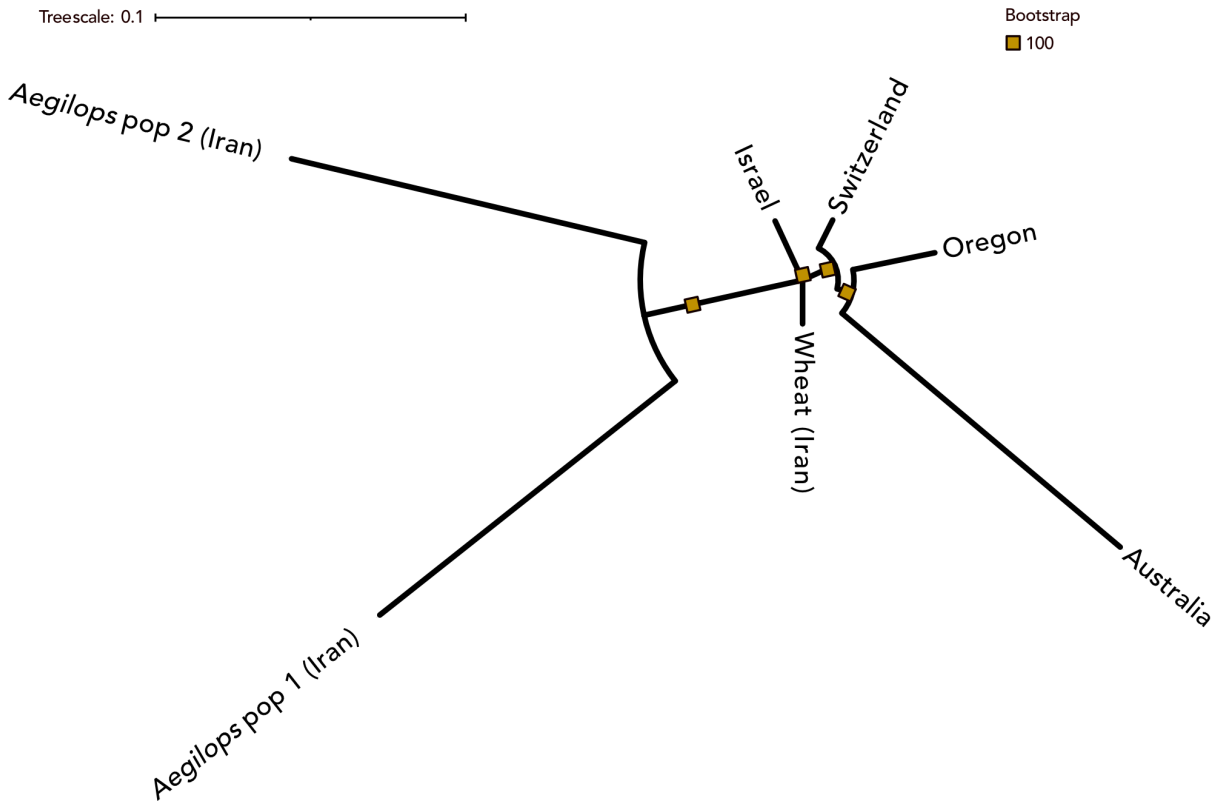


Figure S6. Maximum-likelihood (ML) tree based on intraspecific *Z. tritici* polymorphisms. ML tree was constructed based on genome-wide biallelic SNPs with no missing genotypes across the different *Z. tritici* populations using the software PoMo (De Maio et al., 2015; Schrempf et al., 2016) and 100 bootstraps.

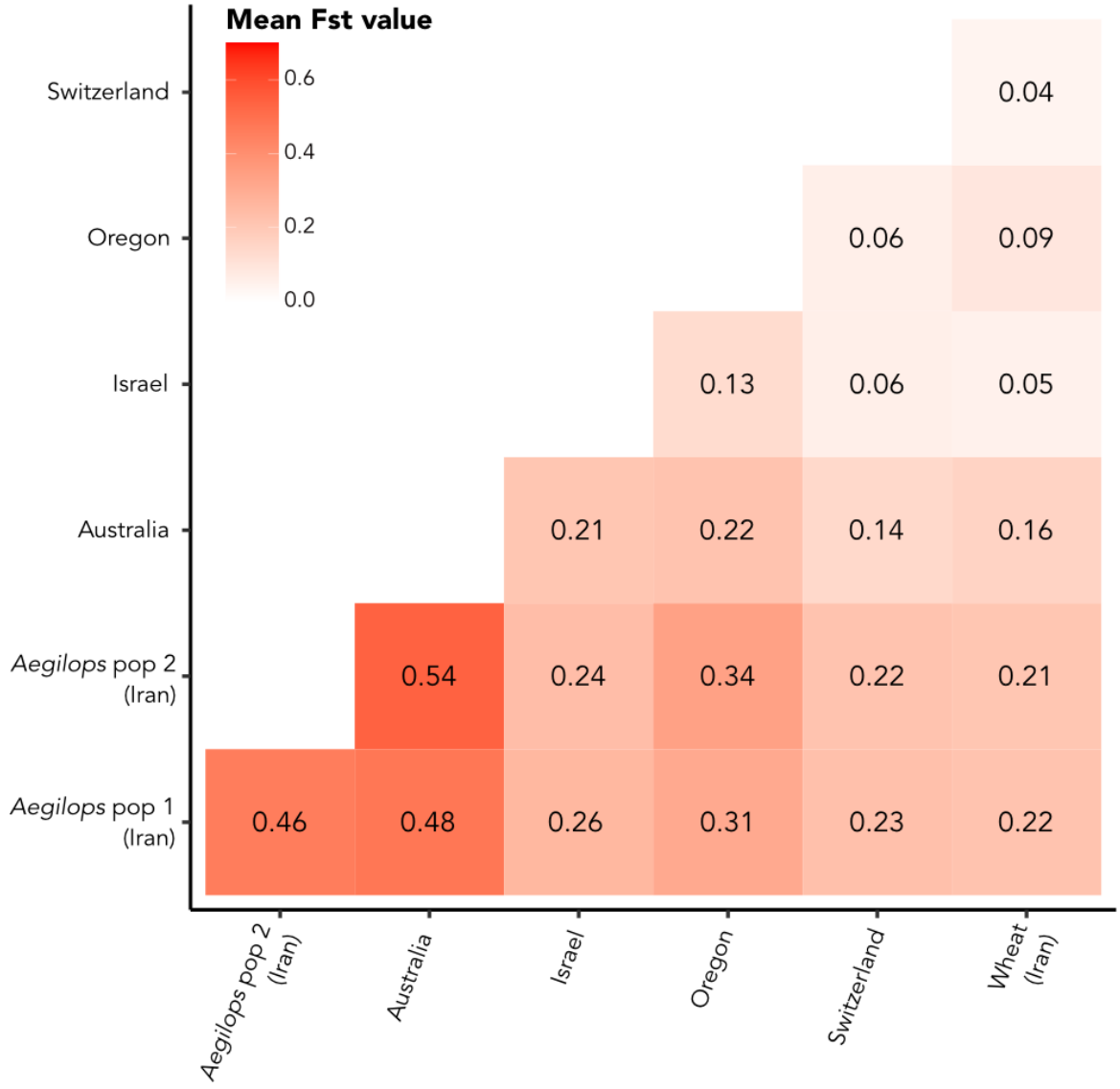


Figure S7. Pairwise divergence values (Fst) between *Z. tritici* populations. Heatmap showing genome-wide mean divergence values based on the Weir & Cockerham's Fst fixation indices (Weir & Cockerham, 1984). Mean Fst values were calculated using biallelic SNPs with no missing genotypes across the different *Z. tritici* populations. Darker colors indicate higher divergence between pairs of populations.

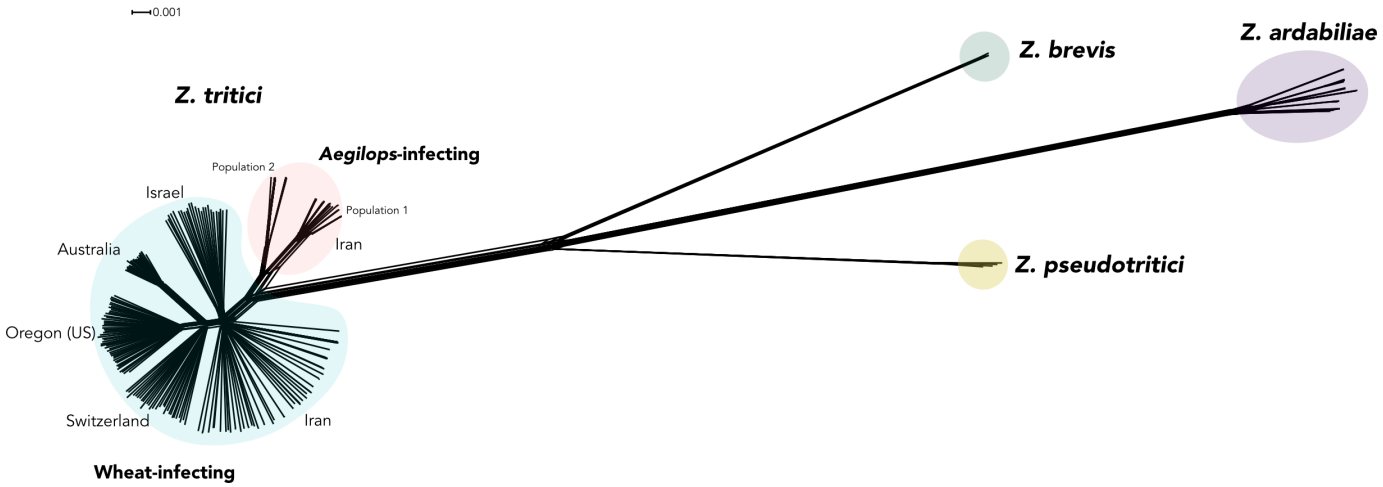


Figure S8. Phylogenetic network between *Zymoseptoria* species. Neighbor-net network based on whole genome distances. In the *Z. tritici* cluster, red ellipse represents the two *Aegilops*-infecting *Z. tritici* populations while the blue elliptical shape represents worldwide wheat-infecting *Z. tritici* populations. Geographical locations of origin are also indicated. Scale bar indicates branch distances.

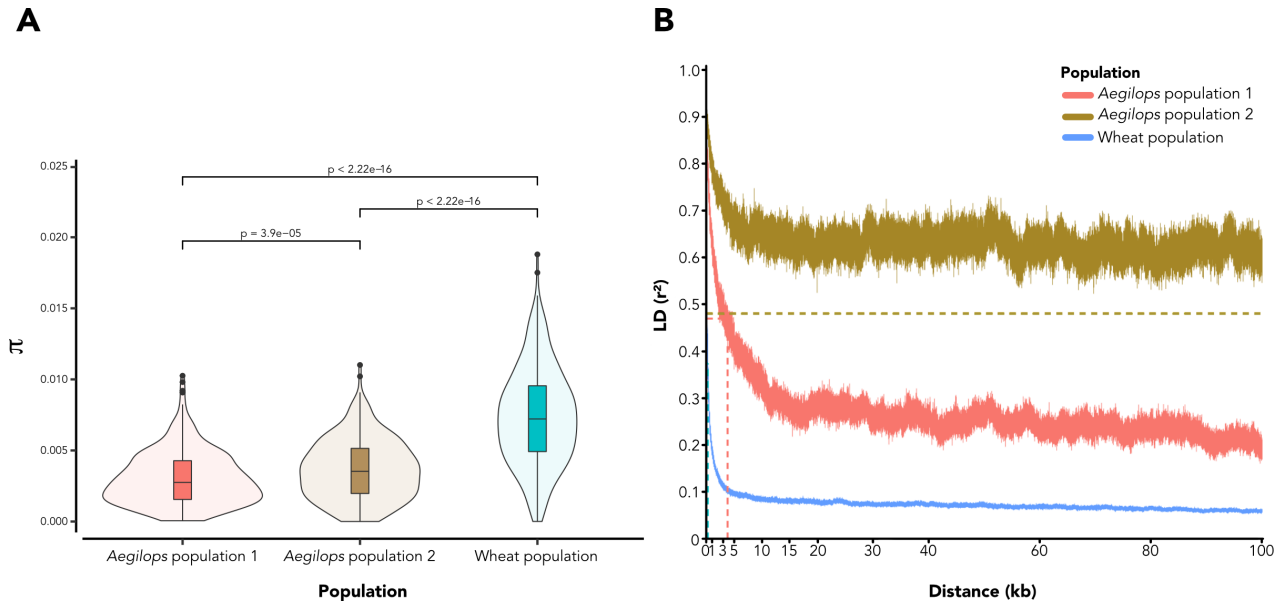


Figure S9. Nucleotide diversity and linkage disequilibrium decay in host-diverging *Z. tritici* populations. (A) Genome-wide nucleotide diversity (π) calculated for each *Z. tritici* population individually. P-values were calculated using pairwise Wilcoxon rank sum tests. **(B)** Linkage disequilibrium decay across 100 kb of chromosome 1. Top, brown line represents the LD decay for *Aegilops*-population 2; mid, red line represents the LD decay for the *Aegilops* population 1 while the bottom, blue line represents the LD decay for the wheat-infecting population. Dashed perpendicular lines represent distance to reach 50% maximum LD value in each population.

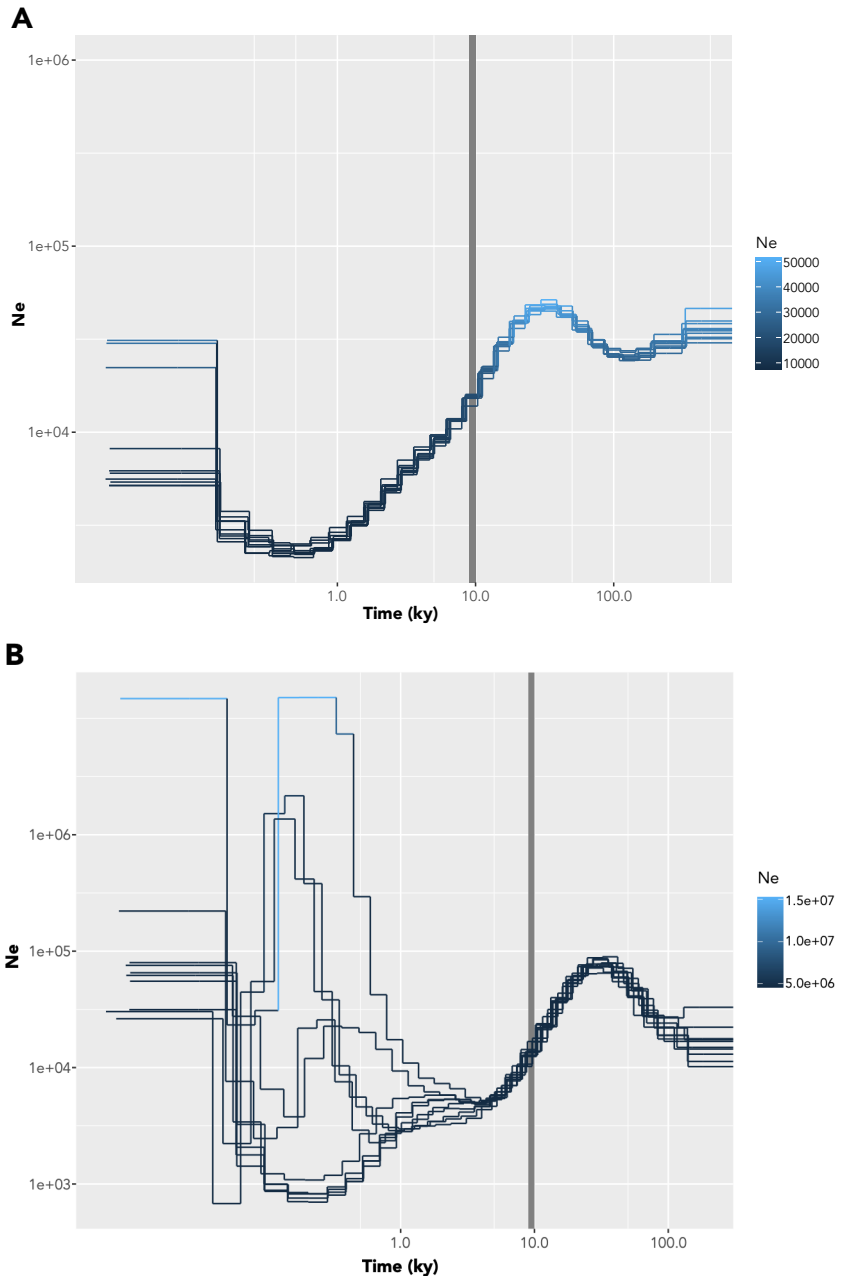


Figure S10. Effective population size (Ne) of the *Aegilops*- and wheat-infecting *Z. tritici* populations. The effective population size was calculated for each population individually using 10 different combinations of 8 haploid genotypes for the wheat-infecting population **(A)** and 10 different combinations of 8 haploid genotypes for the *Aegilops*-infecting population **(B)**. Each line in both plots represents the different genotype combinations. Grey, vertical bars represent timeframe of wheat domestication and *Z. tritici* speciation (~10.000 years ago; Feldman & Levy, 2012; Glémin et al., 2019; Salamini et al., 2002; Stukenbrock et al., 2007). Time is indicated as thousands of years (ky). Calculations were performed using MSMC2 (Schiffels & Durbin, 2014; Schiffels & Wang, 2020) assuming a mutation rate of 3.3e-8 per cell cycle (Lynch et al., 2008) and 1 generation of sexual reproduction per year (Stukenbrock et al., 2011).

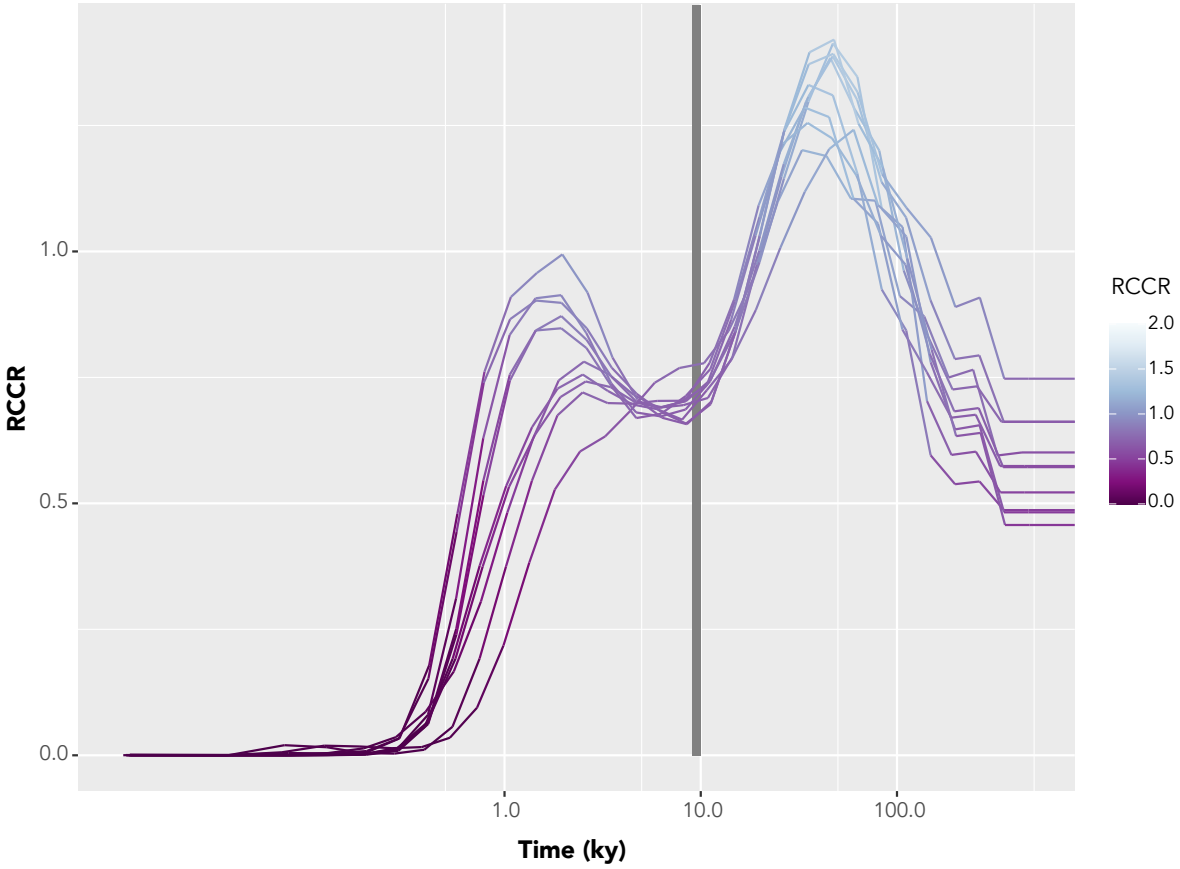


Figure S11. Relative cross-coalescence rate (RCCR) between host-diverging *Z. tritici* populations. The RCCR was calculated using 10 different combinations of 16 haploid genomes, 8 from wheat-infecting and 8 from the *Aegilops*-infecting *Z. tritici* population. Each line represents a different combination of samples between the two populations. Grey, vertical bar represents timeframe of wheat domestication and *Z. tritici* speciation (~10.000 years ago; Feldman & Levy, 2012; Glémin et al., 2019; Salamini et al., 2002; Stukenbrock et al., 2007). Time is indicated as thousands of years (ky). Calculations were performed using MSMC2 (Schiffels & Durbin, 2014; Schiffels & Wang, 2020) assuming a mutation rate of 3.3e-8 per cell cycle (Lynch et al., 2008) and 1 generation of sexual reproduction per year (Stukenbrock et al., 2011).


Aegilops pop 1
Aegilops pop 2
Wheat pop
Zt244
Mock
Aegilops cylindrica
Aegilops tauschii
Triticum aestivum cv. Chinese Spring
Triticum aestivum cv. Riband
Triticum compactum
Triticum dicoccon
Triticum monococcum
Triticum spelta var. alba

Figure S12. Summary of qualitative virulence assay performed by *Z. tritici* population batches. In this experiment, batches of twelve *Z. tritici* isolates based on the identified populations, namely "*Aegilops* pop 1" batch, representing the *Aegilops* population 1, "*Aegilops* pop 2" batch, representing the *Aegilops* population 2 and the "Wheat pop" batch, representing the wheat population were inoculated individually along mock and the reference *Z. tritici* isolate IPO323 (strain Zt244) as control treatments in 31 wheat and wild grasses accessions (Supplementary Table S4). Pictures show inoculated leaves at 21 days post-inoculation (dpi) and represent a subset of the results obtained. Each grey column represents leaves inoculated with a different population batch. Complete description of qualitative virulence results can be found at Supplementary Table S7.

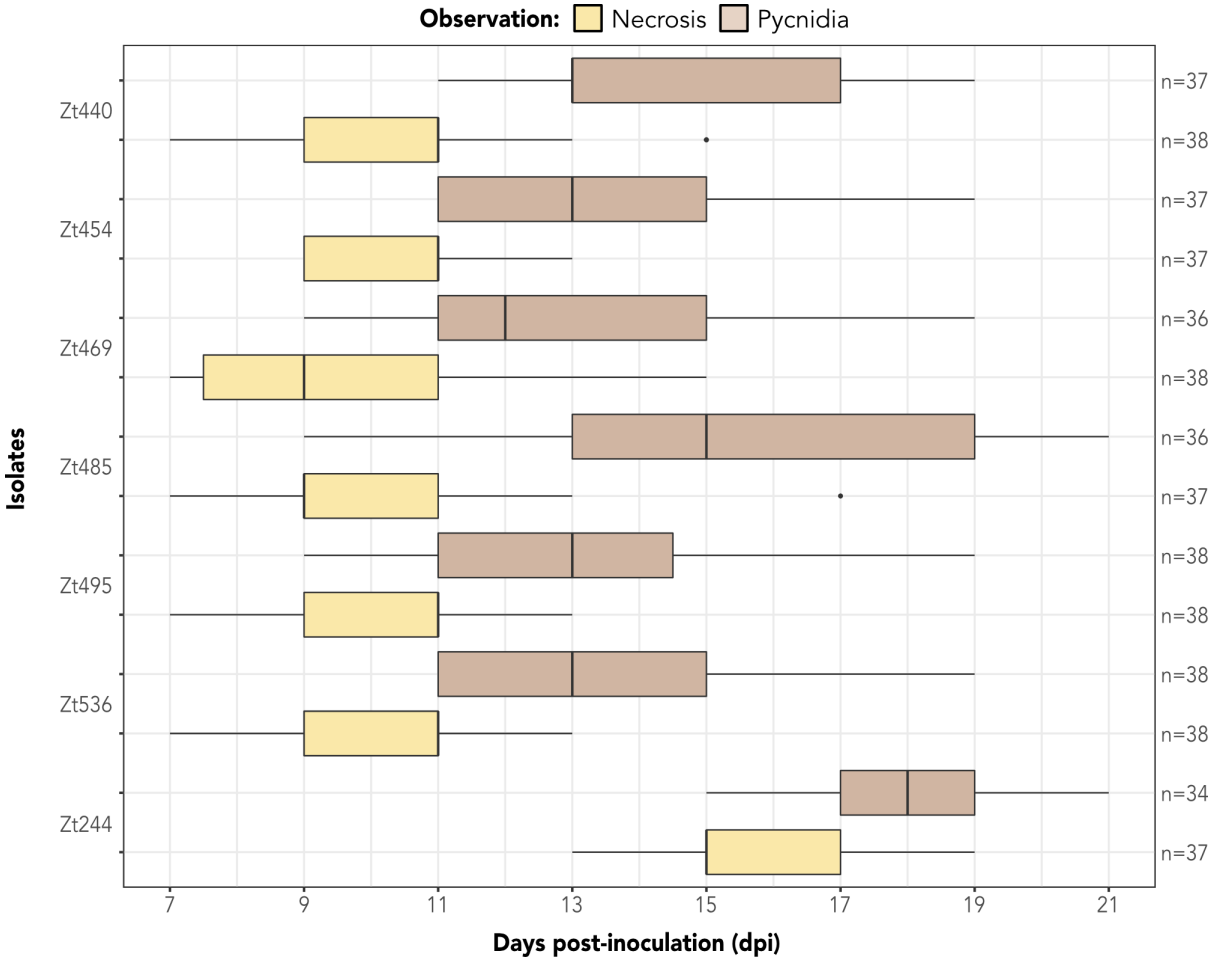


Figure S13. Disease progression analyses on host-diverging *Z. tritici* isolates. To analyze the infection progression, we used a subset of *Aegilops*-infecting *Z. tritici* isolates from *Aegilops* population 1 along the *Z. tritici* IPO323 isolate (strain Zt244) and mock treatments as controls, and inoculated them individually at similar inoculum concentrations on its correspondent hosts: *Aegilops cylindrica*, for the *Aegilops*-infecting *Z. tritici* isolates and *Triticum aestivum* cv. Riband, for the *Z. tritici* IPO323. First signs of necrosis (yellow bars) and pycnidia (brown bars) were recorded every two days. Number of plants showing symptoms (n) are indicated.

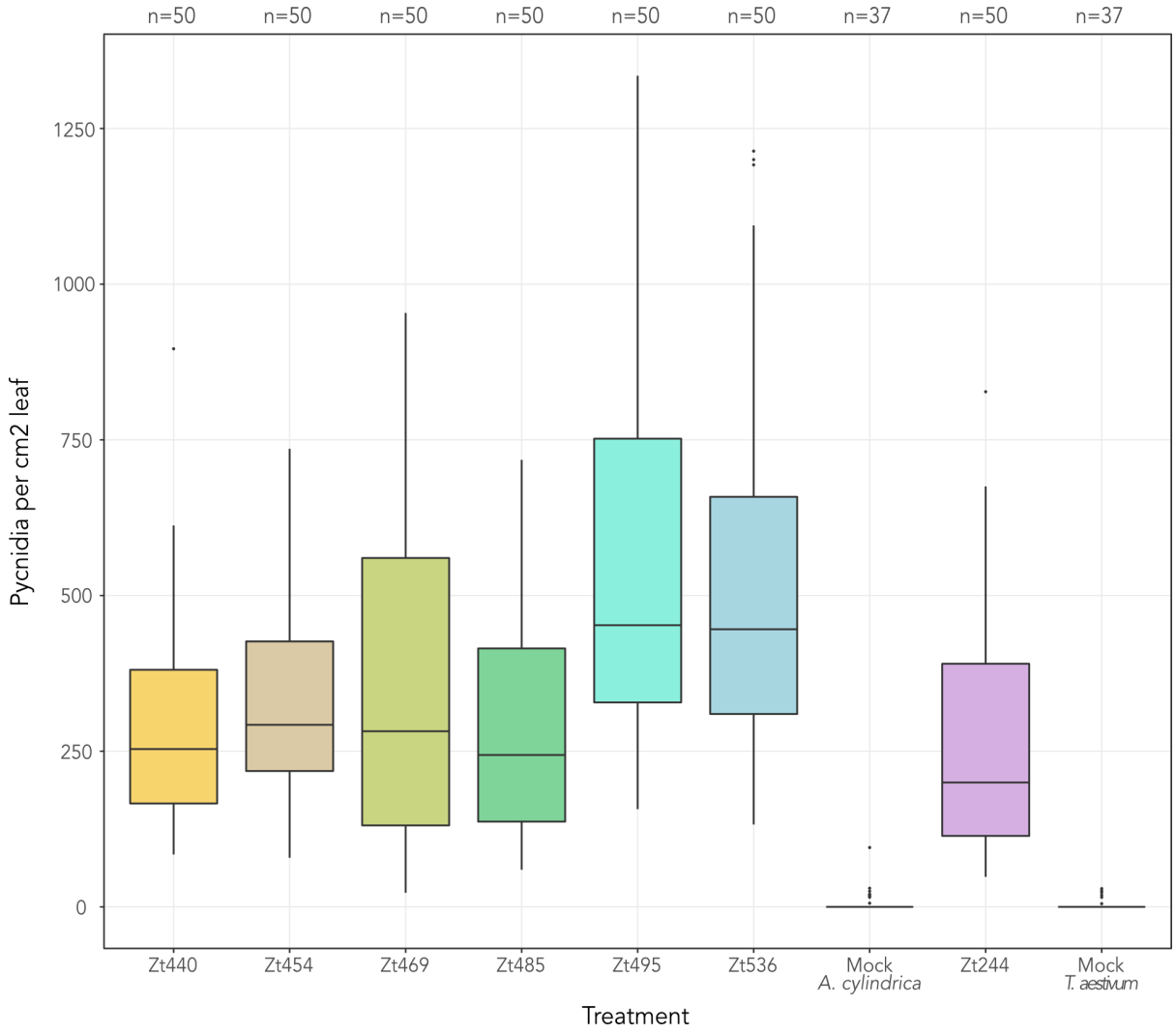


Figure S14. Quantitative analyses of *Z. tritici* virulence in distinct hosts. Leaves of *Aegilops cylindrica* were inoculated with a subset of *Aegilops*-infecting *Z. tritici* isolates from *Aegilops* population 1. As a comparison to a wheat-infecting *Z. tritici* isolate, *Triticum aestivum* cv. Riband plants were inoculated with the *Z. tritici* IPO323 isolate (strain Zt244). Mock treatments were used as controls. Leaves were scanned at 21 dpi and pycnidia density (per cm^2 leaf) was used as a readout of virulence. Number of plants per treatment (n) are indicated.

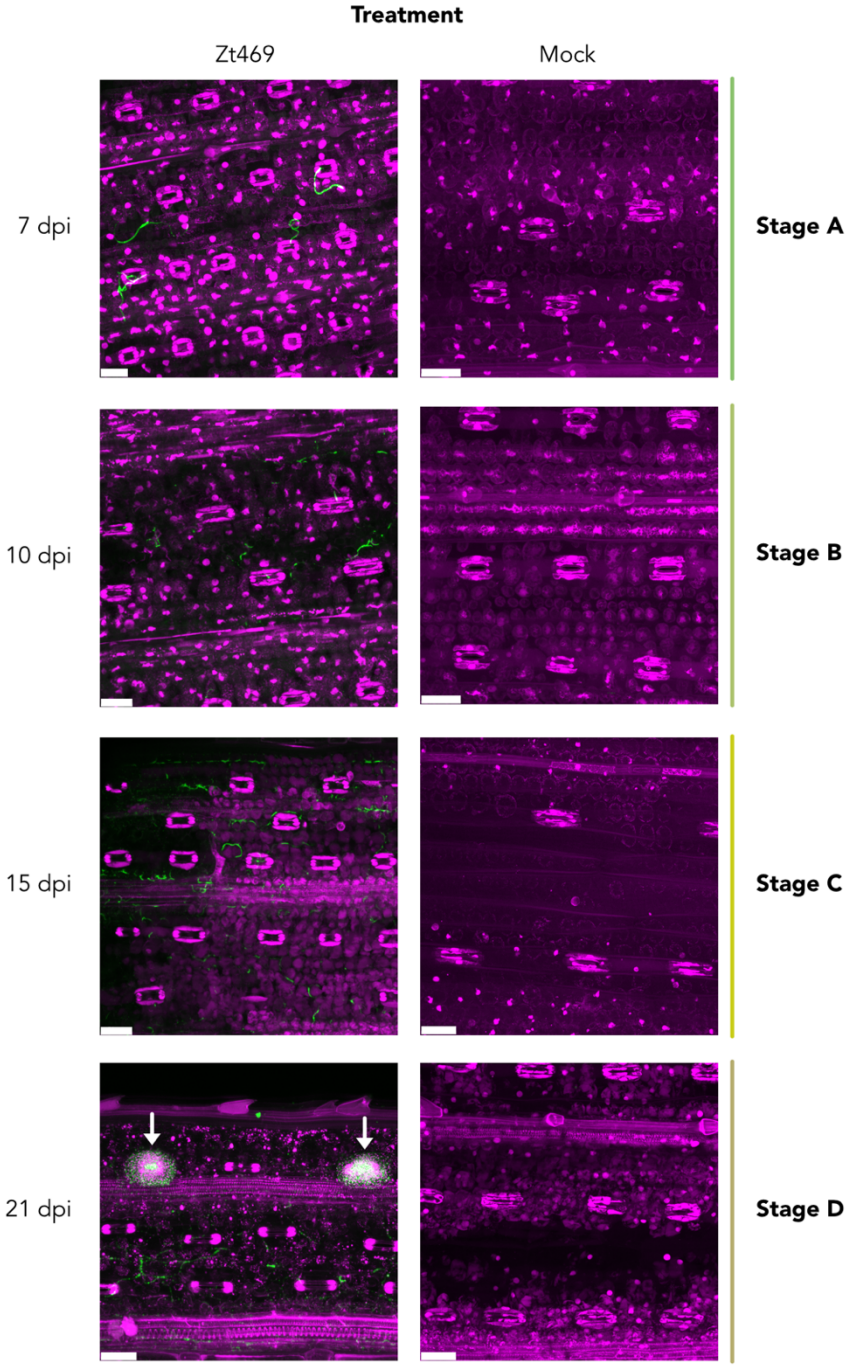
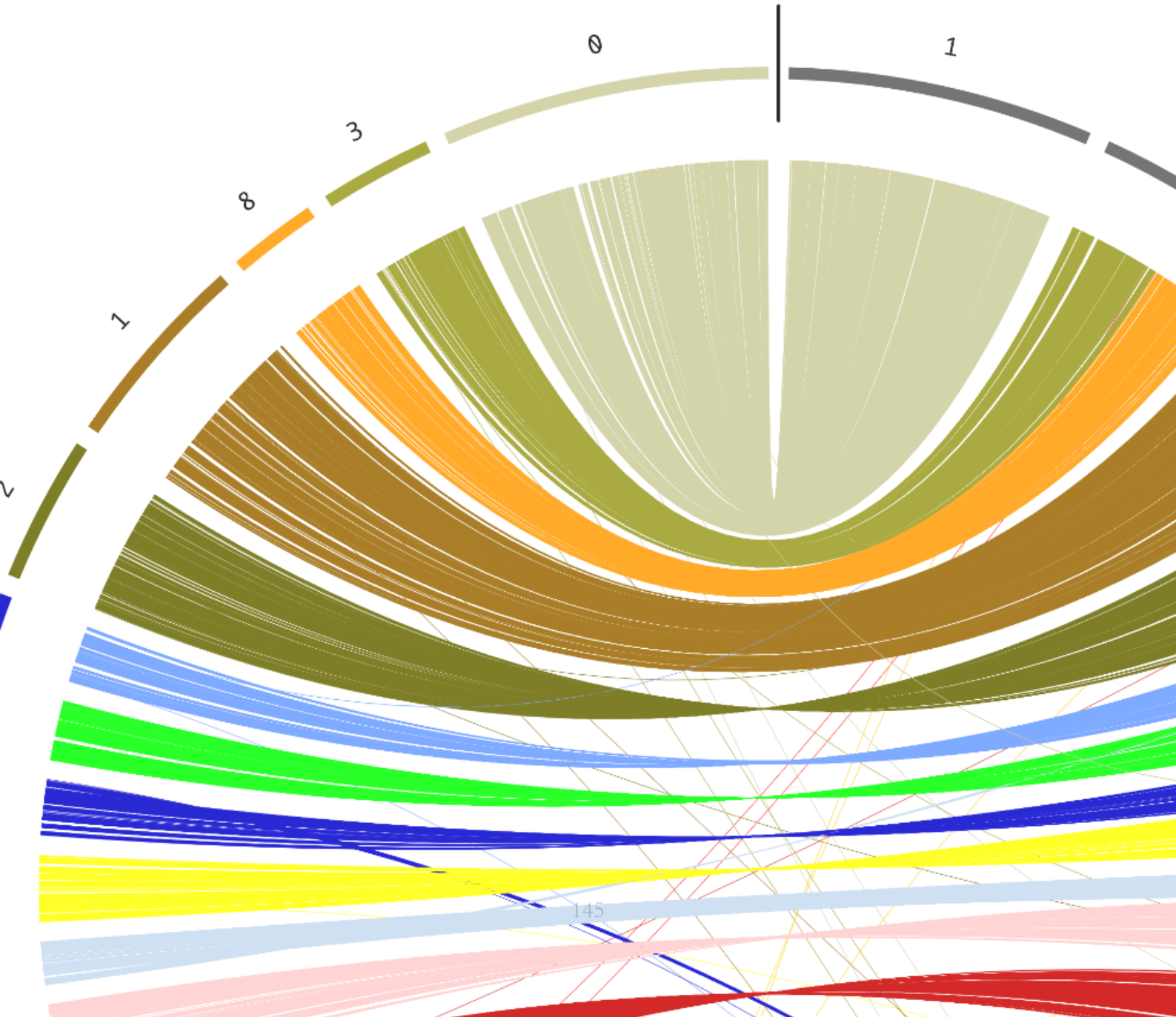


Figure S15. *Aegilops*-infecting *Z. tritici* isolate Zt469 shows four developmental stages during *A. cylindrica* infection. Micrographs obtained by CSLM (maximum projections of confocal image z-stacks) showing the four infection stages (stages "A" to "D") of *Z. tritici* inside *A. cylindrica* leaves. Leaves were collected at the indicated time points (dpi; days post-inoculation). Mock samples (0.1% Tween in sterile water) were used as controls. Nuclei and *A. cylindrica* cells are displayed in purple and fungal hyphae in green. At stage "D", two pycnidia harboring pycnidiospores are shown occupying substomatal cavities (white arrows). Scale white bars represent 50 µm. Animation of z-stack micrographs of fungal-infected leaves in each stage are available in the supplementary USB key.

Chapter III

A new chromosome is described in fungal species infecting wild grasses



Chapter III

A new chromosome is described in fungal species infecting wild grasses

Research paper manuscript

Authors: **Wagner C. Fagundes**, Janine Haueisen, Mareike Möller, Alice Feurtey, Rune Hansen, Fatemeh Salimi, Alireza Alizadeh and Eva H. Stukenbrock

Abstract

Many fungal plant pathogens have dynamic genomic architectures that can contribute to rapid evolution and adaptation to new niches. *Zymoseptoria tritici,* an important fungal pathogen of wheat, has a compartmentalized and rapidly evolving genome. In the reference *Z. tritici* isolate IPO323, 8 of the 21 chromosomes are accessory, comprising around 12% of its haploid genome. In spite of the profound impact on genome organization, the origin of accessory chromosomes in *Z. tritici* is still poorly understood. We explored genomic variation in populations of *Z. tritici* infecting wild wheat relatives from the genus *Aegilops*. Using a multi-omics approach (genomics, transcriptomics and epigenomics), we discovered a new chromosome in these *Z. tritici* isolates, and use this finding as a new opportunity to study the origin of accessory chromosomes. The newly identified chromosome presents similar characteristics to known accessory chromosomes in *Zymoseptoria* species, including presence-absence variation among *Aegilops*-infecting isolates, low gene expression *in vitro* and *in planta*, and enrichment with heterochromatin-associated histone methylation marks (H3K27me3). Interestingly, we found an orthologous chromosome sharing almost 50% of the genes in *Zymoseptoria ardabiliae*, a closely related fungal species also infecting wild grasses. This ortholog chromosome also presents accessory chromosomes characteristics, but lacks the enrichment of heterochomatin-associated methylation marks. Transcriptomic analyses revealed that the orthologous chromosome in *Z. ardabiliae* presents a high expression of transposable elements (TEs) coupled with lower signatures of host-genome defense mechanisms against TE expansion and spread (RIP), indicating that this chromosome is a reservoir of active transposons. We suggest that the chromosome has been exchanged between *Z. tritici* and *Z. ardabiliae* by introgressive hybridization events and propose that hybridization may play an important role in the evolution of new accessory chromosomes.

Introduction

Fungal plant pathogens exhibit remarkable variation on the genome structure and high levels of genome plasticity, even between closely related species (Möller & Stukenbrock, 2017). Sequencing of hundreds of fungal plant pathogens genomes and transcriptomes has provided unprecedented resources to study genome architecture and their analyses have highlighted striking differences in genome structure and gene and transposons regulation. Key observations from these studies involved the identification of distinct and rapidly evolving compartments on genomes of closely related fungal plant pathogen species, including gene-sparse, repeat-rich compartments, AT-rich isochores, clusters of tandem duplicated genes and accessory or lineage-specific chromosomes (Coleman et al., 2009; Dong, Raffaele, & Kamoun, 2015; Dutheil et al., 2016; Faino et al., 2016; Ma et al., 2010; Ohm et al., 2012; Plissonneau, Stürchler, & Croll, 2016; Rouxel et al., 2011; Stukenbrock et al., 2010; Williams et al., 2016). These distinct genome compartments present different levels of mutation and recombination rates that contribute for a high genomic dynamism and a rapid adaptive evolution of fungal pathogens to specific niches or hosts (Dong et al., 2015; Möller & Stukenbrock, 2017; Raffaele & Kamoun, 2012).

Accessory or lineage-specific chromosomes have been shown to contribute to virulence and adaptive evolution in different fungal pathogen species. Overall, fungal accessory chromosomes are generally small (< 2Mb in length), enriched with Transposable Elements (TEs) and have a lower gene density and transcriptional activity when compared to core chromosomes (Bertazzoni et al., 2018; Galazka & Freitag, 2014). Moreover, accessory chromosomes show a non-mendelian segregation, resulting in a presence-absence variation (PAV) among individuals of a population (Croll & McDonald, 2012; Komluski, Stukenbrock, & Habig, 2022). Accessory chromosomes have also been shown to be enriched in heterochromatin-associated histone methylation marks (e.g. H3K27me3 and H3K9me3), correlating with the low transcriptional activity found in these chromosomes (Bertazzoni et al., 2018; Galazka & Freitag, 2014; Schotanus et al., 2015). In some fungal species, accessory chromosomes encode virulence-related genes that can determine the disease outcome and host range of specific fungal species or lineages, as it is the case for *Leptosphaeria maculans* (Balesdent et al., 2013)*, Fusarium solani* (Miao, Covert, & VanEtten, 1991) and *Fusarium oxysporum* f. sp. *lycopersici* (Ma et al., 2010; van der Does et al., 2016). Despite the PAV between fungal individuals, accessory chromosomes have been maintained over long evolutionary periods, raising questions of how these chromosomes are transmitted generation to generation and more importantly, how they originate. Few studies have experimentally addressed these questions in different fungal pathogens (Croll, Zala, & McDonald, 2013; Habig, Kema, &

Stukenbrock, 2018; Ma et al., 2010) but the evolutionary mechanisms of accessory chromosomes origin remain largely unknown.

Zymoseptoria tritici, besides being an important fungal pathogen of wheat, is also an important experimental model on the study of genome evolution of plant pathogens. *Zymoseptoria tritici* causes the *Septoria tritici* blotch (STB) on wheat, one of the most devastating diseases for wheat production across the world (Fones & Gurr, 2015; Torriani et al., 2015). Coalescence analyses suggest that speciation of *Z. tritici* from wild ancestors occurred approximately 11,000 years ago in the Fertile Crescent region in the Middle East and coincided with the domestication of wheat (Stukenbrock, Banke, Javan-Nikkhah, & McDonald, 2007). Population genetics studies using polymorphism data from worldwide *Z. tritici* populations have indicated high levels of genetic variation within and between *Z. tritici* populations and extensive gene flow from regional to continental scales (Hartmann, McDonald, & Croll, 2018; Linde, Zhan, & McDonald, 2002). In addition to high levels of genomic diversity, *Z. tritici* also shows a highly variable and compartmentalized genome (Feurtey et al., 2020; Goodwin et al., 2011). *Zymoseptoria tritici* has one of best assembled fungal genomes with 21 chromosomes sequenced from telomere to telomere on the reference IPO323 isolate, 8 of them being accessory (Goodwin et al., 2011). Accessory chromosomes in *Z. tritici* show hallmarks that distinguish them from the core chromosomes: PAV between isolates; TE-rich; low gene density; enrichment of the heterochromatin-associated methylation mark H3K27me3; and low gene transcription activity not only *in vitro* but also *in planta* (Dhillon, Gill, Hamelin, & Goodwin, 2014; Goodwin et al., 2011; Haueisen et al., 2019; Kellner et al., 2014; Schotanus et al., 2015). Moreover, accessory chromosomes in *Z. tritici* can be lost during meiosis (Croll et al., 2013; Fouché, Plissonneau, McDonald, & Croll, 2018; Wittenberg et al., 2009) as well as during mitosis *in vitro* and *in planta* (Möller, Habig, Freitag, & Stukenbrock, 2018). However, in contrast to other fungal pathogens, accessory chromosomes in *Z. tritici* are not enriched in virulence-related genes and no virulence factors have yet been identified on these chromosomes, even though findings suggest a possible fitness cost in a wheat cultivar-specific manner (Habig, Quade, & Stukenbrock, 2017).

Zymoseptoria species also serve as excellent models to study genome architecture in fungal plant pathogens, particularly in closely related plant pathogen species and lineages that occur in wild and agricultural environments. The closest known relatives of *Z. tritici* namely *Z. brevis*, *Z. pseudotritici* and *Z. ardabiliae* are found associated to different wild grasses in the Middle East as e.g. *Lolium* spp., *Elymus repens*, and *Dactylis glomerata* (Quaedvlieg et al., 2011; Stukenbrock et al., 2012; Stukenbrock, Bataillon, Dutheil, Hansen, et al., 2011). Comparative genomics analyses using long-

read assemblies and gene annotations of different *Zymoseptoria* species have revealed a diverse set of genes and a large distribution of accessory chromosomes within and between species, suggesting that genome compartmentalization is an ancestral trait in the *Zymoseptoria* genus (Feurtey et al., 2020; Grandaubert, Bhattacharyya, & Stukenbrock, 2015).

Genome evolution in *Zymoseptoria* is also impacted by interspecific hybridization. Comparative genome analyses have revealed that gene flow and recurrent interspecific hybridization occurs between *Zymoseptoria* species at the center of origin of these pathogens and can lead to a mosaic pattern of genetic variation in *Zymoseptoria* genomes (Feurtey, Stevens, Stephan, & Stukenbrock, 2019; Möller et al. 2021; Stukenbrock, Christiansen, Hansen, Dutheil, & Schierup, 2012; Wu, Macielog, & Hao, 2017). Considering the sympatry between *Zymoseptoria* species and infection of co-existing grasses in the Middle East, the exchange of genetic material between these fungal species may indeed take place and even facilitate the introgression of particular genes that may be advantageous in specific hosts (Feurtey et al., 2019). However, the impact of hybridization in host specificity and the extent of genomic components transferred between species remain largely unknown in *Zymoseptoria*.

Recently, we have characterized host-specific populations of *Z. tritici* infecting wild grass species of the genus *Aegilops* through comparative infection assays and population genomics analyses (Chapter II). *Aegilops*-infecting *Z. tritici* isolates show high divergence from a worldwide collection of wheat-infecting *Z. tritici* isolates and from closely related *Zymoseptoria* species. Distinct footprints of positive selection were detected between the two host-diverging populations analyzed and demography analyses further suggested a split of the wheat- and *Aegilops*-infecting lineages around 10,000 years ago (Chapter II). The close relatedness of the two distinct lineages of *Z. tritici* provided us an excellent model system to address the evolution of genome architecture and accessory chromosome composition in natural fungal populations.

In this study, we explore the genomic variation in *Z. tritici* isolates infecting *Aegilops* species. We present a new high-quality reference genome based on long read-sequencing for the *Aegilops*-infecting *Z. tritici* Zt469 coupled with gene and Transposable Elements (TEs) predictions, and compare the gene content between this isolate and other *Z. tritici* and closely related *Zymoseptoria* species. Using a combination of genomics, transcriptomics and epigenomics approaches, we identify a unique accessory chromosome in *Aegilops*-infecting *Z. tritici* isolates which has synteny to another accessory chromosome in the closely related *Z. ardabiliae* species. Analyses of the orthologous chromosomes in the two *Zymoseptoria* species reveal differential levels of TE

expression and host-genome defense mechanism activity against TEs besides distinct enrichment of heterochromatin-associated methylation marks. We suggest that this chromosome has been exchanged between *Z. tritici* and *Z. ardabiliae* and that interspecies hybridization may play a role in the evolution of new accessory chromosomes.

Material and Methods

DNA sequencing and whole-genome assemblies

Reference genomes for *Aegilops*-infecting *Z. tritici* isolate Zt469 and *Z. ardabiliae* isolate Za100 were sequenced using the single-molecule real-time (SMRT) PacBio technology. Zt469 was isolated from *Aegilops* spp. in 2018 (Chapter II) and the *Z. ardabiliae* isolate Za100 (STIR11_8.1.1) was isolated in 2011 from an unknown grass species. Fungal cultures were maintained in liquid YMS medium (4 g/L yeast extract, 4 g/L malt extract, 4 g/L sucrose) at 200 rpm and 18 °C and high-quality DNA extractions was performed following a CTAB DNA extraction protocol described previously (Fagundes, Haueisen, & Stukenbrock, 2020). Library preparations and sequencing were performed at the Max Planck-Genome-Centre, Cologne, Germany using a PacBio Sequel II platform. Genomes were assembled *de novo* using the SMRT Analysis software v.5 (Pacific Bioscience) using default and "fungal" parameters as previously described (Feurtey et al., 2020). The genome assemblies with the best quality determined by the software Quast (Gurevich, Saveliev, Vyahhi, & Tesler, 2013) were used for further analysis. A previous version of the Zt469 genome assembly was published recently (Möller et al., 2021) and an improved version of this genome assembly is here presented. In this version, we have filtered the Zt469 assembly to match filtering steps done previously in order to remove poor quality contigs (Feurtey et al., 2020; Plissonneau et al., 2016). This step removed a large number of contigs with excessively high or low coverage, reducing the number of contigs from 128 as published previously (Möller et al., 2021) to only 45 contigs in Zt469, and from 39 to 38 contigs in Za100 (Supplementary Table S1). Telomeric repeats for each contig were identified as previously described (Feurtey et al., 2020; Supplementary Table S1). Genome assemblies and gene annotations based on long-read sequencing (PacBio) of the *Z. tritici* isolates Zt05 and Zt10 and of the closely related species *Zymoseptoria brevis* (Zb87), *Z. passerinii* (Zpa63), *Z. pseudotritici* (Zp13) and *Z. ardabiliae* (Za17) were obtained from a previous study (Feurtey et al., 2020). Genome assembly, gene and repeat annotations from the reference wheat-infecting *Z. tritici* isolate IPO323 were also obtained from previous studies (Goodwin et al., 2011; Grandaubert et al., 2015; Lorrain, Feurtey, Möller,

Haueisen, & Stukenbrock, 2021). Population data from *Aegilops*- and wheat-infecting *Z. tritici* and from additional *Z. ardabiliae* isolates were also obtained from elsewhere (Chapter II; Stukenbrock et al., 2012; Stukenbrock, Bataillon, Dutheil, Stukenbrock, et al., 2011; Stukenbrock & Dutheil, 2018).

Repeat and gene annotations

Gene and TE annotations for Zt469 and Za100 genomes followed a pipeline described previously (Feurtey et al., 2020). Briefly, we first used GeneMark-ES for an *ab initio* prediction using the option "--fungus" (Ter-Hovhannisyan, Lomsadze, Chernoff, & Borodovsky, 2008). Then, using RNA-seq data of *in vitro* growth (described below), we increased the quality of *ab initio* gene predictions. To this end, we mapped the quality-trimmed and masked reads to the newly assembled genomes using HISAT2 (Kim, Langmead, & Salzberg, 2015) and used the BRAKER1 pipeline to predict genes in each genome using the "fungus" flag (Hoff, Lange, Lomsadze, Borodovsky, & Stanke, 2016). In a last step, we assembled the RNA-seq reads into gene transcripts using Trinity (Grabherr et al., 2011) and aligned these gene transcripts using PASA (Haas et al., 2003). EVidence Modeler was then used to produce consensus gene models from the two independent predictions and the *de novo* assembled transcripts (Haas et al., 2008). TE content in both genomes was annotated using the TEdenovo and TEannot tools from the REPET package (https://urgi.versailles.inra.fr/Tools/REPET) (Flutre, Duprat, Feuillet, & Quesneville, 2011; Quesneville et al., 2005) following the developer's recommendations and default parameters as described in (Feurtey et al., 2020). TEs were classified according to the nomenclature defined by Wicker et al. (2007) (Wicker et al., 2007). For visualization in Circos plots (Krzywinski et al., 2009), we calculated gene and TE densities in 100 kb non-overlapping windows using bedtools v2.26.0 (Quinlan & Hall, 2010).

ChIP- and RNA-sequencing of *in vitro* cultures

We used three biological replicates of the *Z. tritici* isolate Zt469 and *Z. ardabiliae* isolate Za100 grown *in vitro* for chromatin immunoprecipitation followed by sequencing (ChIP-seq) and RNA-sequencing. Cells from the same biological replicate were used for both RNA and ChIP DNA extractions. Fungal spores were grown on YMS plates for three to four days, harvested and resuspended in 6 mL of 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4). For RNA extraction, 1 mL of resuspended cells were centrifuged, ground in liquid nitrogen, and total RNA was extracted using TRIzol (Invitrogen, Karlsruhe, Germany) following

the manufacturer's instructions. Cleaned (DNAse-treated) RNA samples were sent to Admera Health (South Plainfield, NJ, USA) for library preparation. For ChIP, the remaining 5 mL of resuspended cells were crosslinked with 0.5% formaldehyde for 15 min at room temperature and quenched by adding 150 µL of 2.5 M glycine. Chromatin immunoprecipitation was performed as previously described (Möller et al., 2019; Soyer et al., 2015) and antibodies against H3K4me2 (#07–030, Merck Millipore), H3K9me3 (#39161, Active Motif) and H3K27me3 (#39155, Active Motif) were used. ChIP-seq libraries were prepared with a modified version of the Next Ultra II DNA Library Prep Kit for Illumina (NEB, #E7645S). Sequencing of ChIP-seq and RNA-seq samples was performed on an Illumina HiSeq 3000, obtaining 2 x 150-nt read pairs by Admera Health (South Plainfield, NJ, USA).

In planta transcriptome sequencing

We isolated total RNA from Zt469-infected and mock-treated *A. cylindrica* leaves using the Direct-zol RNA MiniPrep kit (Zymo Research, Irvine, CA, US) following the manufacturer's instructions. We generated infection stage-specific transcriptome data based on leaf material collected and analysed by confocal laser scanning microscopy (CSLM) from a previous study (Chapter II). Four infection stages were determined ranging from early biotrophy [stage "A"; 7days post-inoculation(dpi)] to late necrotrophy (stage "D"; 21 dpi; Chapter II). Tissue from the same inoculated leaf was used for both RNA isolation and microscopy analyses (Supplementary Figure S1). Per infection stage, we selected the three most representative samples of that time point as biological replicates for transcriptome sequencing. Each biological replicate consisted of material from three inoculated *A. cylindrica* leaves that were pooled and homogenized in liquid nitrogen (Supplementary Figure S1). Preparation of strand-specific RNA-seq libraries including polyA enrichment was performed at the Max Planck Genome Center (MPGC), Cologne, Germany (http://mpgc.mpipz. mpg.de). Library sequencing was also performed at the MPGC on an Illumina HiSeq3000 platform yielding paired-ends reads of 150 nt.

Orthologous genes and genome synteny analyses

We identified orthologous genes between the *Zymoseptoria* genome assemblies using the software PoFF (Lechner et al., 2014) implemented in Proteinortho (Lechner et al., 2011) to account for synteny information. The identified orthogroups were used to visualize synteny between genomes using the software Circos (Krzywinski et al., 2009).

Functional annotations

Distinct tools were used to predict the putative gene functions in the newly assembled Zt469 and Za100 genomes. First, we obtained the aminoacid sequence of each gene model and used the tool Predector to predict secreted proteins and PFAM domains (Jones et al., 2021). Predector runs several tools for fungal secretome and effector analysis, including the softwares SignalP (versions 3, 4, 5, 6) (Armenteros et al., 2019; Dyrløv Bendtsen, Nielsen, von Heijne, & Brunak, 2004; Petersen, Brunak, Von Heijne, & Nielsen, 2011; Teufel et al., 2022), TargetP (version 2.0) (Armenteros et al., 2019), DeepLoc (Armenteros, Sønderby, Sønderby, Nielsen, & Winther, 2017), TMHMM (Krogh, Larsson, Von Heijne, & Sonnhammer, 2001; Sonnhammer, von Heijne, & Krogh, 1998), Phobius (Käll, Krogh, & Sonnhammer, 2004), DeepSig (Savojardo, Martelli, Fariselli, & Casadio, 2018), CAZyme finding (with dbCAN) (Zhang et al., 2018), Pfamscan (Madeira et al., 2022), searches against PHI-base (Urban et al., 2022), Pepstats (Madeira et al., 2022), ApoplastP (Sperschneider, Dodds, Singh, & Taylor, 2018), LOCALIZER (Sperschneider et al., 2017), Deepredeff (Kristianingsih & MacLean, 2021), and EffectorP (versions 1, 2 and 3) (Sperschneider & Dodds, 2022; Sperschneider, Dodds, Gardiner, Singh, & Taylor, 2018; Sperschneider et al., 2016). Predector then gathers the results of each one of these tools and output a summary table of ranked candidates using a learning-to-rank machine learning method (Jones et al., 2021). We also used the software eggnog-mapper to provide additional Clusters of Orthologous Groups (COG), GO (Gene Ontology) and KEGG (Kyoto encyclopedia of genes and genomes) annotations (Huerta-Cepas et al., 2017). Then, in order to analyze the expression of genes involved in host-pathogen interactions, we categorized the gene models into "Effector" and "CAZyme" based on the individual tools score thresholds recommended by Predector (Supplementary Table S2). "Small Secreted Proteins (SSPs)" category also followed the criteria described previously (Grandaubert et al., 2015; Supplementary Table S2). At last, we used Antismash v.6.0 (fungal version) to detect biosynthetic gene clusters (BGCs) in each genome (Blin et al., 2021; Supplementary Table S2). Gene models that did not belong to any of these categories were classified as "Other". The outputs of all tools and transcription *in vitro* and *in planta* (when applicable) for each gene model as well as the genes belonging to each category can be found at Supplementary Tables S3-S5.

ChIP- and RNA-seq data analyses

Sequencing adapters and low-quality reads were trimmed from raw paired-end RNA and ChIP-seq reads using Trimmomatic v 0.39 (Bolger, Lohse, & Usadel, 2014) with the following parameters: LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:50. Low quality nucleotides (Q<20) in RNA-seq reads were further masked using FASTX-toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). For RNA-seq data, quality-trimmed and masked reads were mapped using HISAT2 (Kim et al., 2015) while for ChIP-seq read mapping was performed using Bowtie2 (Langmead & Salzberg, 2012). Conversion, sorting, merging and indexing of alignment files were performed using SAMtools v. 1.7 (Danecek et al., 2021). We used the software HOMER (Heinz et al., 2010) to detect methylation-enriched regions in the ChIP mappings. We called methylation-enriched peaks individually for each replicate and merged the peaks with bedtools v2.26.0 (Quinlan & Hall, 2010). Only enriched regions found in all replicates were considered for downstream analyses. Sequence coverage of methylation-enriched regions per contigs were also calculated using bedtools v2.26.0 (Quinlan & Hall, 2010). For RNA-seq *in vitro* and *in planta*, we accessed gene expression as Transcript Per Million (TPM) as previously described (Feurtey et al., 2020). In summary, raw read counts per gene were first calculated using the tool htseq-count v2.0.2 from HTSeq with "--mode union" (Anders, Pyl, & Huber, 2015) for each replicate individually. TPM was then calculated by normalizing read counts with coding region length resulting in the number of Reads Per Kilobase (RPK). RPK total counts per sample are then divided by 1 million to generate a "per million" scaling factor. The coding region length of each gene was calculated with the GenomicFeatures R package using the function "exonsBy" (Lawrence et al., 2013). For Circos plot representations (Krzywinski et al., 2009), we first calculated the mean TPM between replicates and then the averages per 100 kb non-overlapping windows in each chromosome/unitig in log2(TPM+1) values. For RNA-seq read mapping statistics, we used the tool Qualimap bamQC v.2.2.1 (Okonechnikov et al., 2016). List of all commands and tools used can be found at Supplementary Text S1.

TE expression analysis

We evaluated Transposable Elements (TEs) transcription activity in both Zt469 and Za100 genomes using the *in vitro* and *in planta* RNA-seq data. Raw RNA-seq reads were trimmed, quality filtered and masked as described above. Quality-trimmed and masked reads were mapped to each respective genome assembly using HISAT2 (Kim et al., 2015) with the flag "--no-mixed". Conversion, sorting, merging and indexing of alignment files was also performed using SAMtools

v. 1.7 (Danecek et al., 2021). Then, we accessed the TE transcription *in vitro* and *in planta* for each replicate and each infection stage individually. To this end, read count tables were generated using the TEcount function of TEtranscript pipeline with the "--mode multi" (Jin & Hammell, 2018; Jin, Tam, Paniagua, & Hammell, 2015) and levels of expression were reported using the Transcript per Million (TPM) normalization as described above. The length of each TE was calculated with the GenomicFeatures R package using the function "exonsBy", in which the entire element was treated as an unique exon (Lawrence et al., 2013).

Repeat-induced point mutation (RIP) analyses

We evaluated RIP signatures along the Za100 and Zt469 genomes and among TE copies within each genome. To this end, RIP-like signatures and the Large RIP-affected genomic regions (LRARs) were calculated using the RIPper software (Van Wyk et al., 2019). Genome-wide RIP composite indices were calculated using default parameters of 1000bp windows with a 500bp step size and following the formula: (TpA/ApT) – (CpA þ TpG/ApC þ GpT) (Van Wyk et al., 2019). LRARs were defined as genomic regions (windows) consecutively affected by RIP that are more than 4000bp in length. To access the RIP composite index of each TE copy, we calculated the indices in 50bp nonoverlapping windows using a previously published custom script (Lorrain et al., 2021). Regions are considered to be affected by RIP when the composite index is > 0 (Van Wyk et al., 2019).

Karyotype analysis

We compared the genome structure and chromosome presence-absence variation (PAV) of different *Z. tritici* and *Z. ardabiliae* isolates using two approaches. Firstly, whole-genome sequencing reads were trimmed using Trimmomatic v 0.39 (Bolger et al., 2014) and mapped to the *Z. tritici* reference genome IPO323 (Goodwin et al., 2011) as well as to the newly assembled *Aegilops*-infecting *Z. tritici* Zt469 and *Z. ardabiliae* Za100 genomes using bwa-mem v. 0.7.17 (Li & Durbin, 2009). After mapping, we normalized the coverage of mapped reads by each assembly size using deepTools v.2.0 (Ramírez et al., 2016). Normalized alignment files were then visualized with the Integrate Genome Viewer (IGV) software v.2.8.2 (Robinson et al., 2011). Only the Zt469 and Za100 contigs containing telomeric repeats and larger than 100 kb were kept for visualization. A second approach involved the analysis of chromosome PAV by Pulsed Field Gel Electrophoresis (PFGE) and Southern blot techniques (Southern, 1975). Preparation of non-protoplast plugs and PFGE run settings were followed as described previously (Möller et al., 2018). As a comparison

to Zt469, we also prepared non-protoplast plugs for the reference *Z. tritici* IPO323 isolate (strain Zt244), the wheat-infecting *Z. tritici* isolate Zt10 (strain Zt366; Stukenbrock et al., 2007) and the *Aegilops*-infecting *Z. tritici* isolate Zt501 (Chapter II). We used chromosomal DNA of *Hansenula wingei* (Bio-Rad, Munich, Germany) as standard size marker for mid-size chromosomes (1 to 3.1 Mb). Gels were stained with ethidium bromide solution (1 mg/ml ethidium bromide in H2O) for 30 min and chromosomal bands were detected with the GelDocTM XR+ system (Bio-Rad). We performed Southern blot as described previously (Southern, 1975) using a DIG (digoxigenin)-labeled probe generated by the PCR digoxigenin labeling Mix (Roche, Mannheim, Germany) following the manufacturer's instructions. Probe primers were designed with Primer-BLAST (Ye et al., 2012) and Geneious v.2020.1.2 software (https://www.geneious.com/home/). List of primers used can be found at Supplementary Table S6.

Short-term *in vitro* growth experiment and chromosome-loss PCR screening

To visualize potential chromosome loss during mitotic (vegetative) growth, a short-term *in vitro* growth experiment was performed as previously described (Möller et al., 2018). Briefly, a glycerol stock (-80 °C) of the *Aegilops*-infecting *Z. tritici* isolate Zt469 was diluted (1:10), plated on YMS-agar (4 g yeast extract, 4 g malt extract, 4 g sucrose, and 20 g agar per 1 liter) plate and grown for 5 days at 18 °C. One single colony was picked, suspended in 100 µl of liquid YMS and three replicate cultures were inoculated with 20 µl of this cell suspension. Cells were grown in 25 ml YMS at 18 °C and 200 rpm. After 3-4 days of growth, 900 µl of the cultures were transferred to fresh 25 ml of YMS medium for further growth at same conditions. A total of eight transfers were conducted in a time frame of 4 weeks.

For screening of accessory chromosome loss by PCR, the cells from the eighth transfer of liquid cultures were diluted (1:100) and plated on YMS-agar plates to obtain single colonies. DNA extraction of randomly chosen single colonies was performed using a rapid DNA extraction protocol previously described (Fagundes et al., 2020). Primers targeting the right portion of each accessory contig were designed with Primer-BLAST (Ye et al., 2012) and Geneious v.2020.1.2 software (https://www.geneious.com/home/) and used to detect chromosome loss in multiplex PCR reactions. List of primers and expected amplicon sizes are described in Supplementary Table S6.

Results

A new accessory chromosome discovered in *Aegilops*-infecting *Z. tritici*

As a first step to characterize the genome structure of the *Aegilops*-infecting *Z. tritici* isolates, we analyzed chromosomal synteny between different *Zymoseptoria tritici* genomes. Analyses of homologous genes between the reference wheat-infecting *Z. tritici* isolate IPO323 (Goodwin et al., 2011) and the *Aegilops*-infecting *Z. tritici* isolate Zt469 revealed a high extent of synteny between the thirteen core chromosomes described for IPO323, whereas only three out of eight accessory chromosomes were found to have synteny with chromosomes of Zt469 (unitigs 18, 19 and 20, Figure 1). Surprisingly, this analysis also showed a complete unitig (with telomeric repeats on both ends) in Zt469 with no alignments to any specific chromosome of the reference strain IPO323, namely "unitig 9" (Figure 1). Only 4.3% (13/299) of the genes on unitig 9 are homologous and spread genome-wide in the IPO323 genome. Among these genes, besides a peptidase from family M3 (Zt469_000009F_arrow_0273) and a cytochrome P450 gene (Zt469_000009F_arrow_0095), most of these homologs represented transposons or "selfish elements" as endonucleases, transposases (DDEs) or proteins of unknown function in fungal plant pathogens. To further validate the presence and size of unitig 9 in Zt469, we conducted Pulsed Field Gel Electrophoresis (PFGE) followed by Southern blot analysis, which revealed the presence of this chromosome in the expected size (~1.6Mb) in Zt469 and the absence of the chromosome in the *Z. tritici* IPO323 isolate and in the other two *Z. tritici* isolates analyzed (Supplementary Figure S2).

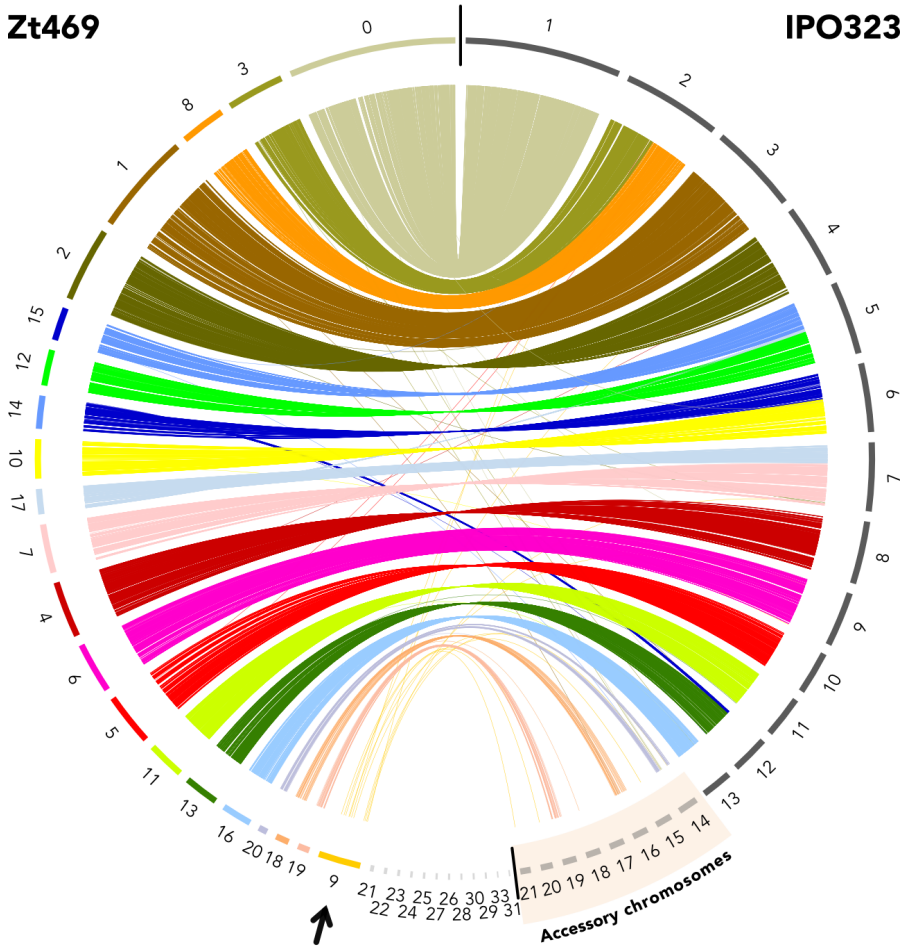


Figure 1. Genome synteny between *Z. tritici* isolates Zt469 and IPO323. Synteny analysis was performed between the reference wheat-infecting *Z. tritici* isolate IPO323 (right) and the *Aegilops*-infecting *Z. tritici* isolate Zt469 (left). Each color represents a different unitig in Zt469. Small unitigs are in light gray. A new chromosome, referred as "unitig 9", is identified in Zt469 and show no synteny to any particular portion of IPO323 genome (black arrow). Links represent homologous genes. Unitigs in Zt469 are ordered following the synteny to the reference IPO323 genome. Accessory chromosomes on the IPO323 genome are indicated.

Unitig 9 is present in some but not all *Aegilops*-infecting *Z. tritici* isolates

In order to examine the presence of unitig 9 in other sympatric *Z. tritici* specimens collected in the Middle East (Chapter II), comparative analyses of chromosome presence-absence variation (PAV) between Iranian *Z. tritici* isolates infecting wheat and *Aegilops* spp. was performed. Using a normalized read coverage approach, mapping results to the wheat-infecting *Z. tritici* IPO323 reference genome showed a large PAV on the accessory chromosomes for all analyzed *Z. tritici* isolates, in which the wheat-infecting isolates exhibited a higher presence of these chromosomes when compared to the *Aegilops*-infecting ones (Figure 2). Using the same mapping approach for the *Aegilops*-infecting Zt469 reference genome, a similar pattern of chromosome PAV was

observed for the three possible accessory unitigs (unitigs 18, 19 and 20) while unitig 9 was completely absent in all wheat-infecting isolates (Figure 3). When analyzing the *Aegilops*-infecting population, we observed that unitig 9 was present in only 11 out of 47 isolates while a higher PAV for the possible accessory unitigs was observed across the population (Figure 3). These results suggest that unitig 9 is a new chromosome present only in *Aegilops*-infecting *Z. tritici* populations.

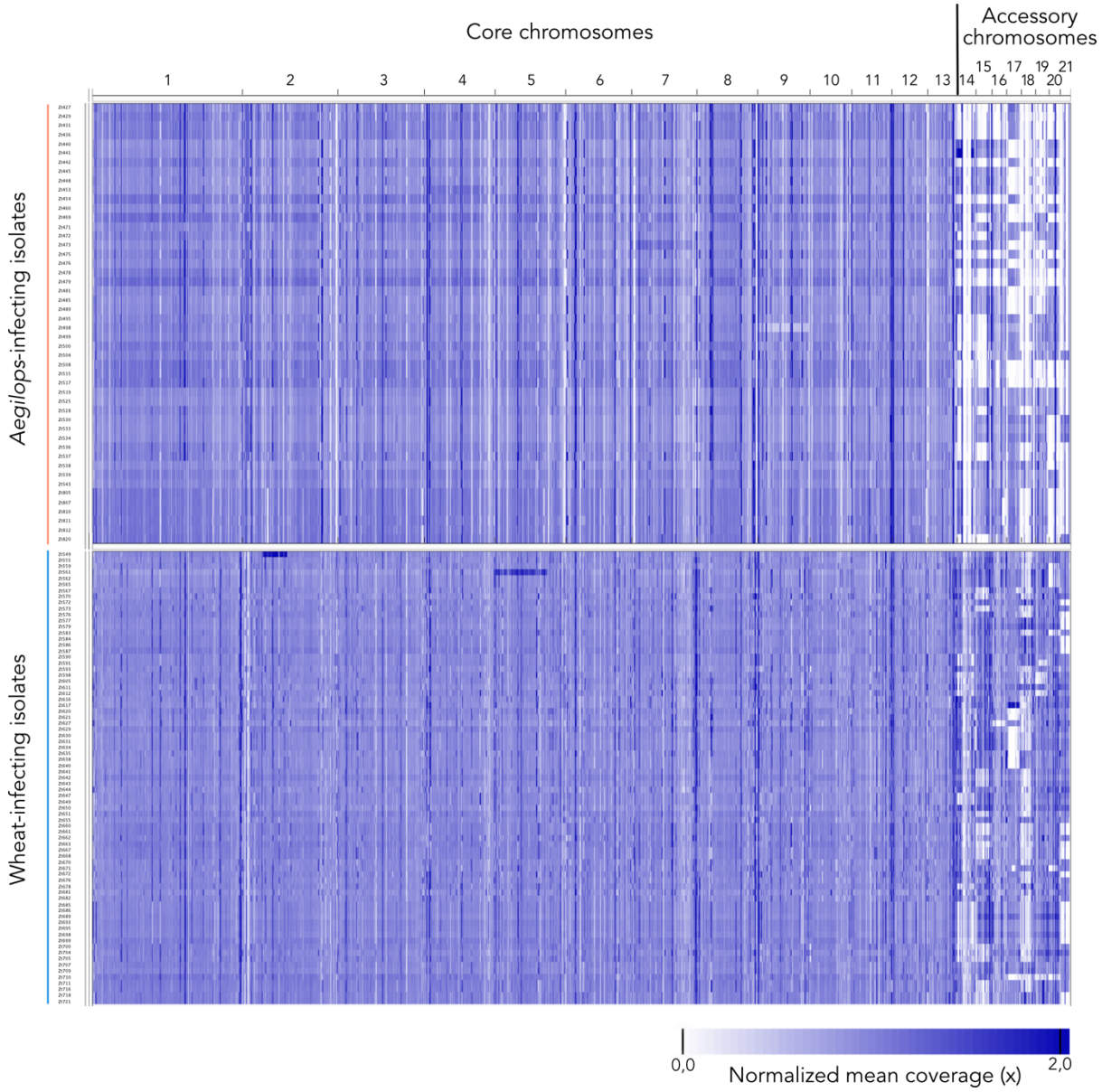


Figure 2. *Aegilops*-infecting *Z. tritici* isolates show high PAV of accessory chromosomes. Chromosome presence-absence variation (PAV) in host-diverging *Z. tritici* populations (Chapter II) was analyzed by read mapping to the *Z. tritici* IPO323 reference genome. Heatmap represents normalized mean coverage of reads mapped to each position of the reference genome. Darker colors represent regions of higher coverage as e.g. repetitive elements or duplications.

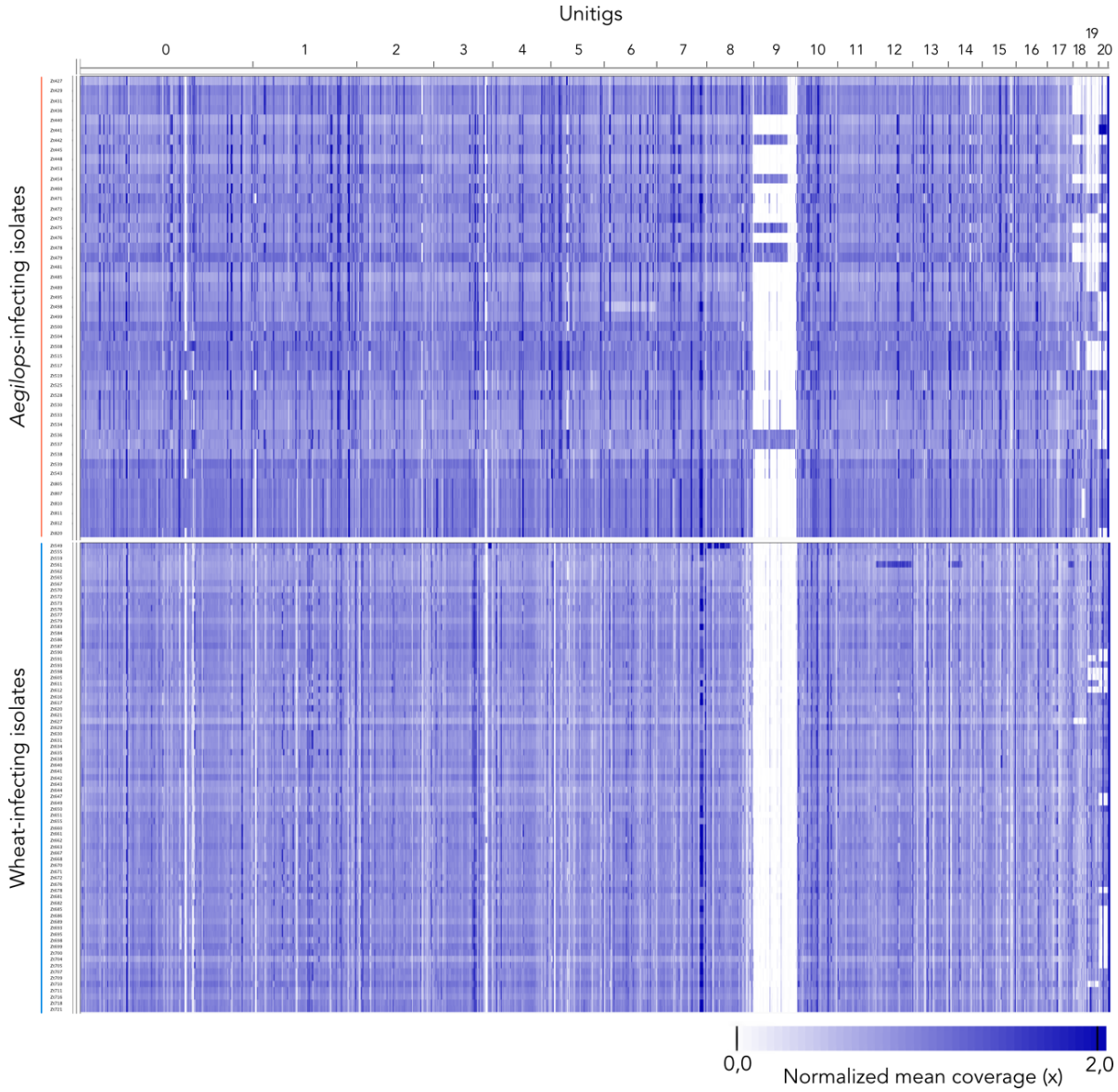


Figure 3. Unitig 9 shows presence-absence variation in *Aegilops*-infecting *Z. tritici*. Chromosome presence-absence variation (PAV) in host-diverging *Z. tritici* populations (Chapter II) was analyzed by read mapping to the *Z. tritici* Zt469 genome assembly. Heatmap represents normalized mean coverage of reads mapped to each position of the genome assembly. Except for unitig 9, only unitigs syntenic to the *Z. tritici* IPO323 reference genome are shown and sorted in descending order of length. Darker colors represent regions of higher coverage as e.g. repetitive elements or duplications.

Unitig 9 contains accessory chromosome hallmarks

Considering the PAV of unitig 9 among *Aegilops*-infecting *Z. tritici* isolates, we hypothesized that unitig 9 represents a new accessory chromosome. Among fungal plant pathogens, including *Z. tritici*, accessory chromosomes are generally enriched with repetitive DNA and have a lower gene density and lower transcriptional activity than the core genome (Goodwin et al., 2011; Möller &

Stukenbrock, 2017). Furthermore, histone H3 lysine 27 trimethylation (H3K27me3) is a prominent histone modification hallmark of accessory chromosomes in distinct fungal species (Galazka & Freitag, 2014; Schotanus et al., 2015) and was shown to directly influence chromosome loss and stability in *Z. tritici* (Möller et al., 2019). Gene and TE annotation as well as transcriptome analyses *in vitro* revealed that unitig 9 contains a low gene density (average of 0.39 genes / kb), high TE density (average of 0.33 TEs / kb) and low transcriptional activity when compared to core unitigs of similar size (e.g. unitig 8; 0.83 genes/kb and 0.28 TEs/kb on average; Figure 4A), consistent with known signatures of other accessory chromosomes in *Z. tritici* (Goodwin et al., 2011; Kellner et al., 2014). The smaller unitigs 18, 19 and 20 also presented similar characteristics (Figure 4A). In line with these results, *in vitro* chromatin immunoprecipitation in combination with high-throughput sequencing (ChIP-seq) targeting three histone methylation marks (H3K4me2 for euchromatin and H3K27me3 and H3K9me3 for facultative and constitutive heterochromatin, respectively) also revealed that unitig 9, together with unitigs 18, 19 and 20, are enriched in H3K27me3, corroborating the low transcriptional activity observed for these unitigs (Figure 4B). We also found the typical signatures of accessory chromosomes on unitig 17 (Figure 4A, B). Interestingly, this unitig is synthetic with the right chromosome arm of chromosome 7 in the *Z. tritici* IPO323 reference genome (Figure 1). It has been previously proposed that chromosome 7 in IPO323 may represent a fusion event of an accessory chromosome to a core chromosome (Schotanus et al., 2015), and our findings add further support to this hypothesis. Altogether, our results suggest that the novel unitig 9 along with unitigs 18, 19 and 20 represent accessory chromosomes in *Aegilops*-infecting *Z. tritici* isolates and carry similar hallmarks to the known accessory chromosomes in other *Zymoseptoria* species.

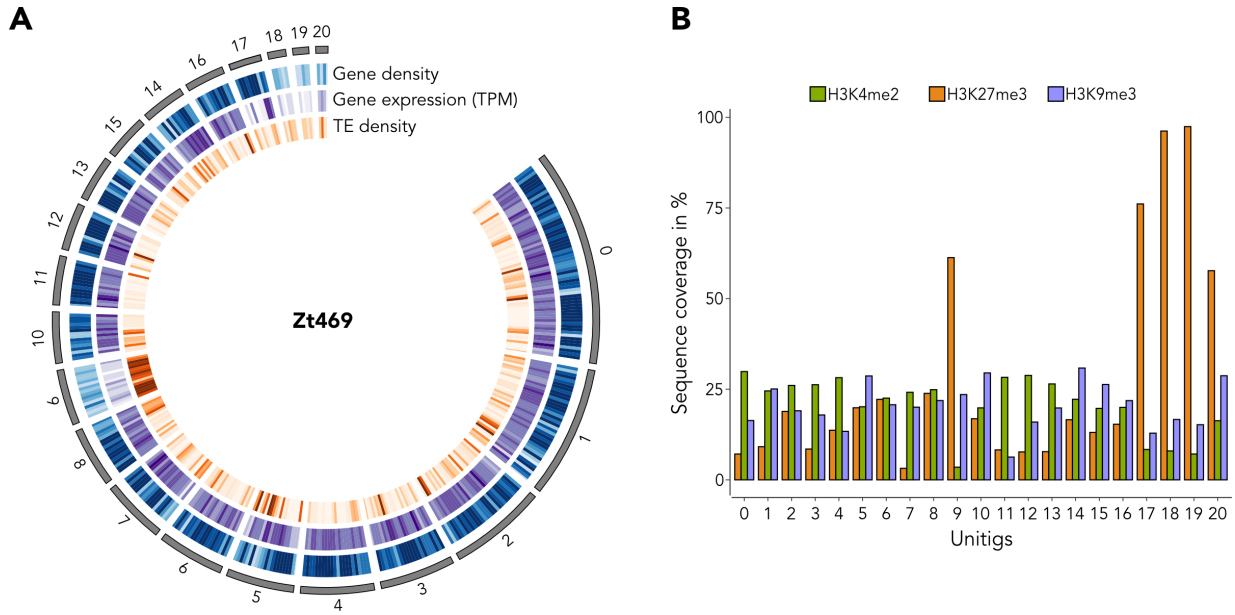


Figure 4. Unitig 9 presents accessory chromosome characteristics. (A) Circos plot representing different features along twenty Zt469 unitigs. Unitigs are represented by the dark gray segments sorted by length. Tracks from outside to the inside are heatmaps in 100kb windows representing respectively: gene density, gene expression *in vitro* in log2(TPM+1) and TE density. Darker heatmap colors indicate higher values. **(B)** Barplot displaying the percentage of sequence coverage of each Zt469 unitig with the heterochromatin methylations H3K27me3 (orange bars) and H3K9me3 (purple bars), and the euchromatin methylation H3K4me2 (green bars) marks relative to unitig length. In both panels, except for unitig 9, only unitigs syntenic to the *Z. tritici* IPO323 reference genome are shown.

Unitig 9 is lost during *in vitro* growth

To further characterize the new accessory chromosome unitig 9, we conducted an *in vitro* short-term growth experiment to access its stability. Previous evolution experiments with *Z. tritici* have revealed that accessory chromosomes can be readily lost during *in vitro* mitotic passages (Möller et al., 2018). After four weeks of mitotic propagation in YMS liquid culture at 18°C, including eight transfers to fresh medium, we estimated chromosome loss in 597 single colonies by PCR assays. We designed primers targeting specific regions in the four accessory unitigs (unitigs 9, 18, 19 and 20) and screened the evolved colonies for the presence and absence of such markers. A chromosome is considered potentially lost if a chromosome marker is absent after PCR amplification. In line with the previous observations of accessory chromosome loss in *Z. tritici* during mitosis (Möller et al., 2018), we also detected potential chromosome loss in three out of the four accessory unitigs among the 597 strains tested. Unitig 9 seemed to be lost in 24.8% strains, while unitig 18 in 36.5% strains. Only one strain showed potential loss of unitig 19 and no potential loss was observed for unitig 20 in all strains tested. Altogether, these results indicate that unitig 9 alongside unitigs 18 and 19 are accessory chromosomes that can be lost during *in vitro* mitotic growth. Additional analyses by PCR targeting different portions of the chromosomes coupled with confirmation by whole-genome sequencing or Pulsed-Field Gel Electrophoresis (PFGE) and Southern blots analyses are necessary to confirm the chromosome loss rates observed.

Almost half of the genes in unitig 9 have orthologs in *Zymoseptoria ardabiliae*

Based on the previous finding that only few genes in unitig 9 have homologous in the reference *Z. tritici* IPO323 isolate, we further inspected the orthology of unitig 9 genes in other *Z. tritici* isolates and *Zymoseptoria* species. We used previously published and annotated whole genome assemblies based on long-read sequencing (PacBio) of the *Z. tritici* isolates Zt05 and Zt10 and of the closely related species *Zymoseptoria brevis* (Zb87), *Z. passerinii* (Zpa63), *Z. pseudotritici* (Zp13) and

Z. ardabiliae (Za17) (Feurtey et al., 2020). In this study, we also used PacBio sequencing to assemble the genome of one more *Z. ardabiliae* isolate (Za100; Supplementary Table S1). We identified 7,703 orthologous genes (i.e. orthogroups) distributed among all nine *Zymoseptoria* genome assemblies analyzed (Figure 5A). Within species, we detected 146 species-specific orthogroups between *Z. tritici* isolates and 429 species-specific orthogroups between the two *Z. ardabiliae* isolates analyzed (Figure 5A). Surprisingly, we also observed that a large number of orthogroups were specifically shared between Zt469 and Za100 (142 orthogroups), just a few orthogroups less than the intersection between all *Z. tritici* genomes (146 orthogroups) (Figure 5A). In fact, comparisons between the *Z. tritici* isolates IPO323 and Zt469 and two *Z. ardabiliae* isolates (Za17 and Za100) revealed a larger number of orthogroups shared between the *Aegilops*-infecting *Z. tritici* isolate Zt469 and the *Z. ardabiliae* isolates (Za17: 59 orthogroups and Za100: 241 orthogroups) as when compared to the intersection between IPO323 and the same *Z. ardabiliae* isolates (Za17: 37 orthogroups and Za100: 81 orthogroups; Figure 5B).

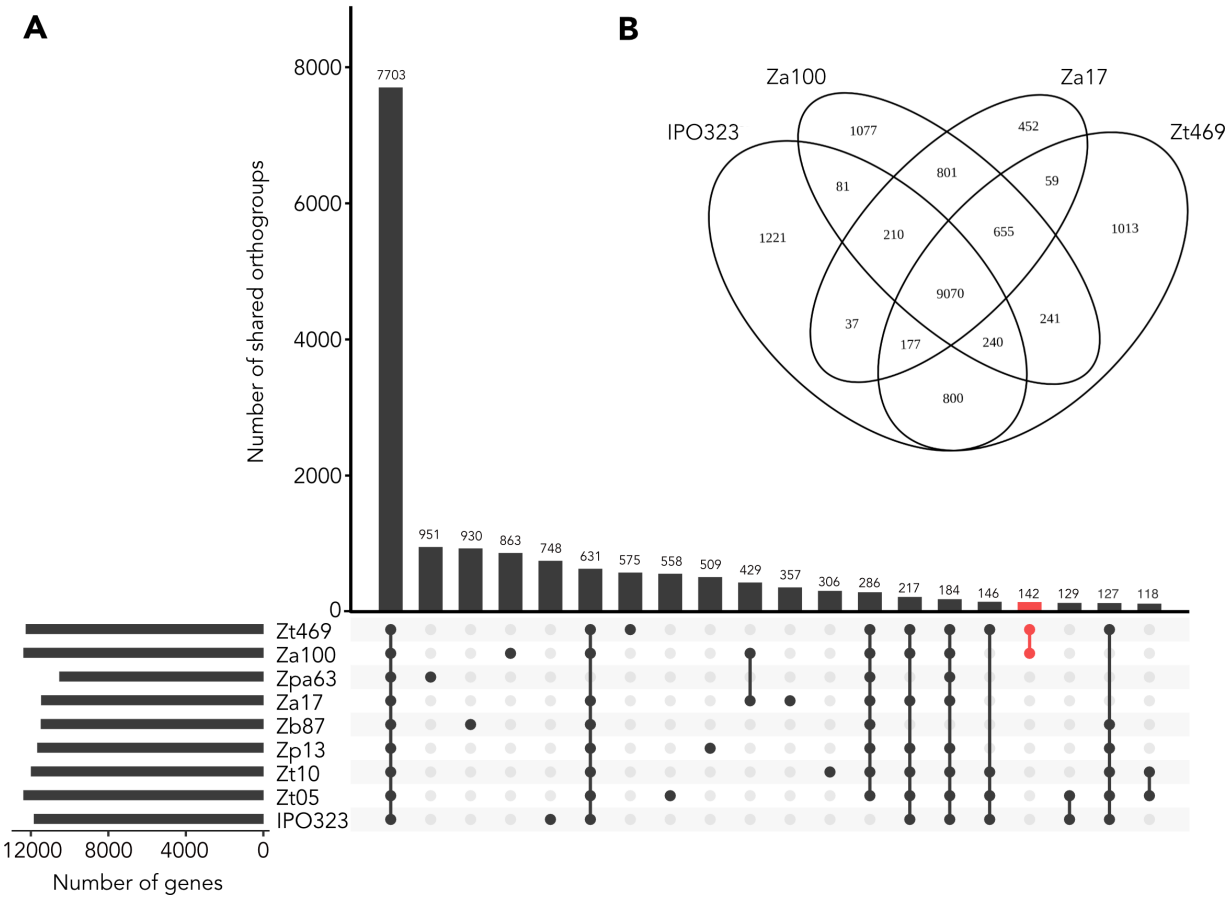


Figure 5. A large number of orthologs is shared between Zt469 and Za100 isolates. (A) Upset plot showing the number of orthogroups shared between the nine *Zymoseptoria* species genomes analyzed. Only intersects with more than 100 orthogroups are displayed. Number of orthogroups shared between Zt469 and Za100 are highlighted in red. **(B)** Venn diagram summarizing the orthogroups exclusively shared between the *Z. tritici* isolates IPO323 and Zt469 and *Z. ardabiliae* isolates Za17 and Za100.

We then focused on the genes present in unitig 9 of Zt469. As previously observed, few orthogroups were observed between unitig 9 of Zt469 and the reference *Z. tritici* isolate IPO323, in which only 4.3% orthogroups from unitig 9 genes were present in this genome (Figure 1 and Figure 6). Similarly, we observed that the proportion of unitig 9 genes present on the other *Z. tritici* genomes as well as on most of the closely related *Zymoseptoria* species analyzed was quite low, ranging from 2% in Zt10 to 4.3 % in Zpa63 (Figure 6). Interestingly, we observed a high proportion of unitig 9 genes with orthologs in *Z. ardabiliae* isolates Za17 and Za100. Almost half of the number of genes in unitig 9 had orthologs in Za100 (47.8%), while Za17 had a lower, but still high proportion of orthologous genes (7%) (Figure 6). This high proportion of orthologs between unitig 9 of Zt469 and the Za100 isolate, in line with the genome-wide orthogroups results observed previously (Figure 5), led us to the hypothesis that potential introgression between *Z. tritici* and *Z. ardabiliae* species may have contributed to the evolutionary history of the *Aegilops*-infecting *Z. tritici* isolates and unitig 9.

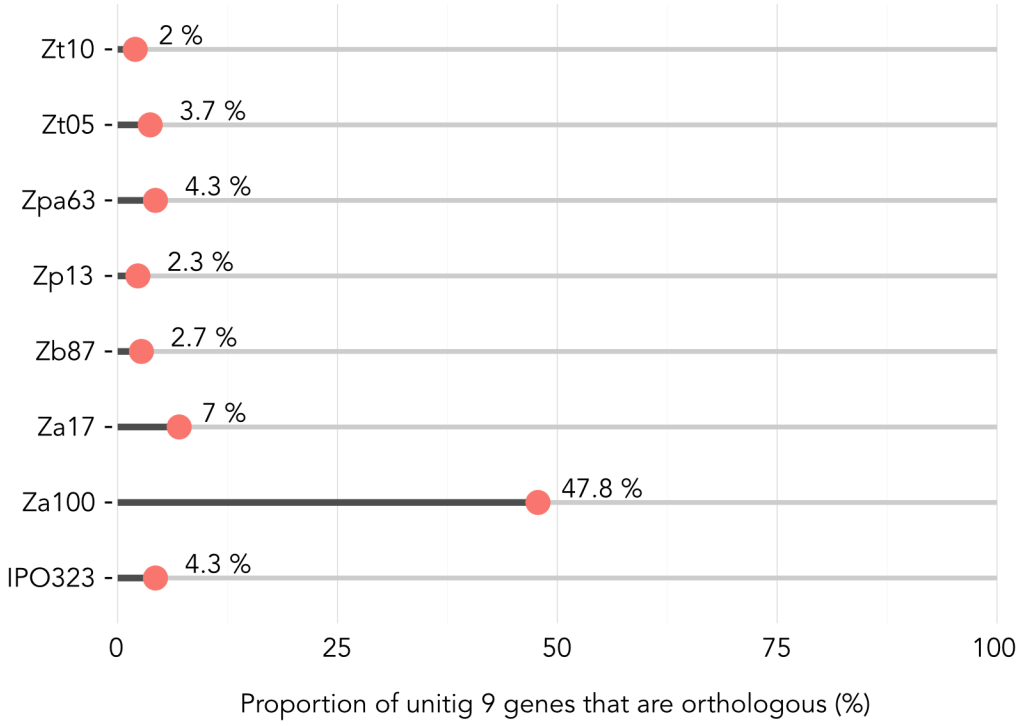


Figure 6. Proportion of unitig 9 genes present in *Zymoseptoria* species genomes. Proportion of genes present in unitig 9 of Zt469 and shared with other *Zymoseptoria* isolates was calculated for each genome individually. Only orthogroups solely shared between unitig 9 and the respective isolate genome are considered.

A full chromosome in Za100 is syntenic with unitig 9 and is in extremely low frequency among *Z. ardabiliae* isolates

The observation that almost half of the proportion of genes in unitig 9 had orthologs in Za100 prompted us to further explore this *Z. ardabiliae* genome. As a first step, we analyzed chromosomal synteny between Za100 and Zt469 using the PacBio long-read sequencing assemblies and gene annotations. The same analysis of gene orthology was performed between the two assemblies and a Circos plot (Krzywinski et al., 2009) was constructed to show the relationship between genes and unitigs. Contrasting to the comparison results observed between the *Z. tritici* isolates Zt469 and IPO323 (Figure 1), unitig 9 was syntenic to unitig 3 in Za100, a presumably completely assembled chromosome showing telomeric repeats in both ends, but significantly larger than unitig 9 in length (~2.6Mb) (Figure 7). In fact, 94.4% (135/143) of the genes of unitig 9 previously observed to have orthologs in Za100 had their orthologs present in unitig 3.

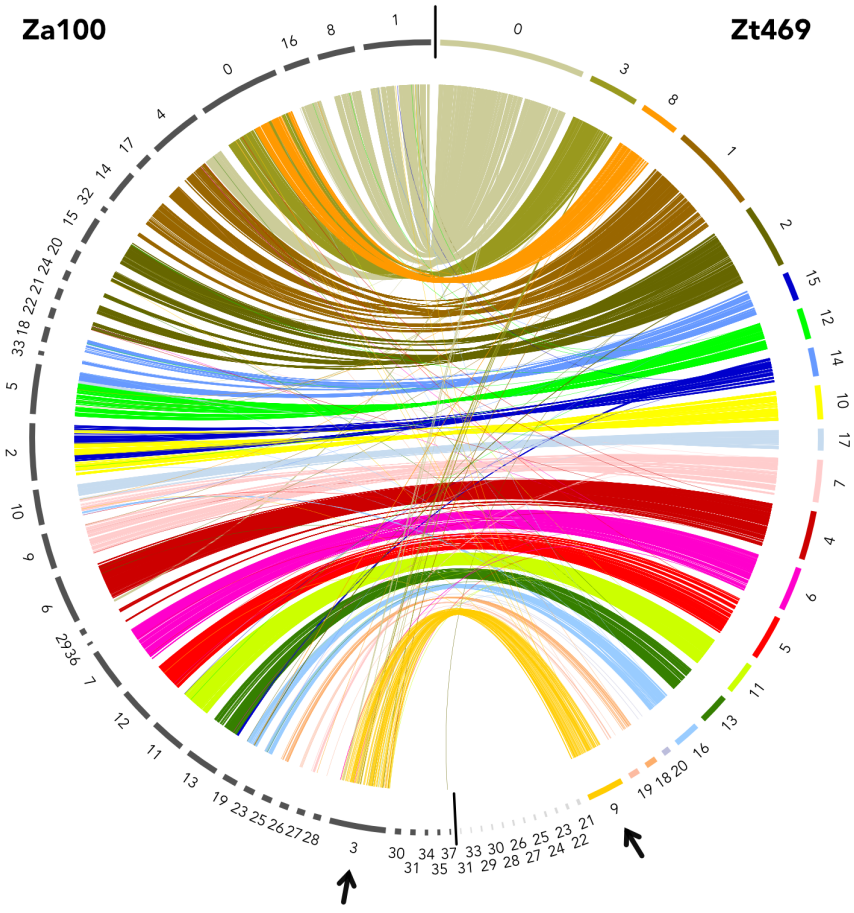


Figure 7. Unitig 3 in Za100 is syntenic to unitig 9 in Zt469. Synteny analysis was performed between the genomes of the *Aegilops*-infecting *Z. tritici* isolate Zt469 (right) and the *Z. ardabiliae* isolate Za100 (left). Each color represents a different contig in Zt469. Small Zt469 contigs are in light gray. Unitig 3 in Za100 shows synteny to unitig 9 in Zt469 (black arrows). Links represent orthologous genes. Contigs in Zt469 are ordered following the synteny to the reference IPO323 genome (Figure 1).

Next, to observe the occurrence of unitig 3 in other *Z. ardabiliae* isolates, we accessed unitig 3 PAV in previously published *Z. ardabiliae* genomes collected at different years and locations (Stukenbrock et al., 2011, 2012; Stukenbrock & Dutheil, 2018). Using the normalized read coverage approach, mapping results to the PacBio Za100 assembly revealed that, besides Za100, only three other isolates of our collection (Za19, Za20 and Za101) had reads mapped to unitig 3 (Figure 8). Similar results were observed when the same *Z. ardabiliae* isolates were mapped to the PacBio Zt469 genome assembly; hereby these four isolates, as expected, showed the presence of unitig 9 reflected by reads mapping to this unitig (Supplementary Figure S3). The low frequency of unitig 3 among *Z. ardabiliae* isolates collected at different years (2004 and 2011) and locations led us to the hypothesis that this chromosome represents an accessory chromosome that is recently introgressed and is in the process of being lost in *Z. ardabiliae* populations.

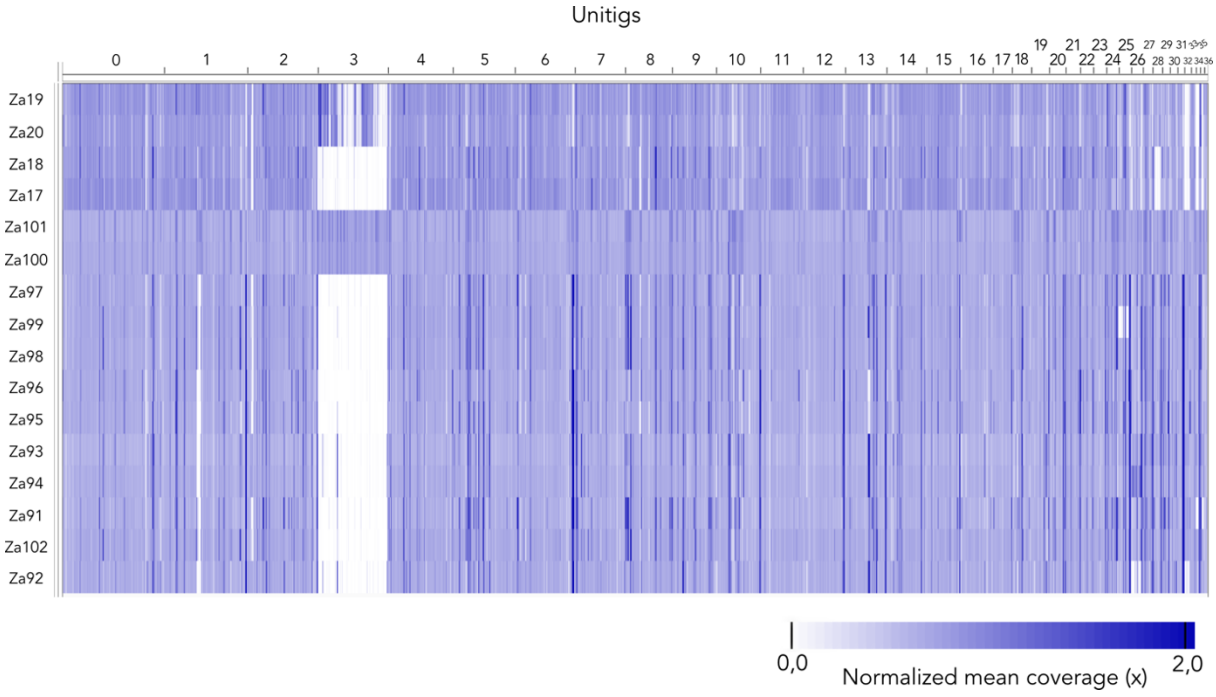


Figure 8. Unitig 3 shows presence-absence variation in *Z. ardabiliae* isolates. Chromosome presence-absence variation (PAV) was analyzed by read mapping to the *Z. ardabiliae* Za100 PacBio genome assembly. Heatmap represents normalized mean coverage of reads mapped to each position of the genome assembly. Only unitigs larger than 100 kb are displayed and sorted by descending order of length. Darker colors represent regions of higher coverage as e.g. repetitive elements or duplications.

Unitig 3 in *Z. ardabiliae* has accessory chromosome hallmarks but it is not enriched in H3K27me3 methylation mark

In order to better characterize unitig 3 in Za100, we analyzed the genomic landscape of this chromosome. Similar to unitig 9 in Zt469, we observed a low gene density (average of 0.44 genes/kb), high TE density (average of 0.34 TEs/kb) and low transcriptional activity when compared to unitigs of similar size (e.g. unitig 4; 0.86 genes/kb and 0.20 TEs/kb on average) (Figure 9A). These signatures are consistent with other accessory chromosomes in *Zymoseptoria* species (Feurtey et al., 2020; Goodwin et al., 2011; Kellner et al., 2014) are also present in smaller unitigs with telomeric repeats in this assembly (unitigs 24 to 37) (Figure 9A). Intriguingly, we observed distinct chromatin patterns across the unitigs, particularly in unitig 3. ChIP-seq results targeting three histone methylation marks (H3K4me2, H3K27me3 and H3K9me3) revealed that unitig 3 is not enriched in the heterochromatin methylation mark H3K27me3 when compared to the other accessory unitigs (unitigs 25 to 37) (Figure 9B). We also observed accessory chromosome signatures on unitig 10, which is syntenic to unitig 17 in Zt469 (Figure 7), which in turn is syntenic to the "accessory arm" of chromosome 7 in *Z. tritici* IPO323 (Figure 1) (Schotanus et al., 2015). The low enrichment of H3K27me3 in unitig 3 suggests that, even though accessory chromosome hallmarks were observed (high TE, low gene density and low transcriptional activity), different histone modifications from the ones observed for other accessory chromosomes in *Zymoseptoria* species are shaping this chromosome.

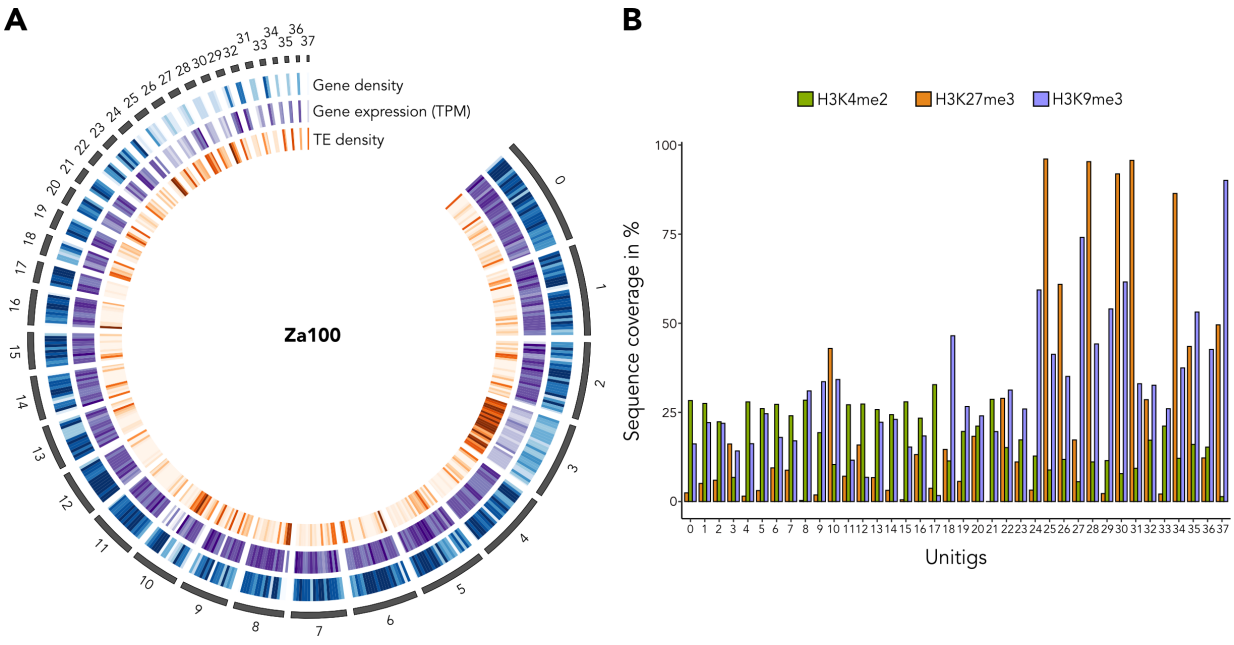


Figure 9. Unitig 3 shows intermediate accessory chromosome hallmarks. (A) Circos plot representing different features along the thirty-eight Za100 unitigs. Unitigs are represented by the dark gray segments sorted by length. Tracks from outside to the inside are heatmaps in 100kb windows representing respectively: gene density, gene expression *in vitro* in log2(TPM+1) and TE density. Darker heatmap colors indicate higher values. **(B)** Barplot displaying the percentage of sequence coverage of each Za100 unitig with the heterochromatin methylations H3K27me3 (orange bars) and H3K9me3 (purple bars), and the euchromatin methylation H3K4me2 (green bars) marks relative to unitig length.

Unitig 3 and unitig 9 show high TEs transcription activity and differential RIP signatures

Considering the differences observed in length and histone methylations marks between unitig 3 in Za100 and unitig 9 in Zt469, we focused on further characterizing the TE landscape of these genomes. TEs can promote intra- and inter-specific variability in terms of genome structure, size and transcriptional regulation, and their proliferation control is tightly linked with DNA methylation and heterochromatin-associated histone modifications (Deniz, Frost, & Branco, 2019; Feschotte & Pritham, 2007; Fueyo, Judd, Feschotte, & Wysocka, 2022; Raffaele & Kamoun, 2012). Analyses of TE content in Zt469 and Za100 revealed that the two genomes vary in the overall proportion of TEs; the genome of Za100 comprises 19,02% of TEs while the genome of Zt469 includes 15,76% of these genetic elements (Supplementary Figure S4A). In both genomes, LTR-retrotransposons represent the largest portion of TEs, followed by TIR-transposons (Supplementary Figure S4B). These results are consistent with previous analyses of TE content in other *Zymoseptoria* genomes, including the previous Zt469 assembly version, which shows that TE proportions of *Z. tritici* genomes are lower than the TE proportions of *Z. ardabiliae* and other closely related *Zymoseptoria* species (Lorrain et al., 2021). Regarding unitig 3 and unitig 9, although both accessory unitigs are enriched in TEs (Figures 4 and 9), unitig 3 in Za100 shows a higher TE proportion than unitig 9 in Zt469, in which the former is covered by 15,36% and the latter is covered by 13.36% of TEs (Supplementary Figure S5A). In unitig 3, retrotransposons from the LINE order make up the largest portion of TEs while in unitig 9 most of the TEs could not have their orders assigned ("noCat"; Supplementary Figure S5B).

Interestingly, not only these unitigs differ in TE proportion but they also differ in the levels of expression of these genetic elements during *in vitro* growth. Analyses of TE expression using RNA-seq data showed that, although both unitigs show a high TE expression (Figure 10A), unitig 3 has significantly higher TE expression levels than unitig 9 (Wilcoxon rank sum test, p-value=0.0016; Figure 10B). In both unitigs, expression levels of TEs also showed to be significantly higher than

the expression of genes (Wilcoxon rank sum test, p-value < 2.22e-16; Figure 10A, B). Based on these results, we hypothesized that unitig 3 and unitig 9 have a reduced efficacy of genome defense mechanisms regulating TE activity.

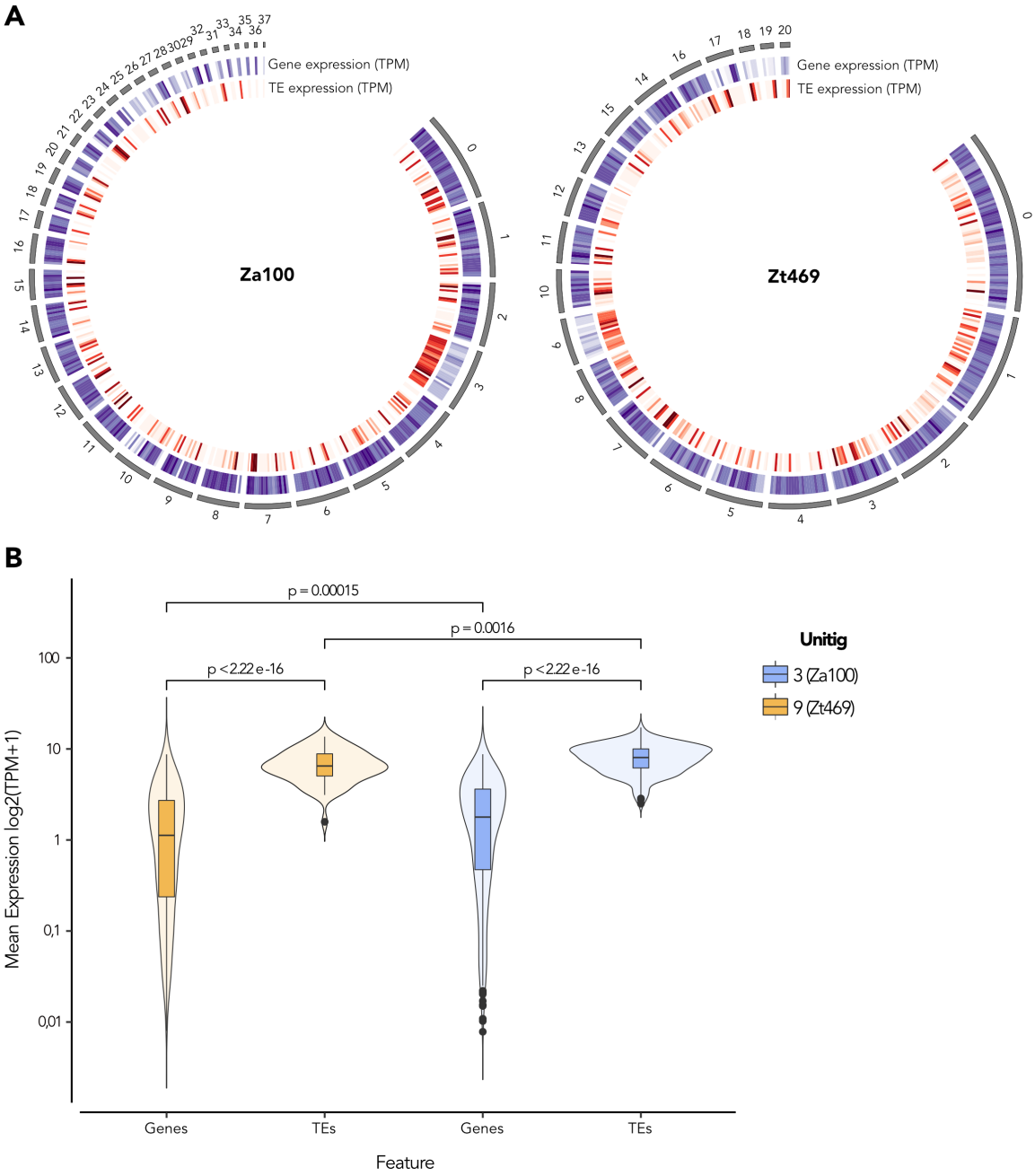


Figure 10. Unitig 3 shows higher transcriptional activity of TEs than unitig 9. (A) Circos plots representing gene and TE expression *in vitro* in log2(TPM+1) along the Za100 (left) and Zt469 (right) genomes. Unitigs are represented by the dark gray segments sorted by length. For Zt469, except for unitig 9, only unitigs syntenic to the *Z. tritici* IPO323 reference genome are shown. Tracks from outside to the inside are heatmaps in 100kb windows representing respectively: gene expression *in vitro* in log2(TPM+1) and TE expression *in vitro* in log2(TPM+1). Darker heatmap colors indicate higher values. **(B)** Violin plots representing the mean expression levels in log2(TPM+1) for both gene and TE features of unitig 3 in Za100 (violet color) and unitig 9 in Zt469 (orange color) during *in vitro* growth. P-values were calculated using pairwise Wilcoxon rank sum tests.

To test this hypothesis, we accessed signatures of another genome defense mechanism, Repeat Induced Point mutations in both Zt469 and Za100 genomes and respective TE copies to get insights into the efficacy of host genome defense mechanisms regulating TE activity. Besides DNA methylation and heterochromatic histone modifications (Deniz et al., 2019; Zemach, McDaniel, Silva, & Zilberman, 2010), fungal species have evolved a fungal-specific mechanism called RIP (Repeat-Induced Point mutation) to specifically mutate and inactivate duplicated sequences as TEs (Cambareri, Jensen, Schabtach, & Selker, 1989; Gladyshev, 2017). We scanned each genome individually by looking at dinucleotide bias mutations in duplicated sequences and calculated the RIP compositive indices and the large RIP affected regions (LRARs) in 1 kbp genomic windows using the RIPper software (Van Wyk et al., 2019). LRARs represent genomic windows consecutively affected by RIP that are more than 4000bp in length (Van Wyk et al., 2019). At the genome level, we observed that RIP signatures were present in 21.85 % of the total genomic content of Za100 and 18.61 % of the total genomic content in Zt469 (Table 1). Large RIP affected regions (LRARs), besides comprising extensive AT-rich regions, had an average size of 23585.43 bps and 18298.63 bps and affected a total of 8.2 Mbp and 6.9 Mbp on Za100 and Zt469 genomes, respectively (Table 1). At the chromosome level, percentage of 1 kbp regions per unitig affected by RIP ranged from 2.72% to 85.52% in Za100 and from 7.22% to 31.5% in Zt469 (Figure 11A, B). Unitig 3 in Za100 was affected in 24.09% by RIP while unitig 9 in Zt469 was more RIPped, totalizing 28.95% of the total unitig content (Figure 11A, B). We also estimated RIP composite indices per 50bp windows of each TE copy in both Zt469 and Za100 genomes and observed that, on average, TE copies were statistically more RIPped in unitig 9 of Zt469 than in unitig 3 of Za100 (Wilcoxon rank sum test, p-value < 2.22e-16; Supplementary Figure S6). The same pattern was observed when looking at TE copies present genome-wide (Supplementary Figure S6). Altogether, these results suggest that, although Za100 genome is more RIPped than Zt469, RIP has a lower efficiency to inactivate TEs in unitig 3 of Za100 than when compared to unitig 9 of Zt469. The higher transcriptional activity of TEs in unitig 3 (Figure 10) also corroborates this assumption. Based on all these findings, we propose that unitig 3 of Za100 is enriched with active TEs.

Table 1. Summary of genome-wide RIP signatures detected in Za100 and Zt469 isolates

Genome	GC content genome (%)	Proportion genome affected by RIP (%)	Number of LRARs	Proportion genome covered by LRARs (Mbp)	Average size of LRARs (bps)	Average GC content of LRARs (%)
Za100	51.34	21.85	346	8.2	23585.43	40.35
Zt469	51.32	18.61	379	6.9	18298.63	41.83

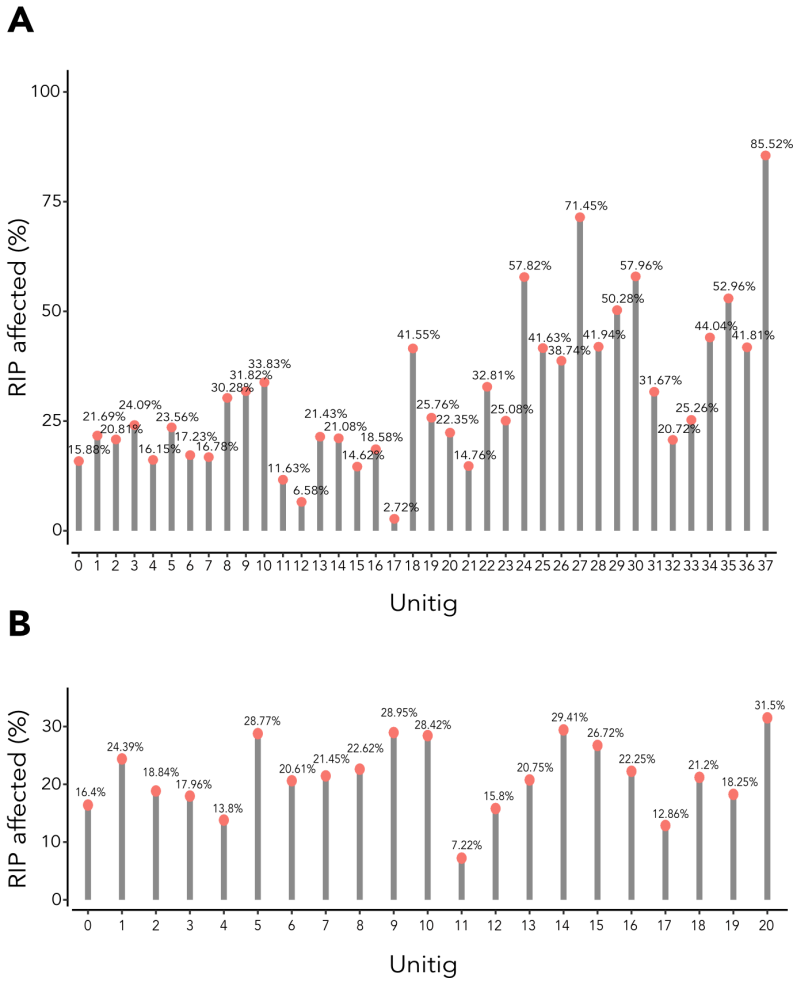


Figure 11. Unitig 3 and unitig 9 are differentially affected by RIP. Lolliplot plots showing the percentage of regions (1 kbp windows) per unitig affected by RIP in Za100 **(A)** and Zt469 **(B)** genomes. Unitigs are sorted by descending order of length. For Zt469 **(B)**, except for unitig 9, only unitigs syntenic to *Z. tritici* IPO323 reference genome are shown.

Gene expression in unitig 9 suggests a dispensable function during infection

We finally asked if unitig 9 in *Z. tritici* Zt469 has a functional relevance by analyses of gene and TE expression patterns during *in vitro* growth and during *in planta* infection. For expression profiling *in planta*, we generated stage-specific RNA-seq datasets based on the four infection stages and correspondent time points identified by confocal microscopy described previously (Chapter II and Supplementary Figures S1). We could not generate *in planta* transcriptome data for Za100 due to the lack of an experimental compatible host. To determine the levels of expression in individual genes and TEs, we performed gene and TE annotations and functionally annotated each gene model using different tools (see Methods and Supplementary Tables S2-S5). Expression levels were compared using normalized read mappings to transcript per million (TPM) in log2(TPM+1). For each condition (*in vitro* growth and per infection stage), we generated transcriptome data of three biological replicates, totalizing an average of 39 (*in vitro*) and 91 million (*in planta*) reads. Percentage of reads that could be mapped to the Zt469 genome were on average of 93.4 % between replicates for *in vitro* growth and ranged from 2.3 % to 76.5 % between replicates and infection stages during *in planta* infection (Supplementary Table S7).

Overall, we observed that Zt469 showed different levels of transcription of genes and TEs when we compare transcription of these features *in vitro* and *in planta* in all infection stages analyzed (Supplementary Figure S7A). Expression of genes between *in vitro* growth and *in planta* infection stages showed to be statistically different genome-wide in Zt469 (Kruskall-Wallis test, $p < 2.2$ e-16; Supplementary Figure S7A). The same pattern was observed when focusing only on TEs (Kruskall-Wallis test, $p < 2.2$ e-16; Supplementary Figure S7A). By looking at the expression levels exclusively in unitig 9, we observed a low transcriptional activity of genes compared to TEs not only *in vitro,* but also *in planta* in all infection stages analyzed (Figures 10; Supplementary Figure S7B and Supplementary Table S8). Focusing only on genes, Kruskall-Wallis test indicate that there is a significant difference in unitig 9 gene expression levels between *in vitro* growth and the four infection stages (Kruskall-Wallis test, $p = 8.3$e-09; Supplementary Figure S7B). Regarding unitig 9 TEs, expression levels of these elements is also statistically different between *in vitro* growth and *in planta* infection stages (Kruskall-Wallis test, $p = 0.00018$; Supplementary Figure S7B). Previous transcriptome studies of *Z. tritici* during *in vitro* growth and *in planta* infection have shown that, on average, the expression levels of genes located on accessory chromosomes were up to 20-fold lower than the expression levels of genes present on core chromosomes (Haueisen et al. 2019; Kellner et al., 2014). Our results corroborate these findings by demonstrating that genes located in unitig 9, and also in the other accessory unitigs 18, 19 and 20, were expressed on average from

1.6- to 2.1-fold lower levels than the genes located in the core chromosomes (unitigs 0 to 17; Supplementary Tables S8 and S9).

Next, we functionally annotated the gene models in the Zt469 genome and focused on the expression profile of specific gene categories (see Methods and Supplementary Tables S2, S3 and S5). Secreted proteins as CAZymes (carbohydrate-degrading enzymes), effectors and other small secreted proteins (SSPs) as well as secondary metabolites encoded by biosynthetic gene clusters (BGCs) play a crucial role in host-pathogen interactions and fungal adaptive evolution (Lo Presti et al., 2015; van der Does & Rep, 2017; Shi-Kunne et al., 2019; Zhao, Liu, Wang, & Xu, 2014), and therefore can provide insights into the potential functional relevance of unitig 9 in *Z. tritici* Zt469, particularly during *in planta* infection. As a comparison to other genes, we grouped all genes that were not classified in these categories as "Other" and also analyzed their expression levels. Between *in vitro* and *in planta* conditions, levels of gene expression showed to be statistically different for all the gene categories analyzed genome-wide in Zt469 (Figure 12A; Kruskal-Wallis tests, $p < 0.02$). Interestingly, we only observed statistical difference in levels of expression between *in vitro* growth and the infection stage "D" for the CAZyme gene category (Wilcoxon rank sum test, $p=0.003$), while for all the other gene categories no statistical difference was between these two conditions (Wilcoxon rank sum test, $p>0.1$; Figure 12A). Levels of expression in genes encoding candidate effectors ("Effector" category) showed to be statistically different between *in vitro* growth and all infection stages, except stage "D" (Wilcoxon rank sum test, $p=0.52$; Figure 12A).

We could not perform the same statistical analyses by gene category when looking at the genes present on unitig 9 in Zt469. This unitig shows a small number of genes belonging to most of the categories analyzed and therefore no robust statistical analyses could be performed. However, we could observe interesting expression patterns among *in vitro* and *in planta* conditions when looking at each individual gene of these categories in unitig 9, particularly at categories playing a role in host-pathogen interactions. Out of 26 genes analyzed, 16 genes showed no transcription (TPM=0) in at least one condition (Figure 12B). Most of the genes analyzed are not transcribed at infection stage "A" (7dpi) followed by a low to intermediate expression ($0.03 < \log2(TPM+1) < 0.030$) at infection stage "D" (21dpi; Figure 12B). Taking all these results together, we suggest that unitig 9 has a potential dispensable function during *in planta* infection. Further analyses on differential gene expression between *in vitro* and *in planta* conditions as well as *in vitro* stress assays and *in planta* infection experiments using Zt469 strains lacking unitig 9 are necessary to fully characterize the

expression profiling of individual genes and ultimately access the functional relevance of this unique unitig to *Z. tritici* fitness and pathogenesis.

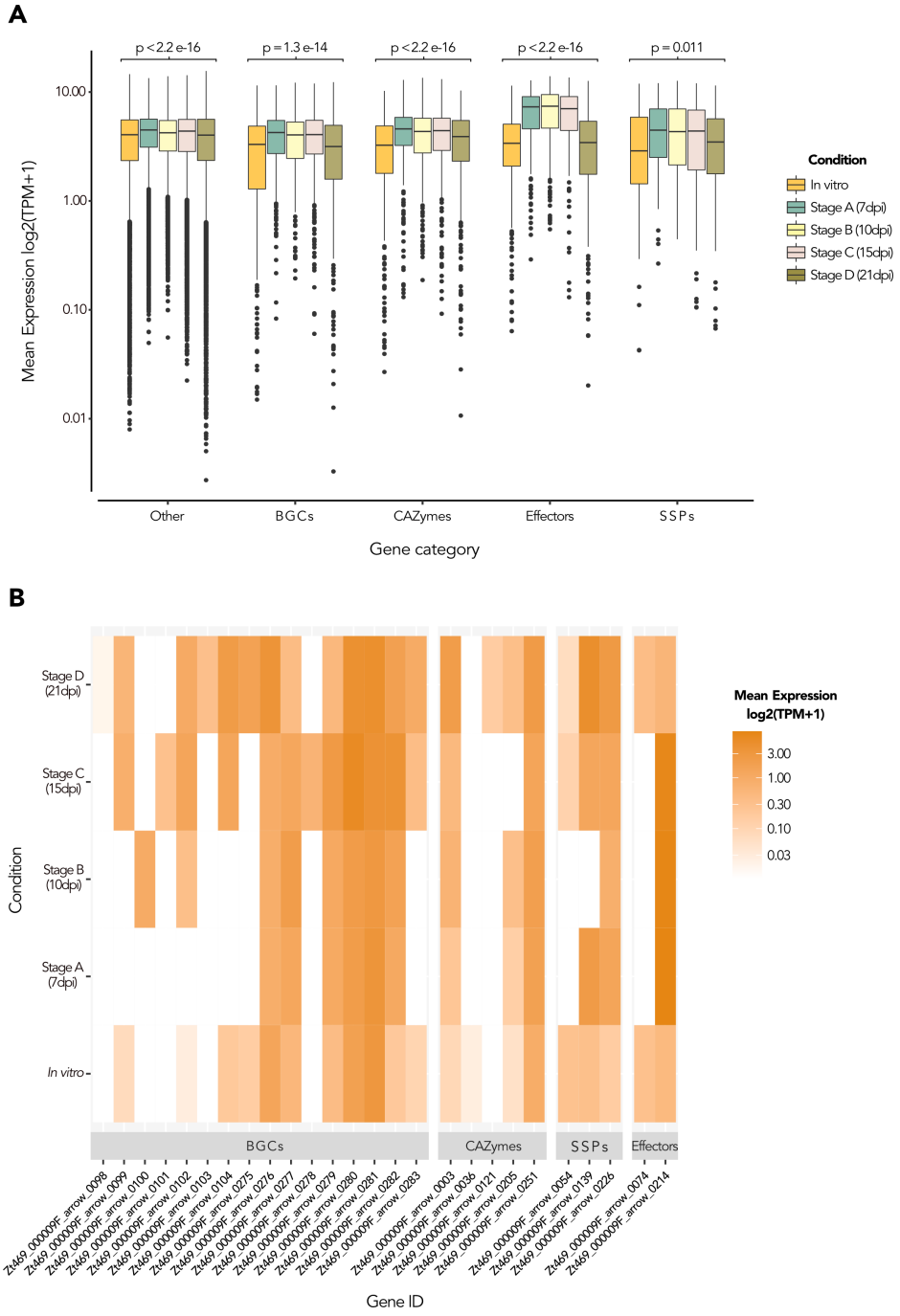


Figure 12. Gene categories involved in host-pathogen interactions show different levels of expression in Zt469 during *in vitro* growth and *in planta* infection. (A) Boxplots representing the mean expression levels in log2(TPM+1) for each gene category genome-wide in Zt469 at different infection stages *in planta* and *in vitro*. P-values were calculated using Kruskall-Wallis tests within each category. **(B)** Heatmaps showing mean expression levels in log2(TPM+1) for each gene individually in unitig 9 belonging to four gene categories analyzed at different infection stages *in planta* and *in vitro*. Only the main gene categories involved in host-pathogen interactions are shown to facilitate visualization. White cells represent no transcription (TPM=0). BGCs (Biosynthetic Gene Clusters); CAZymes (Carbohydrate-degrading enzymes); SSPs (Small Secreted Proteins).

Discussion

Fungal plant pathogens show extensive genomic plasticity and large degrees of genetic variation that contribute to rapid adaptive evolution and overall evolutionary potential (McDonald & Linde, 2002; Möller & Stukenbrock, 2017). In this context, accessory chromosomes provide an optimal ground source of genetic innovations considering the low density of "essential" genes, and therefore a reduced effect of background selection, and the activity of transposable elements (TEs) that can create new genetic variants (Bertazzoni et al., 2018; Croll & McDonald, 2012; Möller & Stukenbrock, 2017). In this study, we characterize new and unique accessory chromosomes in *Aegilops*-infecting *Z. tritici* and in the closely related species *Z. ardabiliae* using a multi-omics approach and high-quality whole genome assemblies, gene and TE annotations. Moreover, we present the first *in planta* transcriptomic analyses of *Z. tritici* infecting *A. cylindrica* plants supported by detailed and infection stage-aware confocal microscopy analyses. Our results suggest another line of evidence that recurrent interspecies hybridization occurs between *Zymoseptoria* species and may serve as a route for accessory chromosomes emergence.

Consistent with previous analyses on the genomic structure of *Zymoseptoria* species (Feurtey et al., 2020; Grandaubert et al., 2015; Haueisen et al., 2019; Kellner et al., 2014; Schotanus et al., 2015), the genome architecture in the analyzed *Zymoseptoria* isolates is also highly compartmentalized. We demonstrated that unitig 9 and the smaller unitigs 18, 19 and 20 in the *Aegilops*-infecting *Z. tritici* isolate Zt469 showed all accessory chromosome hallmarks observed in wheat-infecting *Z. tritici* and closely related *Zymoseptoria* species, including enrichment in heterochromatin methylation mark H3K27me3; PAV between isolates and low levels of gene transcription *in vitro* and *in planta* (Feurtey et al., 2020; Haueisen et al., 2019; Schotanus et al., 2015). Similarly in the *Z. ardabiliae* Za100 isolate, accessory unitigs, including the large unitig 3, showed accessory characteristics that distinguish them from the core unitigs as e.g. PAV between isolates; low gene and high transposon densities; and low transcriptional gene activity. Considering the divergence observed between *Aegilops*- and wheat-infecting *Z. tritici* isolates (Chapter II) and between *Z. tritici* and *Z. ardabiliae* species (Stukenbrock et al., 2007, 2010, 2011 2012), these results corroborate the previous findings that genome compartmentalization and accessory chromosomes are ancestral traits among *Zymoseptoria* species (Feurtey et al., 2020).

We observed interesting findings regarding TE expression control and methylation marks present on unitig 3 of Za100. One of the crucial roles of heterochromatin in eukaryotes is to prevent genome instability by silencing TE replication and spread (Allshire & Madhani, 2018). In several

fungal plant pathogen species, including *Zymoseptoria* species, enrichment of the facultative heterochromatin methylation mark H3K27me3 has been considered a hallmark of accessory chromosomes and can explain the overall transcriptional gene silencing observed for these chromosomes (Galazka & Freitag, 2014; Schotanus et al., 2015; Studt et al., 2016). H3K9me2/3 on the other hand has been demonstrated to be enriched in repeat-rich regions with a direct impact on transposable elements (TEs) control and genome stability (Mikkelsen et al., 2007; Möller et al., 2019; Peters et al., 2001; Zeller et al., 2016). Transcriptional activation of TEs can result in their spread throughout the host genome, causing insertional mutations and genomic instability (Castanera et al., 2016; Hedges & Deininger, 2007; Padeken, Zeller, & Gasser, 2015). Besides histone post-translational modifications and transposon methylation (Freitag, 2017; Moller et al., 2021; Zemach et al., 2010), fungal genomes have developed a fungal-specific mechanism to mutate and silence repetitive sequences known as RIP (Repeat-Induced Point mutations; Cambareri et al., 1989; Gladyshev, 2017). Interestingly, although we observed that unitig 3 shows most of the expected features of an accessory chromosome, this unitig was not enriched with the heterochromatin methylation marks H3K27me3 and H3K9me3. We also observed that unitig 3 has reduced RIP signatures when compared to most small and accessory unitigs in Za100 and to the syntenic unitig 9 in Zt469. In line with these results, we found that the levels of TE transcription in unitig 3 are high and even superior to the levels observed for genes within the same unitig and to the levels observed for unitig 9 in Zt469. Taking all these results together, we suggest that unitig 3 serves as a reservoir of transcriptionally active TEs that lacks a genomic control preventing their spread in Za100. In male *Drosophilla* flies, it has been recently shown that the "toxic" repeat-rich Y chromosomes act as reservoirs of actively transcribed TEs and lead to reduced male fitness by deleterious TE mobilization and ineffective heterochromatin silencing of these repetitive elements (Nguyen & Bachtrog, 2021; Wei, Gibilisco, & Bachtrog, 2020). Further analyses on Za100 TEs activity and their genomic and epigenetic control *in vitro* and potentially *in planta* can shed a light on the mechanisms regulating TE spread and their impact on *Z. ardabiliae* development and fitness.

Introgression can provide a powerful and fast mechanism for adaptive genome evolution in fungal plant pathogens. Exchange of chromosomes between pathogen lineages or species through introgression can generate novel pathotypes and ultimately new pathogen species (Depotter, Seidl, Wood, & Thomma, 2016; Stukenbrock, 2016a, 2016b). Experimental evidence from *Fusarium oxysporum* demonstrates that accessory (or lineage-specific) chromosomes can be vertically transferred between distinct lineages by hyphal fusion and convert non-pathogenic strains into a pathogen in specific hosts (Ma et al., 2010; Vlaardingerbroek et al., 2016). Earlier reports have also

shown that horizontal transfer of chromosomes between fungal strains can occur in *Colletotrichum gloeosporioides* (He, Rusu, Poplawski, Irwin, & Manners, 1998) and in *Alternaria alternata* (Akagi, Akamatsu, Otani, & Kodama, 2009). In *Z. tritici*, it has been suggested that accessory chromosomes have originated via ancient horizontal transfer from an unknown donor followed by extensive recombination events (Goodwin et al., 2011). However, other studies have suggested that other mechanisms may be more likely. One study demonstrated that a *Z. tritici* accessory chromosome emerged by non-disjunction duplication of core chromosomes followed by a degeneration process via breakage-fusion-bridge (BFB) cycles and RIP on duplicated sequences (Croll et al., 2013). Despite these evidences, the role of introgression on the origin of accessory chromosomes in *Zymoseptoria* species remain largely unknown.

In a previous study using a whole-genome Neighbor-net reticulate network approach, we observed that reticulate edges were connecting *Aegilops*-infecting *Z. tritici* isolates to closely-related *Zymoseptoria* species, including *Z. ardabiliae* (Chapter II). The reticulate network analyses suggested that possible evolutionary events as recombination, horizontal gene transfer, or hybridization between taxa were taking place between *Aegilops*-infecting *Z. tritici* isolates and the closely related *Zymoseptoria* species (Huson & Bryant, 2006; Chapter II). In this present study, we report that unitig 9, an accessory chromosome only occurring in *Aegilops*-infecting *Z. tritici* isolates, is syntenic to another accessory chromosome (unitig 3) in the *Z. ardabiliae* isolate Za100, further suggesting that introgression and possible chromosome transfers may have happened between the two closely related species. Thus, considering the results we gathered so far (Figure 14 and Chapter II) and the sympatry of *Z. tritici* and *Z. ardabiliae* species in the Middle East (Stukenbrock et al., 2007, 2010, 2011, 2012), some hypothetical introgression scenarios can be made about the evolutionary origin of the novel accessory chromosomes in the *Z. tritici* and *Z. ardabiliae* species:

In a first scenario, introgression could happened from *Z. ardabiliae* into *Aegilops*-infecting *Z. tritici*. In this scenario, we could infer that unitig 3 is a chromosome that remained from ancestral populations and it is becoming completely lost in the *Z. ardabiliae* species, resulting in the extreme low frequency observed among the analyzed *Z. ardabiliae* isolates. The detrimental effects of Transposable Elements (TEs) for genome integrity (Castanera et al., 2016; Hedges & Deininger, 2007; Padeken et al., 2015) and the high density and transcriptional activity of these elements in unitig 3 also suggest that selection against TEs could eventually lead to the loss of this chromosome in the course of evolution. Moreover, the accessory chromosome hallmarks observed on unitig 9 can further suggest that this chromosome was acquired in *Aegilops*-infecting *Z. tritici* from *Z. ardabiliae* and it has been segregating in this lineage at an intermediate frequency possibly due to

clonal propagation or through "selfish" mechanisms such as meiotic drive (Habig et al., 2018; Wittenberg et al., 2009). The reduced density and transcriptional activity of TEs coupled with higher signatures of RIP and heterochromatin methylation marks in unitig 9 compared to unitig 3 also support that introgression occurred in this direction, particularly when considering the efficiency of genome defenses against TE replication and therefore the higher maintenance of unitig 9 in *Aegilops*-infecting *Z. tritici* populations throughout evolution.

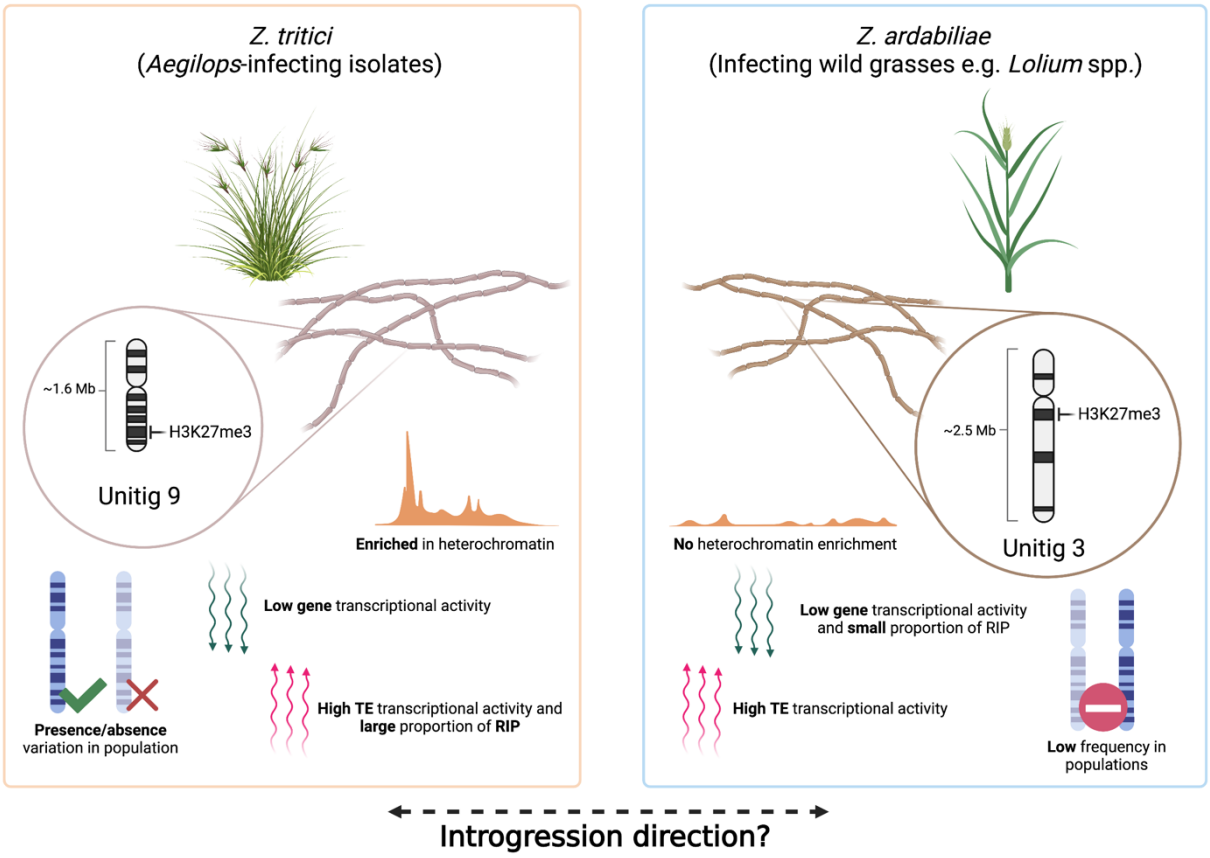


Figure 14. Summary of genomic features in the novel syntenic chromosomes of Zt469 and Za100. Based on the results we gathered so far, we hypothesize that interspecies introgression took place between *Aegilops*-infecting *Z. tritici* isolates and *Z. ardabiliae* species and gave origin to the syntenic accessory chromosomes unitig 9 and unitig 3. Figure created with BioRender.com

In a second introgression scenario, we consider the possibility that a recent chromosomal introgression happened from *Aegilops*-infecting *Z. tritici* isolates into the *Z. ardabiliae* species. This scenario can be supported by the extreme low frequency of unitig 3 observed among *Z. ardabiliae* isolates, suggesting that this chromosome has not been established yet at the *Z. ardabiliae* species level. We also observed high TE transcription activity, low enrichment of the heterochromatin-associated methylation marks (H3K27me3 and H3K9me3) and low RIP signatures in this

chromosome and TE copies that reside within, which could further corroborate that unitig 3 has been recently introgressed and since then acting as a reservoir of transcriptionally active TEs.

Considering these scenarios in light of our findings, we suggest that the first introgression scenario is the most parsimonious explanation for the origin and maintenance of the syntenic accessory chromosomes unitig 9 and unitig 3 in *Z. tritici* and *Z. ardabiliae* species. Further sampling to detect more *Z. ardabiliae* isolates containing unitig 3 is required to better characterize this chromosome through population genomics approaches and to give us support for this hypothetical introgression scenario.

Conclusions

In this study we identified unique, and so far, undescribed accessory chromosomes in *Z. tritici* and *Z. ardabiliae* species using a comparative genomics and multi-omics approach framework. Our findings provide valuable resources to investigate the evolutionary mechanisms and potential impact of interspecies hybridization on the origin of accessory chromosomes in the *Zymoseptoria* genus. The occurrence of these chromosomes only in fungal species infecting wild grasses highlights the importance of wild pathosystems on adaptive genome evolution of fungal pathogens and can further illustrate a possible source of genome novelty for closely related fungal lineages infecting important crops like *Z. tritici* on wheat.

Acknowledgements

The authors would like to thank Cécile Lorrain for support with the TE analyses and the members of the Environmental Genomics group for helpful discussions.

Funding statement

This work was supported by intramural funding of the Max Planck Society and a personal grant from the State of Schleswig-Holstein to Eva H. Stukenbrock. The funders had no role in study design, data collection and analyses, decision to publish, or preparation of the manuscript.

Authors contributions

Conceptualization: WCF, JH, EHS; Zt469 sample collection: FS, AA; Whole-genome DNA extractions: WCF, JH; Genome sequencing and genomics analyses: WCF, AF; ChIP-DNA and *in vitro* RNA isolation and sequencing: MM; *In planta* RNA isolation: WCF, RH; Preparation and writing of manuscript: WCF. Editing of original manuscript: WCF, EHS, JH, RH and MM.

References

Akagi, Y., Akamatsu, H., Otani, H., & Kodama, M. (2009). Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryotic Cell*, *8*(11), 1732–1738. https://doi.org/10.1128/EC.00135-09

Allshire, R. C., & Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology*, *19*(4), 229–244. https://doi.org/10.1038/nrm.2017.119

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, *2*(5), 1–14. https://doi.org/10.26508/lsa.201900429

Armenteros, J. J. A., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics (Oxford, England)*, *33*(21), 3387–3395. https://doi.org/10.1093/bioinformatics/btx431

Armenteros, J. J. A, Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., … Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

Balesdent, M. H., Fudal, I., Ollivier, B., Bally, P., Grandaubert, J., Eber, F., … Rouxel, T. (2013). The dispensable chromosome of *Leptosphaeria maculans* shelters an effector gene conferring avirulence towards Brassica rapa. *New Phytologist*, *198*(3), 887–898. https://doi.org/10.1111/nph.12178

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*(8), 779–788. https://doi.org/10.1094/mpmi-06-17-0135-fi

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Van Wezel, G. P., Medema, M. H., & Weber, T. (2021). AntiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Research*, *49*(W1), W29–W35. https://doi.org/10.1093/nar/gkab335

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cambareri, E. B., Jensen, B. C., Schabtach, E., & Selker, E. U. (1989). Repeat-induced GC to AT mutations in *Neurospora*. *Science*, *244*(4912), 1571–1575.

Castanera, R., Lopez-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., … Ramirez, L. (2016). Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genetics*, *12*(6), e1006108. https://doi.org/10.1371/journal.pgen.1006108

Coleman, J. J., Rounsley, S. D., Rodriguez-Carres, M., Kuo, A., Wasmann, C. C., Grimwood, J., … Vanetten, H. D. (2009). The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics*, *5*(8). https://doi.org/10.1371/journal.pgen.1000618

Croll, D., & McDonald, B. A. (2012). The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens*, *8*(4), e1002608. https://doi.org/10.1371/journal.ppat.1002608

Croll, D., Zala, M., & McDonald, B. A. (2013). Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLoS Genetics*, *9*(6). https://doi.org/10.1371/journal.pgen.1003567

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., … Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/gigascience/giab008

Deniz, Ö., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics*, *20*(7), 417–431. https://doi.org/10.1038/s41576-019-0106-6

Depotter, J. R. L., Seidl, M. F., Wood, T. A., & Thomma, B. P. H. J. (2016). Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. *Current Opinion in Microbiology*, *32*, 7–13. https://doi.org/10.1016/j.mib.2016.04.005

Dhillon, B., Gill, N., Hamelin, R. C., & Goodwin, S. B. (2014). The landscape of transposable elements in the finished genome of the fungal wheat pathogen *Mycosphaerella graminicola*. *BMC Genomics*, *15*(1), 1–17. https://doi.org/10.1186/1471-2164-15-1132

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development*, *35*, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

Dutheil, J. Y., Mannhaupt, G., Schweizer, G., Sieber, C. M. K., Münsterkötter, M., Güldener, U., … Kahmann, R. (2016). A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. *Genome Biology and Evolution*, *8*(3), 681–704. https://doi.org/10.1093/gbe/evw026

Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G., & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*, *340*(4), 783–795. https://doi.org/https://doi.org/10.1016/j.jmb.2004.05.028

Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. (2020). Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow. *Current Protocols in Microbiology*, *59*(1), 1–27. https://doi.org/10.1002/cpmc.128

Faino, L., Seidl, M. F., Shi-Kunne, X., Pauper, M., Van Den Berg, G. C. M., Wittenberg, A. H. J., & Thomma, B. P. H. J. (2016). Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research*, *26*(8), 1091–1100. https://doi.org/10.1101/gr.204974.116

Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, *41*, 331–368. https://doi.org/10.1146/annurev.genet.40.110405.090448

Feurtey, A., Lorrain, C., Croll, D., Eschenbrenner, C., Freitag, M., Habig, M., ... & Stukenbrock, E. H. (2020). Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*. *BMC genomics*, *21*(1), 1-15. https://doi.org/10.1186/s12864-020-06871-w

Feurtey, A., Stevens, D. M., Stephan, W., & Stukenbrock, E. H. (2019). Interspecific gene exchange introduces high genetic variability in crop pathogen. *Genome Biology and Evolution*, *11*(11), 3095–3105. https://doi.org/10.1093/gbe/evz224

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0016526

Fones, H., & Gurr, S. (2015). The impact of *Septoria tritici* Blotch disease on wheat: An EU perspective. *Fungal Genetics and Biology*, *79*, 3–7. https://doi.org/10.1016/j.fgb.2015.04.004

Fouché, S., Plissonneau, C., McDonald, B. A., & Croll, D. (2018). Meiosis leads to pervasive copy-number variation and distorted inheritance of accessory chromosomes of the wheat pathogen *Zymoseptoria tritici*. *Genome Biology and Evolution*, *10*(6), 1416–1429. https://doi.org/10.1093/gbe/evy100

Freitag, M. (2017). Histone methylation by SET domain proteins in fungi. *Annual Review of Microbiology*, *71*(1), 413–439. https://doi.org/10.1146/annurev-micro-102215-095757

Fueyo, R., Judd, J., Feschotte, C., & Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. *Nature Reviews Molecular Cell Biology*, *23*(7), 481–497. https://doi.org/10.1038/s41580-022-00457-y

Galazka, J. M., & Freitag, M. (2014). Variability of chromosome structure in pathogenic fungi-of "ends and odds." *Current Opinion in Microbiology*, *20*, 19–26. https://doi.org/10.1016/j.mib.2014.04.002

Gladyshev, E. (2017). Repeat-Induced Point mutation and other genome defense mechanisms in fungi. In *The Fungal Kingdom* (pp. 687–699). https://doi.org/10.1128/9781555819583.ch33

Goodwin, S. B., M'Barek, S. Ben, Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., … Kema, G. H. J. (2011). Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics*, *7*(6). https://doi.org/10.1371/journal.pgen.1002070

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Grandaubert, J., Bhattacharyya, A., & Stukenbrock, E. H. (2015). RNA-seq-based gene annotation and comparative genomics of four fungal grass pathogens in the genus *Zymoseptoria* identify novel orphan genes and species-specific invasions of transposable elements. *G3;Genes|Genomes|Genetics*,*5*(7), 1323–1333. https://doi.org/10.1534/g3.115.017731

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075.

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., … White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, *9*(1), 1–22. https://doi.org/10.1186/gb-2008-9-1-r7

Habig, M., Kema, G. H. J., & Holtgrewe Stukenbrock, E. (2018). Meiotic drive of female-inherited supernumerary chromosomes in a pathogenic fungus. *ELife*, *7*, e40251. https://doi.org/10.7554/eLife.40251

Habig, M., Quade, J., & Stukenbrock, E. H. (2017). Forward genetics approach reveals host genotype-dependent importance of accessory chromosomes in the fungal wheat pathogen *Zymoseptoria tritici*. *MBio*, *8*(6), 1–16. https://doi.org/10.1128/mBio.01919-17

Habig, M., Stukenbrock, E.H. (2020). Origin, function, and transmission of accessory chromosomes. In: Benz, J.P., Schipper, K. (eds) Genetics and Biotechnology. The Mycota, vol 2. Springer, Cham. https://doi.org/10.1007/978-3-030-49924-2_2

Hartmann, F. E., McDonald, B. A., & Croll, D. (2018). Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Molecular Ecology*, *41*(0), 0–2. https://doi.org/10.1111/mec.14711

Haueisen, J., Möller, M., Eschenbrenner, C. J., Grandaubert, J., Seybold, H., Adamiak, H., & Stukenbrock, E. H. (2019). Highly flexible infection programs in a specialized wheat pathogen. *Ecology and Evolution*, *9*(1), 275–294. https://doi.org/10.1002/ece3.4724

He, C., Rusu, A. G., Poplawski, A. M., Irwin, J. A. G., & Manners, J. M. (1998). Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*. *Genetics*, *150*(4), 1459–1466. https://doi.org/10.1093/genetics/150.4.1459

Hedges, D. J., & Deininger, P. L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *616*(1), 46–59. https://doi.org/https://doi.org/10.1016/j.mrfmmm.2006.11.021

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., … Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*(5), 767–769. https://doi.org/10.1093/bioinformatics/btv661

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Jin, Y., & Hammell, M. (2018). Analysis of RNA-Seq data using TEtranscripts. In Y. Wang & M. Sun (Eds.), *Transcriptome Data Analysis: Methods and Protocols* (pp. 153–167). https://doi.org/10.1007/978-1-4939-7710-9_11

Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, *31*(22), 3593–3599. https://doi.org/10.1093/bioinformatics/btv422

Jones, D. A. B., Rozano, L., Debler, J. W., Mancera, R. L., Moolhuijzen, P. M., & Hane, J. K. (2021). An automated and combinative method for the predictive ranking of candidate effector proteins of fungal plant pathogens. *Scientific Reports*, *11*(1), 1–13. https://doi.org/10.1038/s41598-021-99363-0

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, *338*(5), 1027–1036. https://doi.org/https://doi.org/10.1016/j.jmb.2004.03.016

Kellner, R., Bhattacharyya, A., Poppe, S., Hsu, T. Y., Brem, R. B., & Stukenbrock, E. H. (2014). Expression profiling of the wheat pathogen *Zymoseptoria tritici* reveals genomic patterns of transcription and host-specific regulatory programs. *Genome Biology and Evolution*, *6*(6), 1353–1365. https://doi.org/10.1093/gbe/evu101

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Komluski, J., Stukenbrock, E. H., & Habig, M. (2022). Non-Mendelian transmission of accessory chromosomes in fungi. *Chromosome Research*, *30*(2–3), 241–253. https://doi.org/10.1007/s10577-022-09691-8

Kristianingsih, R., & MacLean, D. (2021). Accurate plant pathogen effector protein classification ab initio with deepredeff: an ensemble of convolutional neural networks. *BMC Bioinformatics*, *22*(1), 1–22. https://doi.org/10.1186/s12859-021-04293-3

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., … Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., … Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, *9*(8), 1–10. https://doi.org/10.1371/journal.pcbi.1003118

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics*, *12*(1), 124. https://doi.org/10.1186/1471-2105-12-124

Lechner, M., Hernandez-Rosales, M., Doerr, D., Wieseke, N., Thévenin, A., Stoye, J., … Stadler, P. F. (2014). Orthology detection combining clustering and synteny for very large datasets. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0105015

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Linde, C. C., Zhan, J., & McDonald, B. A. (2002). Population Structure of *Mycosphaerella graminicola*: From lesions to continents. *Phytopathology*, *92*(9), 946–955. https://doi.org/10.1094/phyto.2002.92.9.946

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., … Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, *66*(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Lorrain, C., Feurtey, A., Möller, M., Haueisen, J., & Stukenbrock, E. (2021). Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. *G3 Genes|Genomes|Genetics*, *11*(4). https://doi.org/10.1093/g3journal/jkab068

Ma, L.-J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., … Rep, M. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. https://doi.org/10.1038/nature08850

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., … Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(April), 276–279. https://doi.org/10.1093/nar/gkac240

McDonald, B. A., & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, *40*(1), 349–379.

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

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), 553–560. https://doi.org/10.1038/nature06008

Möller, M., Habig, M., Freitag, M., & Stukenbrock, E. H. (2018). Extraordinary genome instability and widespread chromosome rearrangements during vegetative growth. *Genetics*, *210*(2), 517–529. https://doi.org/10.1534/genetics.118.301050

Möller, M., Habig, M., Lorrain, C., Feurtey, A., Haueisen, J., Fagundes, W. C., … Stukenbrock, E. H. (2021). Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. *PLoS Genetics*, *17*(3), 1–27. https://doi.org/10.1371/journal.pgen.1009448

Möller, M., Schotanus, K., Soyer, J. L., Haueisen, J., Happ, K., Stralucke, M., … Stukenbrock, E. H. (2019). Destabilization of chromosome structure by histone H3 lysine 27 methylation. In *PLoS genetics* (Vol. 15). https://doi.org/10.1371/journal.pgen.1008093

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, *15*(12), 756–771. https://doi.org/10.1038/nrmicro.2017.76

Nguyen, A. H., & Bachtrog, D. (2021). Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male *Drosophila* with a young Y chromosome. *PLoS Genetics*, *17*(4 April 2021), 1–27. https://doi.org/10.1371/journal.pgen.1009438

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., … Grigoriev, I. V. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathogens*, *8*(12). https://doi.org/10.1371/journal.ppat.1003037

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32(2), 292-294. https://doi.org/10.1093/bioinformatics/btv566

Padeken, J., Zeller, P., & Gasser, S. M. (2015). Repeat DNA in genome organization and stability. *Current Opinion in Genetics & Development*, *31*, 12–19.

Peters, A. H. F. M., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., … Jenuwein, T. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, *107*(3), 323–337. https://doi.org/10.1016/S0092-8674(01)00542-6

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Plissonneau, C., Stürchler, A., & Croll, D. (2016). The evolution of orphan regions in genomes of a fungal pathogen of wheat. *MBio*, *7*(5). https://doi.org/10.1128/mBio.01231-16

Quaedvlieg, W., Kema, G. H. J., Groenewald, J. Z., Verkley, G. J. M., Seifbarghi, S., Razavi, M., … Crous, P. W. (2011). *Zymoseptoria* gen. nov.: A new genus to accommodate *Septoria*-like species occurring on graminicolous hosts. *Persoonia: Molecular Phylogeny and Evolution of Fungi*, *26*, 57–69. https://doi.org/10.3767/003158511X571841

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., & Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Computational Biology*, *1*(2), 0166–0175. https://doi.org/10.1371/journal.pcbi.0010022

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, *10*(6), 417–430. https://doi.org/10.1038/nrmicro2790

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., … Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/nar/gkw257

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., … Howlett, B. J. (2011). Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communications*, *2*, 202. https://doi.org/10.1038/ncomms1189

Savojardo, C., Martelli, P. L., Fariselli, P., & Casadio, R. (2018). DeepSig: Deep learning improves signal peptide detection in proteins. *Bioinformatics*, *34*(10), 1690–1696. https://doi.org/10.1093/bioinformatics/btx818

Schotanus, K., Soyer, J. L., Connolly, L. R., Grandaubert, J., Happel, P., Smith, K. M., … Stukenbrock, E. H. (2015). Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. *Epigenetics and Chromatin*, *8*(1), 1–18. https://doi.org/10.1186/s13072-015-0033-5

Shi-Kunne, X., Jové, R. D. P., Depotter, J. R. L., Ebert, M. K., Seidl, M. F., & Thomma, B. P. H. J. (2019). *In silico* prediction and characterization of secondary metabolite clusters in the plant pathogenic fungus *Verticillium dahliae*. *FEMS Microbiology Letters*, *366*(7), 1–11. https://doi.org/10.1093/femsle/fnz081

Sonnhammer, E. L., von Heijne, G., & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, *6*, 175–182.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, *98*(3), 503–517.

Soyer, J. L., Möller, M., Schotanus, K., Connolly, L. R., Galazka, J. M., Freitag, M., & Stukenbrock, E. H. (2015). Chromatin analyses of *Zymoseptoria tritici*: Methods for chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). *Fungal Genetics and Biology*, *79*(0), 63–70. https://doi.org/10.1016/j.fgb.2015.03.006

Sperschneider, J., Catanzariti, A. M., Deboer, K., Petre, B., Gardiner, D. M., Singh, K. B., … Taylor, J. M. (2017). LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports*, *7*(February), 1–14. https://doi.org/10.1038/srep44598

Sperschneider, J., & Dodds, P. N. (2022). EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Molecular Plant-Microbe Interactions*, *35*(2), 146–156. https://doi.org/10.1094/MPMI-08-21-0201-R

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology*, *19*(9), 2094–2110. https://doi.org/10.1111/mpp.12682

Sperschneider, J., Dodds, P. N., Singh, K. B., & Taylor, J. M. (2018). ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytologist*, *217*(4), 1764–1778. https://doi.org/10.1111/nph.14946

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., … Taylor, J. M. (2016). EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *New Phytologist*, *210*(2), 743–761. https://doi.org/10.1111/nph.13794

Studt, L., Rösler, S. M., Burkhardt, I., Arndt, B., Freitag, M., Humpf, H. U., … Tudzynski, B. (2016). Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*. *Environmental Microbiology*, *18*(11), 4037–4054. https://doi.org/10.1111/1462-2920.13427

Stukenbrock, E. H., Christiansen, F. B., Hansen, T. T., Dutheil, J. Y., & Schierup, M. H. (2012). Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences*, *109*(27), 10954–10959. https://doi.org/10.1073/pnas.1201403109

Stukenbrock, E. H. (2016a). Hybridization speeds up the emergence and evolution of a new pathogen species. *Nature Genetics*, *48*(2), 113–115. https://doi.org/10.1038/ng.3494

Stukenbrock, E. H. (2016b). The role of hybridization in the evolution and emergence of new fungal plant pathogens. *Phytopathology*, *106*(2), 104–112. https://doi.org/10.1094/PHYTO-08-15-0184-RVW

Stukenbrock, E. H., Banke, S., Javan-Nikkhah, M., & McDonald, B. A. (2007). Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, *24*(2), 398–411. https://doi.org/10.1093/molbev/msl169

Stukenbrock, E. H., Jørgensen, F. G., Zala, M., Hansen, T. T., McDonald, B. A., & Schierup, M. H. (2010). Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. *PLoS Genetics*, *6*(12), 1–13. https://doi.org/10.1371/journal.pgen.1001189

Stukenbrock, E. H., Quaedvlieg, W., Javan-Nikhah, M., Zala, M., Crous, P. W., & McDonald, B. A. (2012). *Zymoseptoria ardabiliae* and *Z. pseudotritici*, two progenitor species of the *Septoria tritici* leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella graminicola*). *Mycologia*, *104*(6), 1397–1407. https://doi.org/10.3852/11-374

Stukenbrock, E. H, Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M., … Schierup, M. H. (2011). The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research*, *21*(12), 2157–2166. https://doi.org/10.1101/gr.118851.110

Stukenbrock, E. H, & Dutheil, J. Y. (2018). Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. *Genetics*, *208*(3), 1209–1229. https://doi.org/10.1534/genetics.117.300502

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., & Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research*, *18*(12), 1979–1990. https://doi.org/10.1101/gr.081612.108

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., … Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*. https://doi.org/10.1038/s41587-021-01156-3

Torriani, S. F. F., Melichar, J. P. E., Mills, C., Pain, N., Sierotzki, H., & Courbot, M. (2015). *Zymoseptoria tritici*: A major threat to wheat production, integrated approaches to control. *Fungal Genetics and Biology*, *79*, 8–12. https://doi.org/https://doi.org/10.1016/j.fgb.2015.04.010

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., … Hammond-Kosack, K. E. (2022). PHI-base in 2022: A multi-species phenotype database for Pathogen-Host Interactions. *Nucleic Acids Research*, *50*(D1), D837–D847. https://doi.org/10.1093/nar/gkab1037

van der Does, H. C., Fokkens, L., Yang, A., Schmidt, S. M., Langereis, L., Lukasiewicz, J. M., … Rep, M. (2016). Transcription factors encoded on core and accessory chromosomes of *Fusarium oxysporum* induce expression of effector genes. *PLoS Genetics*, *12*(11), 1–37. https://doi.org/10.1371/journal.pgen.1006401

van der Does, H. C., & Rep, M. (2017). Adaptation to the host environment by plant-pathogenic fungi. *Annual Review of Phytopathology*, *55*(1), 427–450. https://doi.org/10.1146/annurev-phyto-080516-035551

Van Wyk, S., Harrison, C. H., Wingfield, B. D., De Vos, L., Van Der Merwe, N. A., & Steenkamp, E. T. (2019). The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ*, *2019*(8), 1–18. https://doi.org/10.7717/peerj.7447

Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B. J. C., & Rep, M. (2016). Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environmental Microbiology*, *18*(11), 3702–3713. https://doi.org/10.1111/1462-2920.13281

Wei, K. H. C., Gibilisco, L., & Bachtrog, D. (2020). Epigenetic conflict on a degenerating Y chromosome increases mutational burden in *Drosophila* males. *Nature Communications*, *11*(1), 1–9. https://doi.org/10.1038/s41467-020-19134-9

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., … Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. Retrieved from https://www.nature.com/articles/nrg2165.pdf?origin=ppub

Williams, A. H., Sharma, M., Thatcher, L. F., Azam, S., Hane, J. K., Sperschneider, J., … Singh, K. B. (2016). Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of candidate effectors. *BMC Genomics*, *17*(1). https://doi.org/10.1186/s12864-016-2486-8

Wittenberg, A. H. J., van der Lee, T. A. J., M'Barek, S. Ben, Ware, S. B., Goodwin, S. B., Kilian, A., … Schouten, H. J. (2009). Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen *Mycosphaerella graminicola*. *PLoS ONE*, *4*(6). https://doi.org/10.1371/journal.pone.0005863

Wu, B., Macielog, A. I., & Hao, W. (2017). Origin and spread of spliceosomal introns: Insights from the fungal clade *Zymoseptoria*. *Genome Biology and Evolution*, *9*(10), 2658–2667. https://doi.org/10.1093/gbe/evx211

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1), 134.

Zeller, P., Padeken, J., Van Schendel, R., Kalck, V., Tijsterman, M., & Gasser, S. M. (2016). Histone H3K9 methylation is dispensable for *Caenorhabditis elegans* development but suppresses RNA:DNA hybrid-associated repeat instability. *Nature Genetics*, *48*(11), 1385–1395. https://doi.org/10.1038/ng.3672

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, *328*(5980), 916–919.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., … Yin, Y. (2018). DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *46*(W1), W95–W101. https://doi.org/10.1093/nar/gky418

Zhao, Z., Liu, H., Wang, C., & Xu, J. (2014). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, *15*(1), 6. https://doi.org/10.1186/1471-2164-15-6

Supplementary Material

Supplementary Tables

All supplementary tables in ".xlsx" format are deposited on the supplementary USB key.

Table S1. Summary metrics of Zt469 and Za100 genome assemblies, gene and TE annotations.

Assembly metrices were accessed with the software Quast (Gurevich, Saveliev, Vyahhi, & Tesler, 2013).

Table S2. Conditions and thresholds used for gene categories classification.

In order to classify gene models into categories involved in host-pathogen interactions as e.g. candidates effectors and CAZymes, we used individual tools thresholds recommended by the Predector software (Jones et al., 2021). "Small Secreted Proteins (SSPs)" category also followed the criteria described previously (Grandaubert et al., 2015) and Antismash v.6.0 (fungal version; Blin et al., 2021) was used to detected biosynthetic gene clusters (BGCs). Filtering based on these conditions and thresholds was performed independently for each gene category.

Table S3. Summary of gene models, functional annotations and gene expression *in vitro* and *in planta* during four infection stages in Zt469.

Functional annotations were based on several prediction tools as Predector (Jones et al., 2021), eggnog-mapper (Huerta-Cepas et al., 2017) and Antismash v.6.0 (fungal version; Blin et al., 2021). Gene expressions *in vitro* and *in planta* are reported as the mean expression between triplicates in log2(TPM+1).

Table S4. Summary of gene models, functional annotations and gene expression *in vitro* in Za100.

Functional annotations were based on several prediction tools as Predector (Jones et al., 2021), eggnog-mapper (Huerta-Cepas et al., 2017) and Antismash v.6.0 (fungal version; Blin et al., 2021). Gene expression *in vitro* is reported as the mean expression between triplicates in log2(TPM+1).

Table S5. Summary of gene models per category in Zt469 and Za100.

We classified the gene models into categories based on the conditions and thresholds described in Supplementary Table S2.

Table S6. List of primers and expected amplicon sizes used in this study.

Table S7. Summary metrics of RNA-seq data in number of reads and percentage of *Zymoseptoria* genome coverage *in vitro* and *in planta*.

Table S8. Average expression of genes and Transposable Elements (TEs) per unitig in Zt469 during *in vitro* growth and during *in planta* infection stages.

Gene and TE expression *in vitro* and *in planta* were calculated as the mean expression between triplicates in TPM. Averages per unitig and per condition were then calculated as the mean expression of genes and TEs in TPM+1 divided by the total number of these features in each unitig and finally reported as log2(TPM+1). Except for unitig 9, only unitigs syntenic to the *Z. tritici* IPO323 reference genome are shown (unitigs 0 to 20).

Table S9. Average expression of genes and Transposable Elements (TEs) per genome compartment in Zt469 during *in vitro* growth and during *in planta* infection stages.

Gene and TE expression *in vitro* and *in planta* were calculated as the mean expression between triplicates in TPM. Averages per compartment and per condition were then calculated as the mean expression of genes and TEs in TPM+1 divided by the total number of these features in each compartment and finally reported as log2(TPM+1). Except for unitig 9, only unitigs syntenic to the *Z. tritici* IPO323 reference genome are considered (unitigs 0 to 20). Core compartment is composed from unitig 0 to unitig 17 and accessory compartment refers to unitigs 9, 18, 19 and 20.

Supplementary Figures

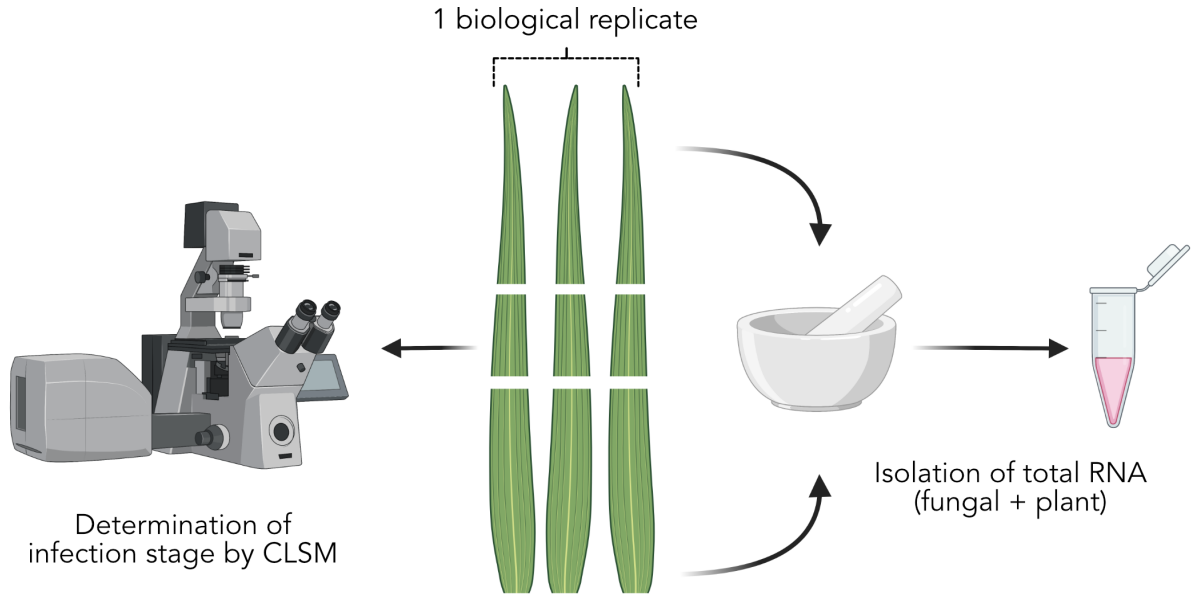


Figure S1. Schematic illustration representing the generation of stage-specific transcriptomes based on confocal microscopy analyses. Central sections (1-2cm) of *Z. tritici*-infected *Aegilops cylindrica* leaves from three independent plants (representing one biological replicate) were stained and analyzed by confocal laser-scanning microscopy in a previous study (Chapter II). The remaining infected leaf material was pooled and ground in liquid nitrogen for total RNA extraction. We chose RNA samples for sequencing based on the morphological infection stage observed in the central leaf section by microscopy in a single-blinded procedure (Chapter II). Illustrations were designed on BioRender.com

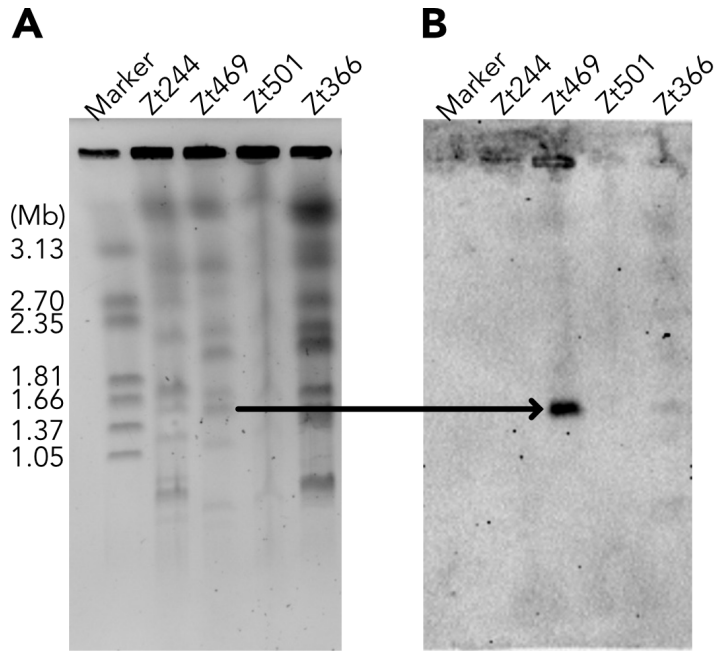


Figure S2. Presence of unitig 9 in Zt469 is confirmed by PFGE and Southern blot analyses. To validate the presence of unitig 9 at the expected assembly size in Zt469, we performed Pulsed-Field Gel Electrophoresis (PFGE) **(A)** followed by Southern blot analyses **(B)** in non-protoplast plugs of this isolate and in the reference *Z. tritici* isolate IPO323 (strain Zt244); in the wheat-infecting *Z. tritici* isolate Zt10 (strain Zt366; Stukenbrock et al. 2007); and in the *Aegilops*-infecting *Z. tritici* isolate Zt501 (Chaper II). A southern blot probe of 1.5 kb was designed to hybridize specifically in unitig 9 of Zt469 (Supplementary Table S6). Probe hybridization was detected at the expected unitig 9 size (~1.6Mb; **B**). Chromosomal DNA of *Hansenula wingei* (Bio-Rad, Munich, Germany) was used as a standard size marker.

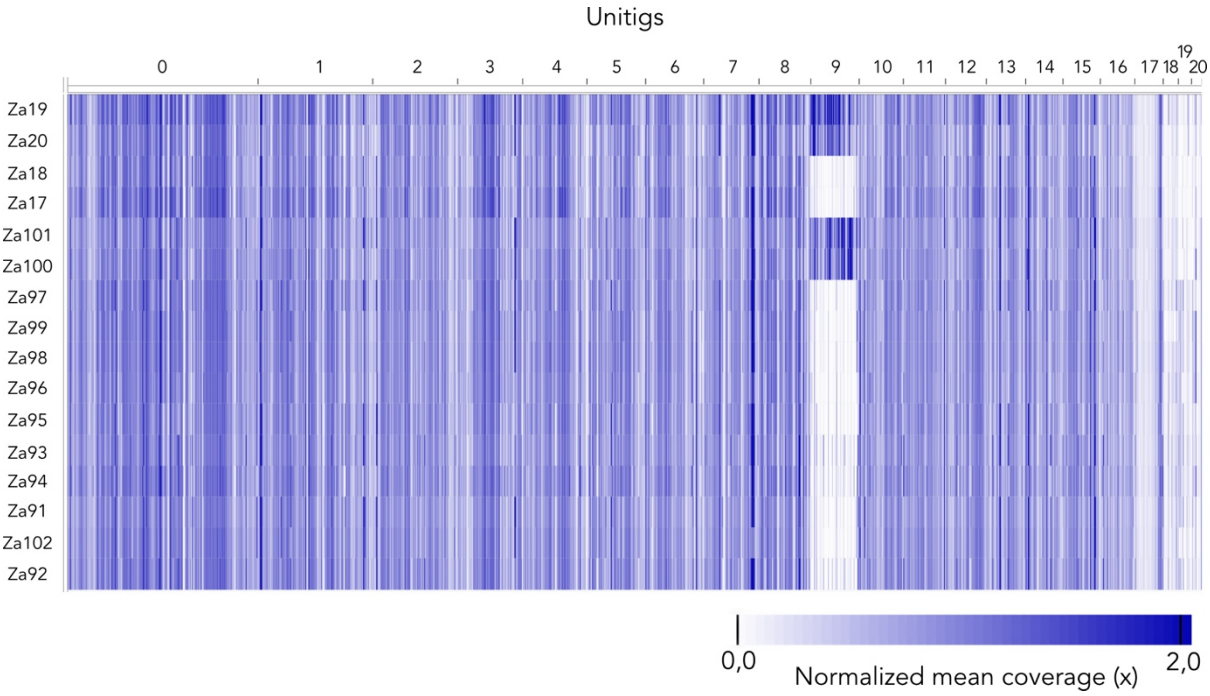


Figure S3. Read mapping to Zt469 genome confirms synteny and PAV of unitig 9 among *Z. ardabiliae* isolates. Chromosome presence-absence variation (PAV) was analyzed by read mapping to the *Z. tritici* Zt469 PacBio genome assembly. Heatmap represents normalized mean coverage of reads mapped to each position of the genome assembly. Except for unitig 9, only unitigs syntenic to the *Z. tritici* IPO323 reference genome are shown and sorted in descending order of length. Darker colors represent regions of higher coverage as e.g. repetitive elements or duplications.

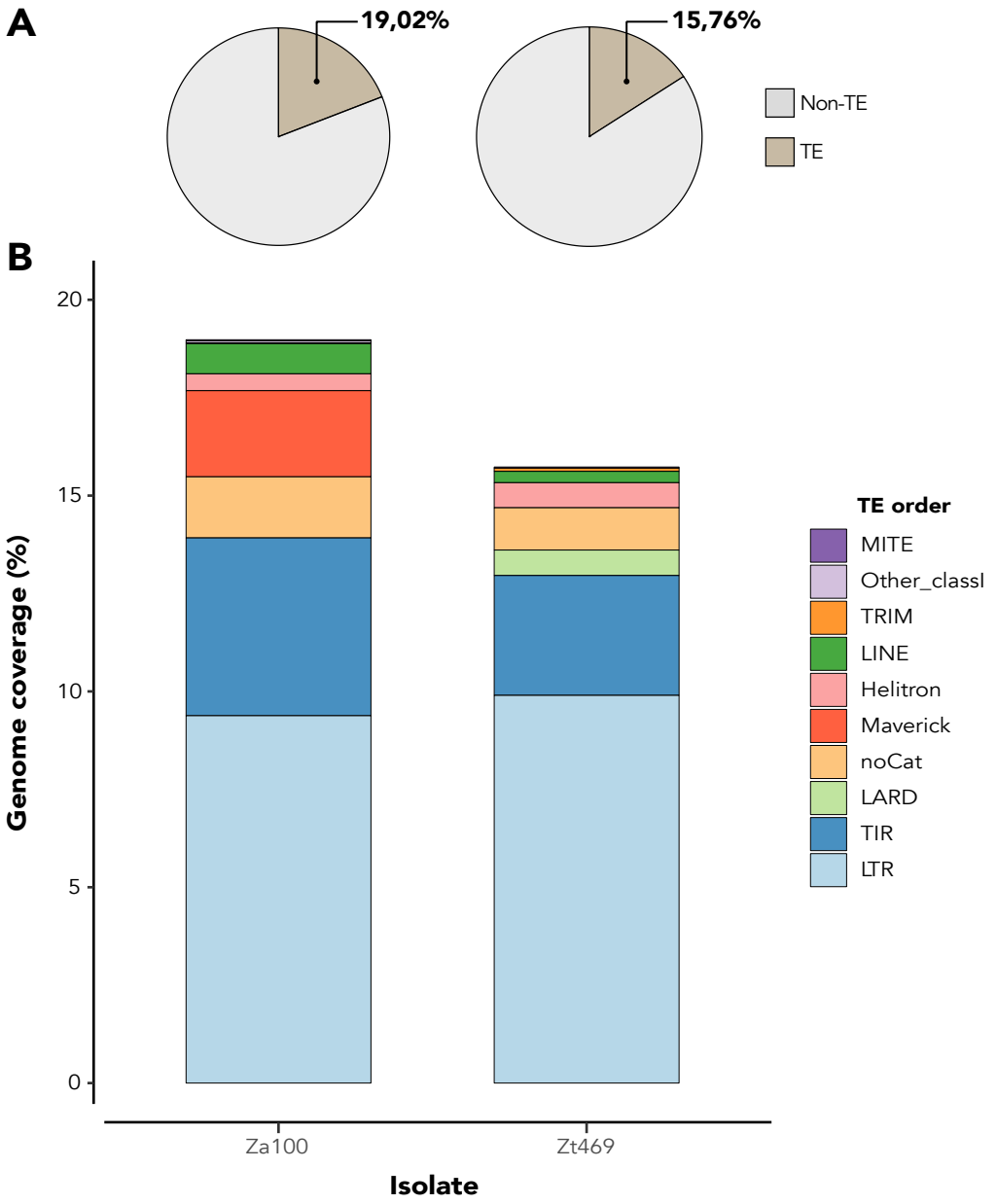


Figure S4. Transposable element (TE) content in Za100 and Zt469 genomes. **(A)** Total percentage of genome covered by TEs in Za100 (left) and Zt469 (right). **(B)** Stacked bar plot showing the TE content (%) per genome. Colors represent TE order coverage with retrotransposons (LTR, LINE, TRIM, LARD and other class I orders) and DNA transposons (TIR, MITE, Helitron and Maverick). TEs that could not have been classified after REPET (Flutre et al. 2011) annotation are indicated as "no category" ("noCat").

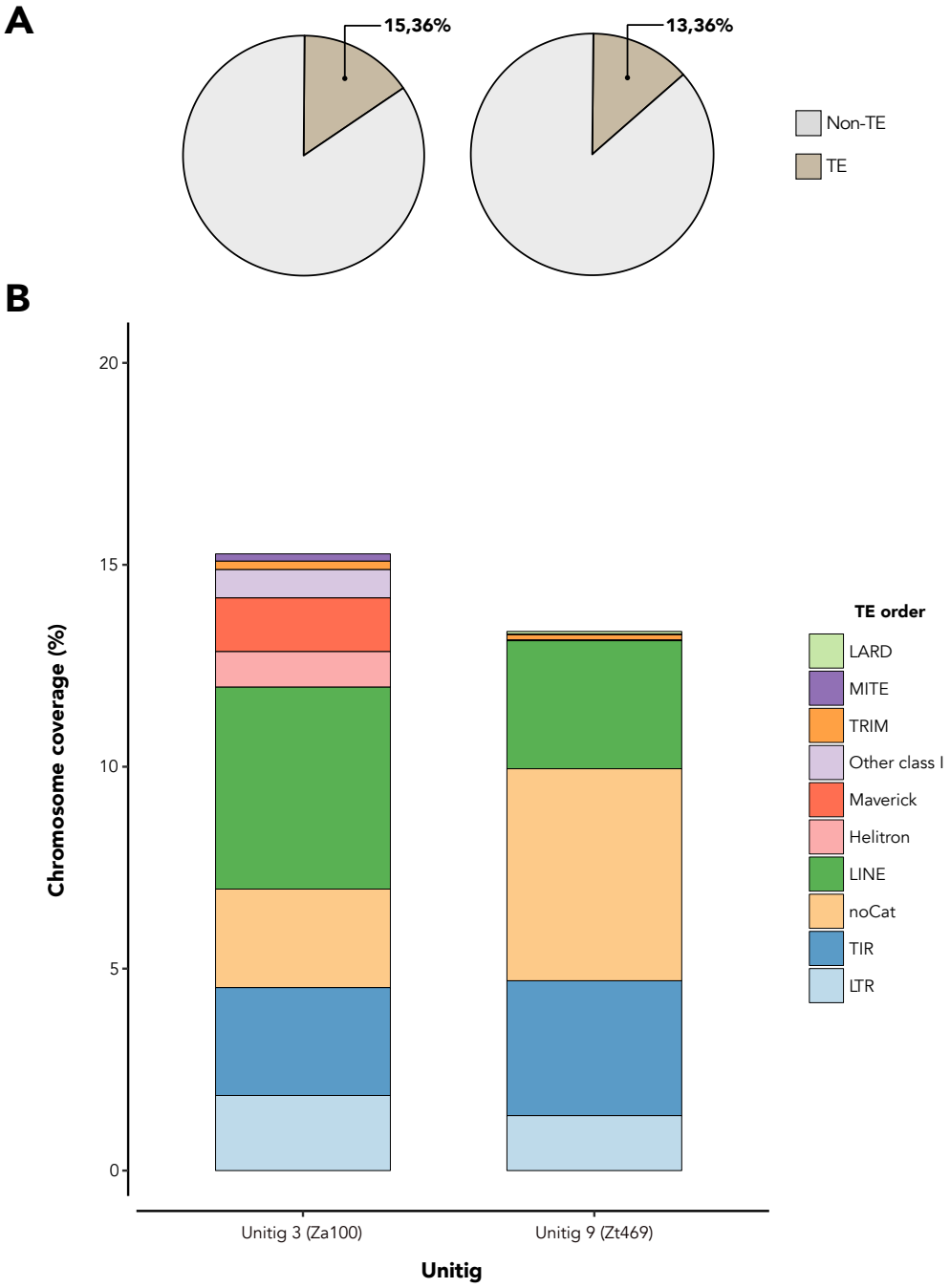


Figure S5. Transposable element (TE) content in the syntenic unitig 3 in Za100 and unitig 9 in Zt469. **(A)** Total percentage of unitig covered by TEs in unitig 3 (left) and unitig 9 (right). **(B)** Stacked bar plot showing the TE content (%) per unitig. Colors represent TE order coverage with retrotransposons (LTR, LINE, TRIM, LARD and other class I orders) and DNA transposons (TIR, MITE, Helitron and Maverick). TEs that could not have been classified after REPET (Flutre et al. 2011) annotation are indicated as "no category" ("noCat").

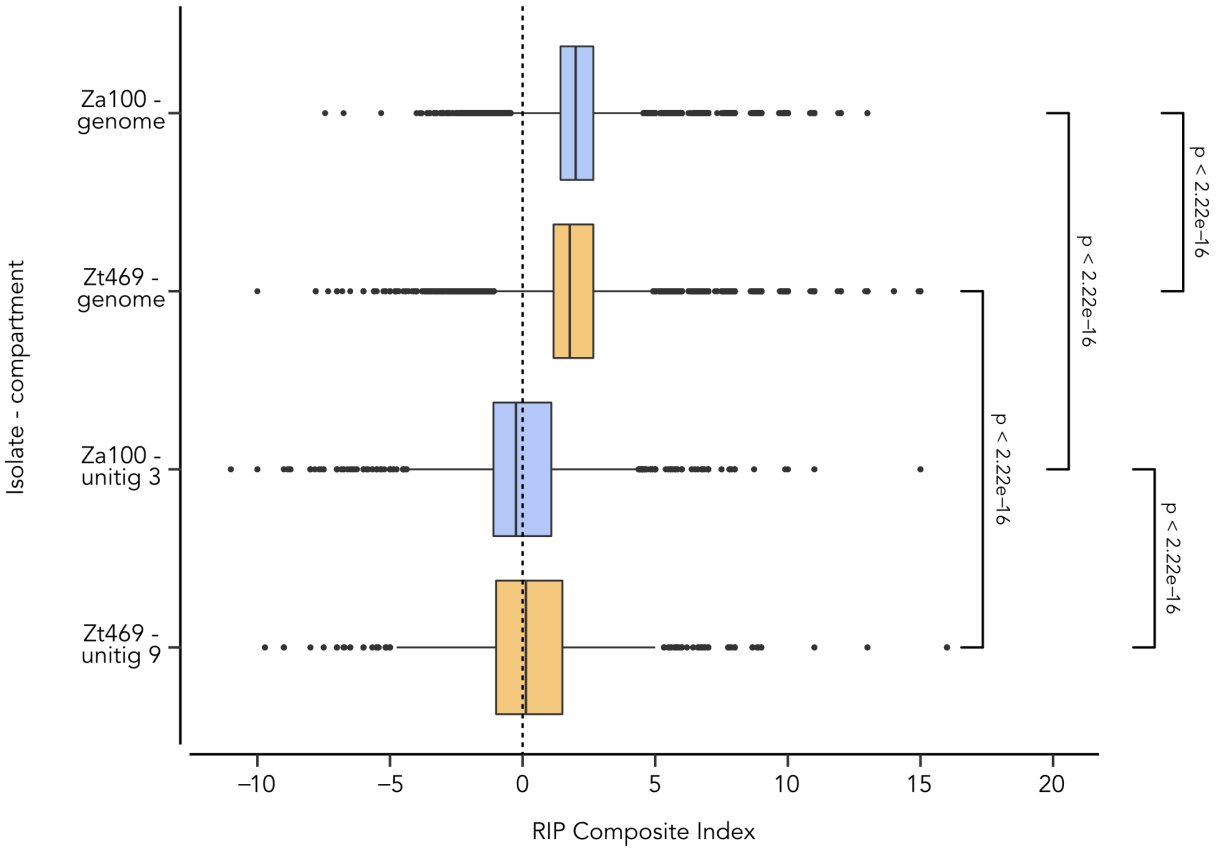


Figure S6. Repeat-Induced Point (RIP) mutation signatures in TE copies of Za100 and Zt469 genomes. Distribution of RIP composite indices calculated using 50bp sliding windows per TE copy genome-wide (top two boxes) and per TE copies present on the syntenic unitigs (bottom two boxes) for *Z. ardabiliae* Za100 (blue color) and *Z. tritici* Zt469 isolates (orange color). Vertical dashed line represents the threshold (0) above which composite indices values indicate RIP signature. P-values were calculated using pairwise Wilcoxon rank sum tests.

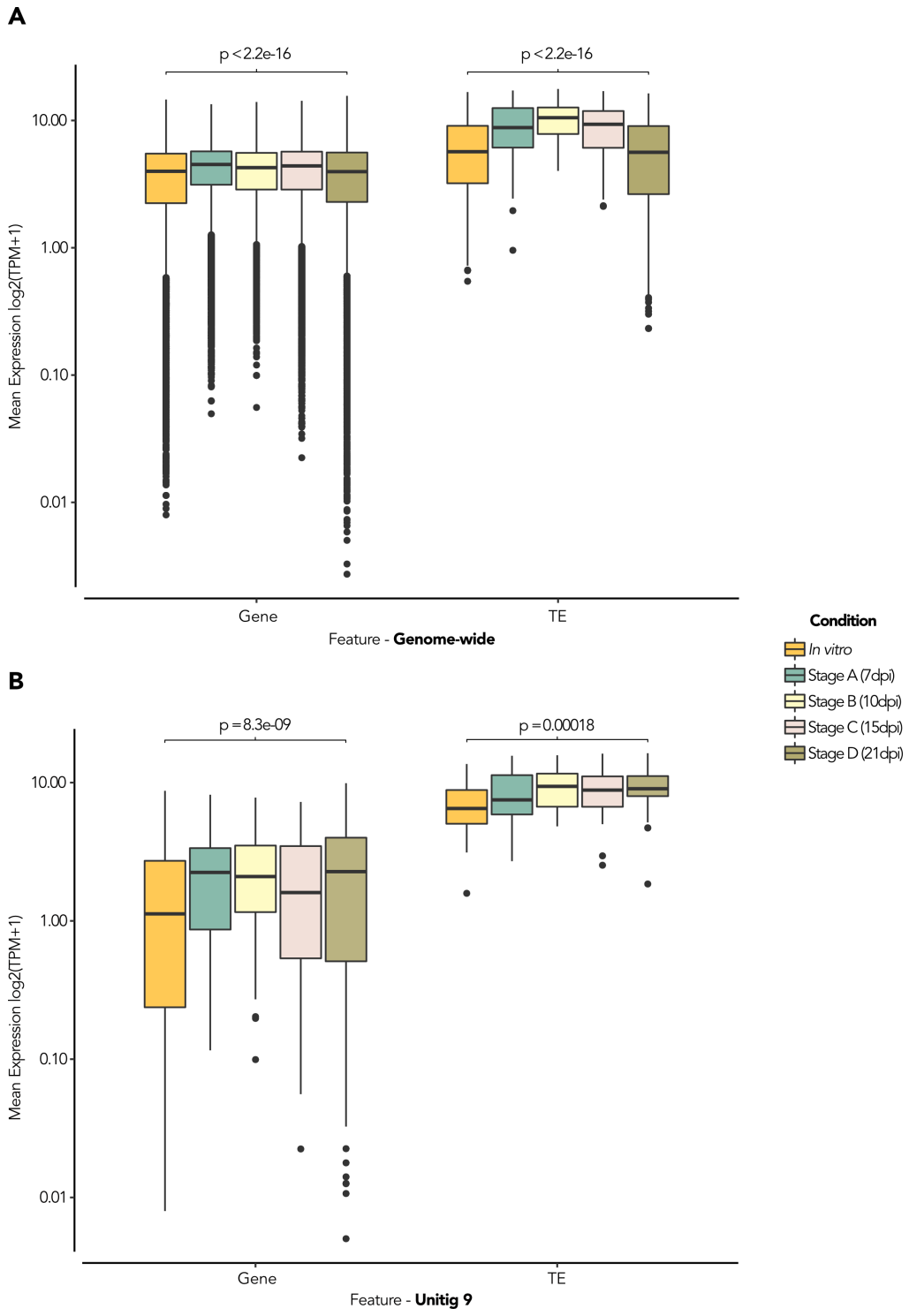
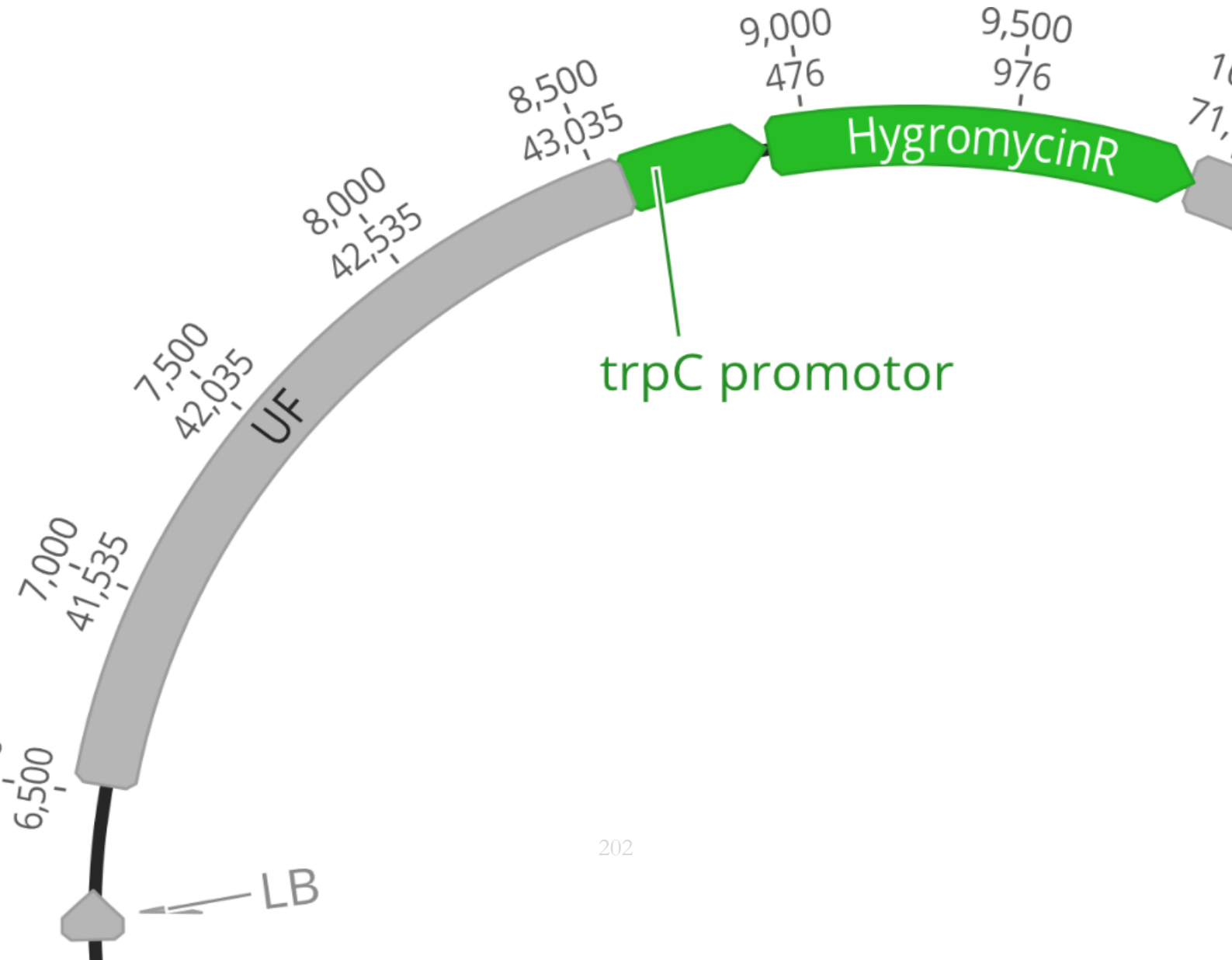


Figure S7. Genes and TEs show different levels of expression in Zt469 during *in vitro* growth and *in planta* infection. Boxplots representing the mean expression levels in log2(TPM+1) for genes and TEs in Zt469 distributed genome-wide **(A)** and in unitig 9 **(B)** at different infection stages *in planta* and *in vitro*. P-values were calculated using Kruskall-wallis tests within each feature and between different conditions.

Chapter IV

Functional genetics of candidate genes under host-driven selection



Chapter IV

Functional genetics of candidate genes under host-driven selection

Research paper manuscript

Authors: **Wagner C. Fagundes**, Janine Haueisen, Frauke Caliebe, Fatemeh Salimi, Alireza Alizadeh and Eva H. Stukenbrock.

Abstract

Host adaptation can lead to strong footprints of selection in fungal plant pathogen genomes. Populations genomics analysis have the power to highlight genes and/or genomic regions that have been under selection and therefore elucidate the possible pathways of rapid adaptation of fungal plant pathogens to new hosts. In a previous study (Chapter II), we identified distinct genomic regions under selective sweep in a newly-characterized *Aegilops*- infecting *Zymoseptoria tritici* population. Among these regions, one unique region comprised three candidate effector genes and one gene encoding a candidate carbohydrate-degrading enzyme (CAZyme), and showed high divergence when compared to wheat-infecting *Z. tritici* isolates. Considering the importance of such gene categories in host-pathogen interactions, we hypothesized that one or more genes in this region play a role for the host adaptation of *Z. tritici* isolates to the wild grass *Aegilops* spp. Here we conduct comparative genomics and functional genetics analyses of this selective sweep locus with the aim to investigate its overall genomic architecture between *Aegilops*- and wheat-infecting *Z. tritici* isolates and to identify potential phenotypic fitness effects *in planta* and *in vitro*. We show that besides polymorphisms, genomic rearrangements and specific transcriptional gene regulations in this genomic region may have played a role in the adaptive evolution of *Aegilops*-infecting *Z. tritici* isolates. Using a reverse genetics approach, we also successfully delete one of the three candidate effector genes present in the selective sweep locus in a high virulent *Aegilops*-infecting *Z. tritici* isolate. We expect that the results generated in this study will contribute to an "end-to-end" analyses of host adaptation in *Z. tritici* species, ranging from the genomic characterization of genes putatively involved on host adaptation to the functional validation of these genes during *in vitro* growth and *in planta* infection.

Introduction

Host adaptation imposes a strong selective pressure on fungal plant pathogen genomes (Giraud, Gladieux, & Gavrilets, 2010; Giraud, Refrégier, Le Gac, de Vienne, & Hood, 2008; Gladieux et al., 2011). In the course of evolution, hosts evolve defense-related mechanisms to recognize pathogen genotypes (Jones & Dangl, 2006). Pathogens in turn also evolve virulence mechanisms to infect and to escape host recognition and downstream host defense responses (Lo Presti et al., 2015; Stergiopoulos & de Wit, 2009). In many plant pathosystems, this constant and dynamic interaction between host defense responses and pathogen virulence mechanisms can be explained by the gene-for-gene (GFG) and "arms-race" coevolutionary scenarios. The GFG model of co-evolution states that for every resistance protein in the host there is a corresponding avirulence protein in the pathogen (Flor, 1971; Stukenbrock & McDonald, 2009). In the antagonistic co-evolutionary "arms-race" scenario, selective pressure on pathogens populations will favor mutations that enable these avirulence (Avr) proteins to escape recognition by the corresponding resistance (R) proteins. The host plants, in turn, will counteradapt by modifying these R proteins to detect the evolved Avr proteins (Tellier, Moreno-Gámez, & Stephan, 2014). Considering that host genotypes and even closely related host species can differ in the repertoire of defense-related genes, selective pressure on pathogens populations, and particularly on virulence-related genes, can be also highly heterogeneous and leave signatures of divergent selection in pathogens genomes (Lo Presti et al., 2015; Stukenbrock & McDonald, 2009; Stukenbrock, 2013).

Filamentous plant pathogens have evolved several mechanisms to infect hosts and evade the host immune system. Virulence-related genes in filamentous pathogens generally encode cysteine-rich, small secreted proteins (SSPs) with no sequence homology known as effectors (Lo Presti et al., 2015; Stergiopoulos & de Wit, 2009; van der Does & Rep, 2017). These small proteins are generally highly expressed during host infection and can modulate host cell structure and plant metabolism through several and diverse functions (Lo Presti et al., 2015; Stergiopoulos & de Wit, 2009; van der Does & Rep, 2017). Besides effectors, different classes of secreted proteins such as plant cell wall–degrading enzymes (PCWDEs) and other carbohydrate-degradation enzymes (CAZymes) as well as non-proteinaceous and proteinaceous secondary metabolites (e.g. toxins) also play an important role in host-pathogen interactions (Bari & Jones, 2009; Kubicek, Starr, & Glass, 2014; Lo Presti et al., 2015). In distinct filamentous pathogen species, divergent selection has been found to affect polymorphisms in effector genes (Dai, Jia, Correll, Wang, & Wang, 2010; Raffaele et al., 2010; Schmidt et al., 2015; van de Wouw et al., 2010); toxins (Stukenbrock & McDonald, 2007); secondary metabolites (Lind et al., 2017; Schardl et al., 2013) and genes encoding PCWDEs

(Brunner, Torriani, Croll, Stukenbrock, & McDonald, 2013). Although these studies have shown the impact of virulence gene mutations on the adaptive evolution of filamentous plant pathogens, the evolutionary mechanisms and the underlying genetics by which pathogen populations evade the host immune system are still poorly understood, particularly in wild (i.e. non-agricultural) pathosystems.

Genes or genomic regions affected by selection can be recognized by analyzing the distribution and frequency of genetic variants among individuals. Positive selection, in particular, can leave "selective sweep" footprints along the genome such as low levels of genetic diversity, high linkage disequilibrium, and an excess of low frequency alleles (Braverman, Hudson, Kaplan, Langley, & Stephan, 1995; Nielsen, 2005; Nielsen et al., 2005; Smith & Haigh, 1974). In host-pathogen interactions, particular interest has been given to identifying signatures of positive selection in pathogens genomes, which in turn may reveal functional traits that have been important for adaptive evolution and virulence (Aguileta, Refregier, Yockteng, Fournier, & Giraud, 2009; Biswas & Akey, 2006; Nielsen, 2005; Nielsen et al., 2005). In fungal plant pathogens, the identification of genomic regions under selection can pinpoint candidate genes involved in pathogenicity (e.g. effectors), fungicide resistance and host specialization, all of which are major goals in the study of adaptive evolution and development of more sustainable control strategies in agricultural settings (Boyd, Ridout, O'Sullivan, Leach, & Leung, 2013; Grünwald, McDonald, & Milgroom, 2016; Vleeshouwers & Oliver, 2014). Coupled with this, the functional characterization of the genetic elements that "stand out" during these selection analyses are essential to depict their role and to further validate their association to particular phenotypes (Grünwald et al., 2016). Selective sweep scans have been previously applied to fungal populations and identified several outlier regions containing genes potentially involved in host adaptation (Badouin et al., 2017) and adaptation to diverse environmental factors (Branco et al., 2017, 2015; Ellison et al., 2011; Hartmann, McDonald, & Croll, 2018; Mohd-Assaad, McDonald, & Croll, 2018). However, the functional role of many of these genes during fungal adaptation to specific hosts and ecological niches remains elusive.

In the fungal wheat pathogen *Zymoseptoria tritici* and closely related *Zymoseptoria* species, different population genetics studies have identified divergent selection in important traits for host adaptation (Grandaubert, Dutheil, & Stukenbrock, 2019; Hartmann et al., 2018; Stukenbrock et al., 2011). Previous coalescent analyses on *Z. tritici* indicated that this species originated from a wild grass-associated ancestor in the Fertile Crescent region in the Middle East around 11,000 years ago, coinciding with the domestication of its host wheat (Banke, Peschon, & McDonald, 2004;

Stukenbrock, Banke, Javan-Nikkhah, & McDonald, 2007). Since then, *Z. tritici* is considered to be a highly specialized and genetically diverse wheat pathogen while the closest known relatives to *Z. tritici*, *Z. pseudotritici* and *Z. ardabiliae* are found associated to different wild grasses in the Middle East (Stukenbrock et al., 2012; Stukenbrock et al., 2011). Several secreted and non-secreted proteins in *Z. tritici* were found to exhibit signatures of positive selection during divergence from the closest relatives *Z. pseudotritici* and *Z. ardabiliae*, suggesting that specialization to wheat has potentially involved the acquisition and fixation of adaptive substitutions in key genes playing a role in virulence and host-pathogen interactions (Stukenbrock et al., 2011). In a following study, Poppe and collaborators (2015) have functionally characterized four of these genes previously identified to have signatures of positive selection (Poppe, Dorsheimer, Happel, & Stukenbrock, 2015; Stukenbrock et al., 2011). Using a reverse genetics approach, the authors independently deleted the four genes in *Z. tritici* and confirmed a virulence-related role in three out of the four genes analyzed, corroborating the hypothesis that these positively selected genes are involved in host specialization and host-specific disease development in *Z. tritici* (Poppe et al., 2015).

In a recent population genomics study, we have identified distinct selective sweep genomic regions in a host-specific population of *Z. tritici* infecting wild grass species of the genus *Aegilops* (Chapter II). *Aegilops*-infecting *Z. tritici* isolates showed high divergence from a worldwide collection of wheat-infecting *Z. tritici* isolates and from closely related *Zymoseptoria* species and demography analyses further suggested a split between the wheat- and *Aegilops*-infecting lineages around 10,000 years ago (Chapter II). Characterization of the selective sweep regions revealed an unique region in chromosome 7 comprising three candidate effector (Zt09_chr_7_00299, Zt09_chr_7_00305 and Zt09_chr_7_00308) and one candidate carbohydrate-degrading enzyme (CAZyme; Zt09_chr_7_00296) gene in close proximity, besides peaks of high divergence when compared to wheat- infecting *Z. tritici* isolates (Chapter II). Considering the importance of such gene categories in host-pathogen interactions, we hypothesize this selective sweep region plays an important role on the adaptive evolution of *Z. tritici* isolates to *Aegilops* species.

In this study, we report functional genetics and comparative genomics analyses of this selective sweep locus with the aim to elucidate its genomic structure and potential role during host adaptation in *Aegilops*-infecting *Z. tritici* isolates. Using synteny and transcriptomic data analyses, we demonstrate that not only polymorphisms but also genomic rearrangements and host- and temporal-specific transcriptional gene regulations in this genomic region may have been under selection during the adaptive evolution of *Z. tritici* isolates infecting *Aegilops* species. We also adopt

a reverse genetics approach and demonstrate the successful deletion of one of the three candidate effector genes in a high virulent *Aegilops*-infecting *Z. tritici* isolate.

Material and Methods

Delimitation of selective sweep region

Considering the large size of the selective sweep region in chromosome 7 identified previously (approximately 62 kb in the *Z. tritici* IPO323 genome; Chapter II), we focused our analyses to a smaller region of approximately 34 kb within the selective sweep locus based on the coordinates of the reference *Z. tritici* IPO323 genome (Chapter II; Goodwin et al., 2011; Grandaubert, Bhattacharyya, & Stukenbrock, 2015). The reasoning behind this choice was to facilitate genomic deletions by homologous recombination in the wild-type *Z. tritici* isolate used in this study (see below). We selected this smaller region based on the high patterns of linkage disequilibrium (LD) in the *Aegilops*-infecting *Z. tritici* population identified previously (Chapter II) and to encompass the three candidate effectors and CAZyme-encoding genes. A schematic illustration of the selective sweep region analyzed is available at Supplementary Figure S1.

Strains and growth conditions

The *Z. tritici* isolate Zt495 was isolated from a leaf sample collected in Iran in 2018 identified as *A. cylindrica* or a closely related species (Chapter II). We chose this isolate for the functional genetics analyses considering its high virulence in *A. cylindrica* during greenhouse infection assays (Chapter II). The Zt495 isolate was inoculated from glycerol stocks onto YMS (yeast-malt-sucrose) medium and incubated at 18°C for 4-5 days prior use as previously described (Fagundes, Haueisen, & Stukenbrock, 2020). For the transformation process of fungal cells, all plasmids were maintained in *E. coli* Top10 cells (Invitrogen, Karlsruhe, Germany) and the *Agrobacterium tumefaciens* strain AGL1 was used for *Agrobacterium tumefaciens* mediated transformation (ATMT). Both bacterial strains were cultivated in dYT (double-yeast-tryptone) medium and grown at 28°C and 37°C conditions for *A. tumefaciens* and *E. coli*, respectively. We supplemented dYT media with 50 µg/ml Rifampicin (Sigma, Taufkirchen Germany) and 100 µg/ml Carbenicillin (Sigma, Taufkirchen, Germany) to maintain the plasmids already present in the AGL1 strain.

Generation of binary plasmids and ATMT of fungal cells

A set of six different plasmid constructs were designed targeting different portions of the selective sweep region in the *Z. tritici* isolate Zt495 (Supplementary Figure S2). This set of binary plasmids was designed to circumvent the challenges in deleting large genomic regions as the selective sweep region (approximately 34 kb) by homologous recombination in *Z. tritici*, and to specifically delete the three candidate effectors genes present in this region (Zt09_chr_7_00299, Zt09_chr_7_00305 and Zt09_chr_7_00308). Plasmid constructs and overlapping primer pairs were designed using Geneious v.2020.1.2 software (https://www.geneious.com/home/). The plasmid pES61, a derivate of the binary vector pNOV-ABCD (Bowler et al., 2010) and which contains a kanamycin resistance for selection of *E. coli* transformants, was used as backbone vector for all plasmid constructs. Transformed *E. coli* cells were selected on dYT media supplemented with 40 μg/ml kanamycin (Sigma, Taufkirchen, Germany). For targeted region deletions, and depending on the size of the regions to be deleted, we amplified two DNA fragments of 1-2 kb upstream and downstream of the selected deletion sites on Zt495 by PCR to serve as flaking regions and facilitate homologous recombination of the selective resistance marker cassettes with the recipient Zt495 genome (1 kb for pES296, pES297, and pES298; 2 kb for pES292, pES294 and pES305). To permit a possible generation of double and triple mutants at a later time, we amplified the ORFs of three different resistant markers for *Z. tritici* and assembled them in different plasmid constructs: the *hph* gene conferring resistance to hygromycin (for pES292, pES294 and pES296; Bowler et al., 2010); the *neo* gene providing resistance to geneticin (G418; for pES297 and pES305; Sidhu et al., 2015), and the *nat* gene which confers resistance to nourseothricin (for pES298; Mehrabi et al., 2015). For each construct, we assembled the DNA fragments up- and downstream of the region to be deleted with the resistance marker cassette and ligated them into the EcoRV-digested plasmid pES61 using Gibson assembly (Gibson et al., 2009).

For ATMT of the *Z. tritici* isolate Zt495, electro-competent *A. tumefaciens* AGL1 cells were transformed with the final plasmid constructs (Supplementary Figure S2) using standard procedures. Plasmid constructs were then introduced into Zt495 following an ATMT protocol previously described (Zwiers & De Waard, 2001). Transformed Zt495 colonies were visible on YMS plates supplemented with 150 μg/mL Hygromycin (Roth, Germany) two weeks after the ATMT procedure. Single-cell colonies were then propagated in YMS medium for further DNA extraction and confirmation of correct homologous recombination by PCR and Southern blot analyses (Southern, 1975). DNA extractions for these steps were performed following a standard phenol-chloroform extraction protocol (Sambrook & Russell, 2001). For confirmation of

transformed strains using PCRs, we amplified the resistant marker cassette and the endogenous locus using the outer primers of the deletion constructs. To generate the Southern blot probes, we amplified the resistance cassette and the downstream flank of the deletion construct using the PCR Digoxigenin (DIG) labeling mix (Roche, Mannheim, Germany) following manufacturer's instructions. List of all primers designed and used in this study can be found at Supplementary Table S1.

Genomic data and synteny analysis

We used previously assembled and annotated reference *Z. tritici* genomes to compare overall synteny of the analyzed selective sweep region. As a reference genome for *Aegilops*-infecting *Z. tritici*, the PacBio genome assembly of the isolate Zt469 was used along with a gene and TE annotation generated previously (Chapter III). The reference genome assembly of the isolate IPO323, published by Goodwin et al. (2011), served as a reference for wheat-infecting *Z. tritici*, including the annotations of genes and transposable elements (TEs) published by Grandaubert et al. (2015) and Feurtey et al. (2020) (Feurtey et al., 2020; Goodwin et al., 2011; Grandaubert et al., 2015). A draft genome assembly was generated for the *Aegilops*-infecting *Z. tritici* isolate Zt495 using SPAdes v. 3.14.1 (Bankevich et al., 2012) and whole-genome sequencing data generated previously (Chapter II). To this end, Illumina paired-end reads of 150bp were trimmed for adapter and sequencing quality using Trimmomatic v 0.39 (Bolger, Lohse, & Usadel, 2014) with the following parameters: LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:50. Quality-trimmed reads were then assembled in SPAdes v. 3.14.1 (Bankevich et al., 2012) using the following parameters: --careful and a kmer range (-k) of '21,33,55,77' as previously performed (Chapter II).

For the Zt495, genome coordinates of the genes in the selective sweep region were annotated by mapping the gene models of IPO323 against the draft genome assembly using nucleotide BLAST in Geneious v.2020.1.2 software (https://www.geneious.com/home/). For the following genes the annotation of Zt469 was used instead of the one for IPO323: Zt09_chr_7_00302, which covers the region containing the two genes Zt469_000007F_arrow_0296 and Zt469_000007F_arrow_0297 in Zt469 (Supplementary Figure S3); Zt469_000007F_arrow_0302, where the corresponding gene in IPO323 (Zt09_chr_7_00305) is 3 bp shorter due to a point mutation in the terminal stop codon (Supplementary Figure S4A); and Zt469_000007F_arrow_0300, which is absent in IPO323.

To identify putative functions, protein BLAST (blastp; Altschul, Gish, Miller, Myers, & Lipman, 1990; Camacho et al., 2009) searches against the National Center for Biotechnology Information (NCBI) protein database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) were performed using the amino acid sequences of the genes annotated in the selective sweep region in Zt495. We also used the SignalP (v. 6.0) software to predict if the analyzed proteins in the Zt495 and IPO323 genomes are secreted (Teufel et al., 2022). For all proteins predicted to be secreted by SignalP, EffectorP (version 3.0) was additionally used to predict if they are classified as candidate effectors (Sperschneider & Dodds, 2022).

Lastly, to compare the selective sweep region in Zt495 to the homologous regions in the IPO323 and Zt469 isolates, a synteny plot of this region was generated with the EasyFig software using default settings (Sullivan, Petty, & Beatson, 2011). We also aligned the nucleotide and protein sequences of each gene and manually recorded the number of aminoacid polymorphisms using Geneious v.2020.1.2 software (https://www.geneious.com/home/).

Gene expression analyses

We compared the expression of the genes localized at the selective sweep region in IPO323 and Zt469 genomes at different conditions. We use previously analyzed RNA-seq datasets of the *Z. tritici* isolates during *in vitro* growth on YMS medium and during *in planta* infection at four different infection stages (Haueisen et al., 2019; Chapter III). For comparison between the studies, the Zt469 RNA-seq data was reanalyzed and the expression of genes reported in FPKM (Fragments per Kilobase of Transcript per Million fragments mapped). To this end, the relative abundance of gene transcripts in FPKM was calculated using the Cuffnorm software from the Cufflinks package v.2.2.1 (Trapnell et al., 2013). Quality-trimming, masking and mapping of RNA-seq reads to Zt469 genome followed the pipeline described previously (Chapter III).

Results and Discussion

Proteins in the selection sweep region of *Aegilops*-infecting Zt495 isolate show no predicted function novelty

As a first step to characterize the selective sweep region, we asked if the proteins present at this locus on the *Aegilops*-infecting Zt495 had similar predicted functions than the ones described in the *Z. tritici* isolate IPO323 (Chapter II; Feurtey et al., 2020). This analysis was based on the

assumption that if mutations have occurred on the genes of this *Aegilops*-infecting isolate, potential new proteins and new functions could have emerged. To this end, blastp searches were conducted for each protein sequence of the selective sweep region in Zt495 against the NCBI protein database. For more than half of the analyzed proteins, no functions were predicted based on homology. For all proteins with potential functions, the best hits were from *Z. tritici* with very low e-values and high query coverages and percentages of identity (Table 1), except for Zt469_000007F_arrow_0296. The query coverage of this protein was rather low, probably due to this gene correspond to only the first part of Zt09_chr_7_00302 annotated in the reference gene models of IPO323 (see Methods and Supplementary Figure S3). Among the functions predicted, we observed a beta-glucosidase (CAZyme) for the protein encoded by Zt09_chr_7_00296 (Grandaubert et al., 2015); an antibiotic biosynthesis monooxygenase for Zt09_chr_7_00297 (Plissonneau, C., unpublished) and a protein kinase for Zt09_chr_7_00306 (Plissonneau, C. unpublished; Table 1).

To further elucidate potential functions of the proteins in the selective sweep region of Zt495, SignalP v6.0 (Teufel et al., 2022) and EffectorP v3.0 (Sperschneider & Dodds, 2022) were used to predict if they are potentially secreted and are classified as effectors, respectively. We also reanalyzed the protein sequences present in the selective sweep region in the reference IPO323 genome as a comparison and to confirm previous functional predictions (Feurtey et al., 2020; Grandaubert et al., 2015). We observed that all three proteins predicted to be secreted and classified as effectors in the *Z. tritici* IPO323 genome also showed to be secreted and act as effectors in the Zt495 genome with high confidence (> 99.9%, Supplementary Table S2). EffectorP v3.0 results also showed that candidate effectors encoded by Zt09_chr_7_00299 and Zt09_chr_7_00305 were predicted to be apoplastic effectors, while Zt09_chr_7_00308 was predicted to be cytoplasmic in both *Z. tritici* genome analyzed (Supplementary Table S2). The CAZyme protein encoded by the gene Zt09_chr_7_00296 was also predicted to be secreted in both Zt495 and IPO323 genomes (67% and 77%, respectively; Supplementary Table S2)

Altogether, these results show that the predicted protein functions of the selective sweep region in Zt495 were equivalent to the ones previously observed in the *Z. tritici* IPO323 isolate (Chapter II; Feurtey et al., 2020). Our assumption that mutations occurring on the genes of the selective sweep region could potentially change the protein functions does not hold true, at least through homology-based and functional prediction analyses.

Table 1. First protein BLAST hits for the genes in the selective sweep region of Zt495

	First BLAST hit					
Gene_ID	**Query coverage**	**Percentage identity**	**Species**	**Function**	**NCBI Accession**	**Reference**
Zt09_chr_7_00296	97%	98.72	*Z. tritici*	-	SMR57635.1	Plissonneau, C., direct submission (17/05/2017)
Zt09_chr_7_00297	100%	99.1	*Z. tritici*	antibiotic biosynthesis monooxygenase	SMR55262.1	Plissonneau, C., direct submission (17/05/2017)
Zt09_chr_7_00298	100%	99.45	*Z. tritici*	isocitrate/ isopropylmalate dehydrogenase	SMR57637.1	Plissonneau, C., direct submission (17/05/2017)
Zt09_chr_7_00299	100%	91.61	*Z. tritici*	probable M. graminicola specific protein	XP_003850686.1	Goodwin et al., 2011
Zt09_chr_7_00300	100%	99.48	*Z. tritici*	insulin-induced protein	SMQ52440.1	Plissonneau, C., direct submission (17/06/2016)
Zt09_chr_7_00301	100%	98.7	*Z. tritici*	contains WD40 domain	XP_003850687.1	Goodwin et al., 2011
Zt469_000007F_arrow_0296 *	72%	99.32	*Z. tritici*	-	SMQ52437.1	Plissonneau, C., direct submission (17/06/2016)
Zt469_000007F_arrow_0297 *	100%	91.23	*Z. tritici*	-	SMQ52438.1	Plissonneau, C., direct submission (17/06/2016)
Zt09_chr_7_00303	100%	96.86	*Z. tritici*	-	SMR55268.1	Plissonneau, C., direct submission (17/05/2017)
Zt09_chr_7_00304	100%	100	*Z. tritici*	-	SMY26081.1	Plissonneau, C., direct submission (06/10/2016)
Zt09_chr_7_00305	99%	95.71	*Z. tritici*	-	SMR55270.1	Plissonneau, C., direct submission (17/05/2017)
Zt09_chr_7_00306	100%	99.5	*Z. tritici*	protein kinase	SMQ52445.1	Plissonneau, C., direct submission (17/06/2016)
Zt09_chr_7_00307	92%	97.42	*Z. tritici*	IGR protein motif	XP_003850690.1	Goodwin et al., 2011
Zt09_chr_7_00308	66%	87.69	*Z. tritici*	-	SMQ52447.1	Plissonneau, C., direct submission (17/06/2016)
Zt469_000007F_arrow_0300 *	34%	98.17	*Z. tritici*	-	SMQ52441.1	Plissonneau, C., direct submission (17/06/2016)

Note: * Zt469_000007F_arrow_296 and Zt469_000007F_arrow_297 correspond to the first and second part of Zt09_chr_7_00302 (as annotated in IPO323), respectively (Supplementary Figure S3). Gene Zt469_000007F_arrow_0300 is only present in *Aegilops*-infecting *Z. tritici* isolates Zt495 and Zt469.

Selective sweep region in *Aegilops*-infecting *Z. tritici* shows extensive structural rearrangements and several aminoacid polymorphisms

To get insights into the genomic architecture of the selective sweep region between *Aegilops*- and wheat- infecting *Z. tritici* isolates, we analyze sequence similarities at the nucleotide and aminoacid level and performed synteny analyses using the Easyfig software (Sullivan et al., 2011). We chose three *Z. tritici* isolates to perform such analyses, namely Zt495, Zt469 and IPO323. The *Z. tritici* isolate IPO323 represents the reference genome of the species (Goodwin et al., 2011), and here also as reference genome for wheat-infecting *Z. tritici* isolates. The *Z. tritici* isolate Zt469 represents the PacBio reference genome of the *Aegilops*-infecting *Z. tritici* isolates (Chapter III) while Zt495 is the *Z. tritici* isolate used for the functional genetics analyses considering its high virulence in *Aegilops cylindrica* plants (Chapter II).

Overall, we observed a wide-range of gene rearrangements in the selective sweep region between the analyzed *Z. tritici* isolates Zt495, Zt469 and IPO323 (Figure 1). The *Aegilops*-infecting *Z. tritici* isolates Zt495 and Zt469 showed a conserved synteny between the genes, while more differences were observed between the Zt495 and the reference wheat-infecting IPO323 isolate (Figure 1). We also observed that the gene and TE (Transposable Element) content of the analyzed region was variable between the isolates. The gene Zt09_chr_7_00304 annotated in IPO323 is absent in the two analyzed *Aegilops*- infecting isolates, spanning a region of around 1.4 kb in length that is deleted in these isolates (gene highlighted in dark pink in Figure 1). In contrast, the gene Zt469_000007F_arrow_0300 annotated in Zt469, which is also present in Zt495, is absent in IPO323, as it lies in a region spanning approximately 1.7 kb inserted in the genome of the two *Aegilops*-infecting *Z. tritici* isolates (genes highlighted in orange in Figure 1). In Zt469 and Zt495, there are two genes annotated where only one gene is annotated in the corresponding region of IPO323, namely Zt469_000007F_arrow_0296 and Zt469_000007F_arrow_0297 in Zt469 and Zt09_chr_7_00302 in IPO323 (genes highlighted in light pink in Figure 1; Supplementary Figure S3).

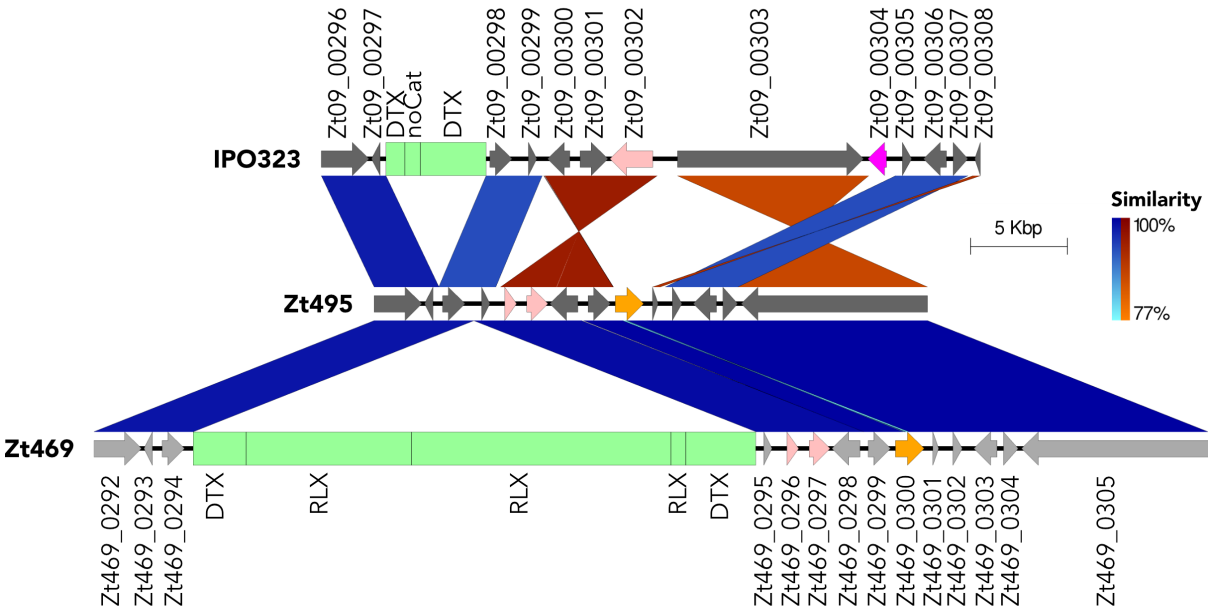


Figure 1. Selective sweep region shows extensive rearrangements between *Aegilops*- and wheat-infecting *Z. tritici* isolates. Plot generated by Easyfig (Sullivan et al., 2011) showing overall synteny and nucleotide similarity of the selective sweep region between the wheat-infecting *Z. tritici* isolate IPO323 and the two *Aegilops*- infecting *Z. tritici* isolates Zt495 and Zt469. Genes are indicated by arrows. Gene names are abbreviated for better visualization and refer to the genes "Zt09_chr_7_XXX" for *Z. tritici* IPO323 gene annotation and "Zt469_000007_arrow_XXX" for *Z. tritici* Zt469 gene annotation. TEs are indicated by green boxes and follow the classification described in Wicker et al. (2007). BLAST similarity is represented by gradient ribbons connecting the genes, in which red-orange colors were used when inversion is present and blue colors when not. Darker ribbon color shades indicate higher nucleotide similarity between genes. DTX = Class II transposons (TIR); RLX = Class I retrotransposons (LTR); noCat = no category assigned.

Regarding TEs, we observed three elements inserted between the genes Zt09_chr_7_00297 and Zt09_chr_7_00298 in the reference wheat-infecting *Z. tritici* isolate IPO323 (Figure 1), in which two belong to the DNA transposon (Class II) TIR order (DTX) and one with no category (noCat) following Wicker's TE classification (Wicker et al., 2007). This TE-rich region of approximately 5 kb in the IPO323 isolate is not present in the two *Aegilops*-infecting *Z. tritici* isolates (Figure 1). For Zt469, TEs were observed between the genes Zt469_000007F_arrow_0294 and Zt469_000007F_arrow_0296 (Figure 1). This TE-enriched region spanned a wide length of approximately 29 kb, contains two TIR transposons (DTX) and three LTR retrotransposons (Class I; RLX) and it is absent in both *Aegilops*-infecting Zt495 and wheat-infecting IPO323 *Z. tritici* genomes (Figure 1).

We also observed several inversions and transpositions between IPO323 and the *Aegilops*- infecting *Z. tritici* isolates. The region comprising the genes Zt09_chr_7_00300, Zt09_chr_7_00301 and Zt09_chr_7_00302 is inverted in Zt495, while the region with the genes Zt09_chr_7_00305,

Zt09_chr_7_00306 and Zt09_chr_7_00307 is translocated in Zt495 in comparison to IPO323 (Figure 1). The two Zt495 genomic regions containing the genes Zt09_chr_7_00303 and Zt09_chr_7_00308 are translocated and inverted in comparison to IPO323 genome (Figure 1).

In order to complement the synteny analyses and visualize the mutations occurring in the genes present in the selective sweep region, we analyzed the sequence similarity of the genes at the aminoacid level between the three *Z. tritici* isolates. To this end, we performed protein sequence alignments of each gene individually and manually recorded the number of aminoacid polymorphisms between the isolates. For all protein sequences analyzed, at least one aminoacid polymorphism was observed between the *Aegilops*-infecting *Z. tritici* isolates Zt469 and Zt495 and the wheat-infecting *Z. tritici* isolate IPO323 (Table 2). Only three out of the 14 protein sequences analyzed showed aminoacid polymorphisms between the two *Aegilops*-infecting *Z. tritici* isolates Zt469 and Zt495 (Table 2). For the two candidate effector proteins Zt09_chr_7_00305 and Zt09_chr_7_00308, aminoacid sequences showed to be identical in the *Aegilops*-infecting isolates Zt495 and Zt469 while polymorphisms were observed when compared to IPO323 (Table 1). Interestingly, for the candidate effector gene Zt09_chr_7_00299, the aminoacid sequence in Zt469 and IPO323 were very similar with only one aminoacid differing, while Zt495 exhibited several polymorphisms to the other two isolates (Table 2). In the protein sequence encoded by the Zt09_chr_7_00303 gene, we observed five additional aminoacids in the *Aegilops*-infecting *Z. tritici* isolates Zt469 and Zt495 when compared to the homologous protein sequence of IPO323 (Table 2). We also observed that in the protein sequence encoded by the candidate effector gene Zt09_chr_7_00305 in the two *Aegilops*-infecting isolates, one aminoacid was added at the end of the sequence when compared to the wheat-infecting *Z. tritici* isolate IPO323 (Table 2). Analyses on the nucleotide sequence alignment of these gene revealed that the additional aminoacid in the protein sequence encoded by Zt09_chr_7_00305 is potentially due to a point mutation leading to a stop codon shift in the two *Aegilops*-infecting *Z. tritici* isolates (Supplementary Figure S4A), while the additional five aminoacids observed in the protein sequence encoded by Zt09_chr_7_00303 in the Zt469 and Zt495 genomes are potentially due to two insertions of 12 bp and 3 bp each in comparison to the *Z. tritici* IPO323 gene sequence (Supplementary Figure S4B).

Table 2. Protein sequence comparison for the genes in the selective sweep region between Zt495, Zt469 and IPO323 *Z. tritici* isolates.

Gene (in IPO323 / in Zt469) **	Number of mutations *			Protein length in aminoacids (in IPO323)	Notes
	Zt495 vs IPO323	Zt495 vs Zt469	Zt469 vs IPO323		
Zt09_00296 / Zt469_0292	8	16	14	800	
Zt09_00297 / Zt469_0293	1	0	1	111	
Zt09_00298 / Zt469_0294	5	0	5	366	
Zt09_00299 / Zt469_0295	12	13	1	143	
Zt09_00300 / Zt469_0299	2	0	2	383	
Zt09_00301 / Zt469_0298	6	0	6	462	
Zt09_00302 / Zt469_0296 ***	-	0	-	203 ****	
Zt09_00302 / Zt469_0297 ***	-	0	-	308 ****	
Zt09_00303 / Zt469_0305	398	4	400	3117	Protein is 5 aa longer in Zt495 and Zt469 than in IPO323
Zt09_00305 / Zt469_0302	8	0	8	164	Protein is 1 aa longer in Zt495 and Zt469 than in IPO323 (1 aa added at end)
Zt09_00306 / Zt469_0303	4	0	4	397	
Zt09_00307 / Zt469_0304	6	0	6	252	
Zt09_00308 / Zt469_0301	10	0	10	98	
NA / Zt469_0300	-	0	-	319 ****	Gene absent in IPO323

Notes: * Number of amino acid mutations between each pair of isolates is given, including point mutations, deletions and insertions. **** Gene names abbreviated. All names for IPO323 refer to "Zt09_chr_7_XXX", while for Zt469 all gene names refer to "Zt469_000007F_arrow_XXX". Candidate effectors genes are highlighted in bold. "NA" = genes not annotated in the respective genome. ******* Zt469_000007F_arrow_0296 and Zt469_000007F_arrow_0297 correspond to the first and second part of Zt09_chr_7_00302 (as annotated in IPO323), respectively. Zt09_chr_7_00302 was not analyzed for IPO323 because it probably consists of two genes and the exact positions of the exons were uncertain. ******** Protein length in Zt469 is given.

Hosts impose strong selective pressure on pathogens genomes, particularly at virulence-related genes. The co-evolutionary fight between host resistance (R) genes and avirulence genes in pathogens genomes favors mutations, deletions or acquisition of new virulence genes to overcome host immune recognition (Jones & Dangl, 2006; Lo Presti et al., 2015; Sánchez-Vallet et al., 2018; Stergiopoulos & de Wit, 2009). Among the virulence genes, genes encoding effectors are generally the most polymorphic and fast-evolving in pathogens genomes (van der Does & Rep, 2017; Win et al., 2012). Previous studies in different fungal and oomycete plant pathogens have associated sequence diversification or complete deletion of effector genes to increased levels of virulence on host genotypes carrying corresponding R-genes (Ghanbarnia et al., 2015; Jiang & Tyler, 2012; Meile et al., 2018; Schmidt et al., 2015). In the blast fungus *Magnaporthe oryzae*, it has been recently shown that a single aspartate to asparagine polymorphism on the APikL2 effector gene could expand the binding spectrum of this otherwise conserved effector to host-target proteins, which suggests a mechanism of adaptation in the APikL effector family to distinct hosts (Bentham et al., 2021).

In light of these findings, our analyses have also shown that genes present in the selective sweep region, particularly the candidate effector genes, showed several aminoacid polymorphisms between and within host-specific *Z. tritici* isolates. We also observed a large number of polymorphisms on the CAZyme-encoding gene Zt09_chr_7_00296 between *Aegilops*-infecting and wheat-infecting *Z. tritici* isolates as well as between the two *Aegilops*-infecting *Z. tritici* isolates analyzed. The differential degrees of polymorphisms observed on these genes, particularly on Zt09_chr_7_00296 and Zt09_chr_7_00299, further suggests that these pathogenesis-related proteins have potentially diversified and adapted to target specific host species and genotypes, as previously demonstrated for the APikL effector family in *Magnaporthe oryzae* (Bentham et al., 2021). Additional analyses on molecular evolution and protein structure using accurate prediction tools as AlphaFold (Jumper et al., 2021) are necessary to investigate the adaptive evolution of these *Z. tritici* proteins to *Aegilops* and wheat host immune machinery.

Even though polymorphisms and signatures of recent positive selection were present on the genes of the selective sweep region analyzed (Chapter II), previous analyses of non-neutral selection using MK (McDonald-Kreitman) tests (McDonald & Kreitman, 1991) did not detect statistically significant shifts on the rate of synonymous versus non-synonymous polymorphisms and substitutions (Chapter II). We argued that this finding is possibly due to the recent signatures of positive selection detected by selective sweep scans and the inability of MK tests to distinguish between past and present selection (Chapter II; McDonald & Kreitman, 1991; Nielsen, 2005;

Pavlidis & Alachiotis, 2017). While these factors can explain the results observed, there is also a growing body of literature suggesting that not only mutations but also genomic rearrangements may be under selection and play a role in the adaptive evolution of eukaryotes and prokaryotes (Brandis & Hughes, 2020; Cao, Brandis, Huseby, & Hughes, 2022; Coghlan, Eichler, Oliver, Paterson, & Stein, 2005; Nair et al., 2007; Yeaman, 2013).

We observed several rearrangements in the selective sweep region of chromosome 7 when comparing the *Aegilops*- and wheat-infecting *Z. tritici* isolates. Chromosomal rearrangements are known to promote genomic novelty in eukaryotic genomes, particularly in fungal plant pathogens (Coghlan et al., 2005; Möller & Stukenbrock, 2017; Raffaele & Kamoun, 2012). Gene gains through duplications and neofunctionalizations as well as gene losses through non-homologous recombination or pseudogenization (Ohno, 1970; Tang & Amon, 2013; Zhang, 2003) facilitate the generation of gene content variation among individuals and have a direct impact on the adaptive evolution of fungal plant pathogens to different ecological niches and hosts (Dong, Raffaele, & Kamoun, 2015; Möller & Stukenbrock, 2017). Moreover, the presence of Transposable Elements (TEs) adjacent to genes can also influence genomic architecture plasticity and further regulate gene expression (Castanera et al., 2016). Gene expansions and contractions have also been associated to host specialization in diverse fungal plant pathogens (Gladieux et al., 2014; Ohm et al., 2012), as already exemplified by studies in *Colletotrichum* species (Baroncelli et al., 2016), in the smut fungus *Melanopsichium pennsylvanicum* (Sharma, Mishra, Runge, & Thines, 2014) and in the grass pathogen *Magnorpothe oryzae* (Couch et al., 2005; Yoshida et al., 2016). In *Z. tritici*, previous reports have also shown that rearrangements and the presence of orphan regions among isolates may also play a role in host specialization and the adaptive evolution of the pathogen in distinct host genotypes (Hartmann, Sánchez-Vallet, McDonald, & Croll, 2017; Meile et al., 2018; Plissonneau, Hartmann, & Croll, 2018). Hence, our results suggest that not only polymorphisms but also structural rearrangements may have contributed to sequence diversification and underlying selection in the selective sweep region in *Aegilops*-infecting *Z. tritici* isolates. Further population and comparative genomics analyses on the genes of this selective sweep region using additional *Z. tritici* and closely related *Zymoseptoria* populations are necessary to extensively characterize this region and potentially indicate when such rearrangements emerged in the course of evolution.

Candidate effector and CAZyme-encoding genes in the selective sweep region of *Aegilops*- and wheat-infecting *Z. tritici* isolates show different levels of expression

Next, to gain further insights into the potential role of the analyzed candidate effector and CAZyme-encoding genes for pathogenicity in *Aegilops cylindrica* and wheat hosts, their expression in Zt469 and IPO323 genomes was evaluated during *in vitro* growth and during *in planta* infection. We used previously generated RNA-seq datasets of these *Z. tritici* isolates during growth on YMS medium and during infection at four different stages in *Aegilops cylindrica* (for Zt469; Chapter III) and wheat (for IPO323; Haueisen et al., 2019). In the *Z. tritici* IPO323 genome, analyses of the normalized abundance of transcripts in FPKM (Fragments per Kilobase of Transcript per Million fragments mapped) revealed that the three candidate effectors were found to have higher levels of expression during *in planta* infection when compared to *in vitro* growth (Figure 2). The candidate effector gene Zt09_chr_7_00299 had a constant increase of expression with progressing of infection from 4 days-post inoculation (dpi) to 20 dpi as previously observed (Figure 2; Haueisen et al. 2019). The other two candidate effector genes Zt09_chr_7_00305 and Zt09_chr_7_00308 were not expressed during *in vitro* growth on YMS medium, while *in planta* they were both expressed at much lower levels than Zt09_chr_7_00299, especially Zt09_chr_7_00305 (Figure 2; Haueisen et al. 2019). Interestingly, these two candidate effector genes did not have expression detected at the infection stage "B" on wheat, suggesting a temporal-specific regulation during infection (Figure 2). The CAZyme-encoding gene Zt09_chr_7_00296 also showed to have higher levels of expression during wheat infection than during growth *in vitro*, particularly during the biotrophic infection stages "A" (Figure 2; Haueisen et al. 2019).

In the *Aegilops*-infecting *Z. tritici* isolate Zt469 however, we observed some distinct expression patterns (Figure 2). For the CAZyme-encoding gene Zt09_chr_7_00296, the homologous gene in Zt469 (Zt469_000007F_arrow_0292) showed the highest levels of expression at stage "B" during *A. cylindrica* infection, differing from IPO323 in wheat (Figure 2). For the candidate effector gene Zt09_chr_7_00299, the homologous candidate effector gene in Zt469 (Zt469_000007F_arrow_0295) also showed higher levels of expression *in planta* than *in vitro*, however with a peak of expression at stage "B" and a constant reduction of expression at later stages ("C" and "D"; Figure 2). The candidate effector gene Zt469_000007F_arrow_0301 in Zt469, which is homologous to the candidate effector Zt09_chr_7_00305 in IPO323, only showed expression during *in vitro* growth and stage "D" while for the candidate effector gene Zt09_chr_7_00308, the homologous gene in Zt469 (Zt469_000007F_arrow_0302) only showed expression during *in vitro* growth (Figure 2). Altogether, these results suggest that the homologous

candidate effectors and CAZyme-encoding genes analyzed may have a host-specific and also temporal-specific expression regulation.

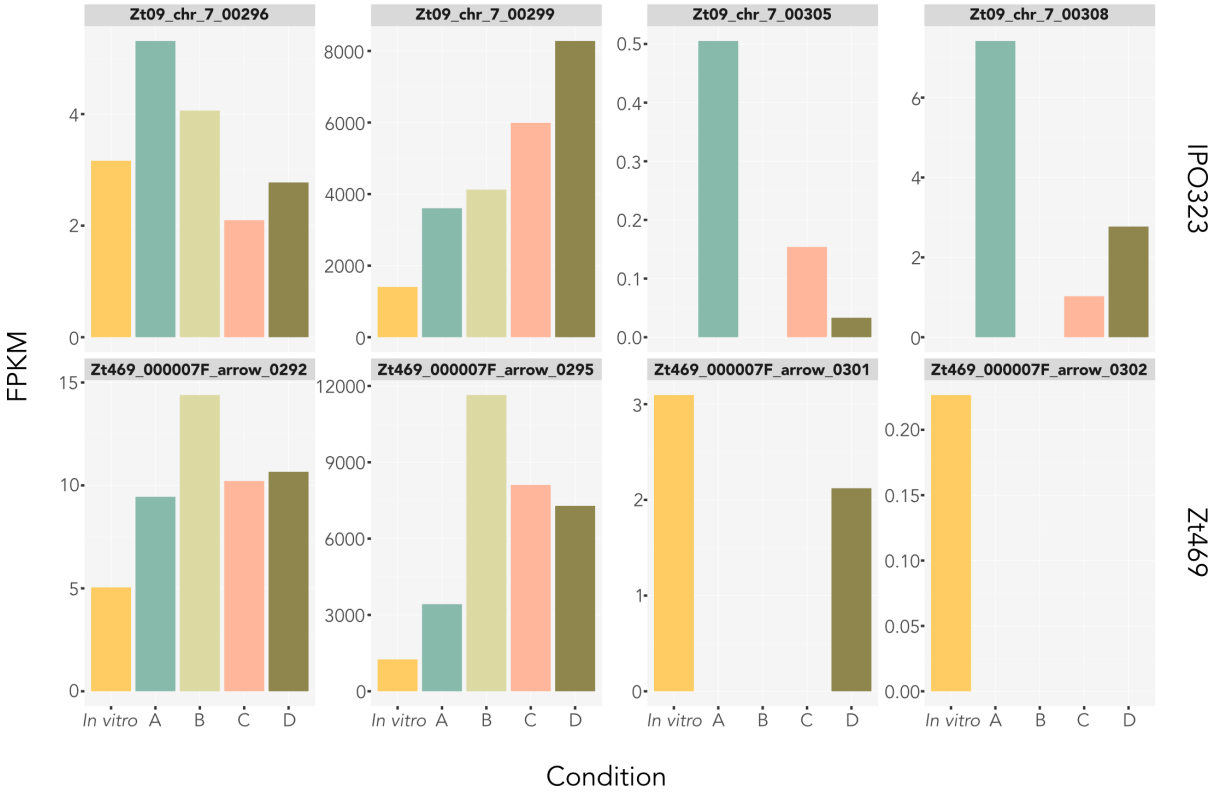


Figure 2. Genes involved in host-pathogen interactions show different levels of expression in Zt469 and IPO323 *Z.tritici* isolates during *in vitro* and *in planta* conditions. Normalized abundance of transcripts in FPKM for the CAZyme-encoding gene (first column from left to right) and the three candidate effector genes in the reference wheat-infecting isolate IPO323 (top row) and their homologs in the *Aegilops*-infecting isolate Zt469 (bottom row). Transcriptome data was obtained from previous studies (Chapter III; Haueisen et al. 2019) and show relative expression of genes during *in vitro* growth (in YMS media) and at four infection stages from early biotrophy (stage A) to late necrotrophy (stage D).

Considering that different hosts can be seen as distinct niches due to the biochemical and morphological properties as R-gene repertoire and cell wall composition, host-driven selective pressure on corresponding fungal pathogen populations can be also highly heterogeneous, even within pathogen species occurring in sympatry (Giraud et al., 2008; Lo Presti et al., 2015; McCoy, 2003; Stukenbrock & McDonald, 2008). Therefore, we hypothesize that host specificity may have influenced not only the polymorphisms and genomic architecture in the selective sweep region analyzed but have also impacted the expression of the genes present at this locus. Functional validation of these genes through reverse genetics approaches are necessary to access their role and potential transcription regulation during *in planta* infection.

Generation of *Aegilops*-infecting *Z. tritici* deletion mutants

At last, to functionally analyzed the impact of the selective sweep region on *Z. tritici* virulence and growth, we designed six independent deletion constructs targeting different portions of the selective sweep region in the *Aegilops*-infecting *Z. tritici* isolate Zt495 (Supplementary Figures S1 and S2). We chose the *Z. tritici* isolate Zt495 as a background for such deletions given its high virulence in *Aegilops cylindrica* plants (Chapter II). Plasmids were constructed using Gibson assembly (Gibson et al., 2009) and mutants were generated using an *Agrobacterium tumefaciens* mediated transformation (ATMT) approach as previously performed (Bowler et al., 2010; Marshall et al., 2011; Poppe et al., 2015; Zwiers & De Waard, 2001). The correct integration of the selective markers cassette by homologous recombination was verified by PCRs and Southern blot analyses (see Methods).

Overall, potentially correct mutants could be generated for the plasmids pES292 (deletion of the complete selective sweep region; Supplementary Figure S2), pES294 (deletion of the first half of the selective sweep region, Supplementary Figure S2), and pES296 (deletion of the candidate effector Zt09_chr_7_00299, Supplementary Figure S2). For pES292, we could identify 29 potentially correct independent *Z. tritici* mutants by PCR. For pES294, PCRs identified 12 potentially correct *Z. tritici* mutants. For both pES292 and pES294 mutants, Southern blot analyses to verify correct integration of selective markers have not yet been possible. However, for the pES296 construct, we could confirm the correct deletion of the candidate effector gene Zt09_chr_7_00299 not only by PCR but also by Southern blot analyses in three independent *Z. tritici* Zt495 mutants (Figure 3B). Although we could not detect the band at expected size in the wild type Zt495 digested DNA used as a control (4895 bp), possibly due to low DNA concentration (Figure 3A), the *Z. tritici* mutants #68, #86 and #98 showed detection at the expected band sizes (2217 bp and 3630 bp; Figure 3B) and therefore can be considered correct independent mutants ready to be used in downstream analyses. The other two potentially correct pES296-transformed mutants #83 and #99 showed a faint additional band at approximately 2 kb, which could indicate ectopic integrations and will be therefore disregarded (Figure 3B).

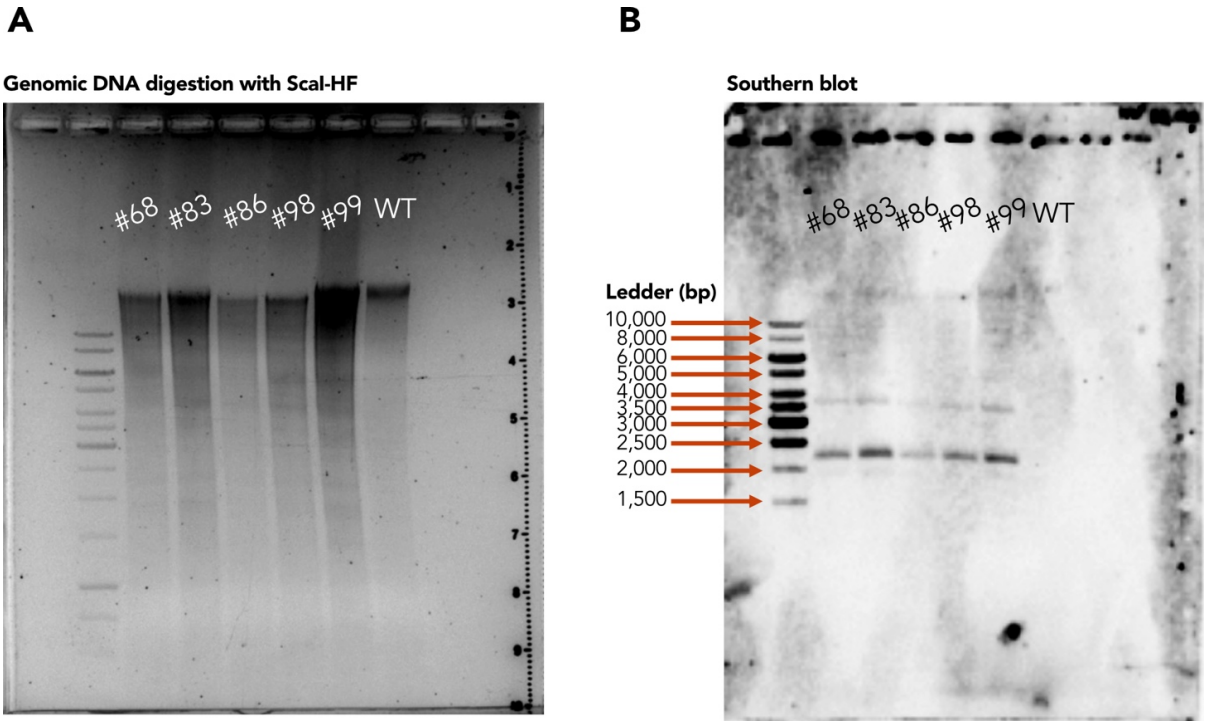


Figure 3. Genomic DNA digestion and southern blot results of pES296 candidate *Z. tritici* mutants. (A) Electrophoresis gel showing the genomic DNA digestion of five potentially correct *Z. tritici* mutants transformed with pES296 plasmid, along with the wild type (WT) Zt495 strain. Genomic DNAs were digested overnight at 37 °C using the ScaI-HF restriction enzyme (New England Biolabs, Frankfurt, Germany). **(B)** Southern blot results using the digestion blot on **(A)** and a probe targeting part of Hygromycin resistance cassete and the downstream flank of the Zt09_chr_7_00299 gene on Zt495 (see Methods). Based on digestion patterns, expected band size for wild type (WT) strain is 4895 bp and for correct deletion mutants are 2217 bp and 3630 bp.

Agrobacterium tumefasciens-mediated transformation (ATMT) is a well-established tool for functional genetic analyses in *Z. tritici* and it has been used in several studies throughout the years (Bowler et al., 2010; Francisco, Zwyssig, & Palma-Guerrero, 2020; Kema et al., 2018; Kilaru, Schuster, Ma, & Steinberg, 2017; Marshall et al., 2011; Möller et al., 2019; Poppe et al., 2015; Zwiers & De Waard, 2001). However, estimates on the efficiency of gene disruption and frequency of homologous recombination in this fungal species has shown to be highly variable, ranging from 0.5 to 75% (Bowler et al., 2010; Mehrabi, 2006; Zwiers & De Waard, 2001). Among the factors that can contribute for such variability, the position of the target locus in the genome; the chromatin structure of these loci as well as the transcriptional regulation of the genes being disrupted have a strong impact and can all affect the DNA availability for homologous recombination (Bird & Bradshaw, 1997; Bowler et al., 2010; Michielse, Hooykaas, van den Hondel, & Ram, 2005). Moreover, the presence of Tranposable Elements (TEs) in vicinity of the target locus can also influence homologous recombination and be the mediators of ectopic integrations. Our transformation assays in the *Aegilops*-infecting *Z. tritici* isolate Zt495 also showed variable results

in regards of transformation efficiency. As an example, in the pES296-transformation assay, out of the 196 potentially correct *Z. tritici* mutants grown in hygromycin-selective media, only three mutants (1.53%) showed to have the expected insertional pattern in the Southern blot analyses. For pES292 and pES294-transformation processes, we could so far identify 29 and 12 potentially correct independent *Z. tritici* mutants, respectively. However, despite the low transformation rates, our results could show that *Z. tritici* isolates infecting *Aegilops* species are also amenable to transformation processes using well established *Agrobacterium*-mediated transformation protocols and can serve as a starting point for functional genetics analyses on a unique wild pathosystem.

Next steps of this study involve the further validation of the pES292 and pES294 *Z. tritici* mutants by Southern blot analyses and phenotypic characterization of pES296 deletion mutants *in vitro* and *in planta*. For *in planta* virulence assays, we aim to perform quantitative plant experiments under greenhouse conditions to determine the pathogenicity of the *Z. tritici* Zt495 deletion mutants on the susceptible *Aegilops cylindrica* species. Independent *Z. tritici* mutant strains will be individually inoculated on *A. cylindrica* leaves at adjusted blastospore suspensions along the wild type strain (Zt495) and mock (0.1% Tween 20 in sterile water) as control treatments as previously described (Chapter II; Fagundes et al. 2020). We will also inoculate the same strains and treatments on *Triticum aestivum* cv. Riband (common bread wheat) plants to investigate potential host range variation. For *in vitro* phenotypic characterization of the deletion mutants, we will perform *in vitro* stress assays using NaCl, H2O2, Congored, Calcofluor and 28°C temperature stress as previously described (Poppe et al., 2015) using each independent deletion mutant along the wild type Zt495 strain. These assays will indicate any putative non-pathogenicity-related functional role of the deleted genes and access their potential pleiotropic effect on *Z. tritici* growth and *in vitro* fitness under stress.

Conclusions

In this study, we report structural and functional genetic analyses on a selective sweep region potentially involved in the adaptive evolution of *Z. tritici* to *Aegilops* species. Our findings demonstrate that not only polymorphisms but also structural rearrangements may be under host-driven selective pressures, and that the expression of genes involved in host-pathogen interactions in the analyzed region is potentially host- and temporal-specific regulated. These findings, along with the generation of *Aegilops*-infecting *Z. tritici* mutants, will help to elucidate the molecular and genomic mechanisms of host adaptation in *Z. tritici* species that can be further involved in other fungal plant pathogens infecting distinct but closely related host species as *Aegilops* and wheat.

Acknowledgements

The authors would like to thank Graziella Reinhardt and Judith Müller for the support in generating the *Z. tritici* mutants and the members of the Environmental Genomics group for helpful discussions.

Funding statement

This work was supported by intramural funding of the Max Planck Society and a personal grant from the State of Schleswig-Holstein to Eva H. Stukenbrock. The funders had no role in study design, data collection and analyses, decision to publish, or preparation of the manuscript.

Authors contributions

Conceptualization: WCF, JH, EHS; Zt495 sample collection: FS, AA; Plasmid design: WCF, FC, JH; Binary plasmids construction: WCF, FC, JH; *Agrobacterium*-mediated transformation of fungal cells: WCF, FC, JH; DNA extractions and southern blot analyses: WCF, FC; Genomics analyses: WCF, FC; Preparation and writing of manuscript: WCF. Editing of original manuscript: WCF, EHS, JH.

References

Aguileta, G., Refregier, G., Yockteng, R., Fournier, E., & Giraud, T. (2009). Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution*, 9(4), 656–670.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.

Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguileta, G., Snirc, A., … Giraud, T. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology*, 26(7), 2041–2062. https://doi.org/10.1111/mec.13976

Banke, S., Peschon, A., & McDonald, B. A. (2004). Phylogenetic analysis of globally distributed *Mycosphaerella graminicola* populations based on three DNA sequence loci. *Fungal Genetics and Biology*, 41(2), 226–238. https://doi.org/10.1016/j.fgb.2003.09.006

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. a. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bari, R., & Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Molecular Biology*, 69(4), 473–488. https://doi.org/10.1007/s11103-008-9435-0

Baroncelli, R., Amby, D. B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., … Thon, M. R. (2016). Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics*, 17(1), 1–17. https://doi.org/10.1186/s12864-016-2917-6

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

Bird, D., & Bradshaw, R. (1997). Gene targeting is locus dependent in the filamentous fungus *Aspergillus nidulans*. *Molecular and General Genetics*, 255(2), 219–225. https://doi.org/10.1007/s004380050492

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, 22(8), 437–446. https://doi.org/10.1016/j.tig.2006.06.005

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowler, J., Scott, E., Tailor, R., Scalliet, G., Ray, J., & Csukai, M. (2010). New capabilities for *Mycosphaerella graminicola* research. *Molecular Plant Pathology*, 11(5), 691–704. https://doi.org/10.1111/j.1364-3703.2010.00629.x

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., & Leung, H. (2013). Plant-pathogen interactions: Disease resistance in modern agriculture. *Trends in Genetics*, 29(4), 233–240. https://doi.org/10.1016/j.tig.2012.10.011

Branco, S., Bi, K., Liao, H. L., Gladieux, P., Badouin, H., Ellison, C. E., … Bruns, T. D. (2017). Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*. *Molecular Ecology*, 26(7), 2063–2076. https://doi.org/10.1111/mec.13892

Branco, S., Gladieux, P., Ellison, C. E., Kuo, A., Labutti, K., Lipzen, A., … Bruns, T. D. (2015). Genetic isolation between two recently diverged populations of a symbiotic fungus. *Molecular Ecology*, 24(11), 2747–2758. https://doi.org/10.1111/mec.13132

Brandis, G., & Hughes, D. (2020). The SNAP hypothesis: chromosomal rearrangements could emerge from positive selection during niche adaptation. *PLoS Genetics*, 16(3), 1–16. https://doi.org/10.1371/journal.pgen.1008615

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, 140(2), 783–796. https://doi.org/10.1093/genetics/140.2.783

Brunner, P. C., Torriani, S. F. F., Croll, D., Stukenbrock, E. H., & McDonald, B. A. (2013). Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. *Molecular Biology and Evolution*, 30(6), 1337–1347. https://doi.org/10.1093/molbev/mst041

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421.

Cao, S., Brandis, G., Huseby, D. L., & Hughes, D. (2022). Positive selection during niche adaptation results in large-scale and irreversible rearrangement of chromosomal gene order in bacteria. *Molecular Biology and Evolution*, 39(4), 1–11. https://doi.org/10.1093/molbev/msac069

Castanera, R., Lopez-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., … Ramirez, L. (2016). Transposable Elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genetics*, 12(6), e1006108. https://doi.org/10.1371/journal.pgen.1006108

Coghlan, A., Eichler, E. E., Oliver, S. G., Paterson, A. H., & Stein, L. (2005). Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends in Genetics*, 21(12), 673–682.

Couch, B. C., Fudal, I., Lebrun, M. H., Tharreau, D., Valent, B., Van Kim, P., … Kohn, L. M. (2005). Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, 170(2), 613–630. https://doi.org/10.1534/genetics.105.041780

Dai, Y., Jia, Y., Correll, J., Wang, X., & Wang, Y. (2010). Diversification and evolution of the avirulence gene AVR-Pita1 in field isolates of *Magnaporthe oryzae*. *Fungal Genetics and Biology*, 47(12), 973–980. https://doi.org/https://doi.org/10.1016/j.fgb.2010.08.003

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development*, 35, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

Ellison, C. E., Hall, C., Kowbel, D., Welch, J., Brem, R. B., Glass, N. L., & Taylor, J. W. (2011). Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2831–2836. https://doi.org/10.1073/pnas.1014971108

Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. (2020). Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: A laboratory workflow. *Current Protocols in Microbiology*, 59(1), 1–27. https://doi.org/10.1002/cpmc.128

Feurtey, A., Lorrain, C., Croll, D., Eschenbrenner, C., Freitag, M., Habig, M., … Stukenbrock, E. H. (2020). Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*. *BMC Genomics*, 21(1), 588. https://doi.org/10.1186/s12864-020-06871-w

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9, 275–296.

Francisco, C. S., Zwyssig, M. M., & Palma-Guerrero, J. (2020). The role of vegetative cell fusions in the development and asexual reproduction of the wheat fungal pathogen *Zymoseptoria tritici*. *BMC Biology*, 18(1), 1–16. https://doi.org/10.1186/s12915-020-00838-9

Ghanbarnia, K., Fudal, I., Larkan, N. J., Links, M. G., Balesdent, M. H., Profotova, B., … Borhan, M. H. (2015). Rapid identification of the *Leptosphaeria maculans* avirulence gene AvrLm2 using an intraspecific comparative genomics approach. *Molecular Plant Pathology*, 16(7), 699–709. https://doi.org/10.1111/mpp.12228

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345. https://doi.org/10.1038/nmeth.1318

Giraud, T., Gladieux, P., & Gavrilets, S. (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology and Evolution*, 25(7), 387–395. https://doi.org/10.1016/j.tree.2010.03.006

Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D. M., & Hood, M. E. (2008). Speciation in fungi. *Fungal Genetics and Biology*, 45(6), 791–802. https://doi.org/10.1016/j.fgb.2008.02.001

Gladieux, P., Guerin, F., Giraud, T., Caffier, V., Lemaire, C., Parisi, L., … Le Cam, B. (2011). Emergence of novel fungal pathogens by ecological speciation: Importance of the reduced viability of immigrants. *Molecular Ecology*, 20(21), 4521–4532. https://doi.org/10.1111/j.1365-294X.2011.05288.x

Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguileta, G., De Vienne, D. M., … Giraud, T. (2014). Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology*, 23(4), 753–773. https://doi.org/10.1111/mec.12631

Goodwin, S. B., M'Barek, S. Ben, Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., … Kema, G. H. J. (2011). Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics*, 7(6). https://doi.org/10.1371/journal.pgen.1002070

Grandaubert, J., Bhattacharyya, A., & Stukenbrock, E. H. (2015). RNA-seq-based gene annotation and comparative genomics of four fungal grass pathogens in the genus *Zymoseptoria* identify novel orphan genes and species-specific invasions of transposable elements. *G3; Genes|Genomes|Genetics*, 5(7), 1323–1333. https://doi.org/10.1534/g3.115.017731

Grandaubert, J., Dutheil, J. Y., & Stukenbrock, E. H. (2019). The genomic determinants of adaptive evolution in a fungal pathogen. *Evolution Letters*, 299–312. https://doi.org/10.1002/evl3.117

Grünwald, N. J., McDonald, B. A., & Milgroom, M. G. (2016). Population genomics of fungal and oomycete pathogens. *Annual Review of Phytopathology*, 54(1), 323–346. https://doi.org/10.1146/annurev-phyto-080614-115913

Hartmann, F. E., McDonald, B. A., & Croll, D. (2018). Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Molecular Ecology*, 41(0), 0–2. https://doi.org/10.1111/mec.14711

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME Journal*, 11(5), 1189–1204. https://doi.org/10.1038/ismej.2016.196

Haueisen, J., Möller, M., Eschenbrenner, C. J., Grandaubert, J., Seybold, H., Adamiak, H., & Stukenbrock, E. H. (2019). Highly flexible infection programs in a specialized wheat pathogen. *Ecology and Evolution*, 9(1), 275–294. https://doi.org/10.1002/ece3.4724

Jiang, R. H. Y., & Tyler, B. M. (2012). Mechanisms and evolution of virulence in oomycetes. *Annual Review of Phytopathology*, 50, 295–318. https://doi.org/10.1146/annurev-phyto-081211-172912

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323–329. https://doi.org/10.1038/nature05286

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kema, G. H. J., Mirzadi Gohari, A., Aouini, L., Gibriel, H. A. Y., Ware, S. B., van Den Bosch, F., … Seidl, M. F. (2018). Stress and sexual reproduction affect the dynamics of the wheat pathogen effector AvrStb6 and strobilurin resistance. *Nature Genetics*, 50(March), 1–6. https://doi.org/10.1038/s41588-018-0052-9

Kilaru, S., Schuster, M., Ma, W., & Steinberg, G. (2017). Fluorescent markers of various organelles in the wheat pathogen *Zymoseptoria tritici*. *Fungal Genetics and Biology*, 105(May), 16–27. https://doi.org/10.1016/j.fgb.2017.05.001

Kubicek, C. P., Starr, T. L., & Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology*, 52, 427–451. https://doi.org/10.1146/annurev-phyto-102313-045831

Lind, A. L., Wisecaver, J. H., Lameiras, C., Wiemann, P., Palmer, J. M., Keller, N. P., … Rokas, A. (2017). Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biology*, 15(11), 1–26. https://doi.org/10.1371/journal.pbio.2003583

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., … Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K. E., … Rudd, J. J. (2011). Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiology*, 156(2), 756–769. https://doi.org/10.1104/pp.111.176347

McCoy, K. D. (2003). Sympatric speciation in parasites–what is sympatry? *Trends in Parasitology*, 19(9), 400–404. https://doi.org/10.1016/S1471-4922(03)00194-6.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 351(6328), 652–654. https://doi.org/10.1038/351652a0

Mehrabi, R. (2006) Signalling pathways involved in pathogenicity and development of the fungal wheat pathogen *Mycosphaerella graminicola*. PhD Thesis, Wageningen University (ISBN 90-8504-409-x). http://library.wur.nl/wda/dissertations/dis3946.pdf.

Mehrabi, R., Gohari, A. M., da Silva, G. F., Steinberg, G., Kema, G. H. J., & de Wit, P. J. G. M. (2015). Flexible gateway constructs for functional analyses of genes in plant pathogenic fungi. *Fungal Genetics and Biology*, 79, 186–192.

Meile, L., Croll, D., Brunner, P. C., Plissonneau, C., Hartmann, F. E., McDonald, B. A., & Sánchez-Vallet, A. (2018). A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to *Septoria tritici* blotch. *New Phytologist*, 219(3), 1048–1061. https://doi.org/10.1111/nph.15180

Michielse, C. B., Hooykaas, P. J. J., van den Hondel, C. A. M. J. J., & Ram, A. F. J. (2005). Agrobacterium-mediated transformation as a tool for functional genomics in fungi. *Current Genetics*, 48(1), 1–17. https://doi.org/10.1007/s00294-005-0578-0

Mohd-Assaad, N., McDonald, B. A., & Croll, D. (2018). Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. *Genome Biology and Evolution*, 10(5), 1315–1332. https://doi.org/10.1093/gbe/evy087

Möller, M., Schotanus, K., Soyer, J. L., Haueisen, J., Happ, K., Stralucke, M., … Stukenbrock, E. H. (2019). Destabilization of chromosome structure by histone H3 lysine 27 methylation. In *PLoS genetics* (Vol. 15). https://doi.org/10.1371/journal.pgen.1008093

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, 15(12), 756–771. https://doi.org/10.1038/nrmicro.2017.76

Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A. C., … Anderson, T. J. C. (2007). Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Molecular Biology and Evolution*, 24(2), 562–573. https://doi.org/10.1093/molbev/msl185

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, 15(11), 1566–1575. https://doi.org/10.1101/gr.4252305

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., … Grigoriev, I. V. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathogens*, 8(12). https://doi.org/10.1371/journal.ppat.1003037

Ohno, S. (2013). Evolution by gene duplication. Springer Science & Business Media.

Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research*, 24(1), 1–17. https://doi.org/10.1186/s40709-017-0064-0

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(1), 1–16. https://doi.org/10.1186/s12915-017-0457-4

Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. (2015). Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLoS Pathogens*, 11(7), 1–21. https://doi.org/10.1371/journal.ppat.1005055

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., … Kamoun, S. (2010). Genome evolution following host jumps in the irish potato famine pathogen lineage. *Science*, 330(6010), 1540–1543. https://doi.org/10.1126/science.1193070

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10(6), 417–430. https://doi.org/10.1038/nrmicro2790

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, 29(1), 24-26. https://doi.org/10.1038/nbt.1754

Sambrook, J., & Russell, D. W. (2001). Molecular cloning: a laboratory manual., 3rd edn. (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York).

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F. E., Soyer, J. L., Tellier, A., & Croll, D. (2018). The genome biology of effector gene evolution in filamentous plant pathogens. *Annual Review of Phytopathology*, 56(1). https://doi.org/10.1146/annurev-phyto-080516-035303

Schardl, C. L., Young, C. A., Hesse, U., Amyotte, S. G., Andreeva, K., Calie, P. J., … Zeng, Z. (2013). Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genetics*, 9(2). https://doi.org/10.1371/journal.pgen.1003323

Schmidt, S. M., Lukasiewicz, J., Farrer, R., van Dam, P., Bertoldo, C., & Rep, M. (2015). Comparative genomics of *Fusarium oxysporum* f. sp. melonis reveals the secreted protein recognized by the Fom-2 resistance gene in melon. *New Phytologist*, 209(1), n/a--n/a. https://doi.org/10.1111/nph.13584

Sharma, R., Mishra, B., Runge, F., & Thines, M. (2014). Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the smut fungus *Melanopsichium pennsylvanicum*. *Genome Biology and Evolution*, 6(8), 2034–2049. https://doi.org/10.1093/gbe/evu148

Sidhu, Y. S., Cairns, T. C., Chaudhari, Y. K., Usher, J., Talbot, N. J., Studholme, D. J., … Haynes, K. (2015). Exploitation of sulfonylurea resistance marker and non-homologous end joining mutants for functional analysis in *Zymoseptoria tritici*. *Fungal Genetics and Biology*, 79, 102–109. https://doi.org/10.1016/j.fgb.2015.04.015

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23–35.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, 98(3), 503–517.

Sperschneider, J., & Dodds, P. N. (2022). EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Molecular Plant-Microbe Interactions*, 35(2), 146–156. https://doi.org/10.1094/MPMI-08-21-0201-R

Stergiopoulos, I., & de Wit, P. J. G. M. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, 47(1), 233–263. https://doi.org/10.1146/annurev.phyto.112408.132637

Stukenbrock, E. H., & McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions*, 22(4), 371–380. https://doi.org/10.1094/MPMI-22-4-0371

Stukenbrock, E. H. (2013). Evolution, selection and isolation: A genomic view of speciation in fungal plant pathogens. *New Phytologist*, 199(4), 895–907. https://doi.org/10.1111/nph.12374

Stukenbrock, E. H., Banke, S., Javan-Nikkhah, M., & McDonald, B. A. (2007). Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, 24(2), 398–411. https://doi.org/10.1093/molbev/msl169

Stukenbrock, E. H., & Mcdonald, B. A. (2007). Geographical variation and positive diversifying selection in the host-specific toxin SnToxA. *Molecular Plant Pathology*, 8(3), 321–332. https://doi.org/10.1111/j.1364-3703.2007.00396.x

Stukenbrock, E. H., & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, 46(1), 75–100. https://doi.org/10.1146/annurev.phyto.010708.154114

Stukenbrock, E. H., Quaedvlieg, W., Javan-Nikhah, M., Zala, M., Crous, P. W., & McDonald, B. A. (2012). *Zymoseptoria ardabiliae* and *Z. pseudotritici*, two progenitor species of the *Septoria tritici* leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella graminicola*). *Mycologia*, 104(6), 1397–1407. https://doi.org/10.3852/11-374

Stukenbrock, E. H, Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M., … Schierup, M. H. (2011). The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research*, 21(12), 2157–2166. https://doi.org/10.1101/gr.118851.110

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics*, 27(7), 1009–1010. https://doi.org/10.1093/bioinformatics/btr039

Tang, Y. C., & Amon, A. (2013). Gene copy-number alterations: A cost-benefit analysis. *Cell*, 152(3), 394–405. https://doi.org/10.1016/j.cell.2012.11.043

Tellier, A., Moreno-Gámez, S., & Stephan, W. (2014). Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution*, 68(8), 2211–2224. https://doi.org/10.1111/evo.12427

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., … Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*. https://doi.org/10.1038/s41587-021-01156-3

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, 31(1), 46–53. https://doi.org/10.1038/nbt.2450

van de Wouw, A. P., Cozijnsen, A. J., Hane, J. K., Brunner, P. C., McDonald, B. A., Oliver, R. P., & Howlett, B. J. (2010). Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *PLoS Pathogens*, 6(11). https://doi.org/10.1371/journal.ppat.1001180

van der Does, H. C., & Rep, M. (2017). Adaptation to the host environment by plant-pathogenic fungi. *Annual Review of Phytopathology*, 55(1), 427–450. https://doi.org/10.1146/annurev-phyto-080516-035551

Vleeshouwers, V. G. A. A., & Oliver, R. P. (2014). Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Molecular Plant-Microbe Interactions*, 27(3), 196–206. https://doi.org/10.1094/MPMI-10-13-0313-IA

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., … Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973–982. https://doi.org/10.1038/nrg2165

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D. G. O., Yoshida, K., Dong, S., … Kamoun, S. (2012). Effector biology of plant-associated organisms: Concepts and perspectives. *Cold Spring Harbor Symposia on Quantitative Biology*, 77, 235–247. https://doi.org/10.1101/sqb.2012.77.015933

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19). https://doi.org/10.1073/pnas.1219381110

Yoshida, K., Saunders, D. G. O., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., … Terauchi, R. (2016). Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics*, 17(1), 1–18. https://doi.org/10.1186/s12864-016-2690-6

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18(6), 292–298.

Zwiers, L. H., & De Waard, M. A. (2001). Efficient *Agrobacterium tumefaciens*-mediated gene disruption in the phytopathogen *Mycosphaerella graminicola*. *Current Genetics*, 39(5–6), 388–393. https://doi.org/10.1007/s002940100216

Supplementary Material

Supplementary Tables

All supplementary tables in ".xlsx" format are deposited on the supplementary USB key.

Table S1. List of primers used in this study.

Table S2. Functional prediction of secreted and effector genes in the selective sweep region of Zt495 and IPO323 *Z. tritici* genomes.

We used the softwares SignalP v6.0 (Teufel et al., 2022) and EffectorP v3.0 (Sperschneider & Dodds, 2022) to predict if the proteins in the selective sweep region of Zt495 are potentially secreted and classified as effectors, respectively. Values shown are the probability of secretion (signal peptide) predicted with SignalP and the probability to function as apoplastic or cytoplasmic effector predicted with EffectorP. We also reanalyzed the protein sequences present in the selective sweep region in the reference *Z. tritici* IPO323 genome as a comparison and to confirm previous functional predictions (Feurtey et al., 2020; Grandaubert et al., 2015).

Supplementary Figures

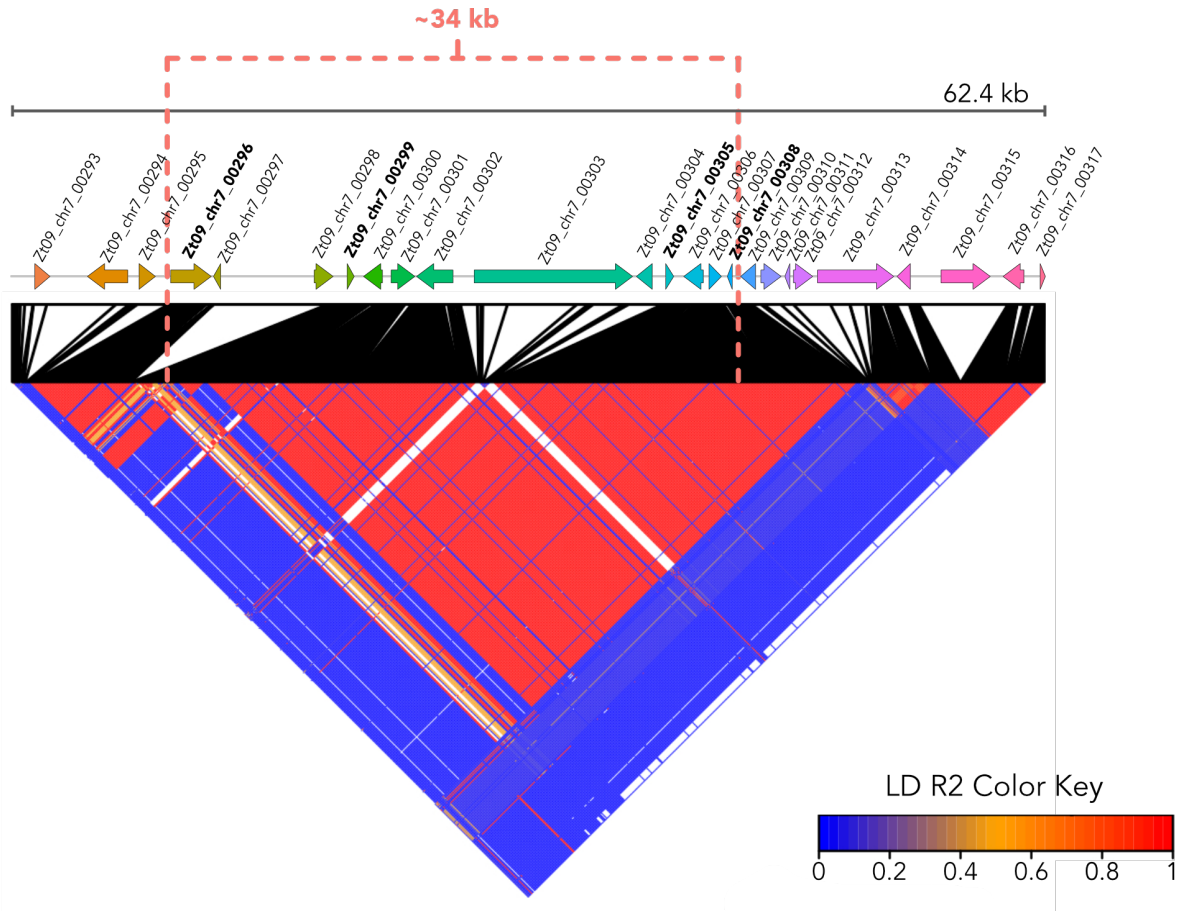
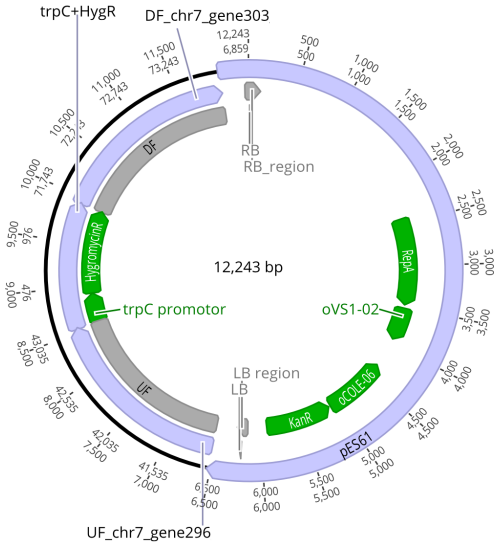


Figure S1. Schematic illustration of the selective sweep region of chromosome 7 analyzed in this study. In order to facilitate genomic deletions by homologous recombination in the wild-type *Z. tritici* isolate, we focused our analyses to a smaller region of approximately 34 kb within the selective sweep locus identified previously (Chapter II) based on the coordinates of the reference *Z. tritici* IPO323 genome (Goodwin et al., 2011; Grandaubert, Bhattacharyya, & Stukenbrock, 2015). Region selected is delimited by dashed red lined. We selected this smaller region to comprise the three candidate effectors and CAZyme-encoding genes (in bold) and based on the high patterns of linkage disequilibrium (LD) in the *Aegilops*-infecting *Z. tritici* population identified previously (Chapter II). Linkage disequilibrium (LD) heatmap was based on coefficient of correlation (r2) between SNPs in the *Aegilops* population 1 (Chapter II). Correspondence between SNPs location in the genomics regions (e.g. genes) and on the heatmap are depicted by black segments.

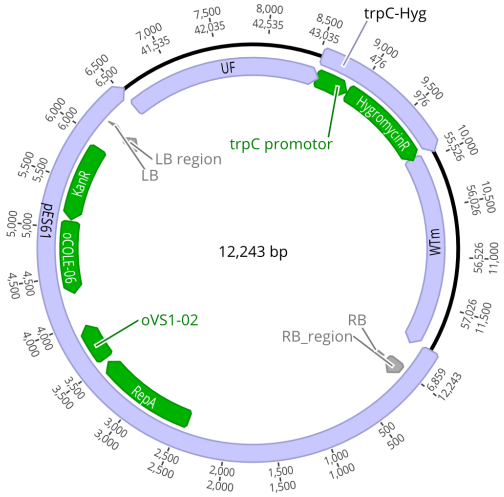
pES292

Deletion of whole selective sweep region



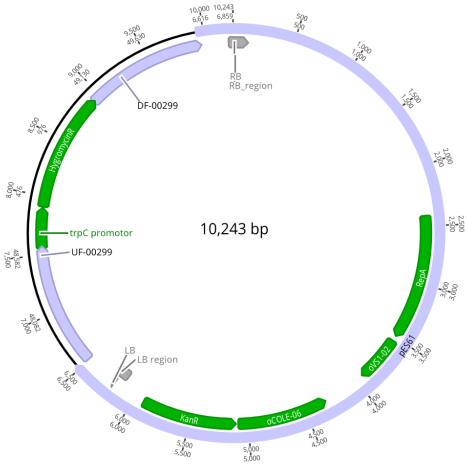
pES294

Deletion of first half of selective sweep region



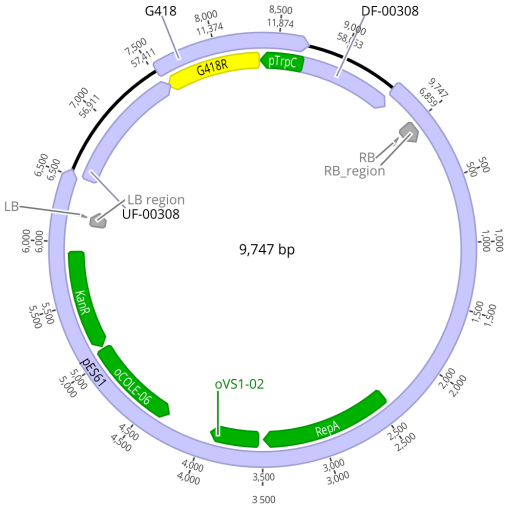
pES296

Deletion of gene Zt09_chr_7_00299



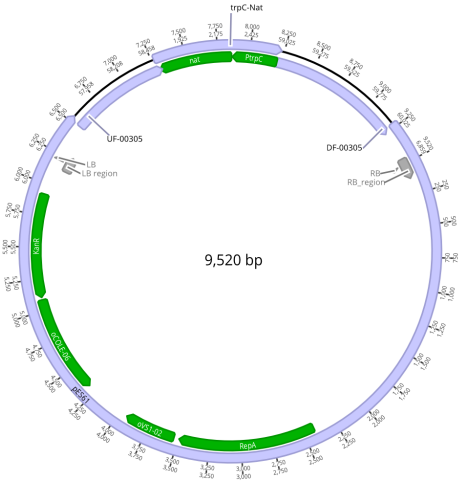
pES297

Deletion of gene Zt09_chr_7_00308



pES298

Deletion of gene Zt09_chr_7_00305



pES305

Deletion of second half of selective sweep region

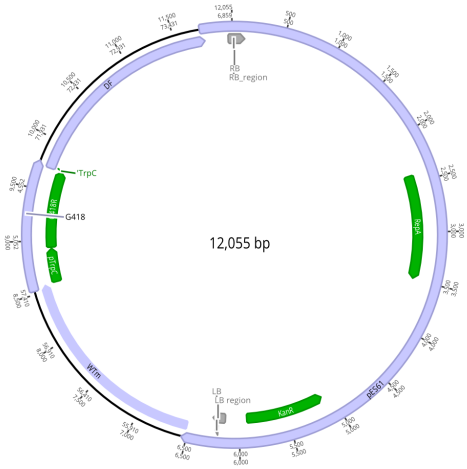
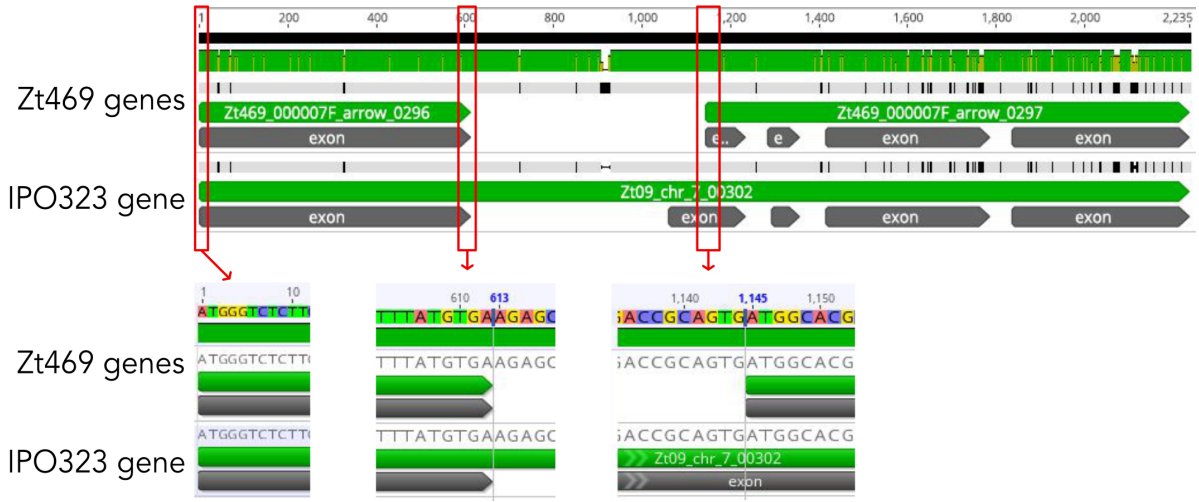


Figure S2. Plasmid constructs used for the functional genetics analyses of the selective sweep region in Zt495. All plasmids were constructed using Gibson assembly (Gibson et al. 2009). We used three resistance marker genes for *Z. tritici* and assembled them in different plasmid constructs: the *hph* gene conferring resistance to hygromycin [for pES292, pES294 and pES296; (Bowler et al., 2010)]; the *neo* gene providing resistance to geneticin [G418; for pES297 and pES305; (Sidhu et al., 2015)], and the *nat* gene which confers resistance to nourseothricin [for pES298; (Mehrabi et al., 2015)]. Target genes / regions of each plasmid construct and plasmid size in bp are indicated. UF = "Upstream Flank"; DF = "Downstream Flank"; LB = "Left Border"; RB = "Right Border"; WTm = "Wild type middle region" (region selected in the wild type isolates between the two halves of the selective sweep region).

A


Zt469 genes
IPO323 gene
Zt469_000007F_arrow_0296
Zt469_000007F_arrow_0297
exon
Zt09_chr_7_00302

B

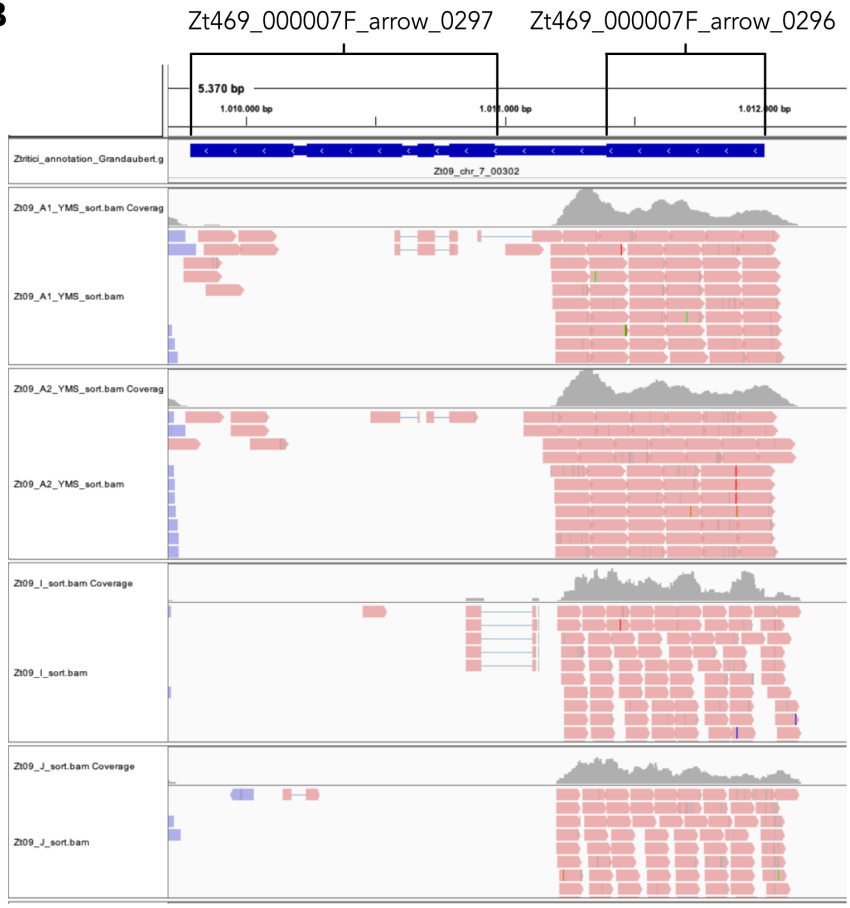

Zt469_000007F_arrow_0297
Zt469_000007F_arrow_0296
5.370 bp
Ztritici_annotation_Grandaubert.g
Zt09_chr_7_00302
Zt09_A1_YMS_sort.bam Coverag
Zt09_A1_YMS_sort.bam
Zt09_A2_YMS_sort.bam Coverag
Zt09_A2_YMS_sort.bam
Zt09_I_sort.bam Coverage
Zt09_I_sort.bam
Zt09_J_sort.bam Coverage
Zt09_J_sort.bam

Figure S3. Evidence for the existence of two instead of one gene in the annotated gene model Zt09_chr_7_00302 of *Z. tritici* IPO323. (A) DNA sequence alignment for the region containing the genes Zt469_000007F_arrow_0296 and Zt469_000007F_arrow_0297 in *Z.tritici* Zt469 genome and the Zt09_chr_7_00302 gene in *Z. tritici* IPO323 genome. Overview of the whole region at the top panel. At the bottom panels, zoom images on the start of Zt469_000007F_arrow_0296/ Zt09_chr_7_00302 genes (left panel); the end of Zt469_000007F_arrow_0296 genes (middle panel); and the start of Zt469_000007F_arrow_0297 (right panel). Like in Zt469, there is a stop codon (TGA) in IPO323 at the end of Zt469_000007F_arrow_0296, and a start codon (ATG) at the start of Zt469_000007F_arrow_0297. Alignment figure created in Geneious v.2020.1.2 software (https://www.geneious.com/home/). **(B)** Visualization of mapped RNA-seq reads on *Z. tritici* IPO323 genome for the region containing Zt09_chr_7_00302. Mapped transcriptome data was retrieved from Haueisen et al. (2019) and represent two biological replicates for the *Z. tritici* IPO323 isolate (strain Zt09) during *in vitro* growth condition on YMS (Yeast-Malt-Sucrose) media (first two tracks) and *in planta* (two bottom tracks; Haueisen et al., 2019). The first exon of Zt09_chr_7_00302, corresponding to Zt469_000007F_arrow_0296, shows a larger number of mapped reads than the other exon corresponding to Zt469_000007F_arrow_0297. Figure generated using the Integrative Genome Browser (IGV; Robinson et al. 2011).

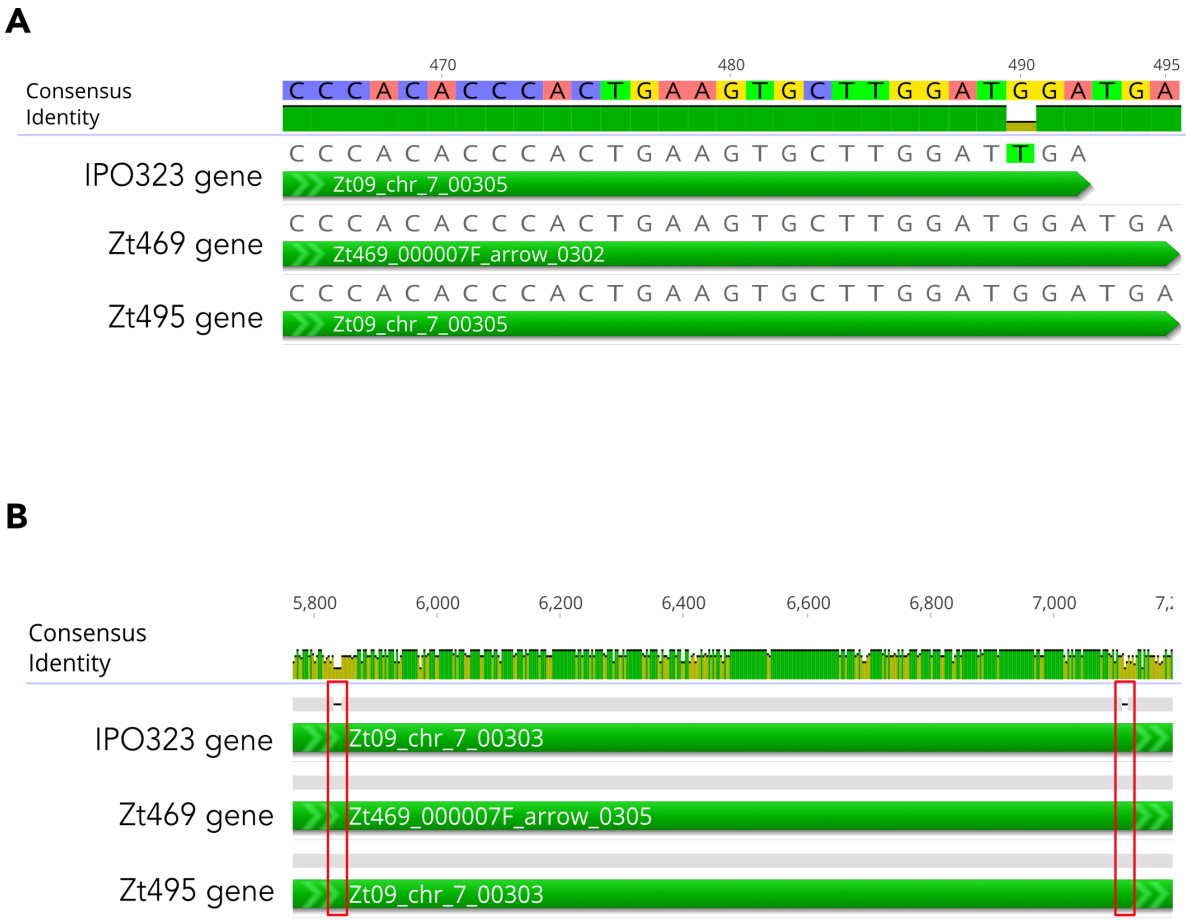


Figure S4. DNA sequence alignment of the genes Zt09_chr_7_00303/ Zt469_000007F_arrow_0305 and Zt09_chr_7_00305/Zt469_000007F_arrow_0302 in the *Z. tritici* IPO323, Zt495 and Zt469 genomes. (A) Zoom image at the end of the Zt09_chr_7_00305/ Zt469_000007F_arrow_0302 gene. In Zt495 and Zt469 genomes, a point mutation (T to G; highlighted) results in a frame shift of the stop codon (TGA) when compared to IPO323. **(B)** Zoom image at the two insertions of 12 and 3 bp each (highlighted in red) on the *Z. tritici* isolates Zt469 (Zt469_000007F_arrow_0305) and Zt495 when aligned to the homologous gene Zt09_chr_7_00303 in IPO323. Alignment figures created in Geneious v.2020.1.2 software (https://www.geneious.com/home/).

General

Conclusions and Perspectives

General Conclusions and Perspectives

Agricultural practices, besides largely impacting human history, have also shaped the evolution of plants and associated pathogens. Since the start of agriculture around 10,000 years ago, anthropogenic influences on the environment as crop domestication and global trades are thought to promote the emergence of new plant diseases (Anderson et al., 2004; Gladieux et al., 2011; Money, 2006; Slippers, Stenlid, & Wingfield, 2005; Woolhouse, Haydon, & Antia, 2005). In fact, agriculture might have favored host specialization, reproductive isolation, and speciation of plant pathogens on new hosts (Hansen, 1987; Kohn, 2005; Stukenbrock & McDonald, 2008). The introduction of domesticated plants and associated pathogens into new areas could simultaneously disturb the population of pathogens already present in these locations, exposing both domesticated and wild plants to new pathogen populations that could potentially shift from wild to domesticated hosts and vice-versa. Wild plant species growing within or nearby agricultural fields, e.g. weeds can represent important reservoirs of new pathogens which can eventually specialize in domesticated hosts and become genetically differentiated from their source population even when they occur at the same location (Stukenbrock & McDonald, 2008). If selection allows, this divergence can lead to sympatric speciation and create a new pathogen species adapted to a different niche than the source population (Giraud, Gladieux, & Gavrilets, 2010; Giraud, Refrégier, Le Gac, de Vienne, & Hood, 2008; Gladieux et al., 2011; Stukenbrock & McDonald, 2008). In this regard, analyses of wild pathogen populations through comparative and population genomics analyses as well the understanding of their host range and infection programs are important allies to access the risk of potential future pathogens in agricultural ecosystems.

The general aim of this work has been to gain insights into the evolutionary, molecular and phenotypic patterns of host adaptation in fungal plant pathogens using host-divergent *Z. tritici* lineages as models of study. Through a plethora of distinct genomic methods and approaches, we demonstrated that fungal individuals isolated from distinct but closely-related and sympatric hosts show a high extent of genomic divergence not only at the nucleotide but also at the structural level, including the occurrence of a newly described accessory chromosome. We also demonstrated particular genomic signatures of selection in the host-diverging *Z. tritici* lineages that could have contributed to the high levels of host specificity observed during infection assays, and that will serve as a starting point for the first functional genetics analyses of the wild *Z. tritici-Aegilops* pathosystem.

In the first chapter, we described a suite of protocols for practical handling of *Z. tritici* under laboratory and greenhouse/phytochamber conditions that has been used throughout all the chapters composing this dissertation. *Zymoseptoria tritici* is a model organism of crop-infecting fungal pathogens and it has been used in several comparative and population genomics analyses. However, *Z. tritici* isolates also show extensive genetic variability and morphology plasticity that can make experimental studies and comparability of results obtained in different laboratories challenging, even when using the same isolate strain. Therefore, the aim of this chapter was to present protocols that can serve either as a baseline for researchers initiating work with *Z. tritici* or as a reference to make experiments more comparable across different research groups. We described steps ranging from isolating, handling and properly storing *Z. tritici* specimens under cryogenic conditions to the inoculation and analyses of *Z. tritici* virulence *in planta*. We also discussed important points that should be addressed in order to avoid the occurrence of aneuploidy and mutation accumulation in laboratory strains and to reduce variability between experiments.

In chapter two, we focused on population genomics and virulence analyses of a unique collection of host-diverging *Z. tritici* isolates and demonstrated that host specificity can shape the population structure and genome evolution of fungal species. Using sympatric Iranian *Z. tritici* specimens isolated from wheat and *Aegilops* spp., we showed evidences that host-specific lineages of *Z. tritici* are co-occurring in the Fertile Crescent region in Middle East since wheat domestication around 10,000-12,000 years ago (Glémin et al., 2019; Salamini, Özkan, Brandolini, Schäfer-Pregl, & Martin, 2002; Stukenbrock, Banke, Javan-Nikkhah, & McDonald, 2007). In contrast to what is observed in other domesticated eukaryotic species as e.g. plants, we demonstrated that *Z. tritici* isolates infecting domesticated wheat plants have a more diverse genome than sympatric wild *Aegilops*-infecting isolates, which can be reasonably explained by the high levels of host specificity found in these isolates and patchy distribution of *Aegilops* species throughout Iran. We could also show that footprints of selective pressure were distinct and wide-spread along the genomes of *Z. tritici* isolates infecting *Aegilops* and wheat plants. Several of the selective sweep regions identified harbored genes with predicted functions potentially involved in host-pathogen interactions such as candidate effectors, CAZymes and polyketide synthase genes, suggesting that they may play a role on *Z. tritici* host specialization. The function and contribution to host specialization of these identified candidate genes will be further elucidated by generation and phenotyping of *Z. tritici* mutants *in vitro* and *in planta*, as we discussed in chapter four. Collaborations with Iranian colleagues and continuous sampling of wild grasses presenting *Zymoseptoria* symptoms in the Middle East will also allow us to further explore the host range and population structure of *Zymoseptoria* species and

to gain insights of possible incipient speciation events that may taking place at center of origin of these fungal pathogens.

Host-diverging *Z. tritici* isolates also differ on the overall genome structure and may explain the origin of accessory chromosomes in *Zymoseptoria* species. In chapter three, we used a multi-omics approach (genomics, transcriptomics and epigenomics) to illustrate how the genome architecture of the Iranian host-diverging *Z. tritici* isolates is organized in comparison to other *Z. tritici* and closely related *Zymoseptoria* species. Corroborating previous findings (Feurtey et al., 2020), we showed that genome compartmentation is an ancestral trait in the *Zymoseptoria* genus, including in the recently described *Aegilops*-infecting *Z. tritici* lineage. We also characterized the occurrence of a new accessory chromosome in *Aegilops*-infecting *Z. tritici* isolates presenting similar characteristics to known accessory chromosomes in *Zymoseptoria* species. Interestingly, we showed that this new accessory chromosome in *Z. tritici* has an orthologous accessory chromosome in *Zymoseptoria ardabiliae*, a closely related fungal species also infecting wild grasses. Considering the sympatry of the *Z. ardabiliae* and *Aegilops*-infecting *Z. tritici* species in the Iran, we proposed that the newly identified chromosome has been exchanged between *Z. tritici* and *Z. ardabiliae* by introgressive hybridization events and suggested that hybridization may play an important role on the origin of new accessory chromosomes in these species. Additional sampling of *Z. ardabiliae* specimens from wild grasses in Iran will allow us to potentially identify more isolates presenting this new chromosome and to characterize gene flow events with *Z. tritici* and other *Zymoseptoria* species through population genomics analyses. Moreover, experimental crosses using *Aegilops*-infecting *Z. tritici* and *Z. ardabiliae* isolates with and without the orthologous chromosome will allow us to gain more insights into the transmission of accessory chromosomes between fungal species and to investigate whether pre- or postmating reproductive barriers may contribute to reproductive isolation between these fungal species and host-specific lineages.

In order to investigate the molecular mechanisms underlying host specialization, functional genetics analyses are ultimately required to associate candidate genes to particular phenotypes. In chapter four, we address this using structural and reverse genetics analyses of an identified genomic region under positive selection (i.e. selective sweep region) that is potentially involved in the adaptive evolution of *Z. tritici* to *Aegilops* species (Chapter II). We showed that several genomic rearrangements are present in this locus under selection when comparing *Aegilops*- and wheat-infecting *Z. tritici* isolates, including inversions and gene gains and losses. We also demonstrated that the expression of genes involved in host-pathogen interactions present in this region is potentially regulated in a host- and temporal-specific manner. At last, we showed the application

of a classical functional genetics tool in a high virulent *Z. tritici* isolate infecting *Aegilops cylindrica* species, successfully deleting one of three candidate effector genes present in the selective sweep region. Taken these results together, we proposed that not only polymorphisms but also structural rearrangements may be under host-driven selective pressures in fungal lineages infecting distinct hosts. With the generation of *Aegilops*-infecting *Z. tritici* mutants and phenotyping of these mutants *in planta* and *in vitro*, we will gain more insights into the relevance of the selective sweep region for host specificity in *Z. tritici* species while also accessing any putative non-virulence-related functional role of these genes for *Z. tritici* growth and *in vitro* fitness under stress. Moreover, additional analyses such as "allele swapping" with the genes present in the same genomic region of wheat-infecting *Z. tritici* isolates can potentially provide interesting findings such as host range expansions or host shifts. These analyses will be particularly relevant for the candidate effector gene Zt09_chr_7_00229 which showed to have high levels of expression in distinct infections stages on both *Aegilops cylindrica*- and wheat-*Z. tritici* pathosystems.

Altogether, this study provided comprehensive analyses of host adaptation in fungal plant pathogens, ranging from evolutionary and virulence mechanisms to the identification and functional analyses of candidate genes possibly involved in pathogenesis and host specificity. The use of unique *Z. tritici* populations isolated from wild and domesticated hosts as models of study widely support the understanding of genomic determinants of speciation in this economically important fungal species as well as bring insights into the impact of host adaptation in fungal genome evolution. Moreover, the possibility of introgressive hybridization coupled with host range expansion in the analyzed pathosystems highlight the important role that wild pathogens may play into future, outbreaking agricultural plant diseases, particularly in regions where distinct pathogen lineages and shared hosts are in close proximity. Thus, the results hereby presented are of relevance not only for evolutionary biologists but also for applied plant pathology fields when generating Integrated Pest Management (IPM) and plant disease surveillance approaches.

References General Introduction & General Conclusions and Perspectives

Agrios, G. N. (2005). Plant pathology. Elsevier.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, 19(10), 535–544. https://doi.org/10.1016/j.tree.2004.07.021

Balter, M. (2007). Seeking agriculture's ancient roots. *Science*, 316(5833), 1830–1835. https://doi.org/10.1126/science.316.5833.1830

Banke, S., Peschon, A., & McDonald, B. A. (2004). Phylogenetic analysis of globally distributed *Mycosphaerella graminicola* populations based on three DNA sequence loci. *Fungal Genetics and Biology*, 41(2), 226–238. https://doi.org/10.1016/j.fgb.2003.09.006

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, 22(8), 437–446. https://doi.org/10.1016/j.tig.2006.06.005

Brown, J. K. M., & Tellier, A. (2011). Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, 49(1), 345–367. https://doi.org/10.1146/annurev-phyto-072910-095301

Couch, B. C., Fudal, I., Lebrun, M. H., Tharreau, D., Valent, B., Van Kim, P., … Kohn, L. M. (2005). Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, 170(2), 613–630. https://doi.org/10.1534/genetics.105.041780

de Vienne, D. M., Hood, M. E., & Giraud, T. (2009). Phylogenetic determinants of potential host shifts in fungal pathogens. *Journal of Evolutionary Biology*, 22(12), 2532–2541. https://doi.org/10.1111/j.1420-9101.2009.01878.x

de Vries, S., Stukenbrock, E. H., & Rose, L. E. (2020). Rapid evolution in plant–microbe interactions – an evolutionary genomics perspective. *New Phytologist*. https://doi.org/10.1111/nph.16458

Depotter, J. R. L., Seidl, M. F., van den Berg, G. C. M., Thomma, B. P. H. J., & Wood, T. A. (2017). A distinct and genetically diverse lineage of the hybrid fungal pathogen *Verticillium longisporum* population causes stem striping in British oilseed rape. *Environmental Microbiology*, 19(10), 3997–4009. https://doi.org/10.1111/1462-2920.13801

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development*, 35, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. (2020). Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow. *Current Protocols in Microbiology*, 59(1), 1–27. https://doi.org/10.1002/cpmc.128

Feurtey, A., Lorrain, C., Croll, D., Eschenbrenner, C., Freitag, M., Habig, M., … Stukenbrock, E. H. (2020). Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*. *BMC Genomics*, 21(1), 588. https://doi.org/10.1186/s12864-020-06871-w

Feurtey, A., Stevens, D. M., Stephan, W., & Stukenbrock, E. H. (2019). Interspecific gene exchange introduces high genetic variability in crop pathogen. *Genome Biology and Evolution*, 11(11), 3095–3105. https://doi.org/10.1093/gbe/evz224

Feurtey, A., & Stukenbrock, E. H. (2018). Interspecific gene exchange as a driver of adaptive evolution in fungi. *Annual Review of Microbiology*, 72(1), 377–398. https://doi.org/10.1146/annurev-micro-090817-062753

Fones, H., & Gurr, S. (2015). The impact of *Septoria tritici* Blotch disease on wheat: An EU perspective. *Fungal Genetics and Biology*, 79, 3–7. https://doi.org/10.1016/j.fgb.2015.04.004

Friesen, T. L., Stukenbrock, E. H., Liu, Z., Meinhardt, S., Ling, H., Faris, J. D., … Oliver, R. P. (2006). Emergence of a new disease as a result of interspecific virulence gene transfer. Nature Genetics, 38(8), 953. https://doi.org/10.1038/ng1839

Gillespie, J. H. (1994). The causes of molecular evolution (Vol. 2). Oxford University Press.

Giraud, T., Gladieux, P., & Gavrilets, S. (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology and Evolution*, 25(7), 387–395. https://doi.org/10.1016/j.tree.2010.03.006

Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D. M., & Hood, M. E. (2008). Speciation in fungi. *Fungal Genetics and Biology*, 45(6), 791–802. https://doi.org/10.1016/j.fgb.2008.02.001

Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J. L. N., Nhani, A., … Fournier, E. (2018). Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *MBio*, 9(1). https://doi.org/10.1128/mBio.01219-17

Gladieux, P., GuÉrin, F., Giraud, T., Caffier, V., Lemaire, C., Parisi, L., … Le Cam, B. (2011). Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology*, 20(21), 4521–4532. https://doi.org/10.1111/j.1365-294X.2011.05288.x

Glémin, S., Scornavacca, C., Dainat, J., Burgarella, C., Viader, V., Ardisson, M., … Ranwez, V. (2019). Pervasive hybridizations in the history of wheat relatives. *Science Advances*, 5(5). https://doi.org/10.1126/sciadv.aav9188

Goodwin, S. B., M'Barek, S. Ben, Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., … Kema, G. H. J. (2011). Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics*, 7(6). https://doi.org/10.1371/journal.pgen.1002070

Grünwald, N. J., McDonald, B. A., & Milgroom, M. G. (2016). Population genomics of fungal and oomycete pathogens. *Annual Review of Phytopathology*, 54(1), 323–346. https://doi.org/10.1146/annurev-phyto-080614-115913

Hansen, E. M. (1987). Speciation in plant pathogenic fungi: the influence of agricultural practice. *Canadian Journal of Plant Pathology*. 9:4, 403-410, https://doi.org/10.1080/07060661.2020.12103320

Islam, M. T., Croll, D., Gladieux, P., Soanes, D. M., Persoons, A., Bhattacharjee, P., … Kamoun, S. (2016). Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biology*, 14(1), 1–11. https://doi.org/10.1186/s12915-016-0309-7

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323–329. https://doi.org/10.1038/nature05286

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990. https://doi.org/10.1038/nature06536

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129), 624–626.

King, J. L., & Jukes, T. H. (1969). Non-Darwinian evolution: most evolutionary change in proteins may be due to neutral mutations and genetic drift. *Science*, 164(3881), 788-798. https://doi.org/10.1126/science.164.3881.788

Kohn, L. M. (2005). Mechanisms of fungal speciation. *Annual Review of Phytopathology*, 43(1), 279–308. https://doi.org/10.1146/annurev.phyto.43.040204.135958

Leroy, T., Caffier, V., Celton, J. M., Anger, N., Durel, C. E., Lemaire, C., & Le Cam, B. (2016). When virulence originates from nonagricultural hosts: Evolutionary and epidemiological consequences of introgressions following secondary contacts in *Venturia inaequalis*. *New Phytologist*, 210(4), 1443–1452. https://doi.org/10.1111/nph.13873

Linde, C. C., Zhan, J., & McDonald, B. A. (2002). Population structure of *Mycosphaerella graminicola*: from lesions to continents. *Phytopathology*, 92(9), 946–955. https://doi.org/10.1094/phyto.2002.92.9.946

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., … Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66(1), 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

McCoy, K. D. (2003). Sympatric speciation in parasites–what is sympatry? *Trends in Parasitology*, 19(9), 400–404. https://doi.org/10.1016/S1471-4922(03)00194-6

McNew, G. L. (1960). The nature, origin, and evolution of parasitism. Plant Pathology: An Advanced Treatise, 2, 19–69.

Menardo, F., Praz, C. R., Wyder, S., Ben-David, R., Bourras, S., Matsumae, H., … Keller, B. (2016). Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. *Nature Genetics*, 48(2), 201–205. https://doi.org/10.1038/ng.3485

Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, 15(12), 756–771. https://doi.org/10.1038/nrmicro.2017.76

Money, N. P. (2006). The triumph of the fungi: a rotten history. Oxford University Press.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428), 96–98. https://doi.org/10.1038/246096a0

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., & Croll, D. (2017). Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. *Frontiers in Plant Science*, 8, 119. https://doi.org/10.3389/fpls.2017.00119

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10(6), 417–430. https://doi.org/10.1038/nrmicro2790

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., & Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*, 3(6), 429–441. https://doi.org/10.1038/nrg817

Scholthof, K. B. G. (2007). The disease triangle: pathogens, the environment and society. *Nature Reviews Microbiology*, 5(2), 152–156. https://doi.org/10.1038/nrmicro1596

Schumann, G. L. (1991). Plant diseases: their biology and social impact. APS press.

Slippers, B., Stenlid, J., & Wingfield, M. J. (2005). Emerging pathogens: fungal host jumps following anthropogenic introduction. *Trends in Ecology & Evolution*, 20(8), 420–421. https://doi.org/10.1016/j.tree.2005.05.002

Stajich, J. E. (2017). Fungal genomes and insights into the evolution of the kingdom. *Microbiology Spectrum*, 5(4). https://doi.org/10.1128/microbiolspec.FUNK-0055-2016

Stergiopoulos, I., & de Wit, P. J. G. M. (2009). Fungal effector proteins. *Annual Review of Phytopathology*, 47(1), 233–263. https://doi.org/10.1146/annurev.phyto.112408.132637

Stukenbrock, E. H., Christiansen, F. B., Hansen, T. T., Dutheil, J. Y., & Schierup, M. H. (2012). Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences*, 109(27), 10954–10959. https://doi.org/10.1073/pnas.1201403109

Stukenbrock, E.H., & McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions*, 22(4), 371–380. https://doi.org/10.1094/MPMI-22-4-0371

Stukenbrock, E. H. (2013). Evolution, selection and isolation: A genomic view of speciation in fungal plant pathogens. *New Phytologist*, 199(4), 895–907. https://doi.org/10.1111/nph.12374

Stukenbrock, E. H., Banke, S., Javan-Nikkhah, M., & McDonald, B. A. (2007). Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, 24(2), 398–411. https://doi.org/10.1093/molbev/msl169

Stukenbrock, E. H., & Bataillon, T. (2012). A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. *PLoS Pathogens*, 8(9), 1–4. https://doi.org/10.1371/journal.ppat.1002893

Stukenbrock, E. H., & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, 46(1), 75–100. https://doi.org/10.1146/annurev.phyto.010708.154114

Stukenbrock, E. H., Quaedvlieg, W., Javan-Nikhah, M., Zala, M., Crous, P. W., & McDonald, B. A. (2012). *Zymoseptoria ardabiliae* and *Z. pseudotritici*, two progenitor species of the *Septoria tritici* leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella graminicola*). *Mycologia*, 104(6), 1397–1407. https://doi.org/10.3852/11-374

Stukenbrock, E. H, Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M., … Schierup, M. H. (2011). The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research*, 21(12), 2157–2166. https://doi.org/10.1101/gr.118851.110

Tellier, A., Moreno-Gámez, S., & Stephan, W. (2014). Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution*, 68(8), 2211–2224. https://doi.org/10.1111/evo.12427

Torriani, S. F. F., Melichar, J. P. E., Mills, C., Pain, N., Sierotzki, H., & Courbot, M. (2015). *Zymoseptoria tritici*: A major threat to wheat production, integrated approaches to control. *Fungal Genetics and Biology*, 79, 8–12. https://doi.org/https://doi.org/10.1016/j.fgb.2015.04.010

Woolhouse, M. E. J., Haydon, D. T., & Antia, R. (2005). Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in Ecology & Evolution*, 20(5), 238–244. https://doi.org/https://doi.org/10.1016/j.tree.2005.02.009

Yoshida, K., Saunders, D. G. O., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., … Terauchi, R. (2016). Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics*, 17(1), 1–18. https://doi.org/10.1186/s12864-016-2690-6

Zhong, Z., Norvienyeku, J., Chen, M., Bao, J., Lin, L., Chen, L., … Wang, Z. (2016). Directional selection from host plants is a major force driving host specificity in *Magnaporthe* species. *Scientific Reports*, 6, 1–13. https://doi.org/10.1038/srep25591

Acknowledgments

"...knowledge emerges only through invention and reinvention, the restless, impatient, continuing, hopeful inquiry human beings pursue in the world, with the world, and with each other."

- Paulo Freire, Pedagogy of the Oppressed (1970)

Such an incredible and challenging journey it has been. I would have never thought 11 years ago that the guy starting his bachelor in biology in south of Brazil would be finishing his PhD in northern Germany. It is a strange and also a fantastic feeling to remember all the pathways that life, and particularly education, has taken me. I feel extremely privileged to have had the opportunity to study and focus my professional career in the diverse and exciting scientific field. In this regard, I would like thank all educators, supervisors and colleagues that contributed to my education and shaped the professional that I am today.

During my PhD, there are also many people that would like to thank and that in one way or another allowed me to learn and grow as a scientist.

First of all, I would like to thank my supervisor Eva Stukenbrock for all her guidance and patience. Thank you so much for providing me the opportunity to work with this fantastic project and for allowing me to grow both professionally and personally. I should also thank you for the incredible respect, professionalism and also kindness that you have always treated me since I came to do an interview with you back in 2018. It has been such a great journey to be part of the Environmental Genomics team and I will never forget these years. For sure you have not only guided me through a completely new set of kills but have also showed me how to be a competent, thoughtful and welcoming supervisor. Thank you very much for everything.

Likewise, I would like to thank my second direct supervisor Janine Haueisen. Thank you very much for teaching me everything you could since my first day in the lab and for always having words of encouragement and kindness. Thank you also for making me feel very welcomed in the group and for always be supportive in different aspects of work and life. Having you here for sure made all the difference – thank you very much!

To my Thesis Advisory Committee (TAC) members, Diethard Tautz and Hinrich Schulenburg, thank you very much for the guidance, wise words and help in different aspects of my PhD. Thank you to thesis examination committee for your commitment, time and interest in my PhD research.

Thank you also to Angela Donner for making all the administrative MPI processes run as smooth as possible and to the really nice and helpful people I shared great moments in the MPI in Plön.

Special thanks also to the International Max Planck Research School (IMPRS) for providing excellent opportunities to grow as a professional and scientist.

I would also like to take the opportunity to thank all colleagues and collaborators that have contributed to the different projects composing my PhD. Thank you all very much for all the questions asked and answered, and for all the objectives that you have helped to accomplish. This PhD for sure would not be possible without your help and advice.

A big thank you to all the current and past colleagues and friends of the EnvGen team for the camaraderie insight and outside the lab, for the very friendly and supportive atmosphere and for always be there for random discussions, questions and good laughs. Special thanks to Lizel, Eli, Danilo, Carolina, Idalia, Victor, Judith, Jovan and Michael for sharing most of the conversations and memorable moments of my PhD. Special thanks also to Andrea for making all bureaucracy easy to deal with and for the very nice atmosphere you bring in. Thank you also to the great students I had the opportunity to supervise: David, Freya, Frauke and Rune. Thank you for allowing me to be part of your education, for allowing me to learn with you, and for contributing to different aspects of this dissertation.

Para toda minha família no Brasil, obrigado por tanto. Obrigado por me permitir crescer e seguir o caminho que estou trilhando hoje. Obrigado por todos os esforços, liberdade, amor e conforto que vocês sempre me deram. Obrigado por acreditarem em mim e por permitirem que eu alcançasse meus sonhos, mesmo que em terras distantes. Essa conquista também é de vocês.

Para todos meus amigos de longa data no Brasil e fora dele, obrigado por fazerem parte do meu dia a dia e por permitirem me fazer parte do dia a dia de vocês, mesmo que por curtas mensagens e ligações. Aos meus queridos amigos de Kiel, obrigado por fazer minha vida aqui mais feliz e por sempre trazerem um pedaço de "lar" com vocês. Com certeza isso fez e faz muito diferença.

Ao meu parceiro Romulo, obrigado por tanto amor e carinho compartilhado. Obrigado por sempre estar comigo nas horas boas e ruins e por sempre me ensinar tanto, dividindo sonhos, angústias e alegrias de dentro e fora do doutorado. Tudo ficou e fica mais colorido e leve contigo por perto. Obrigado também a nossa querida companheirinha Fajka pelo carinho em formas de miados, arranhões e ronronados.

Declaration of Author's Contributions

This dissertation consists of four chapters in the form of published and unpublished manuscripts. Specific author's contributions for each chapter are described below.

Chapter I – Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow

Wagner C. Fagundes, Janine Haueisen and Eva H. Stukenbrock

Chapter I was published on Current Protocols in Microbiology in 2020.

Fagundes, W. C., Haueisen, J., & Stukenbrock, E. H. (2020). Dissecting the biology of the fungal wheat pathogen *Zymoseptoria tritici*: a laboratory workflow. Current Protocols in Microbiology, 59(1), e128. https://doi.org/10.1002/cpmc.128

WCF: Conceptualization; investigation; methodology; validation; visualization; writing-original draft; writing-review & editing. JH: Conceptualization; formal analysis; investigation; methodology; supervision; writing-review & editing. EHS: Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision; writing-review & editing

Chapter II – Host specificity defines a new fungal plant pathogen population

Wagner C. Fagundes, Idalia C. Rojas Barrera, Rune Hansen, Alice Feurtey, Fatemeh Salimi, Janine Haueisen, Alireza Alizadeh and Eva H. Stukenbrock

Conceptualization: WCF, JH, EHS; Samples collection: FS, AA; *Zymoseptoria tritici* isolation: WCF, JH; DNA extractions and clone correction: WCF; Genome sequencing and genomics analyses: WCF, AF; Demography analyses: ICRB; Greenhouse infection assays: WCF, RH; Herbarium amplicon sequencing and analyses: WCF, RH; Preparation and writing of manuscript: WCF. Editing of original manuscript: WCF, EHS, JH and ICRB.

Chapter III – A new chromosome is described in fungal species infecting wild grasses

Wagner C. Fagundes, Janine Haueisen, Mareike Möller, Alice Feurtey, Rune Hansen, Fatemeh Salimi, Alireza Alizadeh and Eva H. Stukenbrock

Conceptualization: WCF, JH, EHS; Zt469 sample collection: FS, AA; Whole-genome DNA extractions: WCF, JH; Genome sequencing and genomics analyses: WCF, AF; ChIP-DNA and *in vitro* RNA isolation and sequencing: MM; Greenhouse infection assays: WCF, RH; Confocal microscopy analyses: RH, JH; *In planta* RNA isolation: WCF, RH; Preparation and writing of manuscript: WCF. Editing of original manuscript: WCF, EHS, JH, RH and MM.

Chapter IV – Functional genomics of candidate genes under host-driven selection

Wagner C. Fagundes, Janine Haueisen, Frauke Caliebe, Fatemeh Salimi, Alireza Alizadeh and Eva H. Stukenbrock

Conceptualization: WCF, JH, EHS; Zt495 sample collection: FS, AA; Plasmid design: WCF, FC, JH; Binary plasmids construction: WCF, FC, JH; *Agrobacterium*-mediated Transformation of fungal cells: WCF, FC, JH; DNA extractions and southern blot analyses: WCF, FC; Genomics analyses: WCF, FC; Preparation and writing of manuscript: WCF. Editing of original manuscript: WCF, EHS, JH.

Affidavit

I hereby declare that this dissertation

- regarding content and design is the product of my own work under guidance of my supervisor Prof. Dr. Eva H. Stukenbrock. I used no other tools or sources beyond the ones cited. Contributions of other authors are listed in the "Declaration of author's contribution" section of this dissertation.

- has not been submitted elsewhere either partially or wholly as part of a doctoral degree to another examining body and no other materials are published or submitted for publication than those already indicated in the dissertation.

- has been conducted and prepared following the Rules of Good Scientific Practice of the German Research Foundation.

- No academic degree has been withdrawn.

Wagner Calegari Fagundes

Kiel, 13.12.2022