Population and evolutionary genomics of two economically important fungal plant pathogens

Dissertation

In fulfillment 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

Demetris Taliadoros

November 2023

First examiner (Supervisor): Prof. Dr. rer. nat Eva Holtgrewe Stukenbrock

Second examiner: Prof. Dr. Julien Dutheil

Member of the of the examination committee: Prof. Dr. Michael Seidl

Examination date: 11th of January 2024

Table of Contents

Summary

The Global Report on Food Crisis 2023, issued by the United Nations, states that humanity faces an acute food insecurity deepening over the past four years. In parallel, it is estimated that rapidly increasing human populations will reach 10 billion people by 2050. These factors call for urgent action for a more sustainable and, at the same time, more productive agriculture. Fungal pathogens cause devasting diseases in crop plants and immensely impact crop yields. While contemporary agricultural practices, such as fungicides and resistant varieties, are employed to control fungal disease emergence, catastrophic epidemics with devastating results still occur. Recent examples of such epidemics are the Asian soybean rust (*Phakopsora pachyrhizi*) and the wheat blast outbreaks *Magnoporthe oryzae* that occurred in South America and Asia.

Understanding the factors that drive the evolution and emergence of fungal plant pathogens can provide essential information for better epidemiological surveillance of these diseases. With the rapid increase in genomic global data of major plant pathogens, population genomics studies have deepened our understanding of the processes and mechanisms that give rise to catastrophic crop diseases.

In Chapter 1, I reviewed contemporary evolutionary analyses to predict functionally relevant traits in fungal plant pathogens. I concluded that

evolutionary analyses that include both polymorphism and divergence data can be used to disentangle the type of selection acting along the genome. Furthermore, Plant pathogens can have complex life cycles, and theoretical developments are necessary to improve our power to detect selection and understand pathogen evolution taking into account their particular genetic structures.

In Chapter 2, I analyze the population genetic structure and demographic history of the pathogen that causes barley net blotch, *Pyrenophora teres* f. *teres.* My research highlights the recent emergence of a significant crop pathogen in tandem with the domestication and expansion of its host, underscoring the influence of human activities and agriculture on the evolution and dissemination of new diseases. Furthermore, I scan the genomic landscape to identify regions evolving under strong positive selection. I found an extensive overlap between the regions identified in this study and previously reported virulence factors, highlighting the importance of the host driving the pathogen's evolution.

Chapter 3 analyzes the population genetic structure of the beet cercospora leaf spot agent, *Cercospora beticola.* Using bioinformatic tools, I uncovered a distinct subdivision of *C. beticola* into various lineages, each associated with domesticated or wild host plants, with limited genetic exchanges. We further report multiple historic introductions of the pathogen into the agricultural

ecosystem. Alarmingly, I found alleles associated to resistance to a major fungicide class aboundant at a world-wide scale. These findings contribute to the growing body of research emphasizing the pivotal role of human activities, including domestication, trade, and agricultural practices, in the spread of plant pathogens.

In Chapter 4, I focused on a N. American population of *C. beticola* that showed high levels of fungicide resistance. I used a maximum-likelihood simulation-based method to infer the demographic history of the population. Subsequently, we scanned the genomic landscape for selective sweeps. Our analysis supports that the population has undergone a strong bottleneck recently. Furthermore, genomic loci, associated with resistance to a commonly used fungicide class, were located in regions that have undergone a selective sweep. Taken all together, our results lead to a worrying conclusion: sterol demethylation inhibitor (DMI) fungicide resistance is evolving fast in the *C. beticola* population in N. America.

Zusammenfassung

Der von den Vereinten Nationen herausgegebene Weltbericht über die Ernährungskrise 2023 stellt fest, dass die Menschheit mit einer akuten Ernährungsunsicherheit konfrontiert ist, die sich in den letzten vier Jahren verschärft hat. Gleichzeitig wird geschätzt, dass die schnell wachsende Bevölkerung bis 2050 auf 10 Milliarden Menschen ansteigen wird. Diese Faktoren machen dringende Maßnahmen für eine nachhaltigere und gleichzeitig produktivere Landwirtschaft erforderlich. Pilzliche Krankheitserreger verursachen verheerende Krankheiten bei Nutzpflanzen und beeinträchtigen die Ernteerträge immens. Obwohl moderne landwirtschaftliche Verfahren wie Fungizide und resistente Sorten eingesetzt werden, um das Auftreten von Pilzkrankheiten zu kontrollieren, kommt es immer noch zu katastrophalen Epidemien mit verheerenden Folgen. Jüngste Beispiele für solche Epidemien sind der Asiatische Sojarost (Phakopsora pachyrhizi) und der Ausbruch der Weizenblast Magnoporthe oryzae, die in Südamerika und Asien auftraten.

Das Verständnis der Faktoren, die die Evolution und das Auftreten von pilzlichen Pflanzenpathogenen vorantreiben, kann wichtige Informationen für eine bessere epidemiologische Überwachung dieser Krankheiten liefern. Mit der raschen Zunahme der globalen Genomdaten wichtiger Pflanzenpathogene

haben populationsgenomische Studien unser Verständnis der Prozesse und Mechanismen, die zu katastrophalen Pflanzenkrankheiten führen, vertieft.

In Kapitel 1 haben wir aktuelle evolutionäre Analysen zur Vorhersage funktionell relevanter Merkmale bei pilzlichen Pflanzenpathogenen untersucht. Wir kamen zu dem Schluss, dass evolutionäre Analysen, die sowohl Polymorphismus- als auch Divergenzdaten umfassen, dazu verwendet werden können, die Art der entlang des Genoms wirkenden Selektion zu entschlüsseln. Darüber hinaus können Pflanzenpathogene komplexe Lebenszyklen haben, und theoretische Entwicklungen sind notwendig, um unsere Möglichkeiten zum Nachweis von Selektion und zum Verständnis der Pathogenevolution zu verbessern.

In Kapitel 2 analysieren wir die genetische Struktur der Population und die demografische Geschichte des Erregers der Netzfleckenkrankheit der Gerste, *Pyrenophora teres* f. *teres*. Unsere Forschung zeigt das jüngste Auftreten eines bedeutenden Kulturpflanzenerregers in Verbindung mit der Domestizierung und Ausbreitung seines Wirts und unterstreicht den Einfluss menschlicher Aktivitäten und der Landwirtschaft auf die Entwicklung und Verbreitung neuer Krankheiten. Darüber hinaus durchforsten wir die genomische Landschaft, um Regionen zu identifizieren, die sich unter starker positiver Selektion entwickeln. Wir fanden eine weitreichende Überschneidung zwischen den von uns

identifizierten Regionen und den zuvor berichteten Virulenzfaktoren, was die Bedeutung des Wirts für die Evolution des Erregers unterstreicht.

Kapitel 3 analysiert die genetische Struktur der Population des Erregers der Rüben-Cercospora-Blattflecken, *Cercospora beticola*. Mit Hilfe bioinformatischer Werkzeuge haben wir eine deutliche Unterteilung von C. beticola in verschiedene Linien aufgedeckt, die jeweils mit domestizierten oder wilden Wirtspflanzen assoziiert sind und einen begrenzten genetischen Austausch aufweisen. Außerdem berichten wir von mehreren historischen Einschleppungen des Erregers in das landwirtschaftliche Ökosystem. Schließlich fanden wir alarmierende Hinweise auf eine weit verbreitete globale Resistenz gegen eine wichtige Fungizidfamilie. Diese Ergebnisse tragen zur wachsenden Zahl von Forschungsarbeiten bei, die die zentrale Rolle menschlicher Aktivitäten, einschließlich Domestizierung, Migration und Handel, bei der Verbreitung von Pflanzenpathogenen betonen.

In Kapitel 4 konzentrierten wir uns auf eine nordamerikanische Population von C. beticola, die ein hohes Maß an Fungizidresistenz aufwies. Wir verwendeten eine auf maximaler Wahrscheinlichkeit basierende Simulationsmethode, um die demografische Geschichte dieser Population zu ermitteln. Anschließend untersuchten wir die genomische Landschaft auf selektive Ausbreitungen. Unsere Analyse belegt, dass die Population in letzter Zeit einen starken Engpass

erlebt hat. Darüber hinaus befanden sich genomische Loci, von denen zuvor berichtet wurde, dass sie mit der Resistenz gegen eine häufig verwendete Fungizidklasse in Verbindung stehen, in Regionen, die einem selektiven Sweep unterworfen waren. Alles in allem führen unsere Ergebnisse zu einer alarmierenden Schlussfolgerung: Die Fungizidresistenz gegen Sterol-Demethylierungsinhibitoren (DMI) entwickelt sich bei der C. beticola-Population in Nordamerika rasch weiter.

General Introduction

General Introduction

Fungi in the agroecosystem

Despite the scientific and technological advances in agriculture over the last centuries, the challenge of providing affordable food for the current global population persists. According to the United Nations Global Report on Food Crisis 2023, 252 million people faced high levels of acute insecurity in 2022. In addition, there is a rapid increase in the human population, which is expected to increase by 25% in the next thirty years (reaching 10 billion people) (1). Considering these numbers, sustainable solutions are urgently needed to ensure food security for an ever-growing human population.

A major challenge that farmers and agronomists have to face is plant pathogens. Humanity has witnessed horrific crop epidemics in the past that even led to the collapse of societies, like the Irish potato famine (1845-1852) caused by the potato late blight agent, *Phytophthora infestans.* Unfortunately, crop epidemics are not a thing of the past. Recent examples of such epidemics are the Asian soybean rust (*Phakopsora pachyrhizi*) and the wheat blast outbreaks *Magnoporthe oryzae* that occurred in South America and Asia. A growing body of work shows that using evolutionary theory and recent advances in genomics can provide new insights into the mechanisms and processes that lead to the emergence and successful adaptation of pathogens to specific conditions. These

studies prove that population genomics can be a powerful tool in our efforts to develop efficient disease management strategies to limit disease epidemics(2–4).

Host domestication, migration, and trade

The domestication of plants, along with the expansion of global trade and agriculture, are significant forces driving plant pathogens' emergence (5). Understanding how pathogens adapt and specialize on domesticated plants and spread within the agricultural ecosystem is critical for effectively managing disease-causing agents. The emergence of a novel crop pathogen can result from either the development of the ability to infect a new host (host jump) or a process known as host-tracking, where the pathogen emerges as a conscience of the domestication of the host (6). Population genetic and evolutionary research has been crucial in tracing significant crop pathogens' roots and historical spread. For instance, the fungal wheat pathogen *Zymoseptoria tritici* originated during the early stages of wheat domestication. Subsequently, it spread alongside wheat farming practices, initially during the Neolithic period and later with the migration of European populations (7,8).

Similarly, the probable center of origin for the smut fungus *Ustilago maydis*, which infects maize, is likely to coincide with the center of maize domestication in Central and South America (9). Numerous additional instances underscore the

critical role of domestication and agricultural trade, shaping the emergence and dispersal of crop pathogens (4,6,10). In the recent past, the breeding of new crop species has accelerated the evolution of novel pathogen species (6). A prime example is the emergence of an unknown powdery mildew lineage impacting the hybrid crop triticale (11).

Population genomics provide essential information for better epidemic surveillance

Population genomics has been established as a valuable tool in our endeavors to gain insights into the global-scale evolutionary pathways of economically important plant pathogens. (3,8,12–14). Such investigations yield valuable insights into how pathogen populations adapt to their local environments and host plants. These insights, in turn, provide critical information directly relevant to disease epidemiology. For instance, by studying genetically diverse pathogen populations worldwide, we can formulate and test hypotheses regarding the species' origin, patterns of dispersion, host specialization, and, more recently, the development of resistance to specific fungicide classes (15–17). This knowledge can serve as a guide for shaping strategies aimed at disease management and for implementing measures to control the movement of plant

materials and the application of specific fungicides, all to minimize the occurrence of disease outbreaks.

Here, I employed population genetic theory and bioinformatic tools to unravel the global population structure, demographic history, and genomic elements facilitating the rapid adaptation of two major fungal plant pathogens in the agro-ecosystem. To this end, our methodologies involved extensive global sampling and the whole genome sequencing of pathogen isolates. Evolutionary analyses were predominantly based in high-quality single nucleotide variation (SNVs). By using data on allele frequencies and the density of variation, I explored the evolutionary dynamics shaping the genetic sequence.

The barley-*Pyrenophora* pathosystem

The host

Plant domestication dates back to approximately 12,000 years ago (18). One of the first domesticated crops was barley, which was domesticated in the Fertile Crescent. Subsequently, barley was brought to North Africa and Eurasia by migration waves of Neolithic farmers (19). In addition, North Africa has been proposed as a center of diversity of wild barley (20–23). Over the past few centuries, barley has undergone additional dispersion as it was introduced to the Americas, Australia, and South Africa by European migrants. (24). In general, the

historical dispersal of barley is closely associated with human activities and mirrors the expansion of cereal cultivation along both ancient and contemporary trade routes.

Disease cycle

Pyrenophora teres f. *teres* is a widespread fungal pathogen of barley causing the disease "net form net blotch" (NFNB) (Figure 1). NFNB is categorized as a stubble-borne diseases, as the fungus typically forms ascocarps (pseudothecia) as overseasoning structures on the infected barley remnants left on the surface after harvest (25). In these mature and fertile pseudothecia, club-shaped and bitunicate asci develop, each containing eight ascospores. Mature ascospores are actively released and spread by wind, serving as the primary source of innoculum early in the growing season. Often, seed-borne mycelium and conidia removed from barley residue or another host can also be the primary source of early-season infection. Following the initial colonization, the fungus produces a large quantity of conidia, which are secondary sources of inoculum and infection. These conidia are carried on conidiophores, slightly swollen at the base, and typically emerge singly or in groups of two or three. Conidia production occurs throughout the growing season and can be dispersed locally by strong wind or rain to cause new infections on nearby plants, or they may be transported longer distances, potentially reaching new barley fields. The dispersion, germination,

and successful infection of conidia are significantly influenced by factors such as relative humidity, temperature, and other environmental conditions (25). Multiple secondary infection cycles can occur during the growing season, leading to high disease severity in susceptible plants if environmental conditions are favorable. Towards the end of the growing season, the fungus colonizes senescent tissue and ultimately forms pseudothecia, protective teleomorph structures used for over-seasoning (25).

Infection of susceptible barley leaves occurs by the production of specialized infection structures, appressoria, whereby the fungus penetrates the cuticle and cell wall of epidermal cells. The characteristic net blotch symptoms arise through necrosis, which is induced rapidly after initial pathogen invasion of the host leaf (25). Net form net blotch disease occurs in all barley-producing regions of the world and has been reported in several African countries (26–28), West and East Asia (29,30), Europe (31), North and South America (32,33), and Australia (34). The pathogen has been known to humans for centuries, initially described as *Helminthosporium teres* (Sacc.); however, insights into the population biology and demographic history of *P. teres* f. *teres* are scarce.

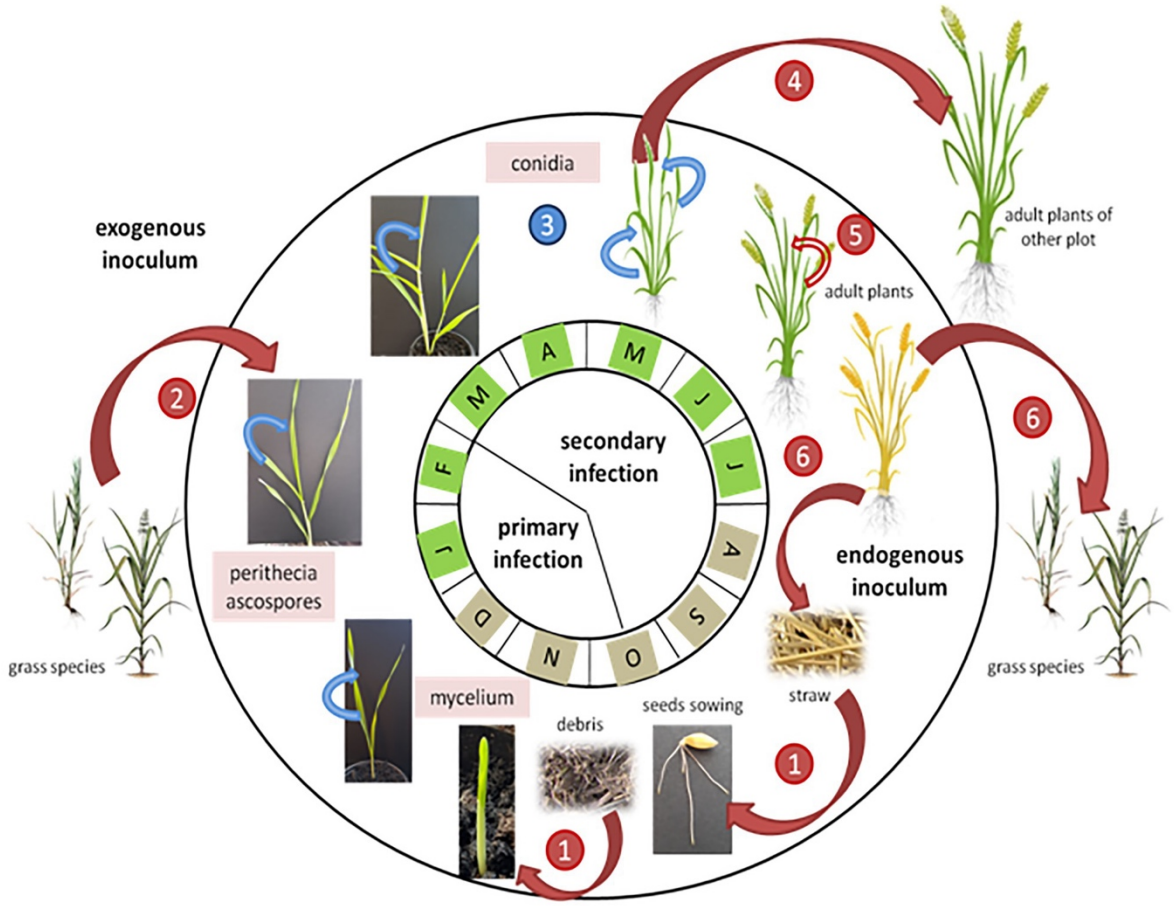


Figure 1. life cycle of *Pyrenophora teres* adapted from Backes et al. 2021. Red arrows signify wind-dispersed infections of P. teres, while blue arrows denote splash-dispersed infections. Months depicted in brown indicate that the inoculum source originates from swarming debris on the soil, whereas months in green suggest the primary inoculum source is from the aerial parts of the plants. The numerical annotations correspond to different stages of the pathogen's infection: 1) infection by P. teres ascospores on infected barley debris; 2) mycelium from grass species infects young barley plants; 3) progression of net blotch disease from the bottom to the top of the barley plant; 4) wind-disseminated conidia contaminating other barley plants; 5) heavily infected crops exhibiting ear abortion; and 6) *Pyrenophora teres* colonizing senescent tissues and generating perithecia on straw and grass species.

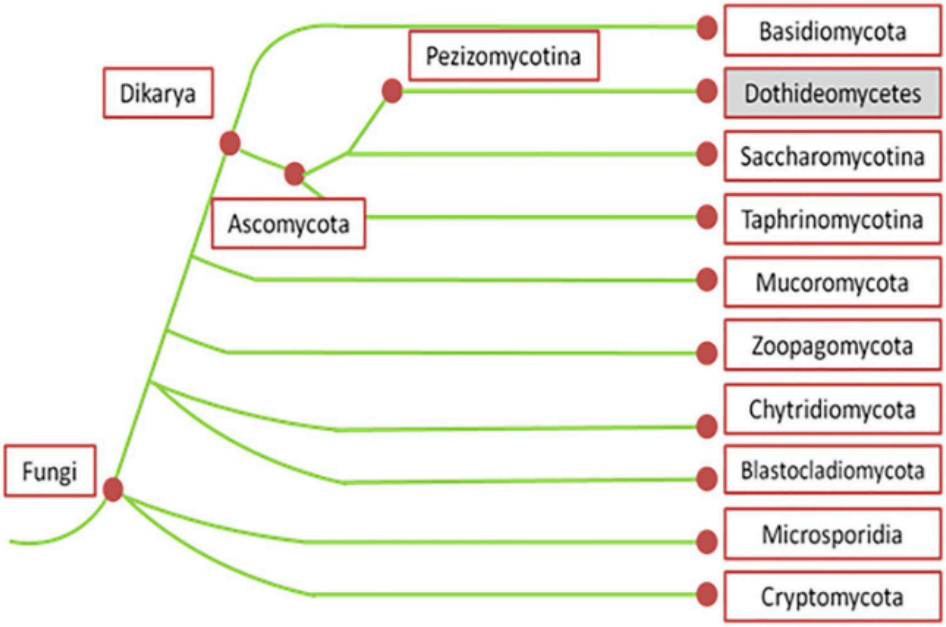


Figure 2: Both *P. teres* f. *teres* and *C. beticola* belong to the ascomycete clase Dothideomycetes. (figure adapted from Backes et al., 2021). They further split in the order level with *P. teres* f. *teres* belonging to pleosporales and *C. beticola* belonging to Capnodiales.

The beet-*Cercospora* pathosystem

The host

As with barley, domesticated beets (*Beta vulgaris* subsp. *vulgaris*) and their wild precursor, the sea beet (*Beta vulgaris* subsp. *maritima*), are closely intertwined with human civilization. It originates from the Mediterranean basin. (35). The earliest documented beet cultivations can be traced back to the ancient Greeks and Romans. Interestingly, the beetroot, the part of the beet plant that we typically associate with beets today, was only developed during the later period

of the Roman Empire. (35). In the 16th century, the cultivation of beets had extended to Northern Europe, and fodder beets had become a common source of animal feed. However, it was not until the 18th century that a German chemist named Andreas Marggraf made a pivotal discovery: he found that sugar could be extracted from beets. This breakthrough laid the groundwork for the development of the crop known as sugar beet (36). This discovery marked a crucial turning point, as beets evolved into a critical source of sugar, especially during the Napoleonic wars. Cultivated beets have spread over the last few centuries due to human migration, particularly to North America. Taken together, there has been a rapid diversification and global dispersal of beets closely associated with human history over the past two millennia.

Disease cycle

Cercospora beticola is the pathogen responsible for causing Cercospora leaf spot (CLS) in *Beta vulgaris* L., which encompasses sugar beet, table beet, Swiss chard, and their wild precursor, the sea beet. This pathogen undergoes several cycles of asexual reproduction within a single growing season (36). Between sugar beet growing seasons, the fungus primarily survives through hyphal structures known as pseudostromata, which exhibit resistance to drying out. These specialized structures are composed of a mix of fungal tissue and remnants of host tissue and are located within the leaf's substomatal cavities. They can maintain viability

for as long as two years. Despite the absence of a recognized sexual form, C. beticola is classified as a heterothallic ascomycete fungus. This classification arises from identifying two alternate mating-type genes (MAT1-1-1 and MAT1-2-1) (37). Effective Cercospora leaf spot (CLS) management is essential for sugar beet production. Modern control strategies involve crop rotation, fungicide application, and the development of resistant varieties. To minimize the risk of CLS inoculum from previous crops, it's common to implement a three-year rotation period in sugar beet fields (36). Additionally, maintaining physical distance between fields can prove beneficial in reducing the spread of CLS among different areas. Despite the economic importance and the broad host range, including wild and domesticated plants, information about host-specialized *C. beticola* lineages is limited (36).

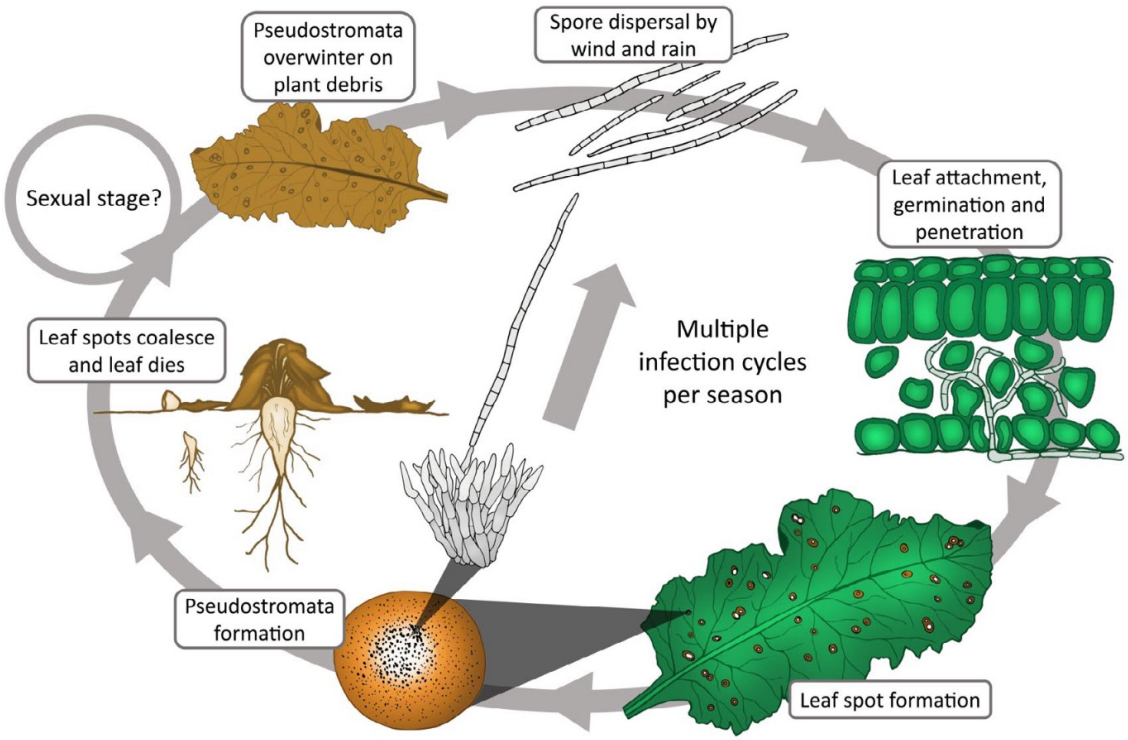


Figure 3. The disease cycle of C. beticola on sugar beet (adapted from Rangel et al. (2020)) commences with the penetration of dispersed conidia through stomata on the leaf surface, leading to intercellular hyphal growth. After seven days of infection, visible leaf spots emerge. Pseudostromata form within these lesions, generating spores through asexual reproduction. These pseudostromata can cause multiple infections throughout a growing season and can also overwinter in the soil, serving as inoculum for the subsequent growing season. It's noteworthy that the sexual stage of C. beticola has not been observed in the field.

Scope of the thesis

Exploring the evolutionary trajectories of crop pathogens can provide essential information in predicting future disease outbreaks. During this thesis, I studied the evolutionary and population histories of two pathogens, one adapted to an early domesticated host and the second adapted to a very recently domesticated

host. The primary goal of this thesis was to uncover the historical and contemporary factors as well as the genomic mechanisms and processes responsible for the emergence and successful adaptation of the *P. teres* f. *teres* and *C. beticola* populations in diverse agro-ecosystems around the world. In light of this knowledge, sustainable strategies can be designed to control pathogen populations to avert disease epidemics.

Chapter 1

"Using evolutionary analyses to predict functionally relevant traits in fungal plant pathogens."

Evolutionary analyses offer a means to uncover genes responsible for traits like virulence and local adaptation, including adaptability to agricultural intervention strategies. This chapter emphasizes the substantial role of evolutionary genomics in discovering traits associated with virulence and examining the ecology and adaptive evolution of plant-pathogen interactions. We specifically addressed the questions:

- What are the state-of-art methods for identifying natural selection in filamentous fungi?
- What are the limitations of these methods?

- Which genomic elements appear to be under selection in filamentous fungi?

Chapter 1 is published in Current Opinion in Microbiology, Volume 73, pages 1 - 17 (2023), doi.org/10.1016/j.mib.2022.102244.

-

Chapter 2

"Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history."

Pyrenophora teres f. *teres* is the causal agent of the net for the net blotch of barley. In this chapter, we present strong evidence that *P. teres* f. *teres* global population structure and demographic history is highly associated with the known history of the host, barley. Furthermore, we report genomic regions evolving under strong selection for each population to uncover the genetic basis of local adaptation of the pathogen. The specific aims of the study were to:

- Describe the global population structure of *Pyrenophora teres* f. *teres.*
- Identify the origin of the pathogen, and address its dispersal history?
- Identify genomic regions evolving under strong selection in different populations of the pathogen.
- Describe the biological relevance of the genomic elements located in these regions.

Chapter 2 is published as a preprint on *biorxiv* doi.org/10.1101/2023.07.28.550921 and is currently under peer review.

Chapter 3

Genome-wide evidence of host specialization in wild and farmland populations of the cercospora leaf spot pathogen, *Cercospora beticola*

Cercospora beticola is the causal agent of Cercospora Leaf Spot (CLS) in domesticated and wild beet. This chapter shows that *C. beticola* is subdivided into multiple lineages associated with domesticated and wild hosts. We further present evidence of limited genetic exchanges between them. Our genetic diversity and phylogenetic analyses provide evidence of a tandem demographic history between the host and the pathogen and multiple introductions of the pathogen to the agro-ecosystem.

- Describe the global population structure of *C. beticola.*
- What is the role of host specialization in population divergence?
- Identify genomic regions that are highly diverging between populations of the pathogen from different hosts.
- Describe the biological relevance of the genomic elements located in these regions.

Chapter 4

DMI Fungicide resistance and specialization to local barley varieties drives the adaptation of the North American *C. beticola* population

North America is one of the leading countries in sugarbeet cultivation. The Red River Valley (RRV) in North Dakota and Minnesota is the largest sugar beet production area in the United States. Demethylation inhibitor (DMI) fungicides are a critical tool in the management of CLS. Resistance to DMI fungicides was reported before in RRV. In this study, I scanned the genomic landscape of *C. beticola* isolates originated from RRV to identify genomic regions that have undergone a selective sweep. My results show that genomic loci associated with fungicide resistance are located in regions that have recently experienced a selective sweep. In this chapter we:

- Inferred the demographic history of an economically important *C. beticola* population.
- Scanned the genomic landscape to identify regions evolving under strong positive selection.
- Assessed the abundance of genomic loci associated with fungicide resistance and host specialization in the genomic regions evolving under strong positive selection.

Chapter 4 is published as parts of broader studies in Genome Biology and Evolution volume 12 issue 9, September 2021, (doi.org/10.1093/gbe/evab209),

and in Molecular Plant Pathology, volume 22, issue 3, March 2021 (DOI: 10.1111/mpp.13026).

Chapter I

Using evolutionary analyses to predict functionally relevant traits in fungal plant pathogens

Chapter I: Using evolutionary analyses to predict functionally relevant traits in fungal plant pathogens

Demetris Taliadoros and Eva H. Stukenbrock

Environmental Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön and Christian-Albrechts University of Kiel, 24118, Germany

This Article was published in Current Opinion in Microbiology, Volume 73, pages 1 - 17 (2023), doi.org/10.1016/j.mib.2022.102244

Abstract

Identifying traits involved in plant-pathogen interactions is one of the major objectives in molecular plant pathology. Evolutionary analyses may assist in the identification of genes encoding traits that are involved in virulence and local adaptation, including adaptation to agricultural intervention strategies. In the past decade, the number of available genome sequences of fungal plant pathogens has rapidly increased, providing a rich source for the discovery of functionally important genes as well as inference of species histories. Positive selection in the form of diversifying or directional selection leaves particular signatures in genome alignments and can be identified with statistical genetics methods. This review summarises the concepts and approaches used in evolutionary genomics and lists major discoveries related to plant-pathogen evolution. We underline the significant contribution of evolutionary genomics in discovering virulence-related traits and the study of plant-pathogen ecology and adaptive evolution.

Keywords: Plant disease, rapid evolution, effectors, natural selection, genome scans, selective sweep, balancing selection

Introduction

A hallmark of human history is extensive migration (Reich, 2018). Consequently, human populations and human-associated organisms, such as crops, ornamental plants, and livestock, have been moved around to foreign environments (Hulme, 2009). Alongside crop plants, plant-associated microorganisms, including pathogens, have been introduced to new environments (Santini et al., 2018). Today, globalisation has dramatically accelerated the speed whereby organisms are moved to new environments, and global trade with agricultural products is the primary driver of pathogen dissemination (Fisher et al., 2012; Santini et al., 2018).

Agricultural ecosystems differ from natural ecosystems; they are characterised by low species diversity and uniform ecological factors. Moreover, for pathogens, available hosts' spatial and temporal distributions differ significantly between natural and agricultural environments (McDonald & Stukenbrock, 2016). The internationalisation of breeding strategies and farming practices led to a homogenisation of the crop genotypes and interventions employed in geographically distant regions. Furthermore, especially in countries with intensified agriculture, there can be extensive input of pesticides, such as fungicides and herbicides, and a regular exchange of crop varieties to minimise disease impacts. Altogether, this creates uniform environments over large geographical regions, facilitating the spread and rapid evolution of crop-associated pathogens (Hulme et al., 2010; Stukenbrock & Mcdonald, 2008).

Plant pathogens exploit their host to obtain nutrients and space for reproduction and completion of their life cycle. They have evolved strategies to

overcome host defences, including physical and chemical barriers and resistance proteins that target molecules produced by the pathogen (Castro-moretti et al., 2020). The suppression of plant defences and the avoidance of recognition are achieved by specific pathogen-produced molecules called effectors. Effectors play a plethora of roles in manipulating host defences and interfering with the cell metabolism but can themselves also act as elicitors of plant defences (Lo Presti et al., 2015). Such antagonistic interactions between plants and their pathogens are major drivers of the diversification and coevolution of the molecules involved (i.e., the defence-related genes on the plant side and the effector genes on the pathogen side) (Möller & Stukenbrock, 2017).

A key objective in molecular plant pathology research is to uncover the underlying components involved in plant-pathogen interactions. As genes encoding these molecules are among the most rapidly evolving genes in plant and pathogen genomes, they show distinct signatures of evolution, such as positive selection that reflect the antagonistic plant-pathogen co-evolution (Castillo & Agathos, 2019; **Ebert et al., 2021; Karasov et al., 2014). Inference of rapid evolution in plant-pathogen genomes can thereby provide us with insights into adaptive processes of ecologically and agriculturally important traits (Beckerson et al., 2019; Guyon et al., 2014; Karasov et al., 2014). Today, whole-genome sequences from many agriculturally important pathogens provide a valuable resource for identifying such selection signatures across genomes and discovering new functionally relevant traits.

Signatures of selection along genomes

Positive selection favours – opposite to purifying selection - particular new mutations, which confer an adaptive advantage over the ancestral form of the sequence. In plant-pathogen interactions, it can, for example, be a mutation that

changes the affinity of an effector protein to its host target or a mutation that allows the escape of immune recognition. In agricultural ecosystems, farmers may introduce a new resistance gene in the field from one year to the other. Such a shift in the genetic composition of the host population can drive adaptive evolution in pathogen populations and favour isolates that carry new mutations in effector genes, allowing the pathogen to overcome or avoid the new host resistance (Derbyshire, 2020). The strength of selection operating on these virulence determinants can be enormous; i.e., in an agricultural field, only pathogen isolates with a particular effector allele can reproduce, and isolates carrying an incompatible allele are removed from the population. This type of evolutionary scenario has been termed “arms race” evolution and essentially refers to directional positive selection of new alleles in the host plant and the pathogen (Tellier et al., 2014).

Balancing selection is the type of selection that favours more than one particular allele in a population. This type of selection results from a “trench warfare” co-evolutionary dynamic (Tellier et al., 2014; Vitti et al., 2013) expected to occur in heterogeneous environments such as natural ecosystems. Here, ecological diversity and variation in the host genetic structure can select for different alleles in spatially separated pathogen populations that are specialised under distinct local conditions.

Positive selection of particular mutations leaves specific signatures in sequence data, which can be detected with statistical genetic analyses (Jensen et al., 2007). Some methods compare the pattern of genetic variation in a sample of sequences, independent of coding or non-coding properties, to identify deviation from the pattern expected under neutral evolution (Tajima, 1989; Fay & Wu, 2000). These methods compute and compare allele frequencies or the site frequency spectrum (SFS), the distribution of sites with distinct mutation

frequencies. One such measure is Tajima's D, which compares the number of pairwise nucleotide differences between isolates (π) and the number of segregating polymorphisms (θ_w) in a sequence alignment. An excess of rare alleles, potentially caused by selection, will inflate the θ_w value in comparison to π leading to a negative Tajima's D value. In contrast, an excess of alleles with intermediate frequencies results in a positive Tajima's D and can be interpreted as balancing selection. Local adaptation or divergent selection between pairs of populations can also be investigated by comparing allele frequencies across genome sequences. Fst is a measure that reflects variation in allele frequencies, whereby a high Fst value indicates population differentiation (Wright, 1949).

Population genomic data can also be used to infer recent positive selection by detecting selective sweeps (Möller & Stukenbrock, 2017). Selective sweeps are distinguished as regions with a local reduction in overall genetic variation, which was locally "reset" by the rapid spread of the selected mutation (**Figure 1**). Therefore, the region linked to the selected locus is characterised by recent mutations in low-frequency and a high extent of linkage disequilibrium. This pattern results from the genetic "hitch-hiking" of polymorphisms in the genomic vicinity of the selected allele. So-called "composite" methods can be used for this type of genome scan as they combine information from a genomic region rather than from a region with few genetic markers, and they integrate multiple statistics to add power to the inference of natural selection (Kim & Stephan, 2002; Kim & Nilsen, 2004). For fungal plant pathogens, genome scans based on the detection of selective sweeps have been used widely to identify functionally relevant traits (**Table 1**), which can also include regulatory sequences that are not recognised by genome annotation tools (e.g. Omrane et al., 2017).

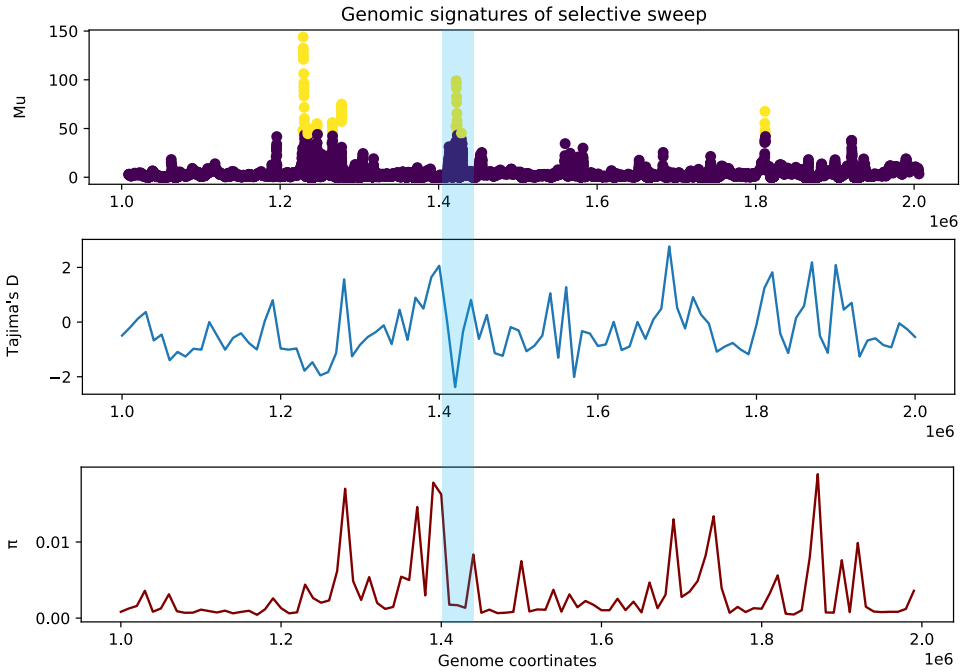


Figure 1: A selective sweep leaves a particular pattern of genetic variation. Population genomic data of the fungus *Cercospora beticola* was used to scan for selective sweeps. In this example a region identified as a selective sweep region by different test statistics are highlighted in blue. (a) Selective sweep map obtained by RaiSD [2]. The μ-statistic values were calculated in windows of 50 Kbp and plotted along chromosome 4 [59]. Significant outlier loci (99.95 quantile) are shown in yellow. RAiSD computes the μ statistic, a composite evaluation test that scores genomic regions by quantifying changes in the SFS, the levels of Linkage Disequlibrium (LD), and the amount of genetic diversity along the chromosome. (b) The Tajima's D statistic calculated in windows of 10 Kbp shows negative values in the region that has undergone a selective sweep. (c) Reduced values of nucleotide variation, here computed as averaged pairwise nucleotide differences (π) in the selective sweep region compared to the genome average. π was calculated in windows of 10 Kbp.

Additional footprints of selection can be unravelled by comparing sequences from distinct species allowing us to infer substitution rates, that is, the rate at which new mutations invade a population. This rate depends on the mutation rate and the strength of selection. Using regions in a gene or genome that are supposedly neutral as a reference, it is possible to infer the direction

and strength of selection at candidate loci. In the case of protein-coding sequences, a straight-forward reference (at least in first approximation) is given by so-called synonymous sites, at which mutations do not alter the encoded amino acid. Selection can then be assessed with codon-based analyses that infer the relative rate of non-synonymous (dN) and synonymous substitutions (dS). A reduced rate, relative to the synonymous rate (dN < dS), thereby reflects purifying selection, while an increased rate reflects positive selection (dN > dS) (Nei & Gojobori, 1986; Z. Yang & Nielsen, 2002).

Substitution rates, however, represent an average estimate over a locus that may have undergone both position and negative selection. The widespread occurrence of negative selection will, in most cases, mask the effect of positive selection. More powerful inference can be performed by comparing sequences at the population and inter-specific levels. The proportion of synonymous and non-synonymous polymorphisms (variable sites among individuals of the same species) or substitutions (fixed differences between species) informs about the type of selection acting on the gene (McDonald and Kreitmann, 1991). The rationale is that adaptive mutations, which are rapidly fixed in a population, will leave no signature in the polymorphism. The ratio of non-synonymous polymorphisms (Pn/Ps) can thus be considered as a measure of purifying selection acting on the gene. It can then be used to test (and quantify) the part of the dN/dS ratio that is due to positive selection only, that is, the proportion of non-synonymous substitutions that is adaptive (Eyre-Walker & Keightley 2009; *Galtier, 2016; McDonald & Kreitmann, 1991; Smith and Walker, 2002). Noteworthily, cases of increased polymorphism compared to divergence (Pn/Ps > dN/dS) may indicate balancing selection.

Table 1. Overview of studies that have used evolutionary analyses to study genome-wide signatures of selection in plant pathogenic fungi, including the distribution of selective sweeps across the genome.

Pathogen	Taxon	Host	Genome size (Mbp)	Statistical parameter used for inference of selection	Nr. of genomes used in the analysis	% of genome affected by selective sweeps	Nr. of genes located in the sweep regions	Reference
Cercospora beticola	Ascomycete	Beets, quinoa, *Chenopodium* spp.	37	ω[1] & μ[2]	89	2,95	0-79	Spanner et al., 2021*
Microbotrium lychnidis-dioicae	Basidiomycete	Members of Caryophyllaceae family	27	CLR[3]	34	17	5 – 33	Badouin et al., 2017*
M. silenes-dioicae	Basidiomycete	Members of Caryophyllaceae family	25	CLR	19	1	3 – 27	Badouin et al., 2017
Rhynchosporium commune	Ascomycete	Barley	55	iHS[4]	14 - 85	1.4 – 4.5	972 across entire genome	Mohd-Assaad et al., 2018
Sclerotinia sclerotiorum	Ascomycete	Several	38	CLR	9 and 11	Authors reported lack of selective sweeps (one sweep throughout genome)	NA	Derbyshire et al., 2019
Parastagonospora nodorum	Ascomycete	Wheat	28	CLR	18 - 181	0.9	2 – 22	Pereira et al., 2021*
Pyrenophora teres	Ascomycete	Barley	47	ω & μ	90	3,9	0 – 48	Unpublished data (Taliadoros, Friesen and Stukenbrock)
Zymoseptoria tritici	Ascomycete	Wheat	35	iHS & CLR	24 - 46	0.5 – 4	57 -1100	Hartmann et al., 2018*

[1] ω statistic (Nielsen et al., 2005) as implemented in OmegaPlus (Alachiotis et al., 2012)

[2] μ statistic (Alachiotis and Pavlidis, 2018)

[3] CLR: Composite likelihood ration (Pavlidis et al., 2013)

[4] iHS: Integrated haplotype score (Voight, 2006) as implemented in R package REHH (Gautier and Vitalis, 2012)

Evolutionary analyses of genome data have proved powerful tools to define new hypotheses and make predictions about functionally relevant traits in fungal plant pathogens. A few studies have demonstrated that a combination of genomic and functional analyses can reveal how variation under selection has relevance for the fitness of the respective organism (in plant pathogens, often traits associated with fungicide resistance and with host specialisation) (e.g. Dong et al., 2014) (**Table 2**). Here we advocate that the availability of thousands of plant-pathogen genomes in public databases provides a unique opportunity to study past and present selection acting on these organisms. With this, we will be able to acquire new insights into the ecology and evolution of plant pathogens and make predictions about the adaptation and dispersal of new virulent genotypes.

"Arms Race" Evolution in crop pathogens

Several studies have applied selective sweep analyses to detect signatures of arms race co-evolution and to identify functionally relevant genes in crop pathogens (*Badouin et al., 2017; **Ebert et al., 2021; Miller et al., 2020; Richards et al, 2019; Yang et al., 2018) (**Table 1**). Positive selection of particular mutations in the effector gene *MGG13871* in the rice pathogen *Magnoporthe oryzae* defines specificity to distinct host species (**Duan et al., 2021). The locus surrounding this "host specificity" effector locates in a region exhibiting the hallmarks of positive selection, including reduced genetic variation and increased LD in a region surrounding *MGG13871*. Intriguingly, different mutations have been fixed in different lineages of *M. oryzae* specialised in rice and non-rice hosts, reflecting the relevance of host genetics in effector evolution.

A genome-wide map of selective sweeps in four geographical populations of the wheat pathogen *Zymoseptoria tritici* revealed signatures of local adaptation with several population-specific selective sweep signatures (*Hartmann et al., 2018). The selective sweep regions include genes encoding small secreted proteins with putative effector function. This indicates that local adaptation may be driven by adaptation to local wheat cultivars used by farmers in Europe, North America, Australia, and the Middle East. Sexual recombination frequently occurs in *Z. tritici,* and previous studies have identified recombination as the primary driver of adaptive evolution (**Grandaubert et al., 2019). This is in accordance with the efficient fixation of beneficial mutations along the genome, leaving signatures of selective sweeps in local populations.

Interestingly, a genome-wide scan of selecive sweeps in the multi-host pathogen *Sclerotinia sclerotiorum* failed to identify any signficant outlier loci (*Derbyshire et al, 2019). This may be due to th fact that *S. sclerotiorum* reproduces mainly clonally whereby detection of selective sweeps based on signatures of unusual LD patterns is prevented.

Table 2. Overview of studies that have identified functionally relevant genes in filamentous plant pathogens, which also exhibit signature of selection.

Pathogen	Taxon	Host	Gene	Mode of selection	Method for detecting selection	Predicted and valitated function	Mode of action	Ref
Cercospora beticola	Ascomycete	Beets, quinoa, *chenopodium* spp.	*CbNip1*	Directional selection	ω[1] & μ[2] statistics	Toxin	Induces death of host cell	Ebert et al., 2021*
Magnaporthe oryzae	Ascomycete	Rice	*MGG13871*	Directional selection	elevated iHS[3], XPEHH[4], and XPCLR[5]	Candidate effector	NA	Duan et al., 2021*
Parastagonospora nodorum	Ascomycete	Wheat	*SnToxA*	Directional selection	SweeD[6]	Effector	Induces host cell death	Richards et al., 2019*
Parastagonospora nodorum	Ascomycete	Wheat	*ToxA*	Directional selection	SweeD	Effector	suppress host defences	Pereira et al., 2021*
Phytophthora infestans	Oomycete	Potato	*avrblb2*	Balancing selection	Elevated Ka/Ks ratio[7]	Effector	suppress host defences	Oh et al., 2009
Phytophthora infestans	Oomycete	Potato	*epiC1*	Directional selection	Elevated Ka/Ks ratio	Protease inhibitor effector	suppress host defences	Dong et al. 2014
Phytophthora infestans	Oomycete	Potato	*Avr3a*	Directional selection	Elevated Gst[8], Fst[9] and negative Tajima's D[10]	RXLR effector	suppress host defences	Yang et al., 2018
Rhyncosporium commune	Ascomycete	Barley	*KES1*	Directional selection	iHS	Oxysterol-binding protein	Hypersensitivity to fungicides	Mohd-Assaad et al., 2018
Rhynchosporium commune	Ascomycete	Barley	*NIP1*	Balancing selection	Presence /abcence variation	Effector	Avirulence gene	Mohd-Assaad et al., 2019*
Sporisorium reilianum	Basidiomycete	Maize and sorghum	*sr10529*	Positive selection	Elevated dN/ds ratio	Protease inhibitor, effector	suppress host defences	Schweizer et al., 2018
Zymoseptoria tritici	Ascomycete	Wheat	*AvrStb6*	Balancing selection	Elevated π[11] value	Effector/Avirulence protein	participates in gene-to-gene interaction with host	Brunner and McDonald, 2018
Zymoseptoria tritici	Ascomycete	Wheat	*Zt89160*	Directional selection	Elevated dN/dS ratio	RCC1 domain	Possibly participates in regulation of virulence-related genes	Poppe et al., 2015
Zymoseptoria tritici	Ascomycete	Wheat	*Zt110804*	Directional selection	Elevated dN/dS ratio	Kinase	possibly involved in protein binding	Poppe at al., 2015

[1] ω statistic (Nielsen et al., 2005) as implemented in OmegaPlus (Alachiotis et al., 2012)
[2] μ statistic (Alachiotis and Pavlidis, 2018)
[3] iHS: Integrated haplotype score (Voight, 2006) as implemented in R package REHH (Gautier and Vitalis, 2012)
[4] XP-EHH: Cross-population extended haplotype homozygosity (Sabeti, 2007)
[5] XP-CLR: Cross-population composite likelihood ration (Chen, Petterson, and Reich, 2010)
[6] CLR: Composit likelihood ration (Pavlidis et al., 2013)
[7] Ka/Ks: Ratio of non-synonymous to synonymous substitutions (Nei & Gojobori, 1986; Z. Yang & Nielsent, 2002)
[8] Gst: Genetic distance measurement generalized for multiple alleles (Nei, 1973)
[9] Fst: Genetic distance measurement (Fst) (wright, 1969)
[10] Tajima's D (Tajima, 1989)
[11] π: Estimator of genetic diversity based on the pairwise nucleotide differences of a genetic region (Nei and Li, 1979)

Virulence factors evolving under a “Trench-warfare” scenario

Examples of elevated polymorphisms in effector genes of plant pathogenic fungi consistent with trench warfare coevolution include the avirulence gene *AvrStb6* of *Z. tritici* (**Brunner & McDonald, 2018). The functional relevance of AvrStb6 is poorly understood. The protein acts as an elicitor of plant defences in wheat cultivars carrying the corresponding resistance genes *Stb6* (Kema et al., 2018; Zhong et al., 2017). Genetic diversity in the *AvrStb6* gene was inferred from 142 isolates collected in four geographical regions and was found to be exceptionally high. *AvrStb6* is located in a recombination hotspot which may contribute to the high allelic variation. The high diversity of *AvrStb6* alleles is predicted to reflect a co-evolutionary dynamics in which alleles are selected to escape recognition while maintaining its function as an effector.

Signatures of “trench-warfare” evolution are also found in other plant-pathogen genes including some RXLR effectors (*Luo et al., 2021; Wang et al., 2019) in oomycete plant pathogens, such as Avrblb2 produced by *Phytophthora infestans* (Oliva et al., 2015) and ATR13 produced by *Hyaloperonospora arabidopsidis* (Allen et al., 2008), as well as the fungal proteins AvrL567 and NIP1 produced by the flax rust pathogen *Melampsora lini* and the barley scald pathogen *Rynchosporium secalis*, respectively (Dodds et al., 2006, Mohd-Assaad et al., 2019). As AvrStb6, these four effectors also act as elicitors of plant defences in hosts with the respective resistance genes. Sequence analyses based on population samples of *P. infestans*, *H. arabidopsis, M. lini* and *R. secalis* indicate that multiple alleles are maintained in the pathogen population, reflecting trench-warfare evolution (Oliva et al., 2015, Mohd-Assaad et al., 2019).

Codon based methods to detect divergent selection between plant pathogen species

Gene alignments and codon-based analyses can provide insights into the processes that drive gene evolution, including the role of positive selection. Different test statistics are available to compare the proportion of non-synonymous to synonymous variants, including maximum likelihood methods that take into account the "evolvability" of different sites and variation in evolutionary rates among branches (Yang, 2007, Vitti et al., 2013). These methods have been applied widely to analyse gene evolution in fungal plant pathogen species and identify those genes that exhibit signatures of positive selection (Aguileta et al., 2012; Dong et al., 2014; Schweizer et al., 2018; Stukenbrock et al., 2011).

In a study by Liang and colleagues, gene evolution was compared in four powdery mildew species of monocot and dicot hosts (*Oidium haveae, Blumeria graminis* f. sp. *Hordei, B. graminis* f. sp. *tritici,* and *Erysiphe negator* (Liang et al., 2018). Based on gene alignments of orthologous effector-encoding genes, the authors found that the vast majority of the single-copy orthologous effector genes (99.71 %) exhibited a Ka/Ks ratio (equivalent to dN/dS defined above) lower than 0.5. Only one CSEP showed a Ka/Ks value significantly higher than 1, indicating positive selection. It is possible that comparisons of average rates of substitutions across genes fail to detect selection. Methods based on codon models which assume that some sites may be under selection while others not (Yang and Nielsen, 2002), could be applied for a more detailed serach for signals of positive selection in the mildew species.

Increased rates of evolution at non-synonymous sites were also compared between genes of core and lineages specific regions in the wilt pathogen *Verticillium dahliae* (Depotter et al., 2019). The genome of *V. dahliae* comprises

highly variable and transposable element-enriched regions showing presence-absence variation between different lineages. These lineage-specific regions encode several predicted effector genes, which interestingly show a higher Ka/Ks ratio than genes encoded by the core genome. This could indicate a higher rate of adaptation of genes located in these particular genomic regions.

Together these findings demonstrate how evolutionary analyses can identify differences in evolutionary rates, not only between members of gene families but also among defined regions in the genome.

Quantifying rates of adaptation

The McDonald-Kreitman test is another codon-based method that contrasts variation within and between species variation, quantifying the rate of adaptation at the molecular level (McDonald and Kreitman, 1991). Extensions of the MK-test consider that not all fixed non-synonymous substitutions are adaptive (*Galtier, 2016; Nick G.C. & Adam, 2002; *Tataru et al., 2017). A population genomic study of *Z. tritici* was used to quantify adaptive evolution using MK-based test statistics and modelling the distributions of fitness effects (DFE) in *Z. tritici* (**Grandaubert et al., 2019). Hereby, the authors showed that overall rates of adaptive evolution (computed as the parameter ω_α) correlate with recombination; genes located in genomic regions with high recombination rates show a higher rate of adaptation. Transposable elements that otherwise are known to drive sequence variability did, on the other hand, not correlate with measures of adaptations. In agreement with the strong selection pressure exerted on effector proteins, genes encoding effectors show a statistically higher adaptation rate than all other genes (**Grandaubert et al., 2019).

Conclusion

The increasing availability of genome data provides a rich resource for studying the evolution of fungal plant pathogens. Many plant pathogen genomes show extensive variation at the structural level, this is mediated by dynamic repertoires of transposable elements, and comparative studies have often used qualitative methods and focused on dynamic genomic regions. Statistical genetics and evolutionary predictions allow more detailed insights into sequence evolution and the impact of natural selection along the genome. Such methods have been applied in several genome studies and have demonstrated their power to identify signatures of positive selection, either in the form of trench-warfare or arms race evolution, in pathogen genomes.

Research of fungal genome evolution relies on applying methods developed typically for organisms with more defined or well-understood life cycles, such as *Drosophila melanogaster* or *Homo sapiens*. Fungi often have more complex lifecycles that include both sexual and asexual propagation. Moreover, the number of sexual cycles per year (the generation time) can vary over time and may, in some cases, involve extended periods with asexual propagation and only rare sexual mating. It is essential to consider these complexities, which may even violate general assumptions implemented in evolutionary models used in some computational methods.

Another confounding factor may be the demographic history of the species. Past bottlenecks and recent population expansions are demographic scenarios that can influence the site frequency spectrum in populations, leading to the wrong interpretation of selection analyses. We must assume that plant pathogens can undergo such demographic scenarios in response to changes in host populations. To account for this, simulations can be used to determine the underlying demographic scenarios responsible for the allele distributions in a

population. In general, theoretical developments in the field are necessary to improve our power to detect selection and understand pathogen evolution.

"Arms race" and "trench warfare" evolution occur in fungal pathogens of both crop and wild plants. Detailed sampling over several years would provide important insights into the dynamics of virulence alleles within pathogen populations. Such temporal sampling has previously documented the exceptional rapidness whereby new virulence alleles can emerge in plant-pathogen populations (e.g. Daverdin et al., 2012; van de Wouw et al., 2010; Miller et al., 2020). Finally, we advocate the immense value of genome data and evolutionary predictions to identify traits under selection in fungal pathogen genomes. Signatures of selection can support the development of new hypotheses regarding gene functions and pathogen biology. Thousands of plant-pathogen genomes are published, and this data still holds essential information about the biology and evolution of these important organisms. Evolutionary analyses are tools to unravel past pathogen histories and predict future pathogen emergence.

Acknowledgements

Research of DT is supported by the DFG-funded Research Training Group TransEvo 2511. The authors are grateful for helpful discussions with members of the EnvGen team and critical input to a previous version of the text by J. Dutheil.

Bibliography

Aguileta, G., Lengelle, J., Chiapello, H., Giraud, T., Viaud, M., Fournier, E., Rodolphe, F., Marthey, S., Ducasse, A., Gendrault, A., Poulain, J., Wincker, P., & Gout, L. (2012). Genes under positive selection in a model plant pathogenic fungus, Botrytis. *Infection, Genetics and Evolution*, *12*(5), 987–996. https://doi.org/10.1016/j.meegid.2012.02.012

Allen, R. L., Meitz, J. C., Baumber, R. E., Hall, S. A., Lee, S. C., Rose, L. E., & Beynon, J. L. (2008). Natural variation reveals key amino acids in a downy mildew effector that alters recognition specificity by an Arabidopsis resistance gene. *Molecular Plant Pathology*, *9*(4), 511–523. https://doi.org/10.1111/j.1364-3703.2008.00481.x

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

This is one of the first studies to have explored the genome-wide abundance and distribution of selective sweeps within and between fungal plant pathogens. The power of CLR methods to infer candidate selective sweep regions is highlighted in this article.

Beckerson, W. C., Rodríguez De La Vega, R. C., Hartmann, F. E., Duhamel, M., Giraud, T., & Perlin, M. H. (2019). Cause and effectors: Whole-genome comparisons reveal shared but rapidly evolving effector sets among host-specific plant-castrating fungi. *MBio*, *10*(6). https://doi.org/10.1128/mBio.02391-19

**Brunner, P. C., & McDonald, B. A. (2018). Evolutionary analyses of the avirulence effector AvrStb6 in global populations of Zymoseptoria tritici identify candidate amino acids involved in recognition. *Molecular Plant Pathology*, *19*(8), 1836–1846. https://doi.org/10.1111/mpp.12662

The study describes the cloning of AvrStb6 in the wheat pathogen *Zymoseptoria tritici* based on a large population genomic dataset. The authors show that the *AvrStb6* gene is highly variable in natural populations of *Z. tritici*

and is located in a genomic hotspot of recombination. The article highlights the effect of co-evolutionary dynamics between avirulence/effector genes and host resistance.

Castillo, J. A., & Agathos, S. N. (2019). A genome-wide scan for genes under balancing selection in the plant pathogen Ralstonia solanacearum. *BMC Evolutionary Biology*, *19*(1), 1–16. https://doi.org/10.1186/s12862-019-1456-6

Castro-moretti, F. R., Gentzel, I. N., Mackey, D., & Alonso, A. P. (2020). Metabolomics as an emerging tool for the study of plant–pathogen interactions. *Metabolites*, *10*(2), 1–23. https://doi.org/10.3390/metabo10020052

Chamary, J. V., Parmley, J. L., & Hurst, L. D. (2006). Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, *7*(2), 98–108. https://doi.org/10.1038/nrg1770

Daverdin, G., Rouxel, T., Gout, L., Aubertot, J.N., Fudal, I., Meyer, M., Parlange, F., Carpezat, J. and Balesdent, M.H., 2012. Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. *PLoS Pathogens*, *8*(11), p.e1003020.

Depotter, J. R. L., Shi-Kunne, X., Missonnier, H., Liu, T., Faino, L., van den Berg, G. C. M., Wood, T. A., Zhang, B., Jacques, A., Seidl, M. F., & Thomma, B. P. H. J. (2019). Dynamic virulence-related regions of the plant pathogenic fungus Verticillium dahliae display enhanced sequence conservation. *Molecular Ecology*, *28*(15), 3482–3495. https://doi.org/10.1111/mec.15168

**Derbyshire MC, Denton-Giles M, Hane JK, Chang S, Mousavi-Derazmahalleh M, et al. (2019) A whole genome scan of SNP data suggests a lack of abundant hard selective sweeps in the genome of the broad host range plant pathogenic fungus *Sclerotinia sclerotiorum*. PLOS ONE 14(3): e0214201.https://doi.org/10.1371/journal.pone.0214201
The authors find an absence of hard selective sweeps in the genome of the multi-host plant pathogen Sclerotinia sclerotiorum. S. sclerotiorum reproduces

mainly clonally which may redice the ability to detect selective sweeps based on signatures of selection such as high LD.

Derbyshire, Mark C. (2020). Bioinformatic Detection of Positive Selection Pressure in Plant Pathogens: The Neutral Theory of Molecular Sequence Evolution in Action. *Frontiers in Microbiology*, *11*(April), 1–14. https://doi.org/10.3389/fmicb.2020.00644

Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Teh, T., Wang, C. I. A., Ayliffe, M. A., Kobe, B., & Ellis, J. G. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8888–8893. https://doi.org/10.1073/pnas.0602577103

Dong, S., Stam, R., Cano, L. M., Song, J., Sklenar, J., Yoshida, K., Bozkurt, T. O., Oliva, R., Liu, Z., Tian, M., Win, J., Banfield, M. J., Jones, A. M. E., Van Der Hoorn, R. A. L., & Kamoun, S. (2014). Effector specialization in a lineage of the Irish potato famine pathogen. In *Science* (Vol. 343, Issue 6170, pp. 552–555). https://doi.org/10.1126/science.1246300

Duan, G., Bao, J., Chen, X., Xie, J., Liu, Y., Chen, H., Zheng, H., Tang, W., & Wang, Z. (2021). Large-Scale Genome Scanning within Exonic Regions Revealed the Contributions of Selective Sweep Prone Genes to Host Divergence and Adaptation in Magnaporthe oryzae Species Complex. *Microorganisms*, *9*(3), 562. https://doi.org/10.3390/microorganisms9030562

**This study shed light on the role of directional selection in host specialisation of the rice blast fungus *Magnaporthe oryzae*. The study identifies 17 candidate effector genes that have undergone differential selective sweep in rice and non-rice infecting isolates.

*Ebert, M. K., Rangel, L. I., Spanner, R. E., Taliadoros, D., Wang, X., Friesen, T. L., de Jonge, R., Neubauer, J. D., Secor, G. A., Thomma, B. P. H. J., Stukenbrock, E. H., & Bolton, M. D. (2021). Identification and characterization of Cercospora beticola necrosis-inducing effector CbNip1. In *Molecular Plant Pathology* (Vol. 22, Issue 3, pp. 301–316). https://doi.org/10.1111/mpp.13026

In this article, the authors reported the discovery and functional characterization of an effector toxin named CbNip1. The phytotoxin cercosporin was previously shown to be responsible for inducing host cell death. In addition, the authors here reported that CbNip1 gene co-localizes with a selective sweep region.

Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol. 269:2097–2108.

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*(3), 1405–1413. https://doi.org/10.1093/genetics/155.3.1405

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, *484*(7393), 186–194. https://doi.org/10.1038/nature10947

*Galtier, N. (2016). Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. *PLoS Genetics*, *12*(1), 1–23. https://doi.org/10.1371/journal.pgen.1005774

This methodological study provides extensions of MK-based test for inference of adaptive protein evolution.

**Grandaubert, J., Dutheil, J. Y., & Stukenbrock, E. H. (2019). The genomic determinants of adaptive evolution in a fungal pathogen. *Evolution Letters*, *3*(3), 299–312. https://doi.org/10.1002/evl3.117

This study reports a difference in the rate of adaptive substitutions of genes encoding and all other genes in the wheat pathogen *Zymoseptoria tritici*. The authors demonstrate that recombination rate correlates with recombination whereby genes showing high rates of adaptive evolution localize in regions with high recombination rates.

Guyon, K., Balagué, C., Roby, D., & Raffaele, S. (2014). Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen Sclerotinia sclerotiorum. *BMC Genomics*, *15*(1). https://doi.org/10.1186/1471-2164-15-336

*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

In this study, a genome-wide map of selective sweeps was generated to identify signatures of local adaptation in four geographical populations of the wheat pathogen *Z. tritici.* The putative selective sweep regions are reported to be enriched with effector genes which may indicate that specialisation to local wheat cultivars plays is a driver of local adaptation.

Hulme, P. E. (2009). Trade , transport and trouble : managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, *1*(46), 10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x

Hulme, P. E., Kühn, I., Wild, J., Arianoutsou, M., & Bacher, S. (2010). Disentangling the role of environmental and human pressures on biological invasions across Europe. *PNAS*, *107*(27), 12157–12162. https://doi.org/10.1073/pnas.1002314107

Ingvarsson, P. K. (2010). Natural Selection on Synonymous and Nonsynonymous Mutations Shapes Patterns of Polymorphism in Populus tremula. *Molecular Biology and Evolution*, *27*(3), 650–660. https://doi.org/10.1093/molbev/msp255

Jensen, J. D., Wong, A., & Aquadro, C. F. (2007). Approaches for identifying targets of positive selection. *Trends in Genetics*, *23*(11), 568–577. https://doi.org/10.1016/j.tig.2007.08.009

Karasov, T. L., Horton, M. W., & Bergelson, J. (2014). Genomic variability as a driver of plant-pathogen coevolution? *Current Opinion in Plant Biology*, *18*(1), 24–30. https://doi.org/10.1016/j.pbi.2013.12.003

Kema, G. H. J., Mirzadi Gohari, A., Aouini, L., Gibriel, H. A. Y., Ware, S. B., Van Den Bosch, F., Manning-Smith, R., Alonso-Chavez, V., Helps, J., Ben M'Barek, S., Mehrabi, R., Diaz-Trujillo, C., Zamani, E., Schouten, H. J., Van Der Lee, T. A. J., Waalwijk, C., De Waard, M. A., De Wit, P. J. G. M., Verstappen, E. C. P., … Seidl, M. F. (2018). Stress and sexual reproduction affect the dynamics of the wheat pathogen effector AvrStb6 and strobilurin

resistance. *Nature Genetics*, *50*(3), 375–380. https://doi.org/10.1038/s41588-018-0052-9

Kim, Y., & Neilsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics*, *167*(3), 1513–1524. https://doi.org/10.1534/genetics.103.025387

Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, *160*(2), 765–777. https://doi.org/10.1093/genetics/160.2.765

Liang, P., Liu, S., Xu, F., Jiang, S., Yan, J., He, Q., Liu, W., Lin, C., Zheng, F., Wang, X., & Miao, W. (2018). Powdery Mildews Are Characterized by Contracted Carbohydrate Metabolism and Diverse Effectors to Adapt to Obligate Biotrophic Lifestyle. *Frontiers in Microbiology*, *9*(December), 1–14. https://doi.org/10.3389/fmicb.2018.03160

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

Luo, M., Sun, X., Qi, Y., Zhou, J., Wu, X., & Tian, Z. (2021). Phytophthora infestans RXLR effector Pi04089 perturbs diverse defense-related genes to suppress host immunity. *BMC Plant Biology*, *21*(1), 1–15. https://doi.org/10.1186/s12870-021-03364-0

McDonald, H. J., & Kreitmann, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *354*, 737–740.

McDonald, B.A. and Stukenbrock, E.H., 2016. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1709), p.20160026.

Miller, M.E., Nazareno, E.S., Rottschaefer, S.M., Riddle, J., Dos Santos Pereira, D., Li, F., Nguyen-Phuc, H., Henningsen, E.C., Persoons, A., Saunders, D.G. and Stukenbrock, E., 2020. Increased virulence of Puccinia coronata f. sp. avenae populations through allele frequency changes at multiple putative Avr loci. PLoS genetics, 16(12), p.e1009291.

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

Mohd-Assaad, N., McDonald, B. A., & Croll, D. (2019). The emergence of the multi-species NIP1 effector in Rhynchosporium was accompanied by high rates of gene duplications and losses. *Environmental Microbiology*, *21*(8), 2677–2695. https://doi.org/10.1111/1462-2920.14583

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

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, *3*(5), 418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410

Nick G.C., S., & Adam, E.-W. (2002). Adaptive protein evolution in Drosophila. *Nature*, *415*(February), 1022.

Oliva, R. F., Cano, L. M., Raffaele, S., Win, J., Bozkurt, T. O., Belhaj, K., Oh, S. K., Thines, M., & Kamoun, S. (2015). A recent expansion of the RXLR effector gene Avrblb2 is maintained in global populations of Phytophthora infestans indicating different contributions to virulence. *Molecular Plant-Microbe Interactions*, *28*(8), 901–912. https://doi.org/10.1094/MPMI-12-14-0393-R

Omrane, S., Audéon, C., Ignace, A., Duplaix, C., Aouini, L., Kema, G., Walker, A.S. and Fillinger, S., 2017. Plasticity of the MFS1 promoter leads to multidrug resistance in the wheat pathogen Zymoseptoria tritici. *MSphere*, *2*(5), pp.e00393-17.

Pereira, D., Oggenfuss, U., McDonald, B. A., & Croll, D. (2021). Population genomics of transposable element activation in the highly repressive genome of an agricultural pathogen. *Microbial Genomics*, *7*(8). https://doi.org/10.1099/MGEN.0.000540

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

Reich, D., 2018. *Who we are and how we got here: Ancient DNA and the new science of the human past*. Oxford University Press.

Richards JK, Stukenbrock EH, Carpenter J, Liu Z, Cowger C, et al. (2019) Local adaptation drives the diversification of effectors in the fungal wheat pathogen Parastagonospora nodorum in the United States. PLOS Genetics 15(10): e1008223. https://doi.org/10.1371/journal.pgen.1008223

Santini, A., Liebhold, A., Migliorini, D., & Woodward, S. (2018). Tracing the role of human ccivilization in the globalization of plant pathogens. *ISME Journal*, *12*(3), 647–652. https://doi.org/10.1038/s41396-017-0013-9

**Schweizer, G., Münch, K., Mannhaupt, G., Schirawski, J., Kahmann, R. and Dutheil, J.Y., 2018. Positively selected effector genes and their contribution to virulence in the smut fungus Sporisorium reilianum. *Genome biology and evolution*, *10*(2), pp.629-645.

This study presents a detailed pipeline for the extraction and comparison of effector genes in smut fungi. The authors perform analyses of gene evolution and find evidence for positive selection in several effector genes during the divergence of smut pathogens.

Smith, N.G. and Eyre-Walker, A., 2002. Adaptive protein evolution in Drosophila. *Nature*, *415*(6875), pp.1022-1024.

Spanner, R., Taliadoros, D., Richards, J., Rivera-Varas, V., Neubauer, J., Natwick, M., Hamilton, O., Vaghefi, N., Pethybridge, S., Secor, G. A., Friesen, T. L., Stukenbrock, E. H., & Bolton, M. D. (2021). Genome-Wide Association and Selective Sweep Studies Reveal the Complex Genetic Architecture of DMI Fungicide Resistance in Cercospora beticola. *Genome Biology and Evolution*, *13*(9), 1–17. https://doi.org/10.1093/gbe/evab209

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., Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M.,

McDonald, B. A., Wang, J., & 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., & Mcdonald, B. A. (2008). The Origins of Plant Pathogens in Agro-Ecosystems. *Annual Review of Phytopathology*, *46*, 75–100. https://doi.org/10.1146/annurev.phyto.010708.154114

Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, *1*(123), 585–595.

*Tataru, P., Mollion, M., Glémin, S., & Bataillon, T. (2017). Inference of Distribution of Fitness Effects and proportion of adaptive substitutions from polymorphism data. *Genetics*, *207*, 1103–1119. https://doi.org/10.1534/genetics.117.300323/-/DC1.1

This methodological study provides extensions of MK-based test for inference of adaptive protein evolution.

Tellier, A., Moreno-Gamez, 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

Tellier, Aurélien, 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

Van de Wouw, A.P., Cozijnsen, A.J., Hane, J.K., Brunner, P.C., McDonald, B.A., Oliver, R.P. and 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), p.e1001180.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, *47*, 97–120. https://doi.org/10.1146/annurev-genet-111212-133526

Wang, S., McLellan, H., Bukharova, T., He, Q., Murphy, F., Shi, J., Sun, S., Van Weymers, P., Ren, Y., Thilliez, G., Wang, H., Chen, X., Engelhardt, S., Vleeshouwers, V., Gilroy, E. M., Whisson, S. C., Hein, I., Wang, X., Tian, Z., … Boevink, P. C. (2019). Phytophthora infestans RXLR effectors act in concert at diverse subcellular locations to enhance host colonization. *Journal of Experimental Botany*, *70*(1), 343–356. https://doi.org/10.1093/jxb/ery360

Wright, S. (1949). The genetical structure of populations. *Ann Eugenics*, *15*, 323–324.

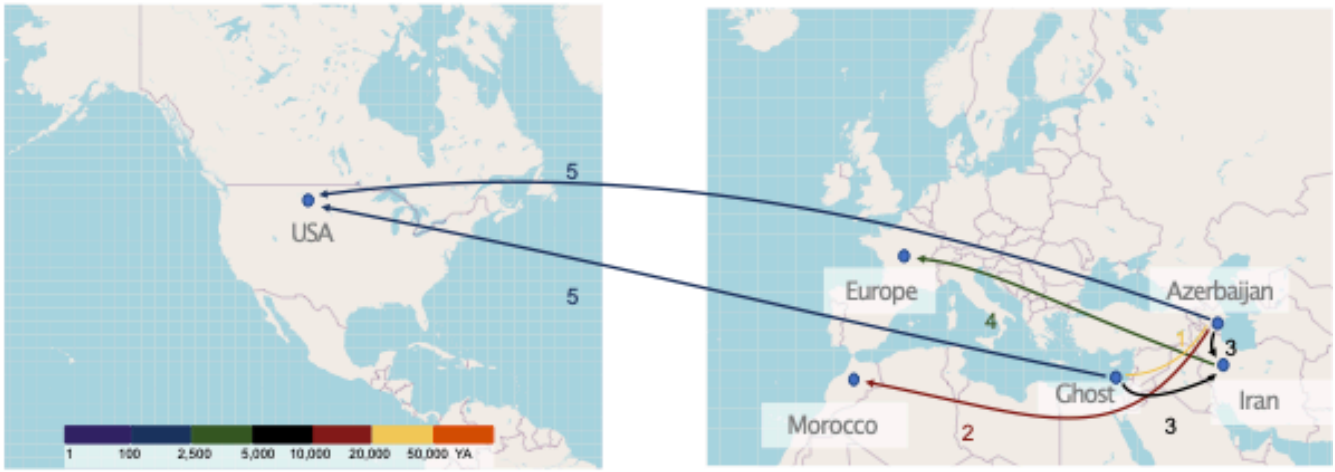
Yang, L., Ouyang, H. B., Fang, Z. G., Zhu, W., Wu, E. J., Luo, G. H., Shang, L. P., & Zhan, J. (2018). Evidence for intragenic recombination and selective sweep in an effector gene of Phytophthora infestans. In *Evolutionary Applications* (Vol. 11, Issue 8, pp. 1342–1353). https://doi.org/10.1111/eva.12629

Yang, Z., & Nielsent, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, *19*(6), 908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148

Zhong, Z., Marcel, T. C., Hartmann, F. E., Ma, X., Plissonneau, C., Zala, M., Ducasse, A., Confais, J., Compain, J., Lapalu, N., Amselem, J., McDonald, B. A., Croll, D., & Palma-Guerrero, J. (2017). A small secreted protein in Zymoseptoria tritici is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. *New Phytologist*, *214*(2), 619–631. https://doi.org/10.1111/nph.14434

Chapter II

Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history



Chapter II: Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history

Authors

Demetris Taliadoros[1,2], Alice Feurtey[1,3,4], Nathan Wyatt[5,6], Benoit Barrès[7] Pierre Gladieux[8], Timothy Friesen[5,6] and Eva Stukenbrock[1,2]

This manuscript is published as a preprint on *biorxiv* (doi.org/10.1101/2023.07.28.550921)

Affiliations

[1] Environmental Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön and

[2] Christian-Albrechts University of Kiel, 24118, Germany

[3] Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, CH-2000 Neuchâtel, Switzerland.

[4] Plant Pathology, D-USYS, ETH Zurich, CH-8092 Zurich, Switzerland.

[5] Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, USDA-ARS, Fargo, ND 58102, USA

[6] Sugar Beet and Potato Research Unit, Edward T. Schafer Agricultural Research Center, USDA-ARS, Fargo, ND 58102, USA

[7] Université de Lyon, Anses, INRAE, USC CASPER, Lyon, France

[8] PHIM Plant Health Institute, Univ Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France

Abstract

Fungal pathogens cause devastating disease in crops. Understanding the evolutionary origin of pathogens is essential to the prediction of future disease emergence and the potential of pathogens to disperse. The fungus *Pyrenophora teres* f. *teres* causes net form net blotch (NFNB), an economically significant disease of barley. In this study, we have used 104 *P. teres* f. *teres* genomes from four continents to explore the population structure and demographic history of the fungal pathogen. We showed that *P. teres* f. *teres* is structured into populations that tend to be geographically restricted to different regions. Using Multiple Sequentially Markovian Coalescent and machine learning approaches we demonstrated that the demographic history of the pathogen correlates with the history of barley, highlighting the importance of human migration and trade in spreading the pathogen. Exploring signatures of natural selection, we identified several population-specific selective sweeps that colocalized with genomic regions enriched in putative virulence genes, and loci previously identified as determinants of virulence specificities by quantitative trait locus analyses. This reflects rapid adaptation to local hosts and environmental conditions of P. teres f. teres as it spread with barley. Our research highlights how human activities can contribute to the spread of pathogens that significantly impact the productivity of field crops.

Keywords: Fungal pathogen evolution, center of origin, plant disease, barley, selective sweeps

Author Summary

Population genetic and genomics studies of several crop pathogens have revealed that human activities, such us domestication, trade, and migration have played a pivotal role in the emergence and spread of plant diseases. In this study, we employed cutting-edge genetic analysis techniques and machine learning tools to shed light on the population structure and historical dispersal patterns of a major fungal pathogen of barley called *Pyrenophora teres* f. *teres*. We found that barley domestication during the Neolithic era potentially gave rise to the pathogen which has since co-evolved with barley to become one of the most devastating barley diseases we face today. In addition, we identified a large number of genomic regions evolving under strong positive selection that were specific to different populations. This finding suggests that populations are evolving fast, becoming well adapted to their local host availability and environmental conditions.

Introduction

Fungi cause devastating diseases in crop plants and can be dispersed across continents by agricultural trade (1,2). Understanding the evolutionary history of fungal pathogens and the mechanisms underlying their emergence and spread is essential in preventing future epidemics in agroecosystems. Notably, information on historic and current evolutionary trajectories can be key in the development of regulation for re-engineer crops and agroecosystem to improve epidemiological surveillance and prevent potential outbreaks (3).

Pyrenophora teres f. *teres* is a widespread fungal pathogen of barley causing the disease “net form net blotch” (NFNB). Infection of susceptible barley leaves occurs by the production of specialized infection structures, appressoria, whereby the fungus penetrates the cuticle and cell wall of epidermal cells. The characteristic net blotch symptoms arise through necrosis which is induced rapidly after initial pathogen invasion of the host leaf (4). Net form net blotch disease occurs in all barley-producing regions of the world, and has been reported in several African countries (5–7), West and East Asia (8,9), Europe (10), North and South America (11,12), and Australia (13). The pathogen has been known to humans for centuries, initially described as Helminthosporium teres

(Sacc.), however insights into the population biology and demographic history of P. teres f. teres are scarce.

Barley was domesticated in the Fertile Crescent approximately 10,000 years ago and was later introduced to North Africa and Eurasia by Neolithic farmers (14). North Africa has been proposed as a center of diversity of wild barley (15–18). In the past few centuries, barley has further been dispersed with European migrants to the Americas, Australia, and South Africa (19). Altogether, the dispersal history of barley is associated with human activities and reflects the spread of cereal cultivation through historical and modern trading routes.

Plant domestication has been associated with the emergence of new fungal pathogens (20). Population genetic and evolutionary studies have been used to track down the origin and dispersal history of important crop pathogens. The fungal wheat pathogen *Zymoseptoria tritici* emerged at the onset of wheat domestication and was dispersed with wheat farming during the Neolithic and much later with European migrants (21,22). Likewise, the center of origin of the maize infecting smut fungus *Ustilago maydis* likely also coincide with the center of maize domestication in Central and South America (23). Several other examples underline the importance of domestication and agricultural trade in shaping the evolution and dispersal of crop pathogens (24–26). In modern times, breeding of new crop species has also driven the rapid evolution of new

pathogen species, such as mildew pathogens affecting the hybrid crop triticale (27).

The dispersal of pathogens across continents may be accompanied by local adaptation to distinct environmental conditions and/or to specific management practices such as local crop varieties or, in more recent times, fungicides (22,28). Signatures of adaptation can be identified in population genomic data as “selective sweeps”, which are genomic regions with low genetic variation and elevated linkage disequilibrium (29,30). Several statistical methods have been developed to distinguish signatures of selective sweeps from other scenarios that can influence patterns of variation along genomes, such as demography, recombination rate variation, and population structure (30). Prime candidates for signatures of strong and recent positive selection in plant pathogens are genes that encode effector proteins (31–34), which are small proteins secreted by pathogens to manipulate their host’s physiology. Effector proteins play determining roles in the suppression of plant immune responses and have been a major focus in molecular plant pathology research (35). Effector genes can be predicted from genome sequences as they typically encode a signal peptide targeting them for secretion into the plant apoplast or translocation into the host cytoplasm. Moreover, effectors are typically cysteine-rich and specifically

expressed during host invasion (36). Several effector genes have been identified by quantitative trait locus analysis (QTL) or genome-wide-association studies (GWAS) underlining the importance of genome data in the discovery of virulence mechanisms (37).

In the present study, we analyzed the haploid genomes of 104 *P. teres* f. *teres* isolates derived from different barley fields worldwide. Based on single nucleotide polymorphisms (SNPs), we characterized the population structure and inferred the demographic history of the pathogen. We specifically investigated patterns of early lineage divergence using different methods for demographic inference. Our analyses provide strong evidence for a recent origin and dispersal of P. teres f. teres, likely coinciding with the domestication and dispersal of barley by Neolithic farmers. We, moreover, investigated signatures of recent natural selection and found an overlap between signatures of selective sweeps and putative virulence factors previously identified by quantitative trait locus (QTL) analyses (38,39). Our study reveals the recent emergence of an important crop pathogen along the domestication and dispersal of its host and underlines the impact of human activities and agriculture on the evolution and spread of new diseases.

Results

Generation of a population genomic dataset of *P. teres* f. *teres*

To study the geographical population genetic structure of *P. teres* f. *teres* and to infer the recent history of this emerging barley pathogen, we generated a population genomic dataset comprising sequence data of 104 isolates from cultivated barley in six countries across four continents (Africa, America, Central Asia, and Europe) (Table S1). The genomes were sequenced with Illumina technology and sequencing reads were mapped to the reference genome of P. teres f. teres (40) to identify SNPs. The average read coverage of genomes was 21X, and we identified a total of 1,092,635 high-quality SNPs among the 104 isolates. Further summary statistics related to the read mapping and variant calling are summarized in Table S2.

Phylogenetic relationship of Pyrenophora species from barley and other grass hosts

Because barley can be infected by other closely related Pyrenophora species, including *P. teres* f. *maculata*, we first reconstructed the phylogenetic relationships between the isolates included in our study and other barley infecting Pyrenophora species to ensure the species identity of our isolates. Our analyses included a set of isolates collected from wild barley exhibiting spot

lesions in the Monterey Peninsula (California), which allowed us to compare genetic diversity in P. teres f. teres populations of cultivated and wild barley.

Using sequence information from four gene loci (ITS, LSU, tub2, and tef1-a) (41,42), we found that the Californian population from wild barley represents a separate lineage or species of Pyrenophora that is more closely related to *P. graminea* than *P. teres* (Figure 1A). The Californian Pyrenophora population thereby provided us with an ideal outgroup for further analyses of the *P. teres* f. *teres* populations.

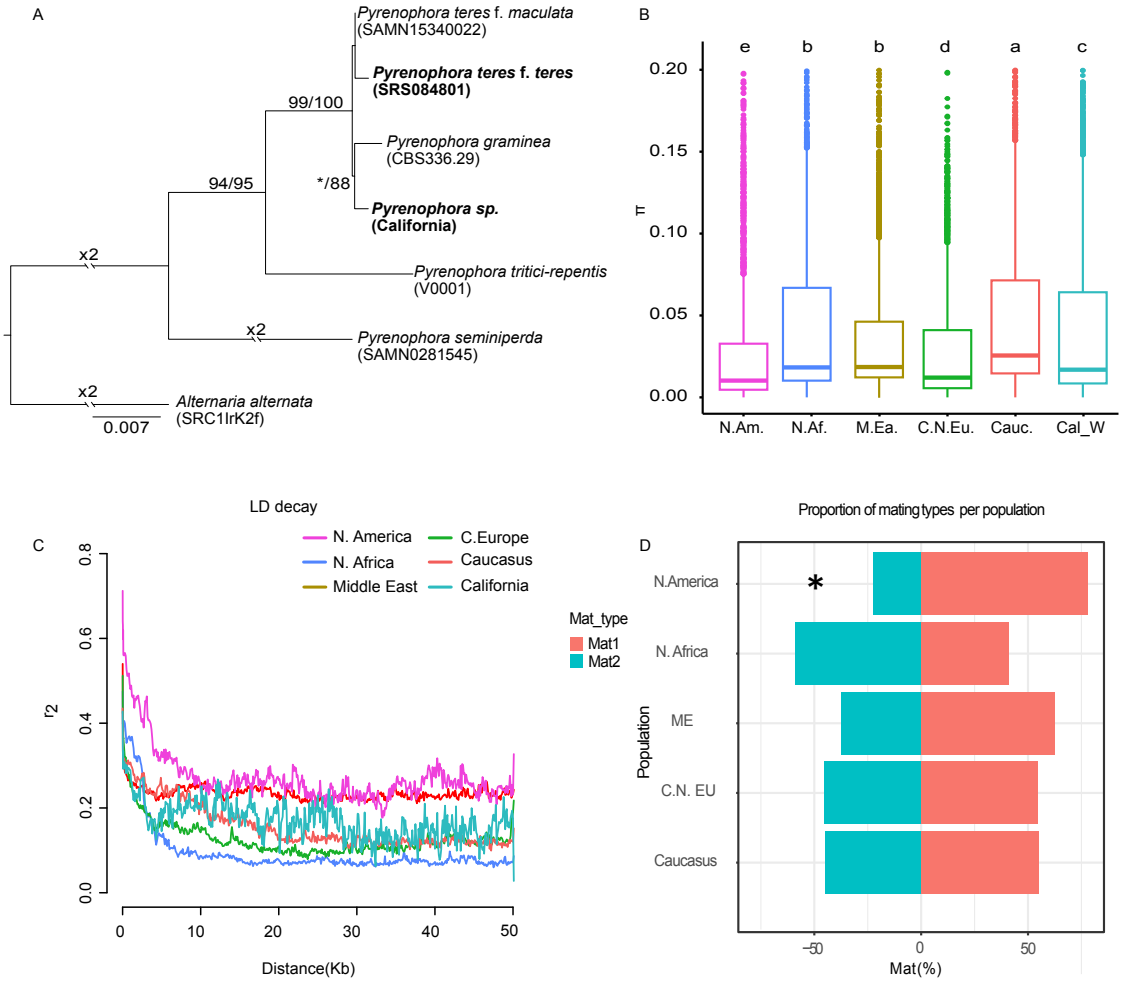


Figure 1: Collection of *Pyrenophora teres* isolates across continents for the inference of pathogen population structure and dispersal. A) Inference of the phylogenetic relationship of closely related Pyrenophora species, including isolates of different species originating from barley. The tree was built with nucleotide sequence alignments of the ITS, tub2, LSU, and tef1 regions obtained from five Pyrenophora species (Maximum likelihood inference, loglikelihood: -12,351.458). Numbers reflect maximum likelihood and maximum parsimony bootstrap, respectively. Alternaria alternata was defined as root. B) Nucleotide diversity of P. teres f. teres populations in each geographic region. Kruskal-Wallis test with post-hoc pairwise Wilcoxon was used to identify significant differences ($p < 0.05$) between the groups (Table S3). C) Linkage disequilibrium decay for each population. D) Percentage of the two mating types occurring in each location. Asterisk indicates significant departure from the 1:1 ratio (chi-squared test, p-value: 0.05).

The Caucasian population of *P. teres* f. *teres* exhibits higher nucleotide diversity

The level of standing genetic variation present in populations can give insight into their demographic history. We compared the nucleotide diversity among the six geographical *P. teres* f. *teres* populations using the Kruskal-Wallis test, which revealed significant differences between the populations (p-value < 2.2e-16). Subsequently, we performed a pairwise Wilcoxon test to assess significant differences between the populations (Table S3). Our analysis showed that the Caucasus population harbours significantly higher genetic diversity (mean π Caucasus = 0.0518) than every other population. Furthermore, even though the

Middle Eastern and North African populations showed significantly lower genetic diversity (mean π: 0.0464) than the Caucasian population, they showed significantly higher diversity than the European and the North American populations. The North American population had the lowest level of nucleotide diversity (mean π: 0.0301) (Figure 1B).

Similar to genetic diversity, we used the Kruskal-Wallis test and the pairwise Wilcoxon test (Table S4) to assess differences in Tajima´s D values. Indeed, we found that Tajima´s D values were significantly different between *P. teres* f. *teres* populations (Kruskal-Wallise test, p-value < 2.2e-16). The Middle Eastern, European, and the North American populations showed significantly higher Tajima´s D values compared to the other populations (mean D: 1.02, 0.90, 0.82 respectively) reflecting an excess of common variants. However, we compute a significantly lower Tajima's D value (0.62) for the North African population which may reflect a population bottleneck. Considering that the Caucasian population exhibits the highest level of diversity with a frequency of genetic variants that reflects a mutation – drift equilibrium, we hypothesize that this population is older than the other populations in our dataset.

Varying extent of linkage disequilibrium (LD) among fungal populations can also inform about the frequency of sexual reproduction, and reflect different ages of populations. More recently founded populations, and populations with lower

frequencies of sexual reproduction, will typically exhibit a greater extent of LD compared to older or sexually recombining populations. We found considerably longer linkage blocks in the North American population, for which the LD statistic r2 was reduced to half of its maximum value at 3.9 Kbp (Figure 1C, Table 1). The long LD blocks observed for the North American population suggest that this population was founded more recently.

Table 1: Summary statistics of the genetic clusters of *P. teres* f. *teres*

Cluster name	No of individuals	Isolate origin	π	Θw	Tajima	Dist r2 < 0.25 (Kbp)	Ne (π)	Ne (θw)
Middle East	7	Iran	0,003	0,001	1,02	3,57	2621,053	1087,316
North Africa	23	Morocco	0,006	0,002	0,62	5,01	4840,351	1915,079
Caucasus	18	Azerbaijan (11), Denmark (3), Iran (2), Morocco (2)	0,005	0,002	0,18	4,28	4607,895	1852,807
Cenral and North Europe	20	France(18), Denmark (1), Azerbaijan (1)	0,005	0,002	0,90	3,80	4414,912	1622,167
North America	20	ND	0,004	0,002	0,82	9,28	3674,561	1450,482
[1] π: Estimator of genetic diversity based on the pairwise nucleotide differences of a genetic region (Nei and Li, 1979)								
[2] Θw: Waterson's Θ (Waterson, 1975)								
[3] Tajima's D (Tajima, 1989)								
[4] Correlation coefficient								
[5]Ne (π): π/2μ								
[6]Ne (θw): θw/2μ								

Pyrenophora teres f. *teres* has a heterothallic mating system implying that mating only occurs between individuals of opposite mating types, Mat1-1 and Mat1-2 (43). To further investigate geographic variation in the frequency of sexual reproduction, we examined the mating type ratio, which is expected to be equal to one under random mating. We used the software SPAdes (44) to de novo assemble genomes and thereby validate and compare the frequency of mating type loci. The null hypothesis of random mating could not be rejected for the P.

teres f. teres populations except the North American population, for which we found a significant departure from the expected 1:1 ratio of mating types (Mat1-1:Mat1-2 = 3.5, Chi-squared test, p = 0.0184) (Figure 1D, Table S5). These analyses suggest that that *P. teres* f. *teres* is regularly undergoing sexual reproduction throughout most of its range. The skewed mating type frequency in the in North American population may reflect a more pronounced contribution of asexual reproduction.

Populations of P. teres f. teres are geographically structured

We characterized the population genetic structure of *P. teres* f. *teres* based on complementary methods using genome-wide SNP data. Firstly, we investigated the extent of clustering using a principal component analysis (PCA) (Figure 2A). The PCA mostly separated isolates according to their geographical origin. We further explored population structure by generating a Neighbour-net network with SPLITSTREE v. 4 and by inferring the extent of shared ancestry using an ADMIXTURE analysis (45). In the ADMIXTURE analysis, the Cross-Validation error used to select the most appropriate number of clusters (K) was minimized at K = 6 (Figure 2B, Table S6). The genetic clusters inferred from the ADMIXTURE analysis corresponded to five clusters mostly circumscribed to North America, North Africa, the Middle East, Europe, and Caucasus (Figure 2B), and a cluster of

three individuals restricted to the Caucasus referred to as the Caucasus-2 cluster. The European cluster, referred to as Europe+, was also present in the Caucasus (one isolate), and the Caucasus cluster, referred to as Caucasus+ was also present in the Middle East, in North Africa, and in Northern Europe. The ADMIXTURE analysis also revealed 15 individuals with shared ancestry in multiple clusters, suggesting admixture. Most of the individuals showing mixed ancestry (9/15) were derived from the Caucasus, where multiple clusters coexist. Furthermore, three isolates with mixed ancestry were derived from Europe, two from North Africa, and one from the Middle East.

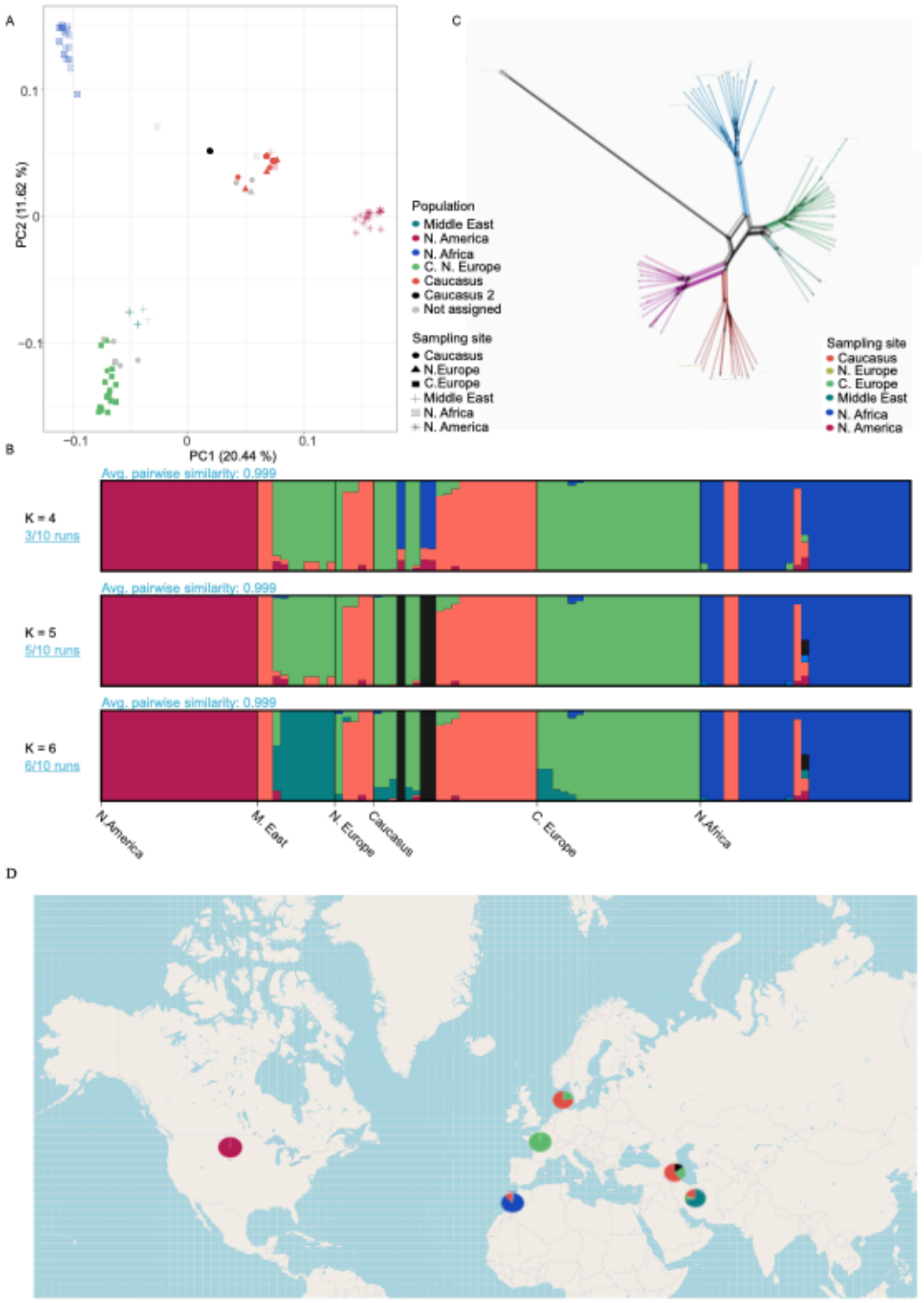


Figure 2: Global population structure of *P. teres* f. *teres*. A) PCA analysis, where shape reflects the origin of the isolate and colour reflects the genetic cluster set with ADMIXTURE at K = 6. B) The program ADMIXTURE was used to compute population structure between the six

geographical populations. Most fit number of hypothetical ancestral groups was identified as six based on the cross-validation method (Table S6). Here we present patterns of four, five, and six hypothetical ancestral groups. C) Neighbour-Net tree generated from SNP data from the P. teres f. teres populations. The branch colour reflects the genetic cluster set with ADMIXTURE at K = 6. D) World map shows the distribution and contribution of the genetic clusters identified by ADMIXTURE at K = 6 at the sampling sites. Underlying map based on OpenStreetMap (OpenStreetMap contributors) data (retrieved from https://www.openstreetmap.org/#map=2/19.1/54.7), freely available under Open Data Licence (https://www.openstreetmap.org/copyright).

The Neighbour-net phylogenetic network essentially revealed the same clusters as the ADMIXTURE analysis. All clusters were connected by reticulations indicating homoplasic mutations caused by incomplete lineage sorting or historical gene flow (Figure 2C). The Caucasus-2 cluster was connected to other lineages by a long, non-reticulated branch, consistent with a relatively long history of isolation from other clusters (Figure S2).

We applied a Mantel test to determine if genetic distance, simply measured as pairwise missmatches across the genome, is correlated with geographic distance between the isolates (46), and indeed confirm that geography explains some of the variation between clusters as spatial and genetic distances are correlated (Figure S3).

Phylogenomic analysis suggests an ancient split of the North African P. teres f. teres population and a Caucasian origin of the North American population

The Middle East, Caucasus, and North Africa are the regions that have the longest history of barley cultivation (16). Domesticated barley was introduced later to Europe and then to America. To test the hypothesis that early dispersal of P. teres f. teres occurred simultaneously with the spread of barley cultivation we inferred the evolutionary relationships between the populations of the pathogen. To this end, we have constructed a population tree, using polymorphism-aware models in IQ-TREE, using the Californian population as root for the P. teres f. teres populations (Figure 3). In this analysis, based on the full complement of polymorphisms, we found two major population splits: one lineage comprising the North African, Middle Eastern, and European population and another lineage comprising the Caucasus and North American populations. Within the former lineage, the branching harbouring the North African population diverged earlier than the branches harbouring the Middle Eastern and European populations. The clustering of Caucasus and North American populations suggested a Caucasian origin of the North American population.

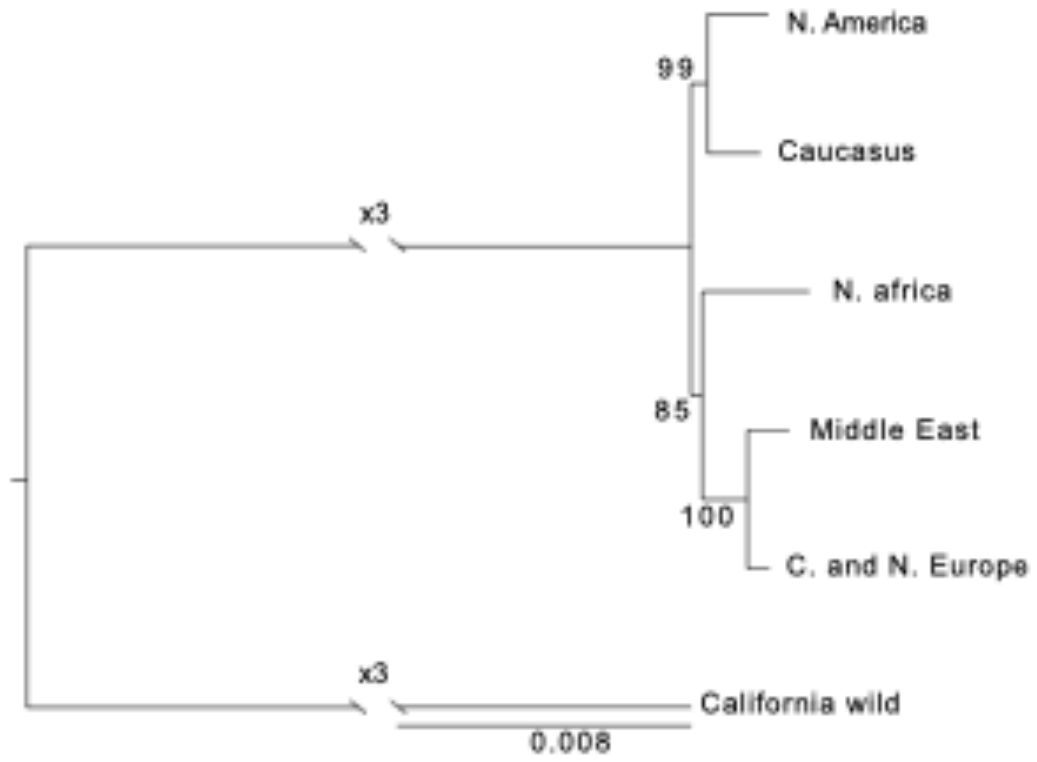


Figure 3: Evolutionary relationship between *P. teres* f. *teres* populations. A) Phylogenetic tree using polymorphism-aware models (PoMo) (Maximum likelihood inference, loglikelihood: -857128.545) to assess the evolutionary relationship between P. teres f. teres populations. Branch numbers reflect maximum likelihood bootstrap values. The tree was rooted using the Californian population. The scale bar represents the expected number of substitutions per site.

The origin and dispersal of *P. teres* f. *teres* correlates with the early history of barley cultivation

We further investigated the demographic history of *P. teres* f. *teres* populations using Approximate Bayesian computations (ABC) with supervised machine learning implemented in the DIYABC Random Forest software (DIYABC-RF) (47).

As the ABC framework requires populations not connected by continuous geneflow, we excluded isolates with shared ancestry in multiple clusters and considered five, non-admixed geographic populations for the analysis. Non-Caucasian and non-European isolates were excluded from the Caucasus+ and European clusters. We compared invasion scenarios in which the origin of each derived population was associated with a demographic bottleneck (see Materials and Methods) (48,49). For each scenario, we assessed the compatibility of the simulated datasets with the observed data using linear discriminant analysis (LDA), by simultaneously projecting simulated and observed data on the first two LDA axes. The overlap between the simulated datasets and the observed data indicated the compatibility of the simulated scenarios and the observed data (Figure S4, S5, S6).

Demographic inference with ABC was performed in three consecutive steps, each step corresponding to a different family of invasion scenarios. The first step considered simple scenarios of the three populations (Caucasus, N. Africa, and Middle East) which we found to be most distant from each other in the split tree analysis (Fig. 2), suggesting a more ancient divergence of these populations. The scenario complexity was then gradually increased as we assessed the evolutionary relationships of the additional populations in each consecutive step. For each family of scenarios, LDA confirmed that the chosen conditions

were suitable for the random forest analysis. To select the most probable hypothetical scenario from each family, we used a random forest classifier with 1,000 trees. Detailed results of DIY-ABC analyses are provided in Table S7 and Figure S4-S6.

Scenarios of family 1: Early divergence of Middle East, Caucasus, North African populations

To elucidate the most ancestral splits, we tested a total of 49 invasion scenarios with different ancestries and branching orders among populations from the Middle East, Caucasus and North Africa, regions where barley was first cultivated. Nine distinct categories of scenarios with similar topologies were considered. The most probable category of scenarios was “group 9” and scenario 45 (posterior probability of 0.778 and 0.401, respectively) (Table S8). Scenario 45 modelled an initial divergence of *P. teres* f. *teres* populations from the Middle East and the Caucasus, and a subsequent emergence of the North African population from the Middle Eastern population (Figure 4A).

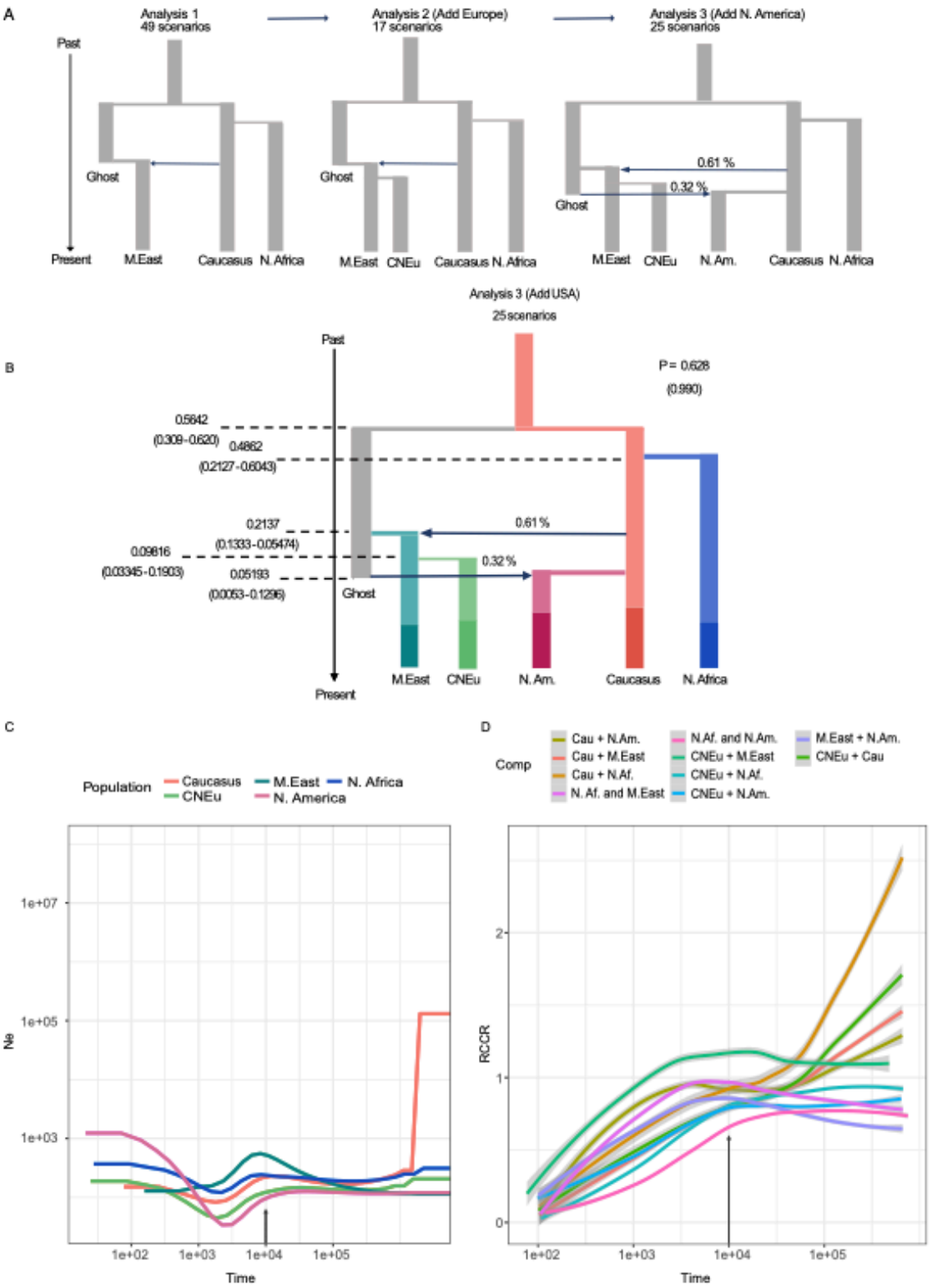

A
Analysis 1
49 scenarios
Analysis 2 (Add Europe)
17 scenarios
Analysis 3 (Add N. America)
25 scenarios
Past
Present
Ghost
M.East
Caucasus
N.Africa
CNEu
N.Am.
0.61 %
0.32 %
B
Analysis 3 (Add USA)
25 scenarios
P= 0.628
(0.990)
0.5642
(0.309-0.620)
0.4862
(0.2127-0.6043)
0.2137
(0.1333-0.05474)
0.09816
(0.03345-0.1903)
0.05193
(0.0053-0.1296)
C
Population
Caucasus
CNEu
M.East
N. America
N. Africa
Ne
Time
D
Comp
Cau + N.Am.
Cau + M.East
Cau + N.Af.
N. Af. and M.East
N.Af. and N.Am.
CNEu + M.East
CNEu + N.Af.
CNEu + N.Am.
M.East + N.Am.
CNEu + Cau
RCCR

Figure 4: Inference of the demographic history of P. teres f. teres populations shows an ancient split coinciding with barley domestication and early migration. A) Development of speciation scenarios across three analyses steps and using approximate Bayesian computation and Random Forest analyses implemented in DIY ABC-RF version 1.0. B) The most probable hypothetical evolutionary scenario of the migration routes of *P. teres* f. *teres* in the Middle East, North Africa, Europe, and North America based on the results of three sequential DIY ABC-RF analyses. The parameter "P" indicates the posterior probability of the most probable scenario. In paratheses are shown the probabilities of the different group containing the most probable demographic scenario (see Methods). We considered a classic invasion scenario for the topology building where each derived population passes through a bottleneck as it gives rise to a new population. Predicted time over the current Ne of the most ancestral population (NAzb) values inferred with random forest are shown. The 90% CI values for each parameter is provided in parenthesis. C) Changes in effective population size for all *P. teres* f. *teres* populations were estimated with MSMC2. The axes were scaled with a mutation rate of 4.5 x 10-7 per site per generation, and one generation per year. We have indicated the estimated time of barley domestication with an arrow. D) Relative cross-coalescence rate for all pairs of populations. Five runs of seven randomly selected individuals per population per run were performed. Shown here, we present the trend line fit between the five runs. Gray area around the line indicates the 95% confidence interval. The arrow indicates the estimated time of barley domestication.

Scenarios of family 2: Founding of the European population through migration from the Middle East

Having determined the branching order among populations from areas of more ancient barley cultivation, we proceeded to examine the more recent history of *P. teres* f. *teres* populations through a second family of scenarios. We compared 17 evolutionary scenarios modeling the relationships among the Central and Northern European populations, and other populations. In line with the population tree presented above, the most probable scenario category was “group 2” and scenario 2 (posterior probability of 0.701 and 0.412, respectively), suggesting that the Central and North European population derived from the Middle Eastern population (Figure 4A).

Scenarios of family 3: Origin of the North American population

The third family of scenarios modeled the origin of the North American population. According to Scenario 9 from group 5, which had the highest posterior probabilities (0.628 and 0.990, respectively), the North American population was established through admixture between the Caucasus population and an unknown "ghost" population (Figure 4A). This analysis also revealed that the earliest divergence was between the Caucasus and ghost population, which suggest that the Caucasus population is the oldest, and that the Middle East population emerged following admixture between the Caucasus and ghost populations.

Parameter inference analysis with DIY-ABC

To estimate the demographic parameters of the scenarios with highest posterior probabilities, we used a random forest with 1000 trees. Time estimates were estimated as the ratio of time over the current effective population size of the predicted most ancestral population, Caucasus population (NAz) (47). Our maximum posterior probability estimate of the divergence time between the Caucasus and unsampled ghost populations was 0.5640 generations/NAz (credibility interval [CI] 0.3090 – 0.6200). The North African population split from the Caucasus population 0.4862 generations/ NAz (CI 0.2127 – 0.6043), and the Middle East population emerged by the admixture of the Caucasus population and the ghost population around 0.2137 generations/ NAz (CI 0.1333 – 0.5474). The Central and Northern European population arose from the Middle Eastern population 0.0982 generations/ NAz (CI 0.0334 – 0.1903), and the North American population emerged 0.05193 generations/ NAz (CI 0.0053 – 0.1296) ago (Table 2, Figure 4B).

Table 2: Results for DIYABC-RF estimation of population divergence times normalized over the most ancestral sampled population N5 under scenario 9 detailed in Figure 4

Parameter	Prediction*	90% CI		Global (prior) NMAE	Local (posterior) NMAE
t1/N5	0,052	0,005	0,130	0,437	0,316
t2/N5	0,098	0,033	0,190	0,440	0,311
t3/N5	0,214	0,133	0,547	0,320	0,275
t4/N5	0,486	0,213	0,604	0,245	0,210
t5/N5	0,564	0,310	0,620	0,200	0,190
Note					

*RF analysis included 20,000 simulated data sets and the number of trees was set to 1000. Global (prior) and local (posterior) error rates were estimated using out-of-bag estimators from a sample of 10,000 data randomly chosen in a training set. CI, credibility interval; NMAE, normalized mean absolute error.

A severe demographic bottleneck in the history of the crop pathogen coincided with the domestication of barley

To further investigate population size variation through time and population split, we applied a Multiple Sequentially Markovian Coalescent (MSMC2) approach (50). Inferences of population size changes were performed using all available individuals (Figure 4C). Furthermore, we estimated the divergence time between populations using five independent runs of 14 randomly selected isolates per population pair (seven isolates per population) (Figure 4D). We found that the effective population size of the Caucasus population was initially the largest, but that the population subsequently experienced a demographic bottleneck. In agreement with the ABC analyses, we also found evidence for an early divergence between Caucasus and the North African population. Considering a mutation rate (μ) of 5.7 x 10-7 per base pair (51) and assuming

here on average one sexual generation per year (52), these events coincide with the domestication of barley about 7,267 to 14,545 years ago (Table S9).

We computed the Relative Cross Coalescence Rate (RCCR) for pairwise combinations of populations to estimate splitting times (Figure 4D). These analyses provided further support for the close evolutionary relationship between Central Europe and Middle East populations, as also observed in the PCA and NeighborNet tree analyses. Furthermore, the RCCR of the North America and Caucasus populations decays slower, which indicates extensive amounts of geneflow after divergence of the populations. The geneflow and late split between North America and Caucasus populations interfered with the RCCR, is in line with the emergence of the North American population from the Caucasian population as inferred by the ABC analysis.

We want to underline that our inference of actual coalescence times is based on assumptions that we considered reasonable for some unknown parameters. For example, the number of sexual cycles of P. teres f. teres is not known, and might even have varied throughout evolutionary times and host shift events. Nevertheless, the relative estimates of population divergence with two independent methods suggest that the Caucasus has been the center of origin of P. teres f. teres. Moreover, both methods applied here, provide evidence for an early divergence of a pathogen lineage in North Africa. In summary, our

inference of pathogen population history suggests a parallel dispersal of the pathogen alongside its host and emphasizes the fundamental importance of early agriculture on pathogen evolution.

Recent positive selection has shaped genomic regions encoding putative virulence-related genes

To identify genomic regions that may have experienced selective sweeps during the spread of P. teres f. teres, we used three methods (SweeD, OmegaPlus, and RAisD) which combine information from the site-frequency-spectrum (SFS) and patterns of LD and π along the genome (53–55). We conducted analyses on each population separately to identify population-specific selective sweeps. These selective sweeps may indicate local adaptation in the pathogen populations.

Demography can greatly impact the distribution of genetic variants along the genome and thereby bias inference of selective sweeps (56). We therefore combined the selective sweep analyses with simulations of genetic variation under different demographic scenarios (see Material and Methods). Multiple regions exhibiting signatures of selective sweeps were identified with the three methods (Table S10 – S15). We compared and combined selective sweeps maps of the three methods to get a final list of candidate regions exhibiting signatures of recent positive selection.

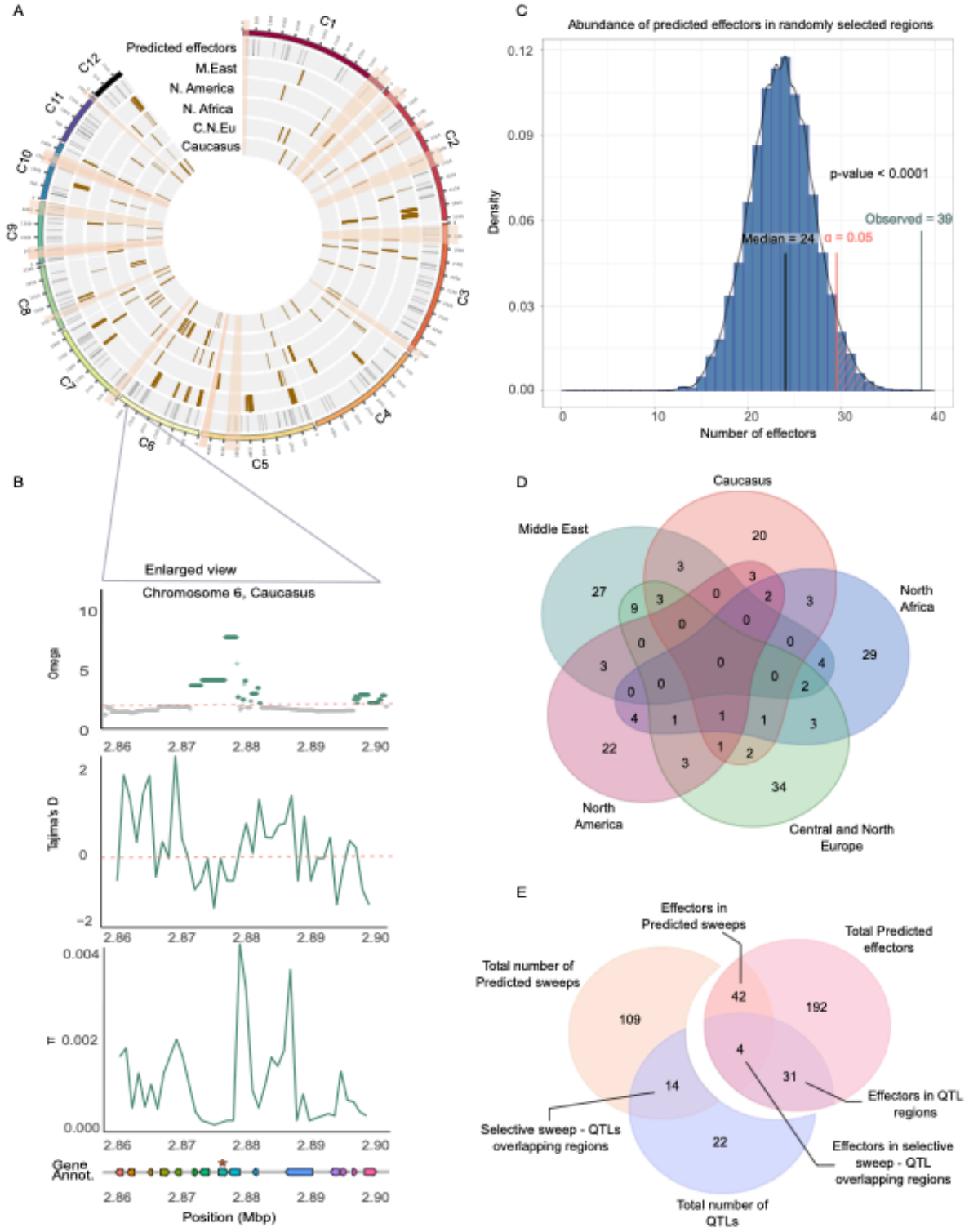

A
Predicted effectors
M.East
N. America
N. Africa
C.N.Eu
Caucasus
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
B
Enlarged view
Chromosome 6, Caucasus
Omega
Tajima's D
π
Gene Annot.
Position (Mbp)
C
Abundance of predicted effectors in randomly selected regions
Density
p-value < 0.0001
Observed = 39
Median = 24
α = 0.05
Number of effectors
D
Caucasus
Middle East
North Africa
North America
Central and North Europe
E
Effectors in Predicted sweeps
Total Predicted effectors
Total number of Predicted sweeps
Effectors in QTL regions
Selective sweep - QTLs overlapping regions
Effectors in selective sweep - QTL overlapping regions
Total number of QTLs

Figure 5: Distribution of selective sweeps across the genome in five P. teres f. teres. populations. A) Genomic map of selective sweeps for each population. The first track shows coordinates of genes encoding predicted effectors (40). Highlighted are the fourteen QTL regions associated with pathogenicity that were identified in previous studies. (38,39). B) OmegaPlus, Tajima's D, and nucleotide diversity (π) analyses across a selective sweep region on chromosome 6. Shown is only the Caucasian+ population. This region was identified to be under selection in all populations except the Middle Eastern. At the bottom, the gene and effector annotation presented. A predicted effector situated in the selective sweep region and another effector located in the genomic area. C) To determine if genes encoding putative effectors are enriched in selective sweep regions, we performed an enrichment analysis based on the distribution of predicted effector abundance in randomly selected genomic regions of the same number and length as the selective sweep regions. As many as 10,000 runs of random resampling of genomic regions was perform to validate that effector genes indeed are enriched in regions that have experienced recent positive selection. D) Venn diagram of selective sweep regions shared and unique to the *P. teres* f. *teres* populations. E) Effector content of the selective sweep regions and QTLs. Effector annotation was obtained from Wyatt et al., (2018), and previously reported QTLs associated with virulence in P. teres f. teres (38,39).

Table 3: Summary of total and unique selective sweep regions, as well as, genes and effectors located in the regions per populations. The number of sweep regions, total genes, and effectors only identified in one population are charachterised as unique.

Population	Total no of Sweep regions	Total no of genes	Total number of predicted effectors	Population-specific sweep regions	Population-specific genes	Population-specific effectors	Selective sweep - QTL overlap	QTLs name	QTL Reference
Middle east	27	253	14	11	198	13	1	VK2	40
N. America	22	107	9	8	75	8	3	VK1, VK2, AvrHar	40, 54
N. Africa	29	91	5	19	67	3	2	VK2, PttTif1	40, 39
C.&N. Europe	34	180	5	21	136	5	1	PttPin1	39
Caucasus	20	101	7	11	69	6	2	VR1, PttCell1	40, 39
California *(P.* sp)	24	155	5	12	108	4	3	VK1, PttCell1, Pttcell2	40, 39

Our final list of sweeps includes a total of 109 regions across all *P. teres* f. *teres* populations, with 20 to 27 selective sweeps per population (Figure 5A, D, Table 3). We identified 42 putative effector genes (40) colocalizing with the 109 selective sweep regions, suggesting that genes encoding virulence related traits such as effectors have been most prone to experience recent positive selection. Some selective sweep regions overlapped while others were unique to distinct populations, possibly representing adaptation to different resistance genes in barley or other local environmental conditions. For example, we found one selective sweep region on chromosome 6 (position 2,856,396 -2,964,731) that is present in all P. teres f. teres populations, except the Middle Eastern population. This region includes a gene encoding a predicted effector, which represents a candidate for future functional studies (Figure 5B, Table 3).

We tested if effector genes were significantly enriched in selective sweep regions. To this end, we performed a permutation test to assess the relative abundance of predicted effector genes in the selective sweep regions compared to the rest of the genome. Indeed, we found that the abundance of effector genes in selective sweep regions is higher compared to randomly sampled regions along the genome (Figure 5C).

We furthermore explored previously generated lists of candidate virulence determinants in *P. teres* f. *teres*. Previous studies have used quantitative trait locus (QTL) analyses to identify determinants of virulence on different barley cultivars (38,39). We found that 14 putative virulence related genes (QTL candidates) co-localized with selective sweep regions(38,39). (Figure 5E, Table 3). The QTL candidate region VK2 on chromosome 6 (39) overlapped with a selective sweep region, which was found in each of the populations in the Middle East, North America, and North Africa. The QTL candidate regions VK1 (39) and AvrHar (57) on chromosomes 3 and 5, respectively, co-localized with selective sweep regions predicted in the North American population. Two putative effectors are in the VK1 region on chromosome 3 (Table 3).

We predict that the selective sweeps in P. teres f. teres reflect recent adaptation to barley and local agricultural environments. We further addressed divergent adaptation in Pyrenophora pathogens on different hosts by comparing selective sweep maps of P. teres f. teres and the Pyrenophora population obtained from wild barley in California. To this end, we considered the windows that showed a composite likelihood ratio (CLR), ω, and μ higher than 99,95 % for significant outliers.

We identified 24 selective sweeps along the genome of the Californian population, including five genes predicted to encode effector genes.

Approximately half of the selective sweeps predicted in the Californian population were shared with the domesticated barley-infecting populations suggesting that the same suite of genes is important for virulence on wild and cultivated hosts. This hypothesis is further supported by the fact that some P. teres f. teres QTL candidates (39) overlap with selective sweep regions in the wild barley pathogen (Table 3). Hereby, also the QTL locus VK1 co-localized with selective sweeps in the wild-barley pathogen.

In summary, the selective sweep analyses identify multiple loci in P. teres f. teres that have experienced recent positive selection. Functional analyses of candidate genes in these regions may shed light on the adaptation of the pathogen to different barley cultivars.

Discussion

Understanding the evolutionary origin of crop pathogens is crucial to predict future epidemics. In this study, we addressed the history of the globally occurring pathogen of barley, *P. teres* f. *teres*. We used a global population sample and extensive genome sequencing to assess the population structure and demographic history of the crop pathogen. Our analyses were based on the hypothesis that P. teres f. teres could have emerged and co-evolved with barley

during early crop domestication. Extensive sampling of the pathogen in geographical regions representing the most ancestral history of barley domestication and cultivation (58) allowed us to dissect the early history of P. teres f. teres. We also characterized the population structure and demography from present-day barley-producing countries, including France, Denmark and the USA. Our detailed population genomic analyses provide evidence for a scenario where P. teres f. teres emerged in the Middle East at the onset of barley domestication, and subsequently dispersed with Neolithic farmers to North Africa and Europe.

We compared measures of nucleotide diversity among the different P. teres f. teres populations and observed higher diversity in the Caucasian and North African populations. Notably the North American population represented an overall low nucleotide diversity indicating either a more recent origin of the population or a recent bottleneck.

Next, we applied two independent methods to infer the population histories of *P. teres* f. *teres*. Both methods provide evidence for a scenario where the most ancestral populations of *P. teres* f. *teres* have originated in the Fertile Crescent region. We note that our inferences of population histories based on the Middle East, Caucasus, and North Africa populations, may have been affected by sampling bias as only ten isolates were available from the Middle East, in

contrast to 21 and 27 from Caucasus and North Africa, respectively. For the parameter inference of ABC-RF, we have calculated time over the effective population size to reduce the error of parameter inference, as suggested in (47). We computed current effective populations sizes based on the SNP data to rescale parameters with a mutation rate of 5.7 x 10-7 (25) and one generation per year (see Materials and methods). Using these values, we estimated that the Caucasian population was founded around 13,000 years ago (CI: 7,267 – 14,546). The history of *P. teres* f. *teres* not only parallels the evolution of barley. It also parallels the history of a small number of other prominent crop pathogens which have emerged and co-evolved with their host during domestication. Other important pathogens that emerged with their host during domestication include the wheat pathogenic fungus *Zymoseptoria tritici* causing the disease septoria tritici blotch (59), the rice blast fungus *Magnaporthe oryzae* (60), and the corn smut fungus *Ustilago maydis* (61). In these studies, coalescence analyses were used to infer the divergence time between wild and crop-infecting populations of the pathogen and to infer major demographic events, such as bottlenecks that coincide with the domestication of the host. The emergence of *P. teres* f. *teres* was also associated with a considerable population bottleneck probably reflecting strong selection on pathogen individuals with the right gene combination necessary to invade a new host niche.

Interestingly, we find evidence for the early emergence of a distinct *P. teres* f. *teres* population in North Africa. Using the above-mentioned scaling the divergence between pathogen populations in North Africa and Caucasus occurred around 11,400 years ago (CI: 4,991 – 14,181). This scenario is in agreement with the introduction of barley into North Africa by neolithic farmers and the early development of distinct barley varieties (62).

Archaeological remains suggest that barley was cultivated in Central Europe from approximately 6,000 (63) to 4,800 years ago (64). We find evidence that the *P. teres* f. *teres* pathogen accompanied the introduction of barley as our population genomic data suggest the emergence of the European pathogen population occurred around 2,300 years ago (CI: 785 – 4,465). More recently, the North American population has split from all other populations. Our analyses suggest the emergence of the North American population approximately 1,200 years ago (CI: 125 – 3,044), which conflicts with the much later introduction of barley to North America by European migrants. We speculate that this inconsistency reflects the uncertainty of our parameter scaling and note that the confidence interval of our estimates still concurs with a European-based introduction to North America a few centuries ago (19). Moreover, our data strongly corroborate a late divergence between the North American and

Caucasian populations. While we included a large sample of isolates, covering a broad geographic region, it is still possible that some pathogen variation was not collected in Europe. Consequently, an intriguing hypothesis emerges: what we recognize as the North American population may have originated from a distinct population in Europe which was not sampled here.

In conclusion, the demographic history of *P. teres* f. *teres* recovered using ABC and MSMC2 reflects the introduction of its host, barley, in different locations, highlighting the significant role of historic trading in the dispersal of crop pathogens (Figure 6).

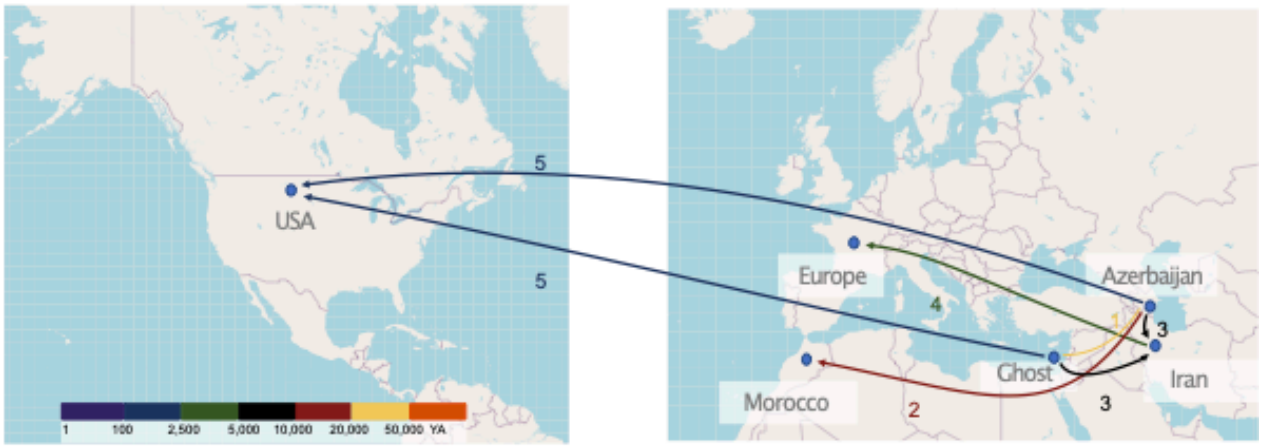


Figure 6: Invasion scenario of P. teres f. teres based on our population and phylogenomic analyses and the known history of the host. Different colours represent the approximate period when the proposed events occurred. (1) The Fertile Crescent is the most plausible center of origin of the pathogen. (2) Ancestral divergence of the North African population is consistent with an early migration of the pathogen to North Africa, possibly with early barley cultivation by neolithic farmers in North Africa. (3) More recent populations have emerged in Europe and (4) North America. Underlying map based on OpenStreetMap (OpenStreetMap

contributors) data (retrieved from https://www.openstreetmap.org/#map=2/19.1/54.7), freely available under Open Data Licence (https://www.openstreetmap.org/copyright).

We used the P. teres f. teres SNP data to compute overall nucleotide diversity in the pathogen populations. Interestingly, *P. teres* f. *teres* showed higher nucleotide diversity (π) compared to other prominent pathogens such as the wheat pathogen *Zymoseptoria tritici* (65), the wheat powdery mildew pathogen *Blumeria graminis* f.sp. *tritici* (25), and the rice blast fungus *Magnaporthe oryzae* (66). Genetic variation is instrumental in rapid adaptation of pathogens and the high nucleotide diversity in *P. teres* f. *teres* may be an important factor in the successful spread of the pathogen. Differences between species may be explained by different extent of sexual recombination and gene flow and can also be highly impacted by past population bottlenecks. Based on our LD analyses and the distribution of mating type frequencies, we find evidence for frequently occurring sexual recombination in P. teres f. teres (67). An exception is the North American population that exhibits a higher extent of clonality. The population genetic structure of the North American population is in agreement with a recent founder event of the population, but may also reflect the large-scale monocropping systems in North America that may favour clonal spread of the pathogen over large spatial scales.

We identify six genetic clusters that largely correlate with geographic origin. However, some isolates appear to co-exist and pertain to distinct clusters with little evidence of introgression. This observation may indicate local adaptation and possibly some limits to gene flow, for example between isolates adapted to distinct barley cultivars.

To explore signatures of local adaptation, we used different methods to identify selective sweeps. In total we identify 109 selective sweep regions. We asked how many, and which selective sweeps would be common among the barley-infecting populations; considering that these loci could represent important host specificity loci. Intriguingly, most selective sweeps are population-specific or shared among a small number of populations. Only 23 regions are shared among all *P. teres* f. *teres* populations. Several selective sweeps are shared with the Californian population of P. teres occurring on wild grasses. Possibly, genes in these “conserved” sweep regions represent fundamentally important virulence traits.

For the population-specific sweeps, we speculate that the pathogen undergoes strong selection from its local environment, including barley cultivars utilized in different countries. We find that selective sweeps are enriched with predicted effector genes in our regions, which may highlight how host genetics is a main driver of rapid evolution in this pathogen. In addition, as many as 15 out of the

109 selective sweep regions were previously identified as candidate loci in QTL studies aiming to identify virulence determinants in *P. teres* f. *teres* (38,39).

Conclusion

A growing body of evidence suggests that human activities play a major role in the emergence and dispersal of plant pathogens (25,61,68). Here, we employed population genetic approaches with statistical and simulation tools to unravel the population structure and dispersal history of a major fungal barley pathogen. Our results sediment the conclusion that crop domestication in the Neolithic was accompanied by the emergence of several new plant pathogens; pathogens which co-evolved and spread with their hosts and presently represent some of the most important crop diseases we have. Wild relatives of domesticated plants represent important resources of genetic resistances. Likewise, they may be hosts to “wild populations” of pathogens. Exploring genetic variation in natural plant-pathogen systems holds a large potential for the discovery of new crop resistances as well as pathogen virulence determinants.

Material and methods

Genome data

124 *P. teres* whole genomes were sequenced using Illumina technology (PRJNA923641) (69). The sequenced isolates were sampled from barley fields on four different continents. Twenty isolates were obtained from two North Dakota State University experimental fields in Fargo and Langdon, North Dakota, USA. 27 isolates were collected from six locations in Morocco, North Africa. Ten strains were isolated in Iran, and 21 isolates were sampled from five locations in Azerbaijan, South Caucasus. Twenty-one isolates were sampled in Central and Northern Europe and five in Denmark, Europe. Finally, we included a collection of isolates from a wild barley species collected in California, USA. The Californian isolates were included with the purpose of identifying recently diverged genomic features in the barley-infecting populations of *P. teres* f. *teres* (Table S1).

Read mapping and variant calling

A pipeline was developed to filter and map Illumina reads to a reference genome and extract high-quality single nucleotide polymorphisms (SNPs). In brief, the program Trimomatic version 0.38 (Bolger et al., 2014) was used to filter and trim sequencing adapters, nucleotide bases, based on sequencing quality (PHRED 33), and read length (reads shorter than 30 bp were discarded). Overlapping

reads were merged using PEAR version 0.9.11 (Zhang et al., 2014). Burrows-Wheeler Aligner (BWA) version 0.7.17 (72) and Stampy v. 1.0.20 were used (73) to map individual reads to the reference genome of P. teres f. teres (0-1 P. teres f. teres genome GCA_000166005.1) obtained from the NCBI (40) (Table S2). Haplotyping and genotyping procedures were performed with the GATK HaplotypeCaller version 4.2.18 (74), providing a final VCF file with the raw SNP calls.

We next conducted filtering of SNPs based on the following criteria: (1) The call quality divided by the depth of sample reads should be larger than 2, (2) the depth per position should be higher than 8, (3) mapping quality of reads supporting each SNP should be higher than 40, (4) allele-specific rank sum test for mapping qualities of the reference (REF) versus alternative (ALT) reads should be higher than -12.5, (5) allele-specific rank sum test for relative positioning of REF versus ALT allele within reading must be higher than -8. (6) each genome has to have an average read coverage of at least 2. For the application of these filters, GATK VariantFiltration version 4.0.11 was used (74). After applying these hard-filtering criteria, 1,092,635 SNPs were kept, and we further refer to this dataset as the “full high-quality dataset”.

We note that additional subsets or filtering steps were added for specific analyses. Most clustering analyses assume that the markers used are

independent. Therefore, for these analyses, we filtered the full high-quality dataset based on the linkage disequilibrium (LD) decay patterns considering a distance of at least 3,12 Kbp (distance of r2/2 averaged across populations) between SNPs (Table 1). After filtering for LD, a dataset of 465,963 SNPs was retained. We refer to this dataset as the "independent SNP dataset".

We also generated a dataset of SNPs exclusively located in non-coding, presumably neutrally evolving genome regions. For this, we excluded all SNPs located in predicted gene regions and in 500 bp gene-flanking regions, both upstream and downstream, in a third filtering step based on LD. We obtained the coordinates of the genes from the 0-1 reference annotation file (40). After this filtering step, 160,472 high-quality independent and presumably neutrally evolving biallelic SNPs were kept. We refer to this as the "neutral dataset", which we used to infer the demographic history of the species.

Population genetic structure

Population genetic structure was inferred using three different approaches: a principal component analysis (PCA), ADMIXTURE version 1.3 (75), and Neighbour-Net analyses (76), each of them based on the "independent SNP dataset." The PCA was here applied to reveal genetic clustering among the isolates and was created using the R package SNPRelate v. 1.6.4 and visualized

with the R package ggplot2 (77). Population structure was further characterized using a maximum likelihood approach implemented in ADMIXTURE. Ten replicate runs for a range of K-values (1-10) were performed. The best K value was determined using the ADMIXTURE software to estimate the cross-validation error(75) (Table S6). Finally, we explored the genetic structure and reticulation patterns among lineages, using the distance-based method for constructing Neighbour-Net networks as implemented in the program Splitstree 4 version 4.15.1 (78). A mantel test was performed to assess the correlation of geographic and genetic distance using the R package ape v. 5.6-2 (79)

Genetic diversity, neutrality tests and linkage disequilibrium

We further used the "full high-quality dataset" to compute and compare genetic variation among populations. ANGSD v.0.939 (80) was used to estimate the genetic diversity for each population as the nucleotide diversity (π) and the number of segregating sites ($W\theta$), as well as values of Tajima's D. Furthermore, to assess the effects of sampling bias on the genetic diversity estimations, we recalculated and visualize π if five independent rans using from two to seven individuals per population (figure S7). The tool PopLDdecay (81) was used with default parameters to estimate the linkage disequilibrium (LD) decay for each genetic cluster. VCFtools v. 0.1.17 (82) was used to calculate the fixation index

Fst (83) between the populations. A Kruskal-Wallis test with post-hoc pairwise Wilcoxon was used to identify significant differences ($p < 0.05$).

Mating types

MAT1-1and MAT1-2 mating type sequences were obtained from GenBank (accession no. HM121994 and HM122006, respectively) (84). Assemblies were created using SPAdes (44) with default parameters for each isolate. Subsequently, the mating type of each assembly was assessed by blasting the mating type sequences against them. To that end, blastn (85) with default parameters was used.

Phylogenetic reconstruction

A Maximum likelihood approach was applied to assess phylogenetic relationships between the population isolated from wild barley in California and six others Pyrenophora species. Sequences of four DNA loci (ITS, LSU, tub2, and tef1-a) were extracted from Alternaria alternata (SRC11RK2F), *Pyrenophora teres* f. *maculata* (SAMN15340022), *Pyrenophora teres* f. *teres* (SRS084801), Pyrenophora graminea (CBS 336.29), *Pyrenophora tritici-repentis* (V0001), and *Pyrenophora seminiperda* (SAMN02981545) from NCBI. Californian isolate CAWB5, showing the highest raw read count among the Californian isolates was

selected to represent the group. Subsequently, it was assembled with the software SPAdes (44), following the process described under the “Mating types” section, and the sequences of ITS, LSU, tub2, and tef1-a were extracted. Consensus sequences of the four individual loci were aligned with MAFFT v 7.490 (86) using default parameters, manually adjusted using Unipro Ugene v. 43.0 (87), and concatenated using SeqKit (88). The concatenated alignment was then subjected to maximum-likelihood (ML) analysis using Iqtree version 2.0.3 (89). The best-fitting substitution model was chosen based on the Bayesian Information Criterion (BIC) using the ModelFinder algorithm implemented in Iqtree version 2.0.3 (90). Moreover, 1000 bootstrap replicates were performed to obtain branch support values using the bootstrap approximation option of Iqtree (91). Further support for the phylogenetic inference was provided by a maximum-parsimony (MP) analysis using MPBoot (92). Similar to ML, MP analysis was performed, including 1000 bootstrap replicates. Alternaria alternata (GCF_001642055.1) was selected as the outgroup taxon for both ML and MP analyses. The resulting trees were edited in FigTree 1.4.4.

IQ-TREE polymorphism-aware models (PoMo) (93) were used to reconstruct relationships between *P. teres* f. *teres* populations. For the preparation of the input file, the FastaVCFtoCount.py script provided with the PoMo software was used. Similar to the previous phylogenetic analysis, the best-fitting substitution

model was chosen based on the Bayesian Information Criterion (BIC) using the ModelFinder algorithm implemented in Iqtree version 2.0.3. Again, 1000 bootstrap replicates were performed to obtain branch support values using the bootstrap approximation option of Iqtree. The Californian P. teres population was used as an outgroup in this analysis.

Inference of the demographic history of *P. teres* f. *teres* populations

The demographic history of *P. teres* f. *teres* populations was inferred using approximate Bayesian computation (ABC) with a supervised machine learning algorithm implemented in DIYABC-RF version 1.0 (47). Since the ABC framework requires populations without continuous gene flow, the five non-admixed, well-populated clusters revealed by the population structure analyses were used in this analysis. Due to the limited number of individuals in the cluster (only three isolates from Caucasus), this second Caucasus cluster was excluded from the inference. Out of the five remaining clusters, three were entirely consistent with the geographical origin of the populations: North Africa, Middle East, and North America. For the remaining two genetic clusters, the composition of isolates did not reflect on a single geographic location but rather a mixture of isolates from different locations, although these clusters originated primarily from Europe and Azerbaijan, Caucasus (Table 1). Most of the isolates (19/20) of the fourth cluster

originated from Europe (France and Denmark). Similarly, the majority (12/18) of the fifth cluster isolates originated from Caucasus. For the inference of the demographic history, we only kept the 19 isolates from Europe, representing the fourth genetic cluster, and the 12 isolates originated from Caucasus, representing the fifth cluster.

Since the records about *P. teres* f. *teres* invasion history are scarce, the phylogenetic analyses obtained with PoMo were incorporated as the starting point to construct hypothetical evolutionary scenarios. First, the three populations (North Africa, Middle East, Caucasus) that were further apart from each other on the tree, indicating ancient split and isolation among these, were selected as a starting point. Three sequential DIYABC-RF analyses were performed as follows: For the first analysis, 49 scenarios were tested, describing cases where (1) either of the single population or (2) an admixture event between two populations gave rise to the other populations (figure S4).
Considering a wider geographical distribution of *P. teres* f. *teres* not covered by our sampling, we also included scenarios that tested an unsampled (referred to as a "ghost") population as the putative ancestral population. We included scenarios where either a present-day sampled population was derived from a ghost population or the present-day population emerged by admixture from a

ghost population with another sampled population. These scenarios were analyzed individually and combined in groups of similar scenarios (47). For the combined groups, scenarios were joined into thirteen groups based on the population that was most ancestral (Table S8): Groups 1,2, and 3 consist of scenarios considering the Caucasus, North Africa, and Middle East population as the origin, respectively. Groups 4,5 and 6 consider North America, Middle East, and Caucasus to have been established through an admixture event of the other two populations, respectively. Group 7 considers that all three populations diverged at the same time. Groups 8 to 13 consider a ghost population to be parental to one of the sampled populations. A detailed description of the scenarios and scenario families can be found in the Supplementary material S1 and Table S7.

In the second analysis, 17 scenarios were considered to assess the emergence and relationship of the European population in relation to the putatively ancestral populations from Caucasus, Middle East, and North America (Figure S5). Group 1 consisted of scenarios where Europe and Caucasus have the same ancestral population (scenarios 1,8,9,16). Group 2 considers Europe and Middle East share a common ancestor (scenarios 2,13). Group 3 considers Europe and North Africa share a common ancestor (scenarios 3,6,14). Group 4 considers that Europe and the “ghost” population share a common ancestor (scenarios 4,7).

Group 5 (scenarios 10, 11, 12) consider the European population to be the product of admixture of two other sampled populations. Group 6 considers scenario 17, where the Europe population is the “ghost” population identified in step 1. The best scenario, selected by random forest in analysis two, was used as the base for analysis 3.

In the third analysis, we assessed the relationship of the population originating from the North America with the rest of the population (Figure S6). As many as 26 scenarios were tested. As in the previous steps, group 1 consisted of scenarios where North America and Caucasian have the same ancestral population (scenarios 1,8,9,16). Group 2 considers North America and Middle East to share a common ancestor (scenarios 2,13). Group 3 considers North America and Europe to share a common ancestor (scenarios 3,7,10,23,24). Group 4 considers North America and North Africa share a common ancestor (scenarios 4,8,11,13,25). Group 5 considers the North America and the “ghost” population to share a common ancestor (scenarios 5,9,12,14,15). Group 6, consisting of scenarios 6,7,8,10,11,13 considers the North America population to be the product of admixture of two other sampled populations. Group 7, consisting of scenario 26, considers the North America population as the “ghost” population identified in step 1.

A uniform prior distribution for all effective population sizes in the three analyses ranged from 500 and 5,000,000. The changes in effective population sizes were split into recent changes, where the uniform prior distribution was between 10 and 20,000 years ago, and ancient, where the distribution was between 10 and 1,000,000 years ago. The later distribution was used for all splitting and admixture times with uniform probability. For the scenarios that consider the origin of a population through admixture, the prior distribution of the contribution of each parental population was set to be between 0.01 and 0.99. The random forest analysis for the model choice and the parameter inference was performed with 1000 decision trees and default values of DIYABC-RF. Furthermore, DIYABC-RF uses Hudson's ms simulator (94) with the "-s" parameter to introduce a fixed number of segregating sites under each scenario. Mutation parameters are, in this case, not needed for the simulations (47). Eventually we have scaled the time values by estimating the current effective population sizes based on Watterson's theta and a mutation rate of 4,5 x 10-7 (25) (Table S9).

MSMC2 v2.1.1 (50) was applied to infer changes in the effective population size through time. MSMC2 uses a backward-in-time algorithm to build back genome lineages. The MSMCtools bamCaller script was used for the preparation of mask

files for the low-coverage regions and to “diploidize” the haploid vcf files. After that, the script generate_multihetsep.py included in the MSMC2 software package was used to create the input files for the analysis. Changes in effective population size were inferred using all the isolates available for each population (from 7 for Middle East to 21 for North Africa). Subsequently, the cross-coalescent rate between the populations was estimated using 100 iterations. To this end, five independent runs of 14 randomly selected isolates per population pair (seven isolates per population) were performed for each pair of populations. The coalescence rate within populations and the cross-coalescence rate between populations was re-calculated based on the 14 randomly selected haplotypes in each run.

Genome scans for selective sweeps

Three independent approaches were applied to identify signatures of selective sweeps along the *P. teres* f. *teres* genome. Hereby we used the programs SweeD (53), 2), OmegaPlus v. 3.0.3 (54), and 3) RaiSD v 2.9 (55). The analyses were conducted with the full high-quality dataset. SweeD v. 3.0 uses the Site Frequency Spectrum (SFS) patterns of SNPs to estimate a composite likelihood ratio (CLR) test for detecting complete sweeps (Pavlidis et al., 2013). We used SweeD, OmegaPlus, and RaisD individually for each genetic cluster and each of

the 12 chromosomes and a grid size equal to the number of SNPs present in each chromosome (28,698 – 77,617 points). OmegaPlus is a scalable implementation of the ω statistic (95) that can be applied to whole-genome data. It uses a maximum likelihood framework and utilizes information on the LD between SNPs. For OmegaPlus, the minimum and maximum window sizes were set to 1 kb and 100 kb, respectively. RaiSD computes the μ statistic, a composite evaluation test that scores genomic regions by quantifying changes in the SFS, the levels of LD, and the amount of genetic diversity along the chromosome (55). We used RaiSD with the default window size of 50 kb.

Changes in genetic variation and LD along the genome are also influenced by demography. To account for the effect of demography and determine the significance level of the identified selective sweeps, we simulated 10,000 datasets under the best neutral demographic scenario using the program ms (94), to mimic a population evolving under the same conditions as *P. teres* f. *teres*, but without any effect of selection. We then computed the ω and μ statistics on this data. Setting a significance threshold for the deviation of the ω and μ statistics based on the simulated data sets allowed us to control for the effect of the demographic history of the population on the SFS, LD, and genetic diversity along the genome (53,96). Subsequently, we only kept the selective sweep regions with evidence of selection from at least two methods to control

for false positives. Genome-wide maps of the sweep regions were created using Circos v. 0.69-9 (97).

To test if the abundance of effector genes was different for the predicted selective sweep regions compared to the rest of the genome, we used a permutation test (based on a custom script available in GitHub: https://github.com/Jimi92/Population-genomics-Pyrenophora-teres). In brief, the abundance of predicted effectors was counted in regions of the same size and number equal to the predicted sweep regions. As many as 10,000 replicate runs of random resampled region were performed.

Co-localization of selective sweep regions, QTLs and predicted effectors

QTLs associated with *P. teres* f. *teres* virulence were published in previous works (38,39,101). In addition, effector prediction was performed in a previous work (102)We have compared the candidate selective sweep regions obtained through our analyses to the reported QTL and predicted effector coordinates using BEDTools version 2.27.1 (103)

Data Availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. Genome sequences are accessible through

Zenodo under (DOI: 10.5281/zenodo.8183372). Custom scripts and workflows are available at https://github.com/Jimi92/Population-genomics-Pyrenophora-teres.

Acknowledgements

Asieh Vasighzadeh kindly provided isolates from Iran which were used in this study. The authors are grateful to Idalia Rojas Barrera, Danilo Pereira, Wagner Fagundes and other members of the Environmental Genomics group. DT's work was supported by the DFG Research Training Group RTG 2501. The research was further supported by the U.S. Department of Agriculture, Agricultural Research Service through USDA project 3060-22000-051-000D (Friesen).

References

1. Fellers JP, Sakthikumar S, He F, McRell K, Bakkeren G, Cuomo CA, et al. Whole-genome sequencing of multiple isolates of Puccinia triticina reveals asexual lineages evolving by recurrent mutations. G3: Genes, Genomes, Genetics. 2021;11(9).

2. Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. BMC Biol. 2016;14(84):1–11.

3. Ali S, Gladieux P, Leconte M, Gautier A, Justesen AF, Hovmøller MS, et al. Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. tritici. PLoS Pathog. 2014;10(1):1–13.

4. Liu Z, Ellwood SR, Oliver RP, Friesen TL. Pyrenophora teres: Profile of an increasingly damaging barley pathogen. Mol Plant Pathol. 2011;12(1):1–19.

5. Douiyssi A, Rasmusson DC, Roelfs AP. Responses of barley cultivars and lines to isolates of Pyrenophora teres. Plant Dis. 1998;82(3):316–21.

6. Louw JPJ, Victor D, Crous PW, Holz G, Janse BJH. Characterization of Pyrenophora Isolates Associated with Spot and Net Type Lesions on Barley in South Africa. Journal of Phytopathology. 1995;143(3):129–34.

7. Zeleke T. Evaluation of host reaction and yield performance of malt barley cultivars to net blotch, Pyrenophora teres in Bale Highlands, Ethiopia. Journal of Plant Sciences. 2017;5(1):43–7.

8. Arabi MIE, Al-Safadi B, Charbaji T. Pathogenic variation among isolates of Pyrenophora teres, the causal agent of barley net blotch. Vol. 151, Journal of Phytopathology. 2003. p. 376–82.

9. Sato K, Takeda K. Net blotch resistance in wild species of Hordeum. Euphytica. 1997;95(2):179–85.

10. Arabi MI, Barrault G, Sarrafi A, AlbertiniL. Variation in the resistance of barley cultivars and in the pathogenicity of Drechslera teres f. sp. maculata and D. teres f. sp. teres isolates from France. Plant Pathol. 1992;41(2):180–6.

11. Tekauz A. Characterization and distribution of pathogenic variation in Pyrenophora teres f. teres and P. teres f. maculata from western Canada. Canadian Journal of Plant Pathology. 1990;12(2):141–8.

12. Moya P, Girotti JR, Toledo AV, Sisterna MN. Antifungal activity of Trichoderma VOCs against Pyrenophora teres, the causal agent of barley net blotch. J Plant Prot Res. 2018;58(1):45–53.

13. Cromey MG, Parkes RA. Pathogenic variation in Drechslera teres in New Zealand. New Zealand Plant Protection. 2003;56:251–6.

14. van Zeist W, Bakker. View of Archaeobotanical studies in the Levant: I. Neolithic sites in the Damascus Basin: Aswad, Ghoraife, Ramad. Palaeohistoria. 1982;24:165–256.

15. Åberg E. Hordeum agriocrithon nova sp., a wild six-rowed barley. Ann R Agric Col Swed. 1938;6:159–216.

16. Molina-Cano JL, Fra-Mon P, Salcedo G, Aragoncillo C, de Togores FR, García-Olmedo F. Morocco as a possible domestication center for barley: biochemical and agromorphological evidence. Theoretical and Applied Genetics. 1987;73(4):531–6.

17. Molina-Cano JL, Moralejo M, Igartua E, Romagosa I. Further evidence supporting Morocco as a centre of origin of barley. Theoretical and Applied Genetics. 1999;98(6–7):913–8.

18. Tanno K, Taketa S, Takeda K, Komatsuda T. A DNA marker closely linked to the vrs1 locus (row-type gene) indicates multiple origins of six-rowed cultivated barley (Hordeum vulgare L.). Theoretical and Applied Genetics. 2002;104(1):54–60.

19. Riehl S. Barley in Archaeology and Early History. Oxford Research Encyclopedia of Environmental Science. 2019. 1–26 p.

20. McDonald BA, Stukenbrock EH. Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016;371(1709).

21. Banke S, Mcdonald BA. Migration patterns among global populations of the pathogenic fungus Mycosphaerella graminicola. Mol Ecol. 2005;14(7):1881–96.

22. Feurtey A, Lorrain C, McDonald MC, Milgate A, Solomon PS, Warren R, et al. A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. Nat Commun. 2023;14(1).

23. Munkacsi AB, Stoxen S, May G. Ustilago maydis populations tracked maize through domestication and cultivation in the Americas. Proceedings of the Royal Society B: Biological Sciences. 2008;275(1638):1037–46.

24. Stukenbrock EH, Mcdonald BA. The Origins of Plant Pathogens in Agro-Ecosystems. Annu Rev Phytopathol. 2008;46:75–100.

25. Sotiropoulos AG, Arango-Isaza E, Ban T, Barbieri C, Bourras S, Cowger C, et al. Global genomic analyses of wheat powdery mildew reveal association of pathogen spread with historical human migration and trade. Nat Commun. 2022;13(1):1–14.

26. Santini A, Liebhold A, Migliorini D, Woodward S. Tracing the role of human civilization in the globalization of plant pathogens. ISME Journal. 2018;12(3):647–52.

27. Menardo F, Praz CR, Wyder S, Ben-David R, Bourras S, Matsumae H, et al. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. Nat Genet. 2016;48(2):201–5.

28. Thierry M, Charriat F, Milazzo J, Adreit H, Ravel S, Cros-Arteil S, et al. Maintenance of divergent lineages of the Rice Blast Fungus *Pyricularia oryzae* through niche separation, loss of sex and post-mating genetic incompatibilities. PLoS Pathog. 2022 Jul 1;18(7).

29. Derbyshire MC. Bioinformatic Detection of Positive Selection Pressure in Plant Pathogens: The Neutral Theory of Molecular Sequence Evolution in Action. Front Microbiol. 2020;11(April):1–14.

30. Tellier A, Moreno-Gamez S, Stephan W. Speed of adaptation and genomic signatures in arms race and trench warfare models of host- parasite coevolution. Evolution (N Y). 2014;68(8):73–8.

31. Ebert MK, Rangel LI, Spanner RE, Taliadoros D, Wang X, Friesen TL, et al. Identification and characterization of Cercospora beticola necrosis-inducing effector CbNip1. Vol. 22, Molecular Plant Pathology. 2021. p. 301–16.

32. Duan G, Bao J, Chen X, Xie J, Liu Y, Chen H, et al. Large-Scale Genome Scanning within Exonic Regions Revealed the Contributions of Selective Sweep Prone Genes to Host Divergence and Adaptation in Magnaporthe oryzae Species Complex. Microorganisms. 2021;9(3):562.

33. Richards JK, Stukenbrock EH, Carpenter J, Liu Z, Cowger C, Faris JD, et al. Local adaptation drives the diversification of effectors in the fungal wheat pathogen Parastagonospora nodorum in the United States. Vol. 15, PLoS Genetics. 2019. 1–32.

34. Pereira D, Oggenfuss U, McDonald BA, Croll D. Population genomics of transposable element activation in the highly repressive genome of an agricultural pathogen. Microb Genom. 2021;7(8).

35. Rovenich H, Boshoven JC, Thomma BPHJ. Filamentous pathogen effector functions: Of pathogens, hosts and microbiomes. Curr Opin Plant Biol. 2014;20:96–103.

36. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2 . 0. Mol Plant Pathol. 2018;9(201 8):2094–110.

37. Demirjian C, Vailleau F, Berthomé R, Roux F. Genome-wide association studies in plant pathosystems: success or failure? Trends Plant Sci. 2022;xx(xx):1–15.

38. Koladia VM, Richards JK, Wyatt NA, Faris JD, Brueggeman RS, Friesen TL. Genetic analysis of virulence in the Pyrenophora teres f. teres population BB25 × FGOH04Ptt-21. Fungal Genetics and Biology [Internet]. 2017;107(July):12–9. Available from: http://dx.doi.org/10.1016/j.fgb.2017.07.003

39. Shjerve RA, Faris JD, Brueggeman RS, Yan C, Zhu Y, Koladia V, et al. Evaluation of a Pyrenophora teres f. teres mapping population reveals multiple independent interactions with a region of barley chromosome 6H. Fungal Genetics and Biology [Internet]. 2014;70:104–12. Available from: http://dx.doi.org/10.1016/j.fgb.2014.07.012

40. Wyatt N, Richards J, Brueggeman R, Friesen T. Reference Assembly and Annotation of the Pyrenophora teres f . teres Isolate 0-1. GENES, GENOMES, GENETICS. 2018;8(1):1–8.

41. Bakonyi J, Justesen AF. Genetic relationship of Pyrenophora graminea, P. teres f. maculata and P. teres f. teres assessed by RAPD analysis. Journal of Phytopathology. 2007;155(2):76–83.

42. Wingfield BD, Berger DK, Coetzee MPA, Duong TA, Martin A, Pham NQ, et al. IMA genome F17: Draft genome sequences of an Armillaria species from Zimbabwe, Ceratocystis colombiana, Elsinoë necatrix, Rosellinia necatrix, two genomes of Sclerotinia minor, short read genome assemblies and annotations of four Pyrenophora teres isolate. IMA Fungus. 2022;13(1).

43. Ni M, Feretzaki M, Sun S, Wang X, Heitman J. Sex in Fungi. Annu Rev Genet. 2011;45:405–30.

44. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology. 2012;19(5):455–77.

45. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.

46. Mantel N. The Detection of disease clustering and a generalized regression approach. Cancer Reaserch. 1967;27(2):209–20.

47. Collin FD, Durif G, Raynal L, Lombaert E, Gautier M, Vitalis R, et al. Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. Mol Ecol Resour. 2021;21(8):2598–613.

48. Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability in populations. Evolution (N Y). 1975;29(1):1–10.

49. Slatkin M, Excoffier L. Serial founder effects during range expansion: A spatial analog of genetic drift. Genetics. 2012;191(1):171–81.

50. Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, et al. A genomic history of Aboriginal Australia. Nature. 2016;538(7624):207–14.

51. Komluski J, Habig M, Stukenbrock EH. Repeat-Induced Point Mutation and Gene Conversion Coinciding with Heterochromatin Shape the Genome of a Plant-Pathogenic Fungus. mBio. 2023 Jun 27;14(3).

52. Liu Z, Ellwood S, Oliver R, Friesen TL. Pyrenophora teres: profile of an increasingly damaging barley pathogen. Mol Plant Pathol. 2010;12(1):1–19.

53. Stamatakis A, Alachiotis N, Pavlidis P, Daniel Z. SweeD : Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. Mol Biol Evol. 2013;30(9):2224–34.

54. Alachiotis N, Stamatakis A, Pavlidis P. OmegaPlus : a scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics. 2012;28(17):2274–5.

55. Alachiotis N, Pavlidis P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Commun Biol [Internet]. 2018;1(1). Available from: http://dx.doi.org/10.1038/s42003-018-0085-8

56. Stephan W. Selective Sweeps. 2019;211(January):5–13.

57. Weiland JJ, Steffenson BJ, Cartwright RD, Webster RK. Identification of molecular genetic markers in Pyrenophora teres f. teres associated with low virulence on “Harbin” barley. Phytopathology. 1999;89(2):176–81.

58. Poets AM, Fang Z, Clegg MT, Morrell PL. Barley landraces are characterized by geographically heterogeneous genomic origins. Genome Biol. 2015;16(1):1–11.

59. Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via sympatric speciation. Mol Biol Evol. 2007;24(2):398–411.

60. Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, Van Kim P, et al. 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. 2005;170(2):613–30.

61. Schweizer G, Haider MB, Barroso G V., Rössel N, Münch K, Kahmann R, et al. Population Genomics of the Maize Pathogen Ustilago maydis: Demographic History and Role of Virulence Clusters in Adaptation. Genome Biol Evol. 2021;13(5):1–17.

62. Fuller QD, Weisskopf A. Encyclopedia of Global Archaeology, Barley: Origins and Development. 2014. 763–766 p.

63. Jones H, Civá P, Cockram J, Leigh FJ, Smith LMJ, Jones MK, et al. Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces. BMC Evol Biol. 2011;11(1):1–12.

64. Kirleis W, Klooß S, Kroll H, Müller J. Crop growing and gathering in the northern German Neolithic: A review supplemented by new results. Veg Hist Archaeobot. 2012;21(3):221–42.

65. Hartmann FE, McDonald BA, Croll D. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. Vol. 27, Molecular Ecology. 2018. p. 2725–41.

66. Zhong Z, Chen M, Lin L, Han Y, Bao J, Tang W, et al. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. ISME Journal. 2018;12(8):1867–78.

67. Dahanayaka BA, Vaghefi N, Knight NL, Bakonyi J, Prins R, Seress D, et al. Population Structure of Pyrenophora teres f. teres Barley Pathogens from Different Continents. Phytopathology. 2021;111(11):2118–29.

68. Stukenbrock EH, Banke S, McDonald BA. Global migration patterns in the fungal wheat pathogen Phaeosphaeria nodorum. Mol Ecol. 2006;15(10):2895–904.

69. Li J, Wyatt NA, Skiba RM, Kariyawasam GK, Effertz K, Rehman S, et al. Pathogen genetics identifies avirulence/virulence loci associated with barley 1 chromosome 6H resistance in the Pyrenophora teres f. teres-barley. Available from: https://doi.org/10.1101/2023.02.10.527674

70. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

71. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30(5):614–20.

72. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.

73. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011. Genome Res. 2011;21:936-939(3):936–9.

74. McKenna A, Hanna M, Banks E, Sivachenko AY, Cibulskis K, Kernytsky AM, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

75. Alexander D, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;(19):1655–64.

76. Bryant D, Moulton V. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Mol Biol Evol. 2004;21(2):255–65.

77. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Journeal of Statistical Software. 2017;80(August):1–4.

78. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254–67.

79. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526–8.

80. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014;15(1):1–13.

81. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35(10):1786–8.

82. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.

83. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution (N Y). 1984;38(6):1358–70.

84. Lu S, Platz GJ, Edwards MC, Friesen TL. Mating type locus-specific polymerase chain reaction markers for differentiation of Pyrenophora teres f. teres and P. teres f. maculata, the causal agents of barley net blotch. Phytopathology. 2010;100(12):1298–306.

85. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. BMC Bioinformatics. 2009;10:1–9.

86. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

87. Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, et al. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166–7.

88. Shen W, Le S, Li Y, Hu F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016;11(10):1–10.

89. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.

90. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.

91. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.

92. Hoang DT, Vinh LS, Flouri T, Stamatakis A, Von Haeseler A, Minh BQ. MPBoot: Fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evol Biol. 2018;18(1):1–11.

93. Schrempf D, Minh BQ, De Maio N, von Haeseler A, Kosiol C. Reversible polymorphism-aware phylogenetic models and their application to tree inference. J Theor Biol. 2016;407:362–70.

94. Hudson RR. Generating samples under a Wright–Fisher neutral model ofgenetic variation. Bioinformatics. 2002;18(2):337–8.

95. Kim Y, Neilsen R. Linkage disequilibrium as a signature of selective sweeps. Genetics. 2004;167(3):1513–24.

96. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Rese. 2005;15(11):1566–75.

97. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.

98. Cantalapiedra CP, Herṇandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825–9.

99. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–14.

100. Ge SX, Jung D, Jung D, Yao R. ShinyGO: A graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36(8):2628–9.

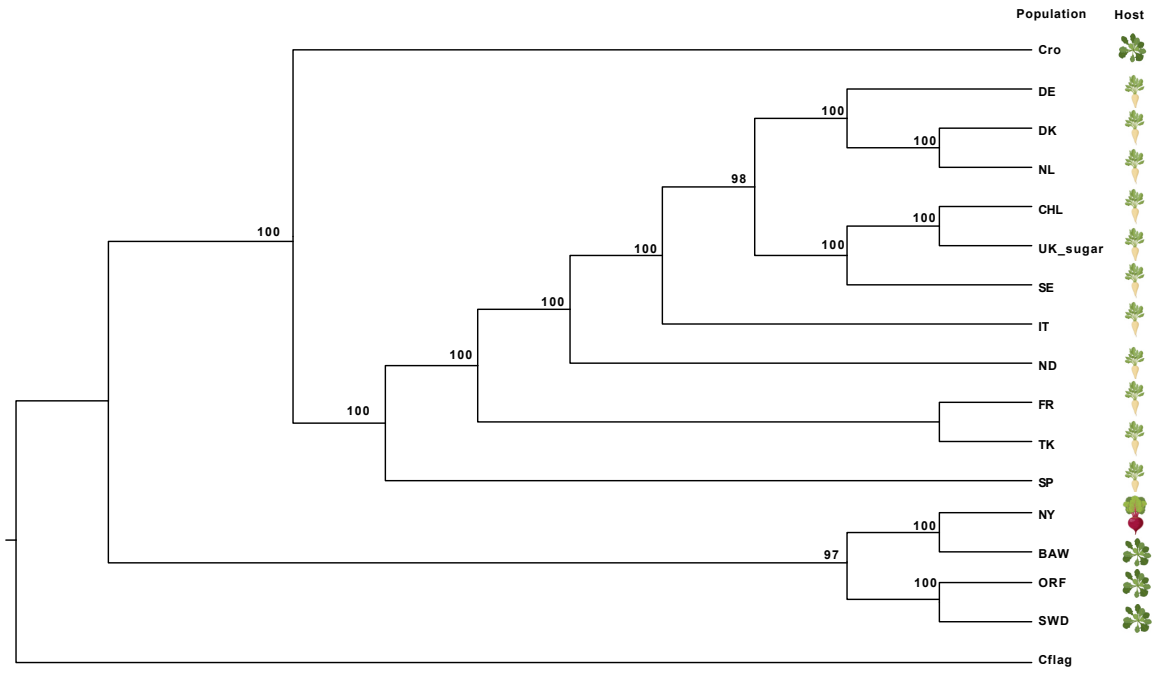
101. Clare SJ, Wyatt NA, Brueggeman RS, Friesen TL. Research advances in the Pyrenophora teres–barley interaction. Mol Plant Pathol. 2020;21(2):272–88.

102. Wyatt NA, Richards JK, Brueggeman RS, Friesen TL. Reference assembly and annotation of the Pyrenophora teres f. teres isolate 0-1. G3: Genes, Genomes, Genetics. 2018;8(1):1–8.

103. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

Chapter III

Genome-wide evidence of host specialization in wild and farmland populations of the cercospora leaf spot pathogen, *Cercospora beticola*



Chapter III : Genome-wide evidence of host specialization in wild and farmland populations of the cercospora leaf spot pathogen, *Cercospora beticola*

Demetris Taliadoros[1,2], Lizel Potgieter[3], Idalia rojas[1], Nathan Wyatt[5], Christian Jung[4], Melvin Bolton[5], and Eva Stukenbrock[1,2]

Affiliations

[1] Environmental Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön and

[2] Christian-Albrechts University of Kiel, 24118, Germany

[3] Swedish University of Agricultural Sciences, Department of Plant Breeding

[4] Plant Breeding Institute, Christian-Albrechts University of Kiel, 24118, Germany

[5] USDA-ARS, Edward T. Schafer Agricultural Research Center, Fargo, ND, United States of America

Abstract

Domestication of plants has been a continuous human activity for over 11,000 years. Modern examples of plant domestication efforts include forest berries like blueberries and the rocket plant. One of the most recently domesticated plants is the sugar beet, only domesticated around 200 years ago in Germany. In many crop-pathogen pathosystems, the adaptation to the agro-ecosystem and the domesticated host drives the evolutionary trajectories of pathogen lineages, resulting in novel epidemics. Here we characterized the population structure of the cercospora leaf spot agent, *Cercospora beticola,* a pathogen that is increasingly important in beet production worldwide, and fungicide resistance is a growing concern. To this end, we used genome sequences of 326 isolates from four continents and four closely related beet subspecies (three domesticated and one wild). As caused by *C. beticola*, leaf spot disease is also a common pathogen of the wild progenitor of sugar beet, *B. vulgaris* ssp. *maritima*, we assessed the extent of host specialization in the species and adaptation to the agro-ecosystem. We found that *C. beticola* populations in the agro-ecosystem were established by wild populations from the Mediterranean Sea and Great Britain on two independent occasions. Our results highlight the ability of *C. beticola* to invade the agro-ecosystem and establish new populations, showing

the rapid adaptation potential of the species. Furthermore, structure analysis revealed a very host-driven population structure and abundance of DMI fungicide resistance associated loci in farmland populations around the world.

Keywords: Fungal pathogen evolution, plant disease, beet, selective sweeps, population divergence

Background

Population genetic have been proven to be a valuable tool in our efforts to understand the evolutionary trajectories of economically significant plant pathogens at a global scale (Feurtey et al., 2023; Pereira et al., 2020; Taliadoros & Stukenbrock, 2023; Thierry et al., 2022). These studies provide valuable insights into the adaptation of pathogen populations to local conditions and hosts, which, in turn, offer crucial information with implications for disease epidemiology. For instance, hypothesis about the origin of the species, patterns of dispersal, host specialized lineages, and more recently development of resistance to specific fungicide classes classes are made by studying genetically distinct pathogen populations around the world (Ali et al., 2014; Gladieux, Bradford, et al., 2018; R. Spanner et al., 2021). This knowledge can, guide the

development of strategies for disease management and measures for regulating the movement of plant materials and the use of certain fungicides to minimize disease outbreaks.

The phenomenon of host-specificity among plant-pathogenic fungi has long been a major subject of interest for plant pathologists (Borah et al., 2018; J. Li et al., 2020). Host-specificity refers to the ability of certain fungal species, or certain members of a fungal species, to induce disease in particular plant species or specific individuals within a plant species. Various hypotheses have been developed to elucidate the foundation of host-specificity, such as the gene-for-gene hypothesis formulated by Henry H. Flor (1971) for flax and flax rust interactions. The hypothesis states that incompatible interactions arise when resistance (R) proteins in certain plants recognize specific avirulence (AVR) proteins of the fungus, leading to successful plant defense against the pathogen. More recently, molecular studies have provided mechanistic insights into the interaction of plant and pathogen proteins (P. N. Dodds et al., 2006; P. Dodds & Thrall, 2009). All interactions between plants carrying a particular R gene and pathogens carrying the corresponding AVR gene are deemed incompatible. Another key advancement in this understanding is the identification of effectors, which are small proteins are secreted by pathogens to manipulate the host's

physiology, particularly by suppressing plant immune responses. The recognition that effectors can possess both virulence and avirulence (AVR) functions elegantly links these two concepts and accounts for the numerous species-specific effector genes observed on the pathogen side, as well as the multitude of receptor genes serving as R proteins on the plant side, in the context of an ongoing evolutionary arms race (Möller & Stukenbrock, 2017). Host specific lineages have been reported in many major fungal plant pathogens such as the rice blast fungus (*Pyricularia oryzae*, syn: *Magnaporthe oryzae*), for which cereal- and grass- specialized lineages were reported diverging depside the geneflow between them (Gladieux, Condon, et al., 2018). Other examples include cultivar-driven divergence of the wheat pathogen *Zymoseptoria tritici* (Hartmann et al., 2018). Effectors have some particular molecular properties that allow for prediction of effector genes from genome sequences. First, effector genes typically contain a signal peptide for secretion into the plant apoplast or translocation into the host cytoplasm. Moreover, these genes tend to be cysteine-rich.. Recent genomic studies showed that effector genes are often subjected to strong positive selection in plant pathogens (Sperschneider & Dodds, 2022).

Cercospora beticola is the causal agent behind Cercospora leaf spot (CLS) in *Beta vulgaris* L., which includes sugar beet, table beet, and Swiss chard, and their wild ancestor, the sea beet.

The pathogen undergoes multiple rounds of asexual reproduction within a single growing season (Rangel et al., 2020). During the periods between sugar beet growing seasons, the fungus primarily survives in the form of hyphal structures called pseudostromata that are resistant to desiccation. These specialized structures, consist of a combination of fungal tissue and remnants of host tissue, are situated within the leaf substomatal cavities and they can remain viable for up to two years. While a known sexual form is absent, *C. beticola* is classified as a heterothallic ascomycete fungus. This classification arises from the identification of two oppositive mating-type genes (MAT1-1-1 and MAT1-2-1). Previous studies found the two mating types to be present at equal frequencies in certain regions of Europe and North America providing evidence for the occurrence of sexual reproduction. Despite the wide host range of the fungus including wild and domesticated plants, information about host specialized fungal lineages is scarce (Rangel et al., 2020).

Domesticated beets (*Beta vulgaris* subsp. *vulgaris*) and their wild ancestor, the sea beet (*Beta vulgaris* subsp. *maritima*) have a long history, associated with the human history. Archaeological and historic evidence suggests that the wild

ancestor of beets, the sea beet originates from the Mediterranean basin (Biancardi et al., 2020). The first reports of beet cultivations come from the ancient Greeks and Romans, which were cultivating beets for it leaves. Beetroot, the part we commonly associate with beets today, was only developed during the late stages of the roman empire (Biancardi et al., 2020). By the 16th century, beet cultivation had spread to Northern Europe and fodder beets were widely used for animal feed. In the 18th century, a German chemist named Andreas Marggraf discovered that sugar could be extracted from beets, laying the foundation for sugar beet production (Rangel et al., 2020). This marked a significant turning point, as beets became a vital source of sugar during the Napoleonic wars were sugar cane became scarce. Eventually, cultivated beets have further dispersed over the last few centuries by human migration to North America. Altogether, there was a rapid diversification and a world-wide dispersal of beets associated with the human history over the past two millennia.

Furthermore, the global spread of plant species is often accompanied with the undesirable spread of associated pathogens (Sotiropoulos et al., 2022; Stukenbrock et al., 2006). Pathogen dispersal can lead to new adaptations allowing pathogens to establish in new local environments, eventually with specific management strategies, e.g. the use of local hosts or locally used fungicides (Croll & McDonald, 2017). Evidence of such adaptation can in some

cases be evident as "selective sweeps" in the genome where particular alleles at certain loci have been subject to strong positive selection. Typical signatures of selective sweeps are genomic regions with reduced genetic variation and increased linkage disequilibrium (Stephan, 2016).

In the present study, we conducted an analysis of the haploid genomes of 326 *C. beticola* isolates obtained from diverse beet fields and natural sea beet habitat across the globe. Our objective was to infer the global population structure of *C. beticola* and identify the factors shaping the structure. In particular, to assess the extent of host specialization and local adaptation within the species. This is the first study with such an extensive dataset originate from both crops and wild plants. We to utilized single nucleotide polymorphisms (SNPs), to explored the population structure and host specialization of the pathogen populations. Our findings strongly support a host driven divergence of *C. beticola* populations. Furthermore, we scanned the genomic landscape to detect signs of recent natural selection and observed a significant overlap between these signatures and potential virulence factors. Finally, we found previously reported loci associated with fungicide (Spanner et al., 2020) resistance to be under strong selection in UK farmlands. Our study reveals a host-driven divergence of an important plant pathogen highlighting the urge for disease control strategies to

avoid the emergence of highly pathogenic fungicide resistant lineages in the near future.

Material and Methods

Genome data

326 *C. beticola* whole genomes were sequenced using Illumina technology. The sequenced isolates were sampled from barley fields on four different continents (Table 1 and S1). Eighty isolates were obtained from two North Dakota State University (NDSU) experimental sugar beet fields in Fargo and Foxhome, North Dakota, USA, North America. Twenty-four isolates were collected from two table beet fields in New York, USA, North America. Seven isolates were collected from a fodder beet field in Chile, South America. Twenty-two isolates from a sugar beet field, in the UK, North Europe. Furthermore, 12 isolates were collected from a sugar beet field in Sweden. Isolates were also collected from single sugar beet fields in six continental European countries. Specifically, 12 isolates were collected from a field in Denmark, 18 from Germany, 11 from France, and 19 from Spain. Additionally, 21 from five fields in Italy. Moreover, we included 18 isolates from Turkey, Asia. To capture the natural variation of the species in continental and Northern Europe, we included 23 isolates from two natural habitats of sea beets in Croatia and forty-three and three natural sea beet locations (Southwold, Bawdsey, and Ordorf) in UK (Table 1).

Read mapping and variant calling

We filtered and mapped Illumina reads to a reference genome and extract high-quality single nucleotide polymorphisms (SNPs). In brief, we used the program Trimomatic version 0.38 (Bolger et al., 2014) to filter and trim sequencing adapters, and nucleotide bases based on sequencing quality (PHRED 33) and read length (minimum read length 30 bp). To avoid overrepresentation artifacts, we merge overlapping reads using PEAR version 0.9.11 (Zhang et al., 2014). Subsequent, we mapped individual raw reads on the 09-40 *C. beticola* genome (De Jonge et al., 2018), using Burrows-Wheeler Aligner (BWA) version 0.7.17 (H. Li & Durbin, 2010) and Stampy v. 1.0.20 (Lunter & Goodson, 2011) (Table S2). Haplotyping and genotyping were performed with the GATK HaplotypeCaller version 4.2.18 (McKenna et al., 2010), providing a final VCF file with the raw SNP calls.

To discard low-quality SNPs from our dataset, we performed "hard-filtering" as the GATK “best practice guidelines” suggests. To this end, the filtering was performed with the following criteria: (1) The call quality divided by the depth of sample reads must be larger than 2, (2) the depth per genome must be higher than 8, (3) mapping quality of reads supporting each SNP must be higher than 40, (4) allele-specific rank sum test for mapping qualities of the reference (REF)

versus alternative (ALT) reads must be higher than -12.5, (5) allele-specific rank sum test for relative positioning of REF versus ALT allele within reading must be higher than -8. (6) Each genome has to have an average read coverage of at least 2. For the application of these filters, GATK VariantFiltration version 4.0.11 was used (McKenna et al., 2010). After applying these hard-filtering criteria, we obtained a dataset of 1,129,645 SNPs (from which 893,266 biallelic). We further refer to this dataset as the "full high-quality dataset."

Additional filtering steps were added according to assumptions for the different analyses. Most clustering analyses assume that the markers used are independent. Therefore, for these analyses, we filtered the full high-quality dataset based on the linkage disequilibrium (LD) decay patterns, considering a distance of at least 2 Kbp (distance of $r^2/2$ averaged across populations) between SNPs (Table 1). After filtering for LD, a dataset of 44,229 SNPs was retained. We refer to this dataset as the "independent SNP dataset."

Population genetic structure

We infer population genetic structure using two different approaches: a detrended correspondence analysis (DCA) (Hill & Gauch, 1980) and an ADMIXTURE version 1.3 (D. Alexander et al., 2009). These analyses were based on the "independent SNP dataset." We applied the DCA to reveal genetic

clustering among the isolates and created it using the R package vegan v. 2.6.2. We visualized the DCA using the R package ggplot2 (Wickham, 2017). We further characterized population structure using a maximum likelihood approach implemented in ADMIXTURE. We ran ADMIXTURE for a range of K-values (1-10). The most fit K value was assessed based on the cross-validation error (Alexander et al., 2009) (Table S5). Finally, to address the effect of geographic distance and individual host to the population structure, we employed an analysis of molecular variance (AMOVA) (Excoffier et al., 1992) performed on the same dataset. To this end, the isolates were grouped geographically according to the country and the continent of their origin, and according to the host they were isolated from. The R package poppr (Kamvar et al., 2015) was used to perform the AMOVA analysis. Since our dataset only encompass isolates from different hosts only from UK, we could specifically assess the significance of geographic isolation and host specialisation in two independent runs. To assess the significance of geographic isolation, we used all the available sequences, while for the assessment of the significance of the host as reproductive isolation factor we focused on the UK populations.

Genetic diversity, neutrality tests, and linkage disequilibrium

We used the "full high-quality dataset" to compute and compare genetic variation among populations. The software pixy was used to estimate the genetic diversity for each population as the nucleotide diversity (π) and the population divergence statistics Fst and Dxy in non-overlapping windows of 10 Kbp . The tool PopLDdecay (C. Zhang et al., 2019) was used with default parameters to estimate the linkage disequilibrium (LD) decay for each genetic cluster. VCFtools v. 0.1.17 (Danecek et al., 2011) was used to calculate the Tajima's D for each population. A Kruskal-Wallis test with post-hoc pairwise Wilcoxon was used to identify significant differences ($p < 0.05$) in the level of genetic diversity and Tajima's D values.

Mating types

MAT1-1 and MAT1-2 mating-type sequences were obtained from GenBank (accession number DQ264736.1 and JN863091.1, respectively) (Bolton et al., 2012; Groenewald et al., 2006). Complete de novo assemblies of genomes were created using SPAdes (Bankevich et al., 2012) with default parameters for each isolate. Subsequently, the mating type of each assembly was assessed by blasting the mating type sequences against them. To that end, blastn with default parameters was used (Camacho et al., 2009).

Evolutionary relationships using polymorphism-aware models

We employed IQ-TREE polymorphism-aware models (PoMo) (Schrempf et al., 2016) to reconstruct relationships among *C. beticola* populations. To convert our fasta to the count file format, we utilized the FastaVCFtoCount.py script provided by the PoMo software. We selected the best-fitting substitution model based on the Bayesian Information Criterion (BIC) using the ModelFinder algorithm within Iqtree version 2.0.3. to derive branch support values, we perform 1000 bootstrap replicates, employing Iqtree's bootstrap approximation option. This analysis used the sister species C. flagellaris as an outgroup.

Genome scans for selective sweeps and population divergence

The software pixy was used to estimate the genetic diversity for each population as the nucleotide diversity (π) and the population divergence statistics Fst and Dxy in non-overlapping windows of 10 Kbp.

We further used computational tools to identify candidate effector genes. Genes encoding putative secreted proteins were identified as proteins carrying a signal peptide using the SignalP-5 server (Juan et al., 2019). SignalP-5 predicts the presence of a signal peptide and the location of its cleavage sites. TMHMM v. 2.0 was used for prediction of transmembrane helices in the selected proteinsc (Krogh et al., 2001). Proteins that were predicted to have transmembrane helices

were excluded and the remaining genes were next used for further prediction analysis with the use of EfectorP v.3 software allowing us to generate a dataset of putative apoplasic and cytoplastic effectors (Sperschneider & Dodds, 2022). Furthermore, to test if the abundance of effector genes differed for the highly diverging regionspredicted selective sweep regions compared to the rest of the genome, we used a permutation test (based on a custom script available on GitHub: https://github.com/Jimi92/. In brief, the abundance of predicted effectors was counted in regions of the same size and number equal to the expected sweep regions. As many as 10,000 replicate runs of random resampled regions were performed. Genome-wide maps of the sweep regions were created using Circos v. 0.69-9 (Krzywinski et al., 2009).

Results

Generation of *C. beticola* population genomic datasets

To study the geographical structure of *C. beticola* populations and to infer the recent history of this emerging sugar beet pathogen, we generated a population genomic dataset comprising sequence data of 326 isolates from cultivated sugar beet in six countries across four continents (Africa, America, Central Asia, and Europe) (Table S1). The genomes were sequenced with Illumina technology, and sequencing reads were mapped to the reference genome of *C. beticola* (De Jonge

et al., 2018) to identify SNPs. The average read coverage across the 326 genomes was 21X. We identified a total of 1,068,644 high-quality SNPs among the 326 isolates. Further summary statistics related to the read mapping and variant calling are summarized in Table S1.

Table 1: Multi-continental sampling and genetic diversity, Tajima's D, mating types ratio and linkage disequillibrium decay.

Sampling location			-	-	π		Tajima's D		-	-	-
Continent	Country	Population	Host	No of isolate	Mean	St.dev	Mean	St.dev	Mat1:Mat2 r	Private varia	Dist r2/2 (Kb
Asia											
	Turkey	TK	Sugar	18	0,0015	0,0035	0,1	1,22	2	4555	2,2
Europe											
	Sweden	SW	Sugar beet	12	0,0014	0,0031	0,06	1,23	0,5	2280	3,1
	Denmark	DK	Sugar beet	13	0,001	0,0023	-0,64	0,98	2,25	561	2,3
	Netherlands	NL	Sugar beet	15	0,0012	0,0032	0,05	-0,17	0,5	595	2,3
	Germany	GE	Sugar beet	14	0,0014	0,0033	0,04	1,23	0,75	1320	4,3
	France	FR	Sugar beet	11	0,0012	0,0039	0,05	1,13	0,83	725	2,9
	Spain	SP	Sugar beet	20	0,0021	0,0032	-0,38	1,25	0,83	56980	3,1
	Italy	IT	Sugar beet	21	0,0012	0,003	0,03	1,22	1,75	657	3,9
	Croatia	CR	Sea beet	24	0,002	0,0034	-0,06	1,25	1,4	28657	0,7
	UK	BB	Sugar beet	24	0,001	0,0032	0,52	1,25	1	417	3,2
		OR	Sea beet	14	0,0006	0,0028	-0,36	0,77	0,08	3376	-
		BD	Sea beet	10	0,0016	0,0037	0,47	1,13	0,43	1655	2,8
		SW	Sea beet	18	0,0018	0,0033	0,51	1,12	1,57	15607	2,3
North America											
	USA	ND	Sugar beet	80	0,0013	0,0027	0,12	1,5	0,8	11886	3,1
		NY	Table beet	24	0,0013	0,0032	0,24	1,25	0,5	7164	3
South America											
	Chile	CH	Fodder beet	8	0,0012	0,0035	0,12	1,17	0,6	157	3,9
Total	-	-	-	326	0,0042	0,0046	-0,34	1,49	0,99	-	

Rich genetic diversity in wild and Mediterranean populations

The level of standing genetic variation present in populations can give insight into the recent and ancetral demographic history. To explore genetic variation in *C. beticola*, we first compared the nucleotide diversity among the 16 geographical populations using the Kruskal-Wallis test, which revealed significant differences in mean values of the parameter π (p-value < 2.2e-16). Subsequently, we performed a pairwise Wilcoxon test to assess significant differences between the populations (Figure 1A, Table S3). Our analysis showed

that two Mediterranean populations (Spain and Croatia) harbor significantly higher genetic diversity (π_{Spain} = 0.0021, $\pi_{Croatia}$ = 0.0020) than every other population. We consider that the higher diversity in western Mediterranean populations could reflects a more ancestral origin of the pathogen in these regions.

Furthermore, populations originating from the wild sea beet plants collected at the sea shore at Southwold and Bawdsey showed significantly higher genetic diversity (mean π_{Swd}= 0.0018, π_{Bawd} = 0.0014) than the rest of the populations, except from the Croatian and Spanish populations.

In contrast, the UK farmland and the Danish population showed significantly lower diversity than the rest (mean π_{UK}= 0.0010, π_{DK} = 0.0010). Intriguingly, a UK population isolated from sea beets in Orford showed the lowest genetic diversity than the rest of the populations (mean π: 0.0006) suggesting a significant variation in diversity among geographically close populations on the wild sea beet host. In addition, the Orford population showed a significantly negative Tajima's D value (Figure 1B). A negative Tajimas D value indicate an excess of low-frequency alleles and may reflect a recent bottleneck. We therefore speculate that the Orford population has a different history compared to the other sea beet populations, and that these to a large extent are not homogenized by gene flow.

We also compared Tajima's D values for the other geographical *C. beticola* populations using the Kruskal-Wallis test, which revealed significant differences among pairs of populations (p-value < 2.2e-16). Subsequently, we performed a pairwise Wilcoxon test to assess significant differences between the populations (Figure 1B, Table S4). In general, sea beet infecting populations show higher levels of Tajima's D, which, in line with the genetic diversity results, suggests a balancing selection occurring in these populations. Most farmland populations showed a Tajima's D value close to 0. In contrast, Spain, Denmark, and Orford showed a significant difference from the rest of the populations and a negative D value. We speculate that the excess of low frequency alleles reflects a recent bottleneck also in Denmark as in Orford as also the level of nucleotide variation is low. However, the high nucleotide variation in Spain and the negative Tajimas D value may also reflect a strong population expansion.

We next addressed the mode of reproduction in the pathogen population, here considering that *C. beticola* can reproduce either clonally or through sexual mating. *C. beticola* has a heterothallic mating system implying that mating only occurs between individuals of opposite mating types, Mat1-1 and Mat1-2. Our mating type analysis provides evidence for a primary role of sexual reproduction in *C. beticola* populations word-wide. We used a Chi-squared test to assess evidence of departure from the 1:1, which resulted in a non-significant result for

all populations except Ordorf. In the case of Ordorf, analysis of matin-type ratio revealed an absence of Mat1-1 from the Ordorf population, suggesting a primarily clonally reproducing population.

We considered the geographic variation in the frequency of sexual reproduction, by examining the mating type ratio, which is expected to be close to one in a population with random sexual mating. To correctly assign a mating type to each isolate, we used the software SPAdes (44) to *de novo* assemble genomes and thereby validate and compare the frequency of mating type loci. The null hypothesis of random mating could not be rejected for the *C. beticola* populations except the Ordorf population, for which we found a significant departure from the expected 1:1 ratio of mating types (Mat1-1-1:Mat1-2-1 = 3.5, Chi-squared test, p = 0.001) (Figure 1C, Table S4). Depside the teleomorph stage of C. *beticola* was never observed, these analyses suggest that *C. beticola* is regularly undergoing sexual reproduction throughout most of its distribution range and on both domesticated and wild hosts.

We also addressed recombination at the genomic level. Hereby we computed the extent of linkage disequilibrium decay for each population. The varying extent of linkage disequilibrium (LD) among fungal populations can be informative about the frequency of sexual reproduction and about the age of populations. Recently found populations, as well as, populations with lower

frequencies of sexual reproduction, will typically exhibit a greater extent of LD compared to sexually recombining and older populations. In line with the analysis of mating types, we found considerably longer linkage blocks in the Orford and Danish populations. The LD statistic r2 was reduced to half of its maximum value at 20 Kbp for the Danish population, while it never dropped to half for the Orford population (Figure 1D, Table 1). Taking into account that the Danish population showed balanced mating type rations, we conclude that the long LD blocks observed suggest that the population was founded more recently. While for Ordorf, we attribute the long LD blocks to a primal role clonal reproduction.

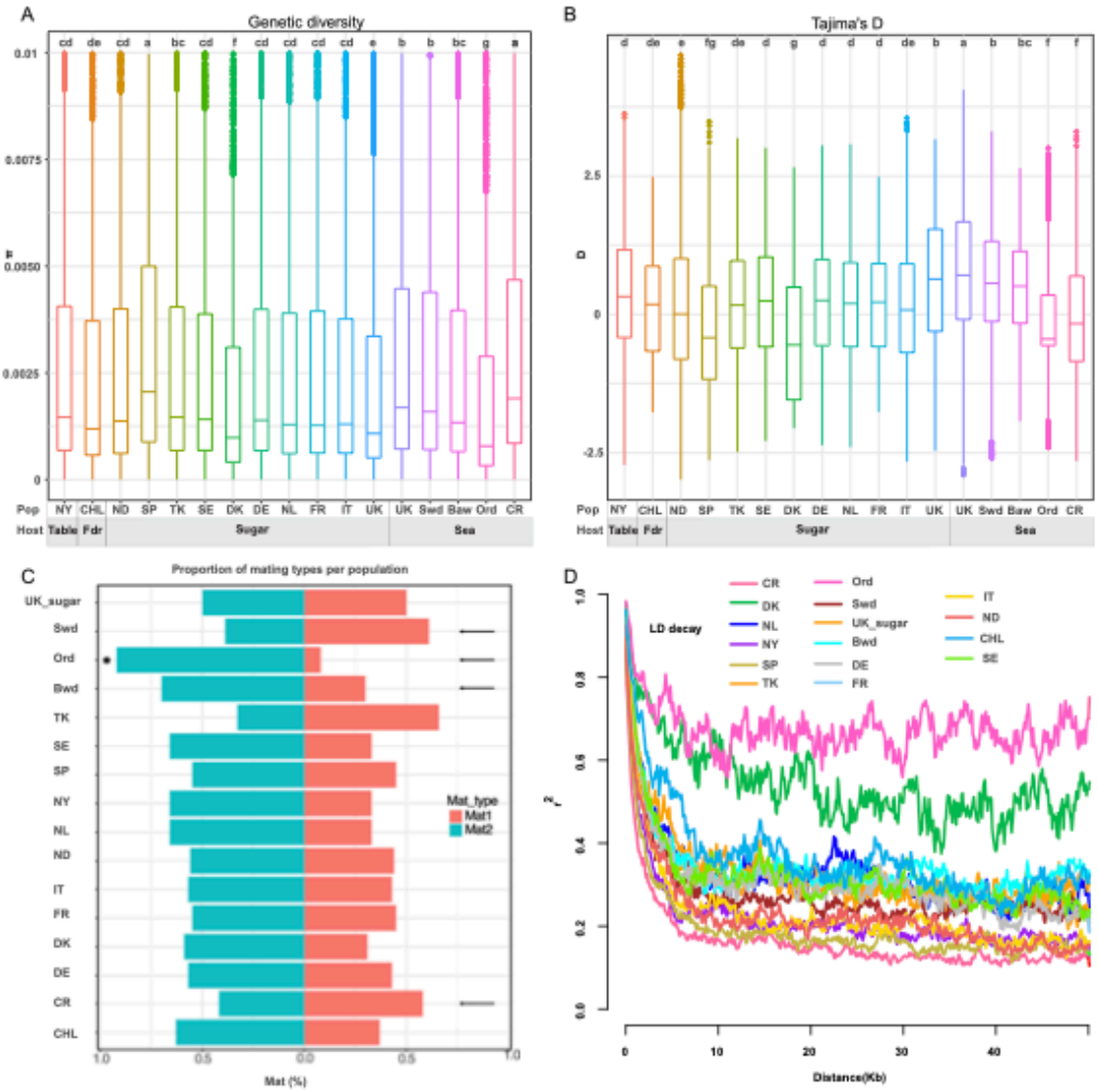


Figure 1: Collection of *Pyrenophora teres* isolates across continents for the inference of pathogen population structure and dispersal. A) Nucleotide diversity of C. beticola populations in each geographic region. Kruskal-Wallis test with post-hoc pairwise Wilcoxon was used to identify significant differences ($p < 0.05$) between the groups (Table 1) Tajima's D of C. beticola populations in each geographic region. Kruskal-Wallis test with post-hoc pairwise Wilcoxon was used to identify significant differences ($p < 0.05$) between the groups (Table 1). C) Percentage of the two mating types occurring in each location. Asterisk indicates significant

departure from the 1:1 ratio (chi-squared test, p-value: 0.05). D) Linkage disequilibrium decay for each population.

The host-driven population structure of *C. beticola*

We characterized the population genetic structure of *C. beticola* based on complementary methods using the "independent SNPs" dataset (see methods). Firstly, we investigated the extent of clustering using a detrended correspondence analysis (DCA) (Hill & Gauch, 1980) (Figure 2A). The DCA mostly separated isolates according to their host, with isolates coming from sea beet and table beet clustering together. We further explored population structure by inferring the extent of shared ancestry using an ADMIXTURE analysis (D. H. Alexander et al., 2009). In the ADMIXTURE analysis, the Cross-Validation error was used to select the most appropriate number of clusters (K). Models with a K value larger than six induce only a small decrease in the Cross-Validation error (Figure 2B, Table S5). At K = 2, the genetic clusters inferred from the ADMIXTURE analysis mainly reflect a separation based on the host, where sea beet and table beet-infecting isolated from UK and NY interestingly are encompassed in one cluster, while sugar and fodder beet-infecting isolated constitute the other cluster. K = 3 further separates the isolates within the sugar-fodder beet cluster, revealing two globally occurred sugar beet infecting clusters. Finally, at K = 6 we observed further sub-clustering within the UK sea beet-infecting isolates, mainly

driven by the geographic origin of the isolates (Bawdsey, Southwold, and Ordorf). In addition, at K = 6 the ND sugar beet-infecting population is divided to two genetic clusters. In line with the analysis of genetic diversity, the Croatian population encompassed the majority of ancestral clusters possibly reflecting a historic admixture of different populations in that region.

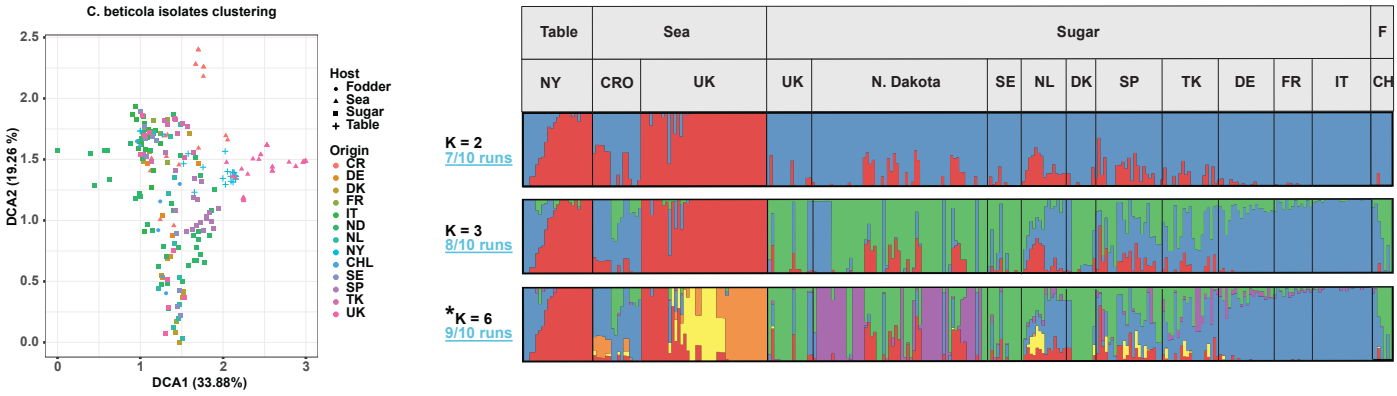


Figure 2: Global population structure of P. teres f. teres. A) PCA analysis, where shape reflects the host of the isolate and the color reflects the origin of the pathogen. B) The program ADMIXTURE was used to compute population structure of the species. Most fit number of hypothetical ancestral groups was identified as six based on the cross-validation method (Table S5). Here we present patterns of two, three, and six hypothetical ancestral groups.

Analysis of Molecular Variance shows that the host type is a significant separating factor of *C. beticola* populations

Host specialization and geographic isolation are two factors that often play a key role in reproductive isolation in filamentous plant pathogens (Möller &

Stukenbrock, 2017). To illuminate the role of geographic separation and host specialisation as the driving factors of *C. beticola* population divergence, we performed two AMOVA analyses (Figure 3, Table 2). To assess the fraction of variation associated with the geographic distance, we use all the available isolates. In this analysis, the major component of genetic variation is within individual populations (77.26 %). An additional 18,98 % of the variation is associated with the genetic variation within the countries. Only 8,89 % of the variation was associated with differences between the countries. Furthermore, we found a negative result for the comparison between continents, which indicates an absence of genetic structure at this level. Overall, the analysis suggested that the genetic variance explained by geography at the country and continent level was not significant (p-value: 0.17 and 0.82, respectfully).

We then tested the effect of host on the population structure of *C. beticola*. For this analysis we focused on the UK, because it was the only country from where we had populations obtained from different hosts. Although, the main contribution of the variance was again due to the variation within samples, the effect of the host plant species was significant (p-value: 0.03). As much as 21.42 % of the genetic variance was found associated with the host. Altogether, the AMOVA suggests the effect of host type on the partition of genetic variance is significant.

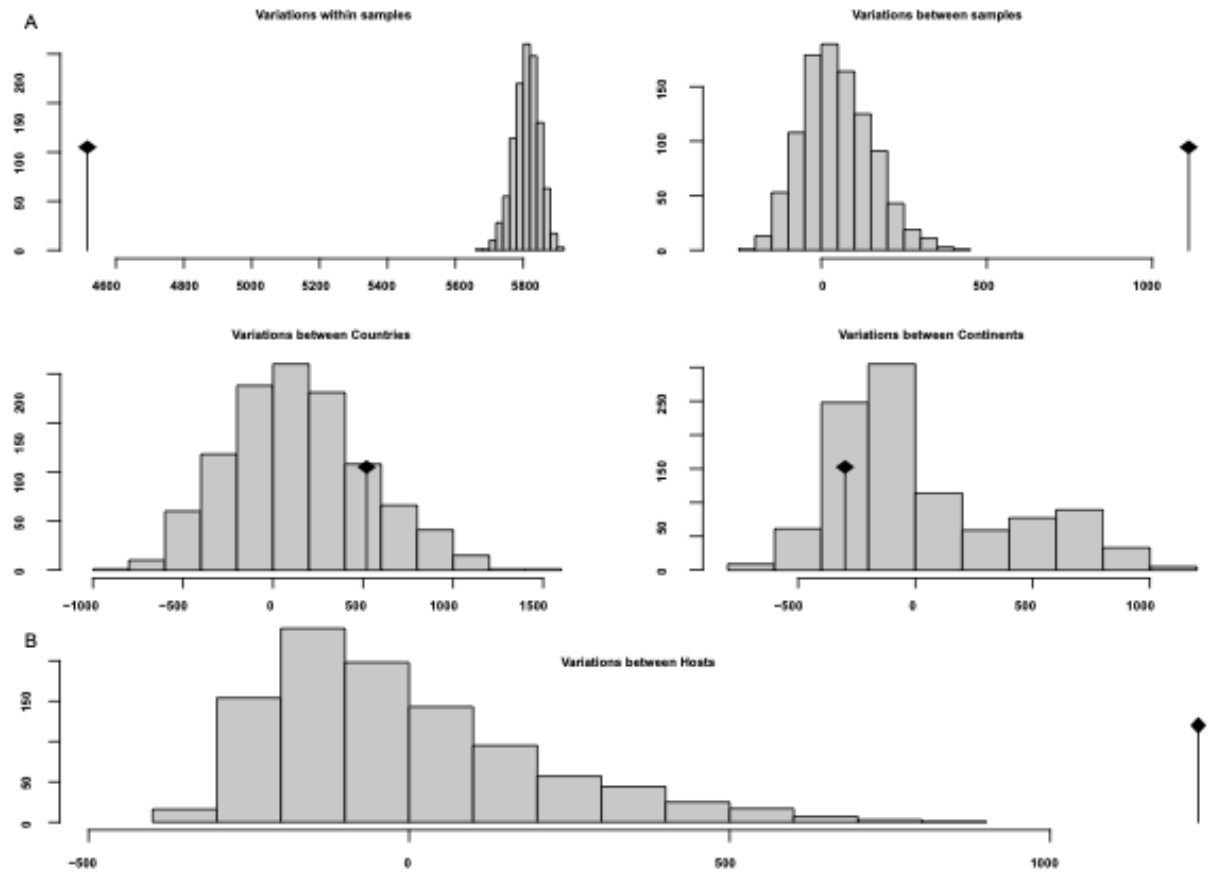


Figure 3: Analysis of Molecular Variance. null distribution of the molecular variance components obtained from random permutations of the matrix of squared interindividual distances (D). A) Assessment of the significance of country and continent as factors for genetic structure. B) Assessment of the significance of Host as factors for genetic structure.

Table 2: AMOVA results for groups according to teir geographic origin and host.

Factor	Source of variation	DF	SS	% var	p-value
Geography					
	Among continents	3	72037,32	-5,13	0,82
	Among countries within continents	10	255615,11	8,89	0,17
	Among samples within countries	15	170422,76	18,98	0,001
	Within samples	260	1174140,96	77,26	0,001
	Total	288	1672216,17	100	
Host					
	Among hosts	1	79184,24	21,42	**0,03**
	Among samples within host	4	73058,52	15,83	0,001
	Within samples	49	318947,12	62,75	0,001
	Total	54	471189,89	100	

Abbreviations: df, degrees of freedom; p-value, obtained fro, 1.000 permutations; SS, sum of squares; % var, percentage of variation explained

Phylogenomic analysis suggests an ancient split of the Mediterranean and UK *C. beticola* population and multiple re-introductions to the agro-ecosystem

The Mediterranean Sea has a long history of beet cultivation (Biancardi et al., 2020). First reports of beet cultivation start at the Greco-Roman time consumed for its leaves. The beet was introduced to Europe during the medieval era and then to America. To test the hypothesis that early dispersal of *C. beticola* occurred simultaneously with the spread of beet cultivation we inferred the evolutionary relationships between the populations of the pathogen. To this end, we have constructed a population tree, using polymorphism-aware models in IQ-TREE, using the sister species *C. flagellaris* as root for the *C. beticola* populations (Figure 4). In this analysis, based on the full complement of polymorphisms, we found two major population splits: one lineage comprising the Mediteranean and sugar and fodder beet-infecting populations around the world, and one lineage comprising the UK sea beet-infecting isolates and the NY Table beet-infecting isolates. Within the former lineage, the branching harbouring the Croatian sea beet-infecting population diverged earlier than the branches harbouring the farmland populations suggesting a common ancestral population that gave rise to of all the farmland populations which emerged after the split of the UK populations. The clustering of the UK sea beet-infecting and

NY red beet populations suggested a UK origin of the NY population. The different origins of the sugar beet-infecting and red beet infecting populations lead to the emergence of an intriguing pattern suggesting that *C. beticola* was introduced to the agro-ecosystem at multiple occasions.

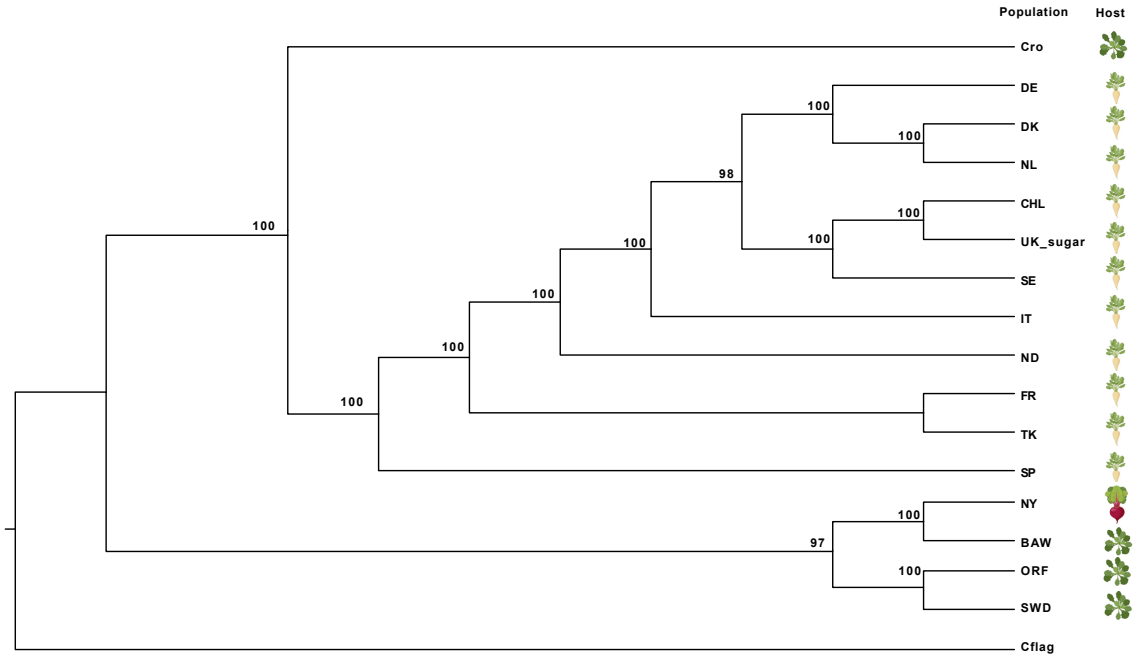


Figure 4: Evolutionary relationship between P. teres f. teres populations. A) Phylogenetic tree using polymorphism-aware models (PoMo) (Maximum likelihood inference, loglikelihood: -857128.545) to assess the evolutionary relationship between C. beticola populations. Branch numbers reflect maximum likelihood bootstrap values. The tree was rooted using the sister species C. flagellaris. Parts of the figure were adapted from “sugar beet”, “table beet”, and “arabidopsis”, by BioRender.com (2020). Retrieved from https://app.biorender.com/biorender-templates

Ongoing host-driven divergence of *C. beticola* in UK

Depside that the UK isolates obtained from farmlands and wild plants were collected from a distance of roughly 150 km apart, our analysis ADMIXTURE suggest that they have different ancestries. Furthermore, there were no signs of admixture between these populations. Indeed, a DCA on the UK data showed a clear separation between the sugar and the sea beet-infecting populations (Figure 5A). AMOVA indicated that some genetic variation that is associated with the host that the isolates originated from. Therefore, we hypothesized that host specialization acts as a factor of sexual isolation. To pinpoint the genomic components associated with host specialization in UK we performed genome-wide scans for highly diverging regions between the three sea and one sugar beet infecting populations all sampled in UK. We computed the population divergence statistic Fst and Dxy for non-overlapping windows of 5 Kbp along the genome. As significantly diverging, we considered the regions showed a statistic value higher than the 99.95% of the rest. To further control for the effects of varying genetic diversity that can influence the Fst results, we focused on the regions that have been identified as significant by both statistics. A map of highly diverging regions between each of the three sea beet and the sugar beet population is shown on figure 5B.

A permutation test was employed to determine whether effector genes showed a significant overrepresentation in regions that were highly diverging. Due to

their determining role in host-pathogen interactions, genes encoding effector proteins often experience strong selective pressure (Möller & Stukenbrock, 2017). To this end, we evaluate the proportion of predicted effector genes within highly diverging regions in comparison to the remainder of the genome. Our results clearly indicated that effector genes are more abundant in the highly diverging regions when compared to randomly selected genomic regions (Figure 5C).

Variation associated with DMI fungicide resistance is under selection in UK and around the world

Another driving factor of populations divergence in plant pathogens is the adaptation to the agricultural environment. In farmlands there is often an extensive use of pesticides, such as fungicides applied to control fungal disease (Stukenbrock & Mcdonald, 2008). Demethilation inhibitor compounds (DMIs) are prominently used for the control of *Cercospora* populations in sugar beet cultivations world-wide. In a previous study, four SNP alleles were found associated with DMI resistance was reported (R. Spanner et al., 2021). To examine if there is any difference of the abundance of DMI resistance alleles we recorded their frequency in the four UK populations (Figure 5D). These mutations associated with fungicide resistance were located at the position

2402041 on chromosome 1, position 849506 on chromosome 4, and positions 1358331 and 1451478 on chromosome 9. All the mutations are located in hypothetical proteins, except the later that is a synonymous mutation within the coding region of the gene encoding for eburicol 14-alpha demethylase CB0940_11379, otherwise known as the gene encoding DMI fungicide target CbCYP51(Bolton, Birla, et al., 2012). Three out of four mutations were not present in the sea beet-infecting populations, while the mutation on chromosome 4 was found in all populations at varying frequencies (15% - 33%). In contrast, the mutations on chromosome 4, and 9 were found in the UK sugar beet infecting populations in high frequencies (75%, 53%, and 53% for chromosome 4, chromosome 9: 1358331, and 1451478, respectively). None of the UK populations carried the mutation on chromosome 1. Both mutations on chromosome 9 are in close vicinity of genomic regions that are significantly diverging between that UK farmland and wild populations (Figure 5E). Altogether, this analysis shows a strong selection for DMI fungicide resistance in UK farmlands, while these mutations are rare or absent on populations on wild beet plants. We further recorded the frequencies of DMI fungicide resistance in *C. beticola* populations around the world obtaining alarming results. Fungicide resistant variation was found present in all populations around the world in various frequencies (Figures S2-S12). A follow-up question arises regarding the

origin and dispersal of the mutations linked to DMI resistance. There are two possibilities to consider: 1) The mutations may have already been present in the wild as part of natural genetic diversity and then became more prevalent in farmland populations due to selective pressures. 2) Alternatively, these could be novel mutations that evolved within farmland populations and were later introduced to wild populations. Our analyses, while unable to definitively discern between these two scenarios, highlight the complexity of the situation. Further research is needed to delve deeper into the emergence and spread of DMI resistance.

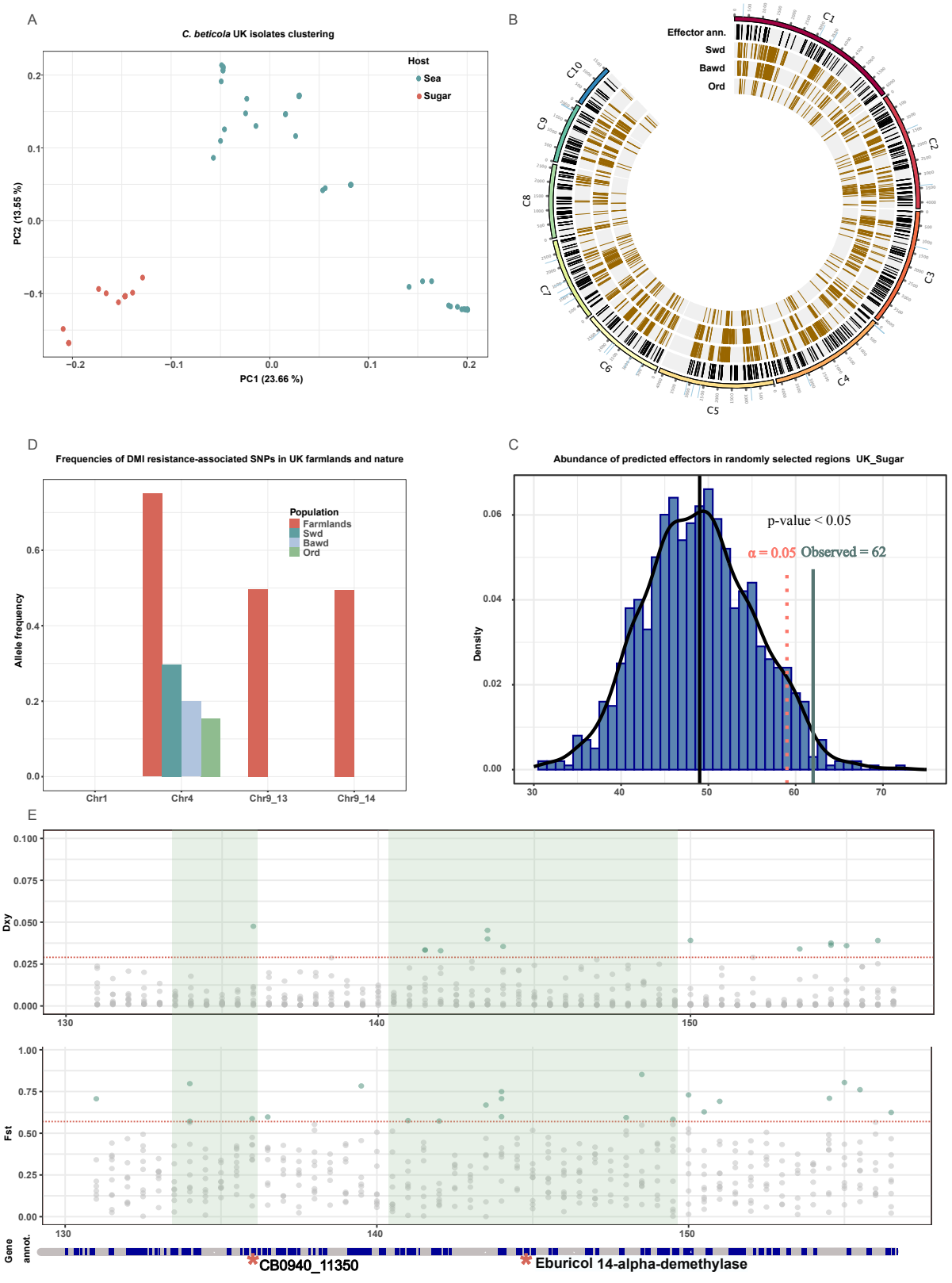


Figure 5: Distribution of highly diverging regions across the genome in C. beticola populations from wild and domesticated hosts. A) DCA analysis. Color reflects the host from which each

isolatewas obtained. B) Genomic map of selective sweeps for each population. The first track shows coordinates of genes encoding predicted effectors. C) To determine if genes encoding putative effectors are enriched in selective sweep regions, we performed an enrichment analysis based on the distribution of predicted effector abundance in randomly selected genomic regions of the same number and length as the selective sweep regions. As many as 10,000 runs of random resampling of genomic regions was perform to validate that effector genes indeed are enriched in regions that have experienced recent positive selection. D) Allele frequencies of four SNPs fund associated with fungicide resistance in UK wild and domesticated host-infecting populations. E) Dxy, and Fst analyses between the british populations of southwold (sea beet infecting), and Bracebridge (sugar beet infecting) across a region on chromosome 9. Highlighted, two genome loci previously identified as associated with fungicide resistance (R. Spanner et al., 2021). At the bottom, the gene annotation presented. The red stars show the two mutations associated with fungicide resistance on chromosome 9, one of them located on the Eburicol 14-alpha methylase gene and the other on a gene producing a hypothetical protein.

Discussion

Lineages of *C. beticola* have recently specialized to a new crop species in agro-ecosystems

Understanding the factors driving the evolutionary trajectories of crop pathogens is crucial to predict future epidemics. In this study, we addressed the extend of host specialization of the globally occurring pathogen of sugar beet, *C.*

beticola. To this end, we used a global extensive sampling of the pathogen from wild and domesticated host to assess the population structure and its evolution on wild and domesticated plants. Our analyses were based on the hypothesis that the agricultural environment and the domestication of beets shaped the evolution *C. beticola* lineages adapting to it. Our detailed population genomic analyses provide evidence that *C. beticola* is subdivided into multiple lineages associated with wild and domesticated hosts. Our analysis of ancestry contribution revealed six main ancestral clusters. Intriguingly isolates coming from the wild showed different ancestries than the isolates coming from farmlands. This separation was even more surprizing for isolates collected in a relatively small distance (within 200 Km) in UK. The UK isolates from farmlands and from wild seabeet, not only showed different ancestries, but also there was no evidence of hybridization between the populations. This indicates some extent of host specialization among closely related lineages of *C. beticola*. Host specialized lineages was also reported for major plant pathogens like in the rice blast fungus (*Pyricularia oryzae,* syn: *Magnaporthe oryzae*) (Gladieux, Condon, et al., 2018) and of the wheat pathogen *Zymoseptoria tritici* (Hartmann et al., 2018). In Hartmann et al., 2018, the authors used statistical methods to identify genes under selection in different host-specific lineages of the pathogens. Signatures of positive selection were identified in known effector genes,

indicating adaptation to distinct wheat cultivars. Moreover, also genes associated with fungicide resistance we identified in selective sweep regions in *Z. tritici* (Hartmann et al., 2018). Highly diverging regions between *C. beticola* populations from the wild and from farmlands were here also found to be enriched with predicted effector genes. Furthermore, genomic variants associated with DMI fungicide resistance was also found in high frequencies in UK *C. beticola* from farmlands but not in the wild reflecting adaptation to agro-ecosystems and the recurrent use of fungicides.

Further evidence of host specific divergence was obtained from an AMOVA analysis. AMOVA results showed that the effect of geographic distance is not significant on the partition of genetic variance of the species. That is in line with previous works that assessed the populations structure of *C. beticola* isolates from farmlands, concluding to a shallow population differentiation (Knight et al., 2019; Vaghefi, Kikkert, et al., 2017; Vaghefi, Nelson, et al., 2017). Our analysis included isolates of *C. beticola* from the wild ancestor of beet, the sea beet. This allowed us to assess the significance of the host in the species population structure and estimate proportion of genetic variance associated with it. The analysis showed that the host had a significant impact explaining 21.42 % of the variation between *C. beticola* populations and support the hypothesis that the

different environments and hosts drive genetic differentiation in pathogen populations.

Sugar beet infecting populations of *C. beticola* have likely emerged in Mediterranean Bassin with subsequent introductions from wild beet species

A primary indicator of the center of origin of a species is the high genetic diversity compared to other locations where the species was introduced later. In the case of *C. beticola*, we observed higher diversity primarily in the East and center of the Mediterranean basin (Spain and Croatia) and secondarily on the wild sea beet plants collected in the UK. A number of studies highlight the effect of host domestication in the emergence of its associated pathogens. For example, pathogens that have emerged with their host during domestication include the wheat pathogenic fungus *Zymoseptoria tritici* causing the disease septoria tritici blotch (Stukenbrock et al., 2007), the rice blast fungus *Magnaporthe oryzae* (Couch et al., 2005), the corn smut fungus *Ustilago maydis* (Schweizer et al., 2021), and the barley net blotch fungus, *Pyrenophora teres* (Taliadoros et al., 2023). The Mediterranean Sea was previously proposed as the centers of diversity and domestication of beets (Biancardi et al., 2020). An

elevated diversity in that region indicates that the emergence of the pathogen is likely associated with the history the host for *C. beticola* as well.

Then we assessed the evolutionary relationships between the populations using polymorphism-aware phylogenetic models. We found a common origin of the sugar beet and fodder beet infecting populations. This populations formed a clade with the Croatian wild population, which was early diverging within the clade. This suggests a Mediterranean origin of the sugar/fodder beet populations. On the second clade, the UK wild and NY table beet populations where clustered. This suggests a UK origin of the NY population, different than the other populations of domesticated beets. The different origins of the sugar beet-infecting and red beet infecting populations lead to the emergence of an intriguing conclusion: *C. beticola* was introduced to the agro-ecosystem in multiple occasions.

In the abovementioned studies, the authors have used coalescence and simulation-based methods to infer the demographic history of the pathogen and assess the parallelism between major demographic events, such as bottlenecks in the inferred history of the pathogen and the known history of the host. Our extended *C. beticola* dataset allows for a detailed exploration of the

demographic history of the species and assessing the impact of host domestication is our future plan.

Diversity and sexual reproduction

We used SNP data to calculate the overall nucleotide diversity within *C. beticola* populations. Interestingly, *C. beticola* exhibited a higher level of nucleotide diversity (π) in comparison to other well-known pathogens like the wheat pathogen *Zymoseptoria tritici* (Hartmann et al., 2018), the wheat powdery mildew pathogen *Blumeria graminis f.sp. tritici* (Sotiropoulos et al., 2022), and the rice blast fungus *Magnaporthe oryzae* (Zhong et al., 2018). Genetic variation plays a crucial role in the rapid adaptation of pathogens, and the elevated nucleotide diversity in *C. beticola* may be a significant factor contributing to the pathogen's successful survival and dispersal. Disparities between species may be attributed to variations in the frequency of sexual recombination, gene flow, and historical population bottlenecks. Our LD analyses, along with the distribution of mating type frequencies, provide evidence for the frequent occurrence of sexual recombination in *C. beticola*. This is in line with previous reports for North American and European populations (Knight et al., 2019; Vaghefi, Nelson, et al., 2017). An exception is observed in the UK population from Ordorf, where clonality is more prevalent. The population genetic structure of the UK

populations reveals the presence of four distinct genetic clusters, three obtained from the nature and one from sugar beet fields. Depside, that these isolates coexist in a relatively small region they belong to distinct clusters with limited signs of geneflow between them. This observation suggests potential local adaptation, reflecting constraints on gene flow, particularly between isolates adapted farmlands and the wild.

Signatures of a host-driven divergence and DMI resistance under selection in UK farmlands

To explore signatures of local adaptation, we used different methods to identify genomic regions that are highly diverging between the one farmland and the three wild populations. Here, to control for the effects of geographic isolation on the population structure, we only used the UK populations from farmlands and the wild. In total we identify 214 highly diverging regions between the populations. Diverging genomic regions between farmland and wild *C. beticola* populations could represent important host specificity loci. Intriguingly, about a third of the diverging regions (73) were shared among all comparisons between the farmland and the wild populations. Furthermore, we found that highly diverging genomic regions were enriched with predicted effector genes, which

further highlight the important of host as a main driver of rapid evolution in this pathogen.

A recent study reported genomic loci associated with DMI resistance (R. Spanner et al., 2021). We recorded the abundance of DMI resistance alleles we recorded their frequency in the four UK populations. We found three out of four variants associated with fungicide resistance present in high frequencies at the UK farmland population. In contrast we found only one of the for variants in low to moderate frequencies present in the wild populations. Furthermore, two of the DMI resistance loci are located in the highly-diverging regions, highlighting a rapid adaptation of the pathogen to the agro-ecosystem. DMI compounds are prominently used for the control of *C. beticola* populations in UK. Furthermore, we found fungicide resistant associated alleles in high frequencies in various populations around the world. Our analysis calls for attention to an ongoing emergence of fungicide resistant crop-specialized lineages of *C. beticola* in UK and around the world.

Conclusion

Using a population genomics framework, we show that *C. beticola* is subdivided into multiple lineages associated with domesticated and wild hosts. We further present evidence of limited genetic exchanges between them. Our genetic diversity and phylogenetic analyses provide evidence of a tandem demographic

history between the host and the pathogen and multiple introductions of the pathogen to the agro-ecosystem. This adds to the growing body of work highlighting the important role of human activities, such as domestication, migration and trade to the emergence and dispersal of plant pathogens. Furthermore, our results enhance our understanding of the evolutionary forces driving the diversification of *C. beticola* and underscore the utility of genomic data for epidemiological surveillance of fungal plant pathogens.

Data Availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. Custom scripts and workflows are available at https://github.com/Jimi92/Cercospora_beticola_pop_gen.

Acknowledgements

The authors are grateful to all members of the Environmental Genomics group. DT's work was supported by the DFG Research Training Group RTG 2501. This work was supported in part by the U.S. Department of Agriculture, Agricultural Research Service through USDA projects 3060-22000-051-000D (Friesen) and 3060-21000-044-000-D.

References

1. Alexander, D., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19, 1655–1664.
2. Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., Enjalbert, J., & de Vallavieille-Pope, C. (2014). Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. tritici. PLoS Pathogens, 10(1). https://doi.org/10.1371/journal.ppat.1003903
3. Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguileta, G., & Snirc, A. (2017). Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. 2041–2062. https://doi.org/10.1111/mec.13976
4. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & 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
5. Biancardi, E., Panella, L. W., & Mcgrath, · J Mitchell. (2020). The Origin of Beets Second Edition.
6. 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
7. Bolton, M. D., Secor, G. A., Rivera, V., Weiland, J. J., Rudolph, K., Birla, K., Rengifo, J., & Campbell, L. G. (2012). Evaluation of the potential for sexual reproduction in field populations of Cercospora beticola from USA. Fungal Biology, 116(4), 511–521. https://doi.org/10.1016/j.funbio.2012.01.011
8. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10, 1–9. https://doi.org/10.1186/1471-2105-10-421

9. Couch, B. C., Fudal, I., Lebrun, M. H., Tharreau, D., Valent, B., Van Kim, P., Nottéghem, J. L., & 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
10. Croll, D., & McDonald, B. A. (2017). The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. Molecular Ecology, 26(7), 2027–2040. https://doi.org/10.1111/mec.13870
11. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., Mcvean, G., Durbin, R., Project, G., & Vcf, T. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330
12. De Jonge, R., Ebert, M. K., Huitt-Roehl, C. R., Pal, P., Suttle, J. C., Spanner, R. E., Neubauer, J. D., Jurick, W. M., Stott, K. A., Secor, G. A., Thomma, B. P. H. J., De Peer, Y. Van, Townsend, C. A., & Bolton, M. D. (2018). Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus Colletotrichum. Proceedings of the National Academy of Sciences of the United States of America, 115(24), E5459–E5466. https://doi.org/10.1073/pnas.1712798115
13. Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. https://academic.oup.com/genetics/article/131/2/479/6007328
14. Feurtey, A., Lorrain, C., McDonald, M. C., Milgate, A., Solomon, P. S., Warren, R., Puccetti, G., Scalliet, G., Torriani, S. F. F., Gout, L., Marcel, T. C., Suffert, F., Alassimone, J., Lipzen, A., Yoshinaga, Y., Daum, C., Barry, K., Grigoriev, I. V., Goodwin, S. B., … Croll, D. (2023). A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. Nature Communications, 14(1). https://doi.org/10.1038/s41467-023-36674-y
15. Flor, H. H. (1971). CURRENT STATUS OF THE GENE-FOB-GENE CONCEPT. www.annualreviews.org

16. Gladieux, P., Bradford, C., Ravel, S., Soanes, D. M., Nhani, A., Chen, L., Leburn, M.-H., Mitchell, T., Farman, M., & Fournier, E. (2018). Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae Pierre. MBio, 9(1), 1–19.
17. Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J. L. N., Nhani, A., Chen, L., Terauchi, R., Lebrun, M. H., Tharreau, D., Mitchell, T., Pedley, K. F., Valent, B., Talbot, N. J., Farman, M., & 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
18. Groenewald, M., Groenewald, J. Z., Harrington, T. C., Abeln, E. C. A., & Crous, P. W. (2006). Mating type gene analysis in apparently asexual Cercospora species is suggestive of cryptic sex. Fungal Genetics and Biology, 43(12), 813–825. https://doi.org/10.1016/j.fgb.2006.05.008
19. 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
20. Hill, M. O., & Gauch, H. G. (1980). DETRENDED CORRESPONDENCE ANALYSIS: AN IMPROVED ORDINATION TECHNIQUE. Vegetatio, 42, 47–58.
21. Juan, J., Armenteros, A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., Heijne, G. Von, & Nielsen, H. (2019). SignalP 5 . 0 improves signal peptide predictions using deep neural networks. Nature Biotechnology, 37(April), 420–423. https://doi.org/10.1038/s41587-019-0036-z
22. Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in Genetics, 6(JUN). https://doi.org/10.3389/fgene.2015.00208
23. Knight, N. L., Vaghefi, N., Kikkert, J. R., Bolton, M. D., Secor, G. A., Rivera, V. V., Hanson, L. E., Nelson, S. C., & Pethybridge, S. J. (2019). Genetic Diversity and Structure in Regional Cercospora beticola Populations from Beta vulgaris subsp. Vulgaris Suggest Two Clusters of Separate Origin. Phytopathology, 109(7), 1280–1292. https://doi.org/10.1094/PHYTO-07-18-0264-R

24. Krogh, A., Larsson, È., Heijne, G. Von, & Sonnhammer, E. L. L. (2001). Predicting Transmembrane Protein Topology with a Hidden Markov Model : Application to Complete Genomes. https://doi.org/10.1006/jmbi.2000.4315
25. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. Genome Research, 19(9), 1639–1645. https://doi.org/10.1101/gr.092759.109
26. 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
27. Lunter, G., & Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011. Genome Research, 21:936-939(3), 936–939. https://doi.org/10.1101/gr.111120.110 [doi]
28. McKenna, A., Hanna, M., Banks, E., Sivachenko, A. Y., Cibulskis, K., Kernytsky, A. M., Garimella, K. V, Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.20
29. Menardo, F., Praz, C. R., Wyder, S., Ben-David, R., Bourras, S., Matsumae, H., McNally, K. E., Parlange, F., Riba, A., Roffler, S., Schaefer, L. K., Shimizu, K. K., Valenti, L., Zbinden, H., Wicker, T., & 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
30. Möller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. Nature Publishing Group, 15(12), 756–771. https://doi.org/10.1038/nrmicro.2017.76
31. Pereira, D., Croll, D., Brunner, P. C., & McDonald, B. A. (2020). Natural selection drives population divergence for local adaptation in a wheat pathogen. Fungal Genetics and Biology, 141(April), 103398. https://doi.org/10.1016/j.fgb.2020.103398
32. Rangel, L. I., Spanner, R. E., Ebert, M. K., Pethybridge, S. J., Stukenbrock, E. H., de Jonge, R., Secor, G. A., & Bolton, M. D. (2020). Cercospora beticola: The intoxicating lifestyle of the leaf spot pathogen of sugar beet. Molecular Plant Pathology, 21(8), 1020–1041. https://doi.org/10.1111/mpp.12962

33. Rovenich, H., Boshoven, J. C., & Thomma, B. P. H. J. (2014). Filamentous pathogen effector functions: Of pathogens, hosts and microbiomes. Current Opinion in Plant Biology, 20, 96–103. https://doi.org/10.1016/j.pbi.2014.05.001
34. Schweizer, G., Haider, M. B., Barroso, G. V., Rössel, N., Münch, K., Kahmann, R., & Dutheil, J. Y. (2021). Population Genomics of the Maize Pathogen Ustilago maydis: Demographic History and Role of Virulence Clusters in Adaptation. Genome Biology and Evolution, 13(5), 1–17. https://doi.org/10.1093/gbe/evab073
35. Sotiropoulos, A. G., Arango-Isaza, E., Ban, T., Barbieri, C., Bourras, S., Cowger, C., Czembor, P. C., Ben-David, R., Dinoor, A., Ellwood, S. R., Graf, J., Hatta, K., Helguera, M., Sánchez-Martín, J., McDonald, B. A., Morgounov, A. I., Müller, M. C., Shamanin, V., Shimizu, K. K., … Wicker, T. (2022). Global genomic analyses of wheat powdery mildew reveal association of pathogen spread with historical human migration and trade. Nature Communications, 13(1), 1–14. https://doi.org/10.1038/s41467-022-31975-0
36. Spanner, R. E., Taliadoros, D., Richards, J., Viviana, R.-V., Neubauer, J., Mari, N., Hamilton, O., Vaghefi, N., Pethybridge, S., Secor, G. A., Friesen, T. L., Stukenbrock, E. H., & Bolton, M. D. (2020). Genome-wide association studies reveal the complex genetic architecture of DMI fungicide resistance in Cercospora beticola. BioRxiv. http://mpoc.org.my/malaysian-palm-oil-industry/
37. Spanner, R., Taliadoros, D., Richards, J., Rivera-Varas, V., Neubauer, J., Natwick, M., Hamilton, O., Vaghefi, N., Pethybridge, S., Secor, G. A., Friesen, T. L., Stukenbrock, E. H., & Bolton, M. D. (2021). Genome-Wide Association and Selective Sweep Studies Reveal the Complex Genetic Architecture of DMI Fungicide Resistance in Cercospora beticola. Genome Biology and Evolution, 13(9), 1–17. https://doi.org/10.1093/gbe/evab209
38. 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
39. 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
40. Stukenbrock, E. H., Banke, S., Javan-Nikkhah, M., & McDonald, B. A. (2007). Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via

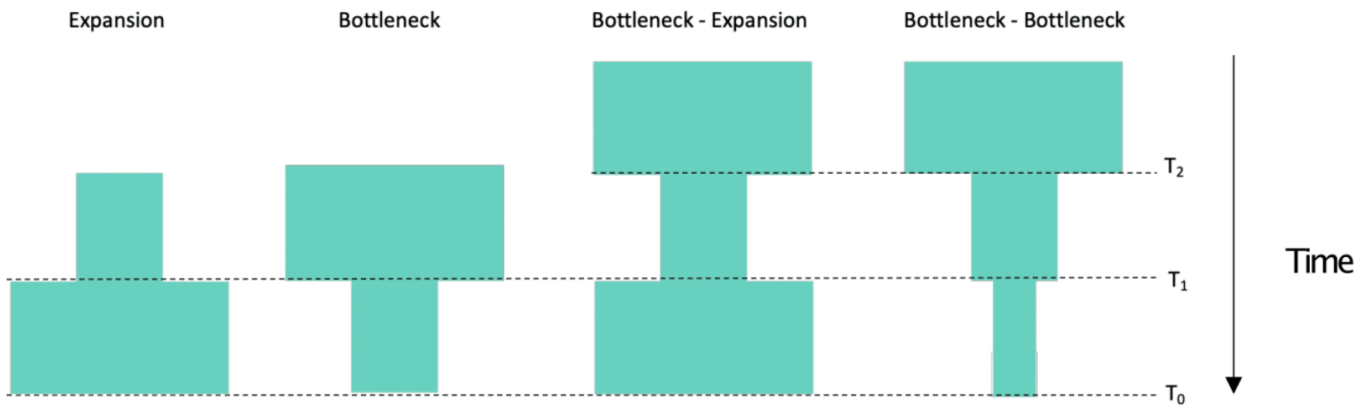
sympatric speciation. Molecular Biology and Evolution, 24(2), 398–411. https://doi.org/10.1093/molbev/msl169

41. Stukenbrock, E. H., & Mcdonald, B. A. (2008). The Origins of Plant Pathogens in Agro-Ecosystems. Annual Review of Phytopathology, 46, 75–100. https://doi.org/10.1146/annurev.phyto.010708.154114
42. Taliadoros, D., Feurtey, A., Wyatt, N., Gladieux, P., & Stukenbrock, E. (2023). Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history. Bioarxiv, 1–21. https://doi.org/10.1101/2023.07.28.550921
43. Taliadoros, D., & Stukenbrock, E. H. (2023). The use of evolutionary analyses to predict functionally relevant traits in filamentous plant pathogens. Current Opinion in Microbiology, 73, 102244. https://doi.org/10.1016/j.mib.2022.102244
44. Thierry, M., Charriat, F., Milazzo, J., Adreit, H., Ravel, S., Cros-Arteil, S., Borron, S., Sella, V., Kroj, T., Ioos, R., Fournier, E., Tharreau, D., & Gladieux, P. (2022). Maintenance of divergent lineages of the Rice Blast Fungus Pyricularia oryzae through niche separation, loss of sex and post-mating genetic incompatibilities. PLoS Pathogens, 18(7). https://doi.org/10.1371/journal.ppat.1010687
45. Vaghefi, N., Kikkert, J. R., Bolton, M. D., Hanson, L. E., Secor, G. A., Nelson, S. C., & Pethybridge, S. J. (2017). Global genotype flow in Cercospora beticola populations confirmed through genotyping-by-sequencing. PLoS ONE, 12(10). https://doi.org/10.1371/journal.pone.0186488
46. Vaghefi, N., Nelson, S. C., Kikkert, J. R., & Pethybridge, S. J. (2017). Genetic structure of Cercospora beticola populations on Beta vulgaris in New York and Hawaii. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-01929-4
47. Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics, 35(10), 1786–1788. https://doi.org/10.1093/bioinformatics/bty875
48. Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593
49. Zhong, Z., Chen, M., Lin, L., Han, Y., Bao, J., Tang, W., Lin, L., Lin, Y., Somai, R., Lu, L., Zhang, W., Chen, J., Hong, Y., Chen, X., Wang, B., Shen, W. C., Lu, G., Norvienyeku, J.,

Ebbole, D. J., & Wang, Z. (2018). Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. ISME Journal, 12(8), 1867–1878. https://doi.org/10.1038/s41396-018-0100-6

Chapter IV

DMI Fungicide resistance and specialization to local barley varieties drives the adaptation of the North American C. beticola Population



Chapter IV: DMI Fungicide resistance and specialization to local beet varieties drives the adaptation of the North American *C. beticola* Population

Demetris Taliadoros and Eva H. Stukenbrock

Environmental Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön and Christian-Albrechts University of Kiel, 24118, Germany

These results were published as parts of two broader studies in the journal Genome Biology and Evolution volume 12 issue 9, September 2021, (doi.org/10.1093/gbe/evab209), and in Molecular Plant Pathology, volume 22, issue 3, March 2021 (DOI: 10.1111/mpp.13026).
Below, I summarize the work that I contributed to the study:

Abstract

Demethylation inhibitor (DMI) fungicides are essential for controlling Cercospora leaf spot, one of the most economically relevant fungal disease affecting sugar beet crops globally. However, there has been a significant rise in the occurrence of DMI-resistant field isolates in many sugar beet production regions. We used population genomic data from field populations of *Cercospora beticola* to conduct a selective sweep analysis to assess if newly reported genomic regions associated with DMI resistance have undergone a selective sweep in a North American population. A demographic history analysis revealed that the

population recently experienced a strong bottleneck. In addition, several of the regions associated found to be associated with DMI fungicide resistance, based on a GWAS analyses, were overlapping with selective sweep regions. Interestingly, we found that a the recently reported *CbNip1* effector gene also co-localizes with a selective sweep in a North American population of the pathogen. Possibly, this reflects specialization to local barley varieties. Insights into the genome evolution of *C. beticola* is essential for the development of effective detection methods and management strategies to control Cercospora leaf spot, thereby contributing to the sustainability of sugar beet production.

Background

Demethylation inhibitor (DMI) fungicides play a crucial role for the management of economically important crop diseases, including Cercospora leaf spot (CLS) in sugar beet (*Beta vulgaris* spp. *vulgaris*) (1). CLS is recognized as the most devastating foliar disease affecting sugar beet crops on a global scale (1). The Red River Valley (RRV) region in North Dakota and Minnesota, United States, is the largest sugar beet production area in the country (NASS 2020). Historically, this region has faced substantial economic losses due to CLS, resulting in significant yield reductions and costs associated with the application of ineffective fungicides, including DMIs(2). Unfortunately, 17 years of recordings

of the prevalence of DMI resistance and the occurrence of resistant isolates in *C. beticola* field populations in the RRV show a stable increased (1).

As discussed in the previous chapters of this thesis, another major driver of pathogen evolution in agricultural systems is the specialization to a specific host or crop cultivar. Fungal plant pathogens employ effectors to suppress the host defences and exploit the host cell resources and mechanisms to its benefit (3). Cercosporin is a toxin that the pathogen employs to induce necrosis of the host cell. The production of cercosporin and the initiation of reactive oxygen species formation triggered by cercosporin are dependent on light. Therefore, cell death induced by this toxin is only fully guaranteed during periods of light. The recently reported effector protein CbNip1, secreted by *C. beticola*, induces increased necrosis even in the absence of light, suggesting that it may enhance necrosis formation alongside light-dependent necrosis (4).

Here we used 190 isolates from RRV to infer the demographic history of this population. Furthermore, we performed genomic scans of 89 fungicide resistant isolates to identify regions that have undergone a selective sweep recently. Our results show that multiple genomic regions associated with DMI resistance have been evolving under strong selection. Furthermore, the recently reported effector protein CbNip1, that induces increased necrosis also in the absence of

light was located in a selective sweep region, indicating that the gene has experience strong selection in N. America recently.

Methods

Inference of Demographic History

Prior to analyses of selective sweep signatures across the *C. beticola* genome, we determined the site frequency spectrum (SFS) and inferred the population history of RRV isolates. Our analysis was based on the fitting of a simulated SFS obtained under four demographic models (depicted in Figure 1) to the observed frequency spectrum (also referred to as unfolded or derived Allele Frequency Spectrum (DAFS)). To assess the shape of a neutral SFS, we aimed to minimize the effect of linkage disequilibrium and selection. To this end, we calculated the DAFS data based on single nucleotide polymorphisms (SNPs) that had a minimum distance of 1 Kbp from predicted coding sequences and a 0.5 Kbp from each other, keeping a total of 47,865 biallelic SNPs. The software Arlequin v. 3.5.2.2 (5) was used to obtain the DAFS. We used the latest assembly of the sister species *Cercospora* cf. *flagellaris* available on NCBI under the project PRJNA503907 to polarize the SNPs. To this end, the ART software (6) was used for *in silico* sequencing of the *Cercospora* cf. *flagellaris* assembly. Then, we proceeded to align the reads generated by the ART software to the *C. beticola*

reference genome. In our analysis, we assumed that the ancestral states of the SNPs were the ones found in the *C.* cf. *flagellaris* outgroup.

FastSIMCOAL2 (7) was used to infer the demographic history of the *C. beticola* population. We used FastSIMCOAL2 to perform coalescent simulations and estimate the likelihood of the data given a specific demographic model and parameter values. We optimized likelihood maximization through a series of Expectation Maximization iterations. To this end we generated: 1) 100,000 simulations to approximate the likelihood with high precision, 2) 40 cycles of the expectation maximization algorithm to ensure that the maximum was reached, and 3) several independent replicate estimations to ensure that the global maximum likelihood was found.

Four models with different population size change scenarios were compared: 1) a recent population expansion, 2) a recent population bottleneck, 3) a bottleneck followed by a population expansion, and 4) a population bottleneck followed by a second recent bottleneck (figure 1). FastSIMCOAL2 requires estimates of i) the effective population sizes (Ne) in number of haploid individuals, ii) the neutral mutation rate and iii) the recombination rate of the species. We set the search range of the present-day Ne based on Watterson's θ and the pairwise nucleotide differences (π), both obtained with the software Arlequin v. 3.5.2.2 (5), and a range of realistic neutral mutation rates (5x10^7, 5x10^8, 3x10^8, 1x10^8)

mutations per site per generation)(8,9). Specifically, the lowest and the highest Ne estimations were used as the lowest and the highest value of the search range. As information about the neutral mutation rate of *Cercospora* spp. is scarce, we performed 25 replicated runs of 100,000 simulations with 40 cycles of the expectation maximization for every combination of all four demographic scenarios and four different mutation rates (5x10^7, 5x10^8, 3x10^8, 1x10^8 mutation per site per generation) in 25 replicated runs per specified mutation rate. We have compared the 16 models using the AIC and choose the neutral mutation rate that showed the lowest Akaike information criterion (AIC) value for our final simulations (supplementary table S1).

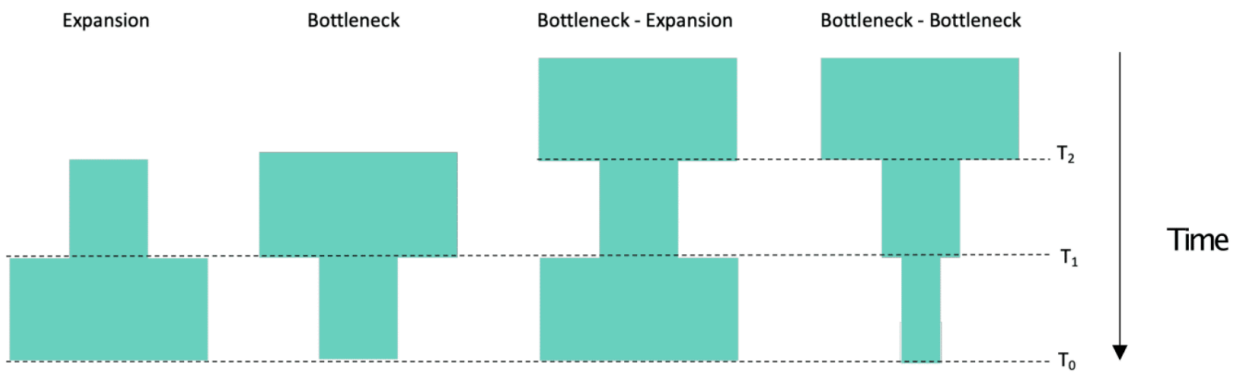


Figure 1. Illustration of the four demographic models that were compared in the demography inference of C beticola. The effective population size is represented by the width of the figure. T0 represents the present time, T1 represents the time of the most recent population size change and T2 a more ancient population size change.

Subsequently, we used the estimated search ranges of the present-day Ne, the best inferred neutral mutation rate and the recombination rate to simulate the four demographic models. For each demographic model, we performed 100,000 simulations, 40 cycles of the expectation maximization, and 50 replicate runs from different random starting values. We recorded the maximum-likelihood parameter estimates that were obtained across replicate runs. Finally, we performed a model selection based on the AIC. The model with lowest Akaike information criterion (AIC) as the demographic model that best fitted the data. We then infer parameter values in a second step by performing 100,000 simulations, 40 iterations of the expectation maximization and 100 replicate runs from different random starting values. Wrong polarization of the SNPs for the calculation of the derived SFS can introduce bias in the demographic history inference. We followed the same methods described above to further infer the demographic history of the population using the folded SFS and compared the models inferred using the folded and unfolded SFS yielding similar results.

Genome Scans for Selective Sweeps

Genomic scans for selective sweeps were performed by two approaches using the programs 1) OmegaPlus v. 3.0.3 (10) and 2) RAiSD v 2.9 (11). OmegaPlus is a scalable implementation of the ω statistic (12) that can be applied to whole-

genome data. It uses information on the LD between SNPs to estimate the likelihood of a certain region to have undergone a selective sweep. Each chromosome was used separately for the OmegaPlus analysis. The grid size was set to be equal to the number of variants contained in each chromosome (28,698–77,617 points). The minimum and maximum window sizes were set to 1,000 and 100,000 bp, respectively.

RAiSD computes the μ statistic, a composite evaluation test that scores genomic regions by quantifying changes in the SFS, the levels of LD, and the amount of genetic diversity along the chromosome (9). Background selection can highly influence RAiSD (11). We therefore assessed the false-positive rate (FPR) due to background selection using simulations. As many as 10,000 full genomes were simulated with a scenario of background selection and no selective sweeps, under the best inferred demographic scenario using the software SFS_CODE (13). Simulations of 37 segments of 1 Mb to resemble our genome data were performed. A population fitness factor (c) was given as the outcome of the effective population size (Ne) and the selection coefficient (S). Here, we employed an array of diverse values for the factor c (NeS): 50, 75, 100, and 200. In these simulations, we utilized the population mutation and recombination rate as estimated from the most probable model in the FastSIMCOAL2 simulations (refer to the "Inference of Demographic History" section).

Subsequently, 1,000 simulations under the previously inferred neutral demographic scenario were performed. Then we ran RAisD on the simulated data to assess the false positive rate caused by different levels of background selection. Following the recommendations of RAisD developers, we used the highest value of falsely identified regions to set the False Positive Rate for our RAisD analysis (11).

The significance of the identified selective sweeps was determined by calculating the ω and μ statistics on 10,000 datasets that were simulated under the best neutral demographic scenario using the ms program. (14) . A total of 37 1-Mb fragments were simulated to represent complete genomes. The inferred scaled mutation rate (θ, 2Neμ) and recombination rate (ρ, 2Ner) from the demographic model described above (under "Inference of Demographic History") served as parameters for these simulations. Establishing a significance threshold for the deviation of the ω and μ statistics based on simulated datasets under the most suitable neutral demographic model enabled us to regulate the influence of the population's demographic history on the Site Frequency Spectrum (SFS), Linkage Disequilibrium (LD), and genetic diversity throughout the genome. (12,15).

In both OmegaPlus and RAiSD, the consecutive windows with significant ω and μ values that overlapped were consolidated. The rightmost and leftmost positions of these windows were employed to delineate the selective sweep

regions. To evaluate the statistical importance of the intersection between GWAS candidates and selective sweeps, a randomization test was executed using 100,000 randomizations in R.

Results

Genome-Wide Scan for Recent Selection in the North American C. beticola Population

We here applied population genomic data to address whether DMI fungicide application had resulted in recent selection signals within the *C. beticola* population. To explore this, our attention was directed towards a set of 89 isolates that were found to display resistance to DMI fungicides (16), and we conducted a comprehensive genome-wide analysis for selective sweeps. To accommodate the influence of demographic events on the distribution of genetic diversity across the genome, we initially deduced the recent population history of the *C. beticola* population. This was achieved through a simulation approach that relied on the unfolded site frequency spectrum (SFS), utilizing the FastsimCOAL2 software (7). The probability of four demographic scenarios was evaluated (see methods). After comparing the Akaike Information Criterion (AIC) values for these four models, we identified strong indications for a single recent bottleneck (supplementary table S2). Additionally, the model involving an

ancient bottleneck followed by a recent expansion exhibited high likelihood, whereas the model featuring population expansion appeared less likely.

Subsequently, we employed a strategy involving 100 replicate runs of 100,000 simulations to deduce parameter values. The parameter values obtained from the best-fitting demographic scenario were then utilized to simulate data under neutrality using the software ms.

Two methods, with varying sensitivity to genomic signals across genome alignments, were employed for the detection of selective sweeps in the *C. beticola* subpopulation exhibiting enhanced fungicide tolerance. OmegaPlus and RAiSD identified 583 and 322 regions of selective sweeps, respectively, distributed throughout the *C. beticola* chromosomes (Figure 1 and supplementary tables S3 and S4). These regions displayed significantly higher values of the μ and ω statistics compared to the highest statistic values obtained from 10,000 simulations under the neutral demographic scenario. Selective sweep regions varied in length, ranging from 0.73 to 104.53 kb for OmegaPlus and from 1.02 to 672.21 kb for RAiSD. The number of included variants within these regions showed a range of 84 to 707 SNPs for OmegaPlus and 52 to 6,208 for RAiSD. Additionally, the number of predicted coding regions within the inferred selective sweeps ranged from zero to 79 for OmegaPlus and zero to 59 for RAiSD. A comparison of the results from the two independent approaches,

OmegaPlus and RAiSD, revealed some overlapping selective sweep regions. In total, 198 overlapping regions between the two selective sweep maps were identified, with a minimum overlap size of 150 bp, spanning from 157 to 7.7 kb (supplementary table S3). These overlapping regions accounted for 2.94% of the genome. Among these, as many as 172 out of the 198 overlapping selective sweeps contained coding sequences, ranging from zero to 44 coding regions per overlapping region, indicating a set of functional traits that recently underwent selection (supplementary table S5).

Signatures of Recent Selective Sweeps Colocalize with DMI Fungicide Resistance Candidates in the *C. beticola* Population

Within chromosome 9, two of the selective sweep regions identified by RAiSD encompassed variants that were also detected in a previously performed GWAS analysis (16), showing a correlation with DMI fungicide resistance). These two RAiSD regions spanned positions 1402834–1412304 and 1495335–1503162 bp of chromosome 9 (supplementary table S5). The first region included intergenic markers 9_1404771 and 9_1405148, as well as marker 9_1403629 located within the coding region of CB0940_11365, which encoded an aldehyde reductase. The second region contained markers 9_1496540 and 9_1497163 within the coding region of CB0940_11398, which encoded a hypothetical

protein. To assess the significance of the overlap between GWAS candidates and selective sweeps, a randomization test was conducted. In the randomization of both sets of features (GWAS candidates and selective sweeps), a significant deficit of overlap was observed, which could be attributed to the non-random distribution of GWAS candidates throughout the genome. Given that several GWAS candidates overlapped with the two sweeps on chromosome 9, the likelihood of observing such a number of SNPs in these regions by chance was assessed by randomizing the GWAS candidates on the corresponding chromosome. The observed numbers were found to be higher than expected by chance (P value = 0.0095 for the first sweep region, P value = 0.0508 for the second sweep region, and P value = 0.0011 when considering both sweeps together, computed with 100,000 randomizations). Coordinates of the predicted genomic features within the selective sweep regions on chromosome 9 are summarized in supplementary table S6.

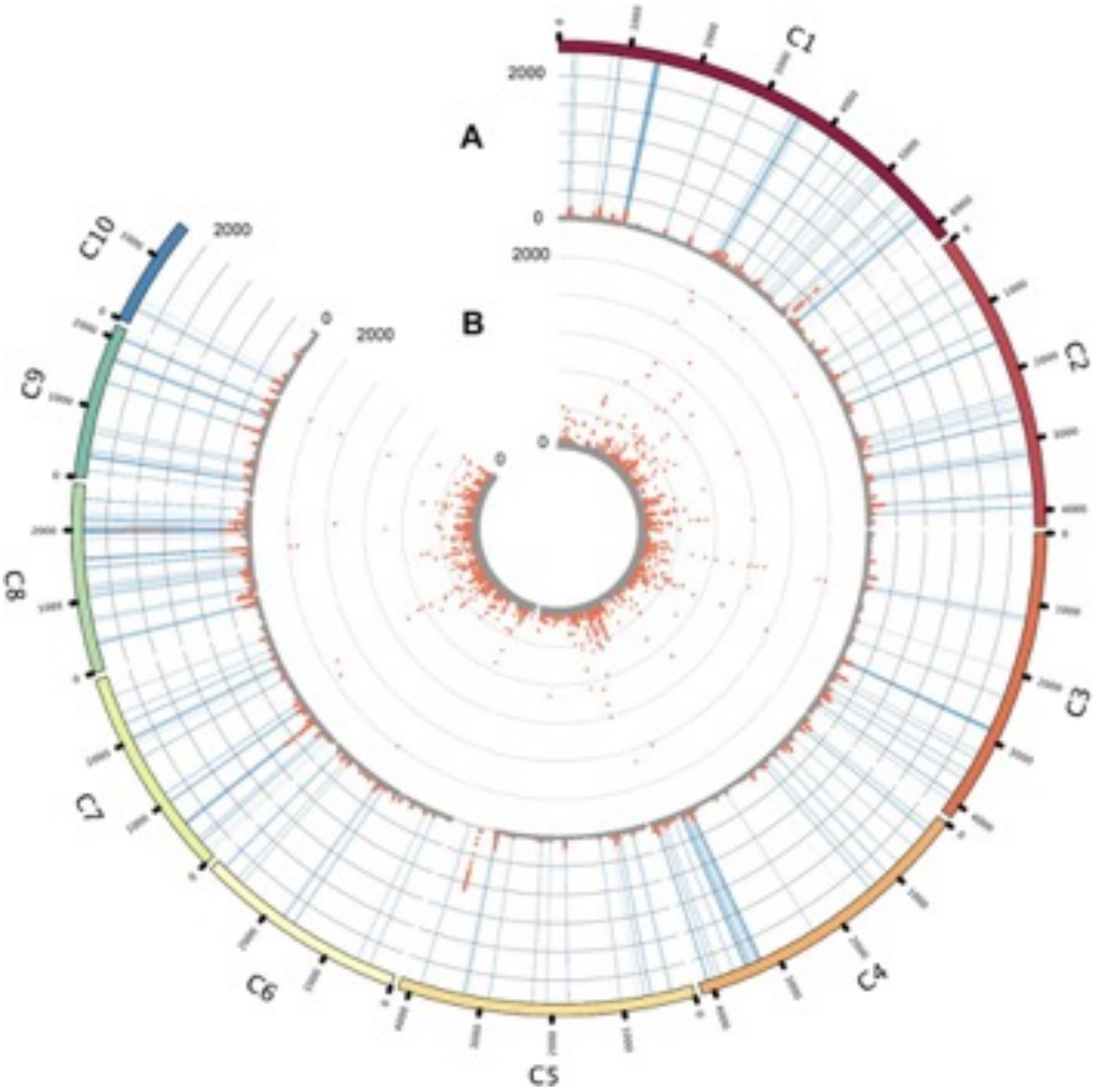


Figure 2. Genomic scans for selective sweeps using RAiSD and OmegaPlus. (A) Selective sweep map obtained by RAiSD; the μ-statistic values were calculated and plotted along the genome. Significant outlier loci are shown in red. (B) Selective sweep map obtained by OmegaPlus; ω-statistic values were plotted across the genome. Significant outlier loci are shown in red. The significance thresholds of the μ and ω statistics were determined with demographic simulations (see Materials and Methods). Blue lines indicate selective sweeps longer than 150 bp detected with both methods.

Genomic region encoding *CbNip1* recently underwent a selective sweep

Cercospora beticola is a hemibiotrophic. After a short biotrophic phase, the fungus switches to the nectotrophic phase to complete its life cicle. Recently, a new toxin named CbNip1 that the *C. beticola* utilizes to cause death to host cells during the necrotic phase was characterized (4). Here, we identified that the gene encoding CbNip1 in *C. beticola* is located in one of the selective sweep regions inferred for the RRV population. Both statistics (ω and μ) for detection of selective sweeps, identified an outlier locus at the end of chromosome 4, which coincided with the region containing the *CbNip1* gene (Figure 3). This particular locus exhibited significantly higher values for both statistics ω (indicating increased linkage disequilibrium on each side of the selective sweep) and μ (indicating signature of a selective sweep on the SFS and reduce genetic diversity in the region) when compared to values obtained from simulations assuming a neutral demographic scenario). The regions identified by these two methods spanned 10,750 bp (OmegaPlus) and 65,192 bp (RAiSD), encompassing 2,754 and 6,118 single nucleotide polymorphisms (SNPs), respectively (Figure 3).

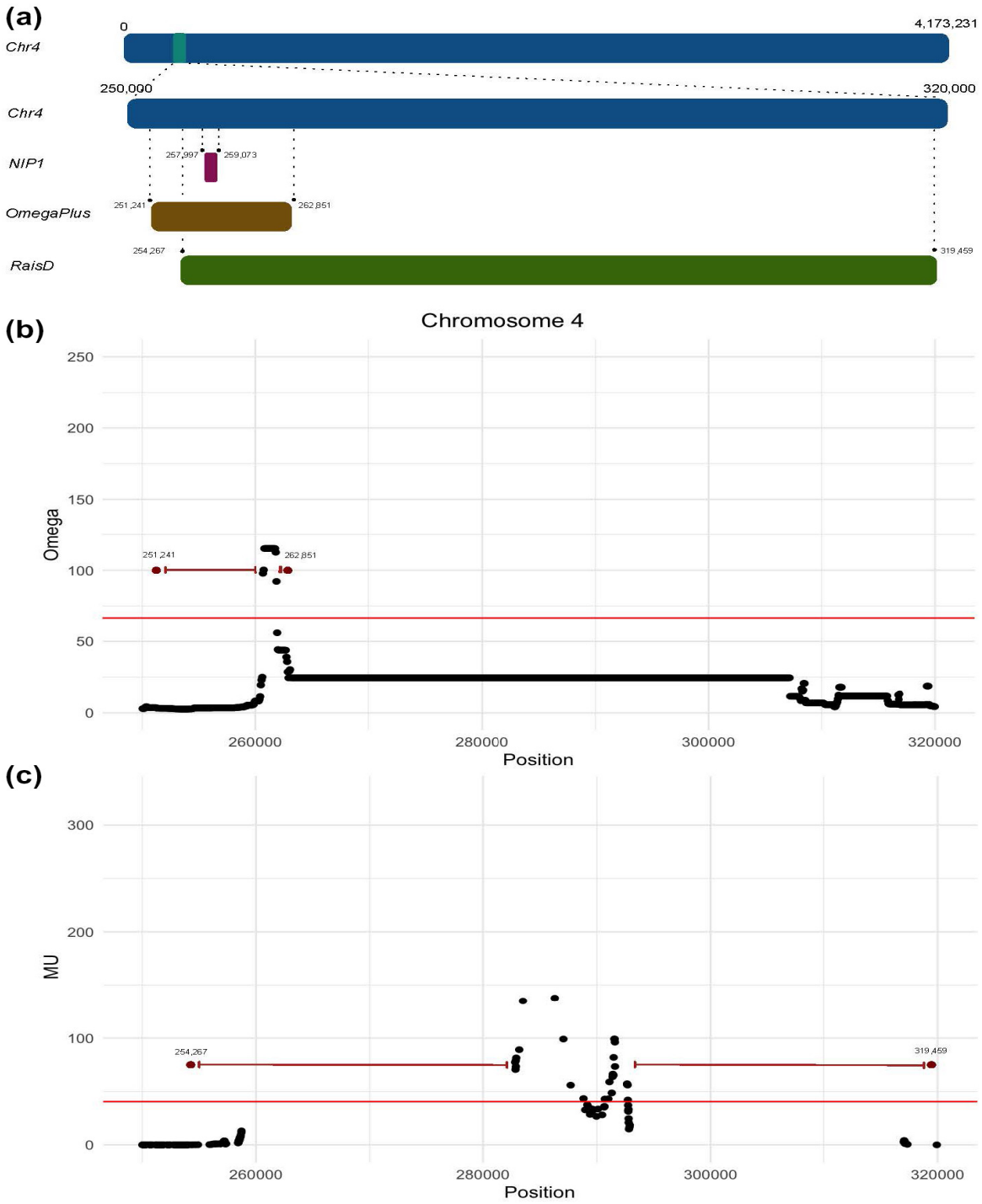


Figure 3. The genomic region encoding CbNip1 colocalizes with a selective sweep. Genome-wide screen of selective sweeps in *Cercospora beticola* revealed a candidate region on chromosome 4 that spans the CbNip1 locus. (a) Position of the CbNip1 gene on chromosome 4 and the regions that were identified by the selective sweep analyses using demographic modelling and methods implemented in the programs Omega and RAiSD. (b) OmegaPlus and

(c) RAiSD candidate regions on the chromosome arm of chromosome 4. The significance thresholds of the μ and ω statistics were determined with demographic simulations. Successive windows that were identified as significant were merged. The points in red represent the most left and the most right variants of the significant successive windows of the two statistics ω and μ computed by OmegaPlus and RAiSD, respectively.

Discussion

Selective sweeps have previously been observed to coincide with pesticide resistance loci in various organisms, including insects (17), weeds (18), and fungi (19). In this study, we used DMI-resistant *C. beticola* isolates, and found that genome-wide selective sweeps overlap with certain loci associated with DMI fungicide resistance. This suggests that the application of these fungicides is a driving factors of the evolution of the pathogen applying selective pressure on the North American population of *C. beticola*. Within putative selective sweep regions, GWAS loci were identified within the hypothetical protein CB0940_11398 and aldehyde reductase CB0940_11365, indicating that alleles at these loci may have provided a significant fitness advantage to *C. beticola*. Further confirmation of their selective advantage requires both functional and genetic assays. Taken all together, our results lead to an alarming conclusion: DMI resistance is evolving fast for the *C. beticola* population in N. America.

Furthermore, we have provided evidence indicating that the *CbNip1* locus is situated in a genomic region that has recently undergone a selective sweep. The presence of *CbNip1* within this selective sweep region highlights the significance of this effector in the population. However, the specific nature of the selection pressure favoring the retention of *CbNip1* remains unclear. It's possible that the selection of this gene variant within the population could be linked to a conserved and specific host target that was widely introduced through sugar beet germplasm. For instance, given the necrotic characteristics of CbNip1, one might speculate that this effector interacts specifically with a host immune receptor introduced through recent introgression, resulting in a reverse gene-for-gene interaction, a phenomenon observed in other systems (20,21). Alternatively, CbNip1 may lack a specific host target, and its nonhost specificity could be attributed to the generally cytotoxic nature of the protein. Further investigations into nonhost responses to CbNip1 are required to elucidate the mechanisms underlying the CbNip1-induced necrotic phenotype.

Conclusion

Extensive application of fungicide and use of resistant varieties are key characteristics of conventional agriculture. Here we showed that a North American population of *C. beticola* has experience strong selection in multiple genomic regions associated with fungicide resistance. We also, identified that an

effector gene had undergone a selective sweep recently, which might indicate an ongoing specialization to a specific host. Our findings provide evidence for the strong directional selection on crop pathogens imposed by intensive agriculture. Genome analyses provide detailed insights into evolutionary mechanisms and population histories and are key for the development of efficient and sustainable management strategies of plant pathogens.

Bibliography

1. Rangel LI, Spanner RE, Ebert MK, Pethybridge SJ, Stukenbrock EH, de Jonge R, et al. Cercospora beticola: The intoxicating lifestyle of the leaf spot pathogen of sugar beet. Mol Plant Pathol. 2020;21(8):1020–41.
2. Bolton MD, Secor GA, Rivera V, Weiland JJ, Rudolph K, Birla K, et al. Evaluation of the potential for sexual reproduction in field populations of Cercospora beticola from USA. Fungal Biol. 2012 Apr;116(4):511–21.
3. Castro-moretti FR, Gentzel IN, Mackey D, Alonso AP. Metabolomics as an emerging tool for the study of plant–pathogen interactions. Metabolites. 2020;10(2):1–23.
4. Huang W, Li L, Myers JR, Marth GT. ART: A next-generation sequencing read simulator. Bioinformatics. 2012 Feb;28(4):593–4.
5. Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa V, Foll M. Robust Demographic Inference from Genomic and SNP Data. PLoS GNETICS. 2013;9(10):1–16.
6. Excoffier L, Lischer H. Windows, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and. Mol Ecol Resour. 2010;10(3):564–7.
7. Badouin H, Gladieux P, Gouzy J, Siguenza S, Aguileta G, Snirc A. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. 2017;2041–62.
8. Alachiotis N, Stamatakis A, Pavlidis P. OmegaPlus : a scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics. 2012;28(17):2274–5.
9. Alachiotis N, Pavlidis P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Commun Biol [Internet]. 2018;1(1). Available from: http://dx.doi.org/10.1038/s42003-018-0085-8
10. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Rese. 2005;15(11):1566–75.
11. Hernandez RD. A flexible forward simulator for populations subject to selection and demography. Bioinformatics. 2008;24(23):2786–7.
12. Hudson RR. Generating samples under a Wright–Fisher neutral model ofgenetic variation. Bioinformatics. 2002;18(2):337–8.

13. Stamatakis A, Alachiotis N, Pavlidis P, Daniel Z. SweeD : Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. Mol Biol Evol. 2013;30(9):2224–34.
14. Weedall GD, Riveron JM, Hearn J, Irving H, Kamdem C, Fouet C, et al. An Africa-wide genomic evolution of insecticide resistance in the malaria vector Anopheles funestus involves selective sweeps, copy number variations, gene conversion and transposons. PLoS Genet. 2020 Jun 1;16(6).
15. Kreiner JM, Sandler G, Stern AJ, Tranel PJ, Weigel D, Stinchcombe JR, et al. Repeated origins, widespread gene flow, and allelic interactions of target-site herbicide resistance mutations. Elife. 2022 Jan 1;11.
16. Hartmann FE, Vonlanthen T, Singh NK, McDonald MC, Milgate A, Croll D. The complex genomic basis of rapid convergent adaptation to pesticides across continents in a fungal plant pathogen. Mol Ecol. 2021 Nov 1;30(21):5390–405.
17. Faris JD, Zhang Z, Lu H, Lu S, Reddy L, Cloutier S, et al. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. Proc Natl Acad Sci U S A. 2010 Jul 27;107(30):13544–9.
18. Friesen TL, Meinhardt SW, Faris JD. The Stagonospora nodorum-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. Plant Journal. 2007 Aug;51(4):681–92.

Conclusion and Perspectives

General Conclusions and Perspectives

Understanding the evolutionary origin of crop pathogens is crucial to predict future epidemics (38–40). This thesis aimed to investigate the evolutionary trajectories of two economically important plant pathogens. Specifically, I sought to identify historical and modern factors that led to the emergence and the successful adaptations of the two fungal pathogens *P. teres* f. *teres* and *C. beticola* populations under different conditions worldwide.

Barley was one of the earliest domesticated crops in the Fertile Crescent approximately 10,000 years ago (41). We showed that the global population structure of the causal agent of barley's net form net blotch, *P. teres* f. *teres,* highly reflects the geographic origin of the isolates. This fact, in turn, suggests that pathogen populations are fast-evolving allowing them to adapt to their local conditions and available hosts. Furthermore, we inferred the demographic history of the pathogen's populations. Our results show that the history of the pathogen is highly associated with the history of barley. This work adds to a growing body of evidence supporting the notion that during the Neolithic period, crops' domestication was accompanied by the emergence of numerous new plant pathogens (2,4,9). These pathogens co-evolved with their host plants and have become some of the most significant crop diseases today, including the

Septoria tritici blotch of wheat (*Zymoseptoria tritici*) (2), and the corn smut (*Ustilago maydis*) (42)

We further aimed to identify the factors that drive the local adaptation of this pathogen. To this end, we employed statistical methods described in Chapter I to identify genomic regions that have undergone a selective sweep. Intriguingly, we found that most of the selective sweep regions we identified were population-specific and enriched with effector genes previously identified through QTL mapping. This result proves that the populations have adapted to different local conditions. It highlights the importance of adaptation to the local hosts as a significant driver of population divergence and local pathogen adaptation.

In contrast to barley's early domestication, sugar beet is one of the latest domesticated plants, being in its present form for only about 300 years (35) Previous studies suggested an extensive geneflow between *C. beticola* populations in North America and Europe (35). These studies primarily focused on populations isolated from farmlands and domesticated hosts. In contrast, in addition to farmland *C. beticola* populations from four different continents, our dataset included populations collected from the wild ancestor of beets, the sea beet in its natural habitat, from three sampling sites in the UK and two sampling

sites in Croatia. This allowed us to assess the extent of isolation and host specialization between populations obtained from farmlands and the wild. Our results confirm an extended geneflow between populations coming from farmlands at a global level.

Furthermore, our analyses prove a host-driven divergence between *C. beticola* populations. Ancestry inference and polymorphism-aware phylogenomic approaches showed a divergence between isolates coming from farmlands and the wild. Interestingly, population divergence without any evidence of hybridization was observed in populations from the wild and farmlands from a very small geographic area in the UK. This highlights the significant impact of the host driving the divergence of C. *beticola* populations.

Additionally, our phylogenomic analysis suggested an early split between the Mediterranean and the British sea beet infecting populations. We observed a close relationship between all sugar beet infecting populations placed in the same clade as the Mediterranean Sea beet infecting. Within the clade, the sea beet infecting population diverged first, suggesting a recent split and diversification of the sugar beet infecting isolates from the Mediterranean wild populations. The other clade contained all the UK sea beet infecting populations and, surprisingly, the N. American table beet infecting population. These results reveal that *C. beticola* farmland populations have different origins, i.e., the

pathogen was introduced to the agricultural environment on multiple independent occasions. Alarmingly, we observed variation associated with DMI fungicide resistance in varying frequencies present in all sugar beet and sea beet-infecting populations worldwide. This thesis adds to a growing concern about the rapidly evolving DMI fungicide resistance in *C. beticola*. It highlights the need to develop management strategies to control the disease and, at the same time, avoid promoting further DMI fungicide resistance.

Taking everything together, this thesis used population genetic approaches and bioinformatic tools to understand the emergence, population history, and local adaptation of two economically important fungal plant pathogens. In both studies, we observed a strong influence of host domestication on the evolutionary trajectory of the pathogen. This evidence adds to a growing body of evidence that suggests that human activities play a significant role in the emergence and dispersal of plant pathogens (25,61,68). Furthermore, in an era of abundance of whole-genome data, this thesis shows the potential of population genomics, when applied to a diverse dataset of natural populations, to enhance our understanding of crop pathogen emergence and local adaptation, assisting in the development of sustainable and efficient strategies for crop disease surveillance.

Future research could incorporate experimental methods to assess the functional relevance of the predicted genes in regions associated with local adaptation identified in the two pathogens. With the rapid increase in genomes high-quality whole genome sequences, questions such as what is the role of structural variation in the adaptation of these pathogens, what are the rates of adaptation of different genomic features, and what does the species' recombination/mutation rate landscape look like. This information can help us further understand the mechanisms and processes that the pathogens employ to adapt and thrive in the agricultural environment.

A major challenge that fugal population genomics faces today is that the research of fungal genome evolution relies on methodologies developed for well-studies life cycles, such as *Drosophila melanogaster* and *Homo sapiens*. Fungi, however, exhibit intricate life cycles encompassing both sexual and asexual propagation. Additionally, the frequency of sexual cycles per year (generation time), can fluctuate and occasionally include prolonged periods dominated by asexual reproduction with infrequent instances of sexual mating. Moving forward, it is crucial to acknowledge that these complexities may even violate general assumptions implemented in some of the computational methods. Therefore, there is a need for the development of improved

evolutionary models that would be able to take into account the life histories of eukaryotic microorganisms like fungal crop pathogen.

Taking a step forward, recent advances in the field of ecology have demonstrated the potential of utilizing population genomics to predict how the ongoing climate change will impact the future of a number of species (43). The same framework can potentially be used to assess the impact of climate change on agriculture and the evolution of crop pathogens.

Additional references

1. Nations U. Our growing population. 2019.
2. Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via sympatric speciation. Mol Biol Evol. 2007;24(2):398–411.
3. Taliadoros D, Feurtey A, Wyatt N, Gladieux P, Stukenbrock E. Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history. bioarxiv [Internet]. 2023;1–21. Available from: https://doi.org/10.1101/2023.07.28.550921
4. Sotiropoulos AG, Arango-Isaza E, Ban T, Barbieri C, Bourras S, Cowger C, et al. Global genomic analyses of wheat powdery mildew reveal association of pathogen spread with historical human migration and trade. Nat Commun. 2022;13(1):1–14.
5. Rout ME, Callaway RM. Interactions between exotic invasive plants and soil microbes in the rhizosphere suggest that “everything is not everywhere”. Vol. 110, Annals of botany. 2012. p. 213–22.
6. Stukenbrock EH, Mcdonald BA. The Origins of Plant Pathogens in Agro-Ecosystems. Annu Rev Phytopathol. 2008;46:75–100.
7. Banke S, Mcdonald BA. Migration patterns among global populations of the pathogenic fungus Mycosphaerella graminicola. Mol Ecol. 2005;14(7):1881–96.
8. Feurtey A, Lorrain C, McDonald MC, Milgate A, Solomon PS, Warren R, et al. A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. Nat Commun. 2023;14(1).
9. Munkacsi AB, Stoxen S, May G. Ustilago maydis populations tracked maize through domestication and cultivation in the Americas. Proceedings of the Royal Society B: Biological Sciences. 2008;275(1638):1037–46.
10. Santini A, Liebhold A, Migliorini D, Woodward S. Tracing the role of human civilization in the globalization of plant pathogens. ISME Journal. 2018;12(3):647–52.
11. Menardo F, Praz CR, Wyder S, Ben-David R, Bourras S, Matsumae H, et al. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. Nat Genet. 2016;48(2):201–5.

12. Pereira D, Croll D, Brunner PC, McDonald BA. Natural selection drives population divergence for local adaptation in a wheat pathogen. Fungal Genetics and Biology [Internet]. 2020;141(April):103398. Available from: https://doi.org/10.1016/j.fgb.2020.103398
13. Badouin H, Gladieux P, Gouzy J, Siguenza S, Aguileta G, Snirc A. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. 2017;2041–62.
14. Thierry M, Charriat F, Milazzo J, Adreit H, Ravel S, Cros-Arteil S, et al. Maintenance of divergent lineages of the Rice Blast Fungus Pyricularia oryzae through niche separation, loss of sex and post-mating genetic incompatibilities. PLoS Pathog. 2022 Jul 1;18(7).
15. Ali S, Gladieux P, Leconte M, Gautier A, Justesen AF, Hovmøller MS, et al. Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. tritici. PLoS Pathog. 2014 Jan;10(1).
16. Gladieux P, Bradford C, Ravel S, Soanes DM, Nhani A, Chen L, et al. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus Magnaporthe oryzae Pierre. mBio. 2018;9(1):1–19.
17. Spanner R, Taliadoros D, Richards J, Rivera-Varas V, Neubauer J, Natwick M, et al. Genome-Wide Association and Selective Sweep Studies Reveal the Complex Genetic Architecture of DMI Fungicide Resistance in Cercospora beticola. Genome Biol Evol. 2021;13(9):1–17.
18. Carey J. Unearthing the origins of agriculture. Proc Natl Acad Sci U S A. 2023 Apr 11;120(15).
19. van Zeist W, Bakker. View of Archaeobotanical studies in the Levant: I. Neolithic sites in the Damascus Basin: Aswad, Ghoraife, Ramad. Palaeohistoria. 1982;24:165–256.
20. Åberg E. Hordeum agriocrithon nova sp., a wild six-rowed barley. Ann R Agric Col Swed. 1938;6:159–216.
21. Molina-Cano JL, Fra-Mon P, Salcedo G, Aragoncillo C, de Togores FR, García-Olmedo F. Morocco as a possible domestication center for barley: biochemical and agromorphological evidence. Theoretical and Applied Genetics. 1987;73(4):531–6.
22. Molina-Cano JL, Moralejo M, Igartua E, Romagosa I. Further evidence supporting Morocco as a centre of origin of barley. Theoretical and Applied Genetics. 1999;98(6–7):913–8.
23. Tanno K, Taketa S, Takeda K, Komatsuda T. A DNA marker closely linked to the vrs1 locus (row-type gene) indicates multiple origins of six-rowed cultivated barley (Hordeum vulgare L.). Theoretical and Applied Genetics. 2002;104(1):54–60.

24. Riehl S. Barley in Archaeology and Early History. Oxford Research Encyclopedia of Environmental Science. 2019. 1–26 p.
25. Liu Z, Ellwood SR, Oliver RP, Friesen TL. Pyrenophora teres: Profile of an increasingly damaging barley pathogen. Mol Plant Pathol. 2011;12(1):1–19.
26. Douiyssi A, Rasmusson DC, Roelfs AP. Responses of barley cultivars and lines to isolates of Pyrenophora teres. Plant Dis. 1998;82(3):316–21.
27. Louw JPJ, Victor D, Crous PW, Holz G, Janse BJH. Characterization of Pyrenophora Isolates Associated with Spot and Net Type Lesions on Barley in South Africa. Journal of Phytopathology. 1995;143(3):129–34.
28. Zeleke T. Evaluation of host reaction and yield performance of malt barley cultivars to net blotch, Pyrenophora teres in Bale Highlands, Ethiopia. Journal of Plant Sciences. 2017;5(1):43–7.
29. Arabi MIE, Al-Safadi B, Charbaji T. Pathogenic variation among isolates of Pyrenophora teres, the causal agent of barley net blotch. Vol. 151, Journal of Phytopathology. 2003. p. 376–82.
30. Sato K, Takeda K. Net blotch resistance in wild species of Hordeum. Euphytica. 1997;95(2):179–85.
31. ARABI MI, BARRAULT G, SARRAFI A, ALBERTINI L. Variation in the resistance of barley cultivars and in the pathogenicity of Drechslera teres f. sp. maculata and D. teres f. sp. teres isolates from France. Plant Pathol. 1992;41(2):180–6.
32. Tekauz A. Characterization and distribution of pathogenic variation in Pyrenophora teres f. teres and P. teres f. maculata from western Canada. Canadian Journal of Plant Pathology. 1990;12(2):141–8.
33. Moya P, Girotti JR, Toledo AV, Sisterna MN. Antifungal activity of Trichoderma VOCs against Pyrenophora teres, the causal agent of barley net blotch. J Plant Prot Res. 2018;58(1):45–53.
34. Cromey MG, Parkes RA. Pathogenic variation in *Drechslera teres* in New Zealand. New Zealand Plant Protection. 2003;56:251–6.
35. Biancardi E, Panella LW, Mcgrath · J Mitchell. The Origin of Beets Second Edition. 2020. 1–42 p.
36. Rangel LI, Spanner RE, Ebert MK, Pethybridge SJ, Stukenbrock EH, de Jonge R, et al. Cercospora beticola: The intoxicating lifestyle of the leaf spot pathogen of sugar beet. Mol Plant Pathol. 2020;21(8):1020–41.

37. Groenewald M, Groenewald JZ, Harrington TC, Abeln ECA, Crous PW. Mating type gene analysis in apparently asexual Cercospora species is suggestive of cryptic sex. Fungal Genetics and Biology. 2006 Dec;43(12):813–25.
38. Giraud T, Gladieux P, Gavrilets S. Linking emergence of fungal plant diseases and ecological speciation. Trends Ecol Evol. 2010;25(7):387–95.
39. Avila-quezada GD, Esquivel JF, Silva-rojas HV, Gerardo S, Garcia-avila CDJ, Quezada-salinas A, et al. Emerging plant diseases under a changing climate scenario : Threats to our global food supply. Emir J Food Agric. 2010;30(6):443–50.
40. Memoona I, Khola R, Sania A, Sobia Z, Fakiha A, Khalid M, et al. Preventing Potential Diseases of Crop Plants Under the Impact of a Changing Environment. In: Emerging Technologies and Management of Crop Stress Tolerance. 2014. p. 193–214.
41. Fuller QD, Weisskopf A. Encyclopedia of Global Archaeology, Barley: Origins and Development. 2014. 763–766 p.
42. Schweizer G, Haider MB, Barroso G V., Rössel N, Münch K, Kahmann R, et al. Population Genomics of the Maize Pathogen Ustilago maydis: Demographic History and Role of Virulence Clusters in Adaptation. Genome Biol Evol. 2021;13(5):1–17.
43. Capblancq T, Fitzpatrick MC, Bay RA, Exposito-Alonso M, Keller SR. Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. 2020; Available from: https://doi.org/10.1146/annurev-ecolsys-020720

Acknowledgements

First, I would like to thank my supervisor, Eva Stukenbrck. I will always be greatful for the opportunities that you gave me, for the lessons that I earned from you, the generosity and the support. I will always remember your kindness, your patience, and our talks.

I also owe a special thanks to the daily advisor of my MSc thesis and first experimental chapter of the present thesis, Alice Feurtey. Your hard work, your brilliance, your kindness and patience have been an inspiration for me. Thank you for been the best teacher. I am growing as a scientist having in mind all the lessons, I have learned working with you and I hope that one day I will be able to inspire my students as much as you have inspired me.

I would like to thank old and new friends/members of the Environmental Genomics Group, in particular members of the Plön team: Ezgi Özgurt, Ben Claassen, Onur Erk Kavlak. Cécile Lorrain, Primrose Boynton, and Rahul Unni for making my time in Plön amazing. Also, the Kiel team, especially: Marco Guerreiro, Elisha Thynne, Lizel Potgieter, Michael Habig, Danilo Pereira, and Janine Haueisen for their insightful comments and great discussions we had.

Furthermore, I would like to thank my TAC committee members Christian Jung and Tal Dagan for their excellent guidance during my PhD. I am also grateful to

my thesis examination committee members Julien Dutheil and Michael Seidl, for taking the time to read and evaluate my thesis.

During my PhD I had the opportunity to interact with some of the most nice and brilliant people I have ever met. With some of them we became great friends. I would like to thank my friend and colleague Amar Dhiman. Amar you are one of the hardest working and positive people I know. It was an honor working with you and I hope that I will work again with you in other projects in the future. Also thank you for being a great friend, and for all the interesting conversations we had. I would also like to thank my friends Loukas Theodosiou for the nice after work walks and discussions, and Michael Barnett for jamming with me. Furthermore, a big thank you to Alec Jacobsen and Louk Seton for the incredible adventures with bikes and Kayaks and for sharing some incredible BBQs with me. I also thank Jun Ishigohoka, Andrea Bours, Carolina Peralta, Beatriz Mourato, Enea Franceschini, Karem Lopez, Stella Kyomen, Felipe Sinhorini, Amor Damatac the II, Gisela Rondriguez, Ian Dewan, Vanesa Perez-Laguna, and Alexandros Vezyrakis for all the moments we shared in Plön.

Lastly, I would like to thank my family and my childhood friend Vasilis Lazouras supporting and believing in me all these years.

Lastly, with this thesis I fulfill a promise that I gave myself in the summer of 2001, when I first watched the National Geographic documentary series with the greek

title “Η μεγάλη εγκυκλοπαίδεια της ζωής” (The big encyclopedia of life), discussing the evolution of life on earth. There and then, fascinated by how all creatures are related my nine-year-old self decided to become an evolutionary biologist. I am happy that, depside the difficulties, I never gave up on that dream and in 2023 I am concluding my studies on this subject.

Declaration of author's contributions

The Ph.D. thesis of Demetris Taliadoros consisted of four chapters in the form of published and unpublished manuscripts. Specific contributions for the chapters are detailed here. Data for chapter 3 are available upon request.

Chapter I: On variant discovery in genomes of fungal plant pathogens

Demetris Taliadoros and Eva H. Stukenbrock.

Chapter I was published in Current Opinion in Microbiology.
Demetris Taliadoros and Eva H. Stukenbrock. " The use of evolutionary analyses to predict functionally relevant traits in filamentous plant pathogens." Current Opinion in Microbiology, Volume 73, pages 1 - 17 (2023), doi.org/10.1016/j.mib.2022.102244

Conceptualization, manuscript writing, and editing: DT, EHS

Chapter II: Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history

Demetris Taliadoros, Alice Feurtey, Nathan Wyatt, Benoit Barrès, Pierre Gladieux, Timothy Friesen and Eva Stukenbrock

Chapter II was published as a preprint on *biorxiv* (doi.org/10.1101/2023.07.28.550921) and is currently submitted to a peer-reviewed journal.

Conceptualization: DT, AF, PG, EHS. Resources: NW, BB, TF, EHS. Formal analysis: DT. Writing original draft: DT and EHS. Writing-reviewing and editing: AF, NW, BB PG, TF, and EHS

Chapter III: Genome-wide evidence of host specialization in wild and farmland populations of the cercospora leaf spot pathogen, *Cercospora beticola*

Demetris Taliadoros, Lizel Potgieter, Idalia rojas, Nathan Wyatt, Christian Jung, Melvin Bolton, and Eva Stukenbrock

Conceptualization: DT, IR, EHS. Resources: LP, NW, CJ, MB, EHS. Formal analysis: DT. Writing original draft: DT and EHS.

Chapter IV: Genome-wide evidence of host specialization in wild and farmland populations of the cercospora leaf spot pathogen, *Cercospora beticola*

Demetris Taliadoros and Eva H. Stukenbrock.

Conceptualization: DT, EHS. Resources: EHS. Formal analysis: DT.

Curriculum Vitae

Demetris Taliadoros

Environmental Genomics Group, Max Planck Institute for Evolutionary Biology, Ploen, Germany

Email: taliadoros@evolbio.mpg.de

Phone: +357 99 803045

Github: https://github.com/Jimi92

Nationality: Cypriot

Home Town: Aglantzia, Cyprus

Current residence: Plön, Germany

Research experience

Ph.D. research, Environmental Genomics Group, Max Planck Institute for Evolutionary Biology (2020-2023)

- Investigating the genetic basis of local adaptation and host specialization of plant pathogenic fungi
- Inferring demographic histories of different fungal plant pathogen populations using simulation-based (e.g., ABC) and Sequential Markov coalescence algorithms
- Identifying genomic regions evolving under strong directional selection through genomic scans for selective sweep signatures within and between populations
- Identifying genomic features relevant for local adaptation, located in selective sweep regions

Relevant skills

- Strong foundation in population genetics
- Proficient in next-generation sequence analysis
- Experienced in awk, R, and python

Teaching experience

2021-2022: Population Genomics (biol-244), Christian Albrechts University of Kiel, Kiel, Germany

Responsibilities:

- Guide students through practical exercises using state-of-art population genomic software and packages on the terminal and in R
- Lecture on genetic effects of population structure: A summary of concepts and tools.

2021-2023: Supervision of internships, BSc, and MSc projects

- Rune Sommerkamp (2022, Intern): Presence/absence variation as an indicator of the demographic history and selection in different plant pathogens
- Rebekah Jolicoeur (2021, Intern): Host-driven divergence of *Cercospora beticola* populations in the UK. The case of crop-specialized lineages

Education

2020-2023: (anticipated):	Ph.D. Evolutionary biology, Kiel University, Kiel, Germany
2016-2019:	M.Sc. Biology, Wageningen University, Netherlands • Specialization: Biodiversity and Evolution • Dissertation: Population genomics and local adaptation of barley net blotch agent, *Pyrenophora teres*
2016-2019:	M.Sc. Forest and Nature Conservation, Wageningen, Netherlands • Specialization: Ecology • Dissertation: Life history traits of *Nasonia vitripennis*: The interplay of larval competition, sex ratio, and emergence time
2012-2016:	B.Sc. Agricultural Sciences, Biotechnology, and Food Science Cyprus University of Technology • Specialization: Plant Science and Technology • Dissertation: Identification and control of nematodes in Cypriot banana plantations. Assessment of a novel nematicide

Referees

1. Prof. Dr. Eva Stukenbrock, Environmental Genomics Group, Max Planck Institute for Evolutionary Biology, Ploen and Christian Alrbechts University, Kiel, Germany, (Ph.D. advisor). email: stukenbrock@evolbio.mpg.de, Tel: +49 431 880 6368
2. Dr. Alice Feurtey, Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, CH-2000 Neuchâtel, and Plant Pathology, D-USYS, ETH Zurich, CH-8092 Zurich, Switzerland (M.Sc. daily supervisor). Email: alice.feurtey@usys.ethz.ch, Tel: +41 44 6323 871

Publications

1. **Taliadoros D**, Stukenbrock EH. The use of evolutionary analyses to predict functionally relevant traits in filamentous plant pathogens. Curr Opin Microbiol. 2023;73:102244.

2. Spanner R, **Taliadoros D**, Richards J, Rivera-Varas V, Neubauer J, Natwick M, et al. Genome-Wide Association and Selective Sweep Studies Reveal the Complex Genetic Architecture of DMI Fungicide Resistance in Cercospora beticola. Genome Biol Evol. 2021;13(9):1–17.

3. Ebert MK, Rangel LI, Spanner RE, **Taliadoros D**, Wang X, Friesen TL, et al. Identification and characterization of Cercospora beticola necrosis-inducing effector CbNip1. Vol. 22, Molecular Plant Pathology. 2021. p. 301–16.

4. Kanetis LI, **Taliadoros D**, Makris G, Christoforou M. A Novel Seimatosporium and Other Sporocadaceae Species Associated with Grapevine Trunk Diseases in Cyprus. Plants. 2022 Oct 1;11(20).

5. **Taliadoros D,** Feurtey A, Wyatt N, Gladieux P, Friesen T, Stukenbrock H. E. Emergence and spread of the barley net blotch pathogen coincided with crop domestication and cultivation history. Available from: https://doi.org/10.1101/2023.07.28.550921 (under peer-review)

Affidavit

I hereby declare that this dissertation

- concerning content and design, is the product of my own work under the supervision of Prof. Dr. Eva Stukenbrock. I have used no other tools or sources beyond the ones cited. Contributions of other authors are listed in the section “Declaration of Author’s Contributions”.
- has been conducted and prepared following the Rules of Good Scientific Practice of the German Research Foundation.
- has not been submitted elsewhere, partially or wholly, as part of a doctoral degree to another examination body, and no other materials are published or submitted for publication than indicated in the thesis.
- No academic degree has been withdrawn

Plön, 20/11/2023

Demetris Taliadoros