Population Genomics of *Cercospora beticola*

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

Lizel Potgieter

March 2021

First examiner: Prof. Dr. rer. nat Eva Holtgrewe Stukenbrock

Second examiner: Prof. Dr. rer. Nat. Tal Dagan

Third Examiner: Prof. Dr. Irene Barnes

Date of oral examination: 13th of April 2021

Table of Contents

Summary

Fungal plant pathogens pose a serious threat to global food safety and security. Infections of crops by phytopathogens can result in significant yield loss. Fungal plant pathogens have evolved with their hosts during the history of crop domestication. While some fungal phytopathogens of modern crops have maintained the ability to infect the wild relatives of these crops, many have evolved host-specificity due to the evolutionary arms race. Co-evolution between plants and their pathogens spans many generations. Therefore, we have employed the pathosystem of the fungal pathogen *Cercospora beticola* and its hosts *Beta vulgaris* ssp. (domesticated beet) and *B. maritima* (sea beet). This pathosystem is exceptional as sugar beet has a relatively short domestication history of ~300 years, compared to several thousand year history of other modern crops. Investigating the effect crop domestication has on fungal evolution in such a short time frame may provide insight into the early processes underlying the evolution of host-specificity.

The availability of whole genome sequencing data for entire populations of fungal plant pathogens has enabled detailed analyses of genomic variation within and among field populations. Using population genomic data, we are able to detect population structure of a phytopathogenic fungus, identify regions that are highly differentiated between isolates, and predict the evolutionary trajectory of disease epidemics. The primary focus of this thesis was to describe the population genomics of the fungus *Cercospora beticola*, and determine the influence of host domestication on recent evolution and population structure of the fungus.

In **Chapter 1**, we have addressed the challenge of assembling and analysing population genomic data of species with structural variation, as is the case for many pathogenic fungi. We compared and contrasted two variant calling methods used in population genomics. We show that the commonly used method of variant calling, reference mapping-based approaches, as well as more recently adapted multiple genome alignment-based methods perform equally well at high sequencing depths in species with variable amounts of repetitive content. However, we also found that reference mapping-based approaches are reliable at average and high sequencing depths, regardless of repetitive content. We offer a perspective of using multiple genome alignment-based

approaches when considering assemblies produced from third-generation sequencing platforms, as well as with genomes where high quality assemblies can be produced with next-generation sequencing platforms in fungal population genomics studies.

In **Chapter 2**, we analyse the population genetic structure of *C. beticola* with the aim of comparing the genetic variation in populations of domesticated and wild beet species. Specifically, we make use of population genomics tools to elucidate whether *C. beticola* isolates from wild and domesticated hosts show strong signals of host specialisation. Sugar beet is comparatively novel crop, and provides insight into the early specialisation processes pathogens of domesticated plants. We collected isolates from wild and domesticated beet from Europe and North America and show that there are not clear populations of *C. beticola* isolates that infect wild or domesticated beet. We show that *C. beticola* isolates are likely a global population, with substantial admixture between individuals from all hosts and locations. While there is admixture between individuals from all locations, isolates from sea beet in the UK showed more differentiation from the isolates from other locations suggesting some barriers to gene flow and distinct population histories of the sea beet isolates. We investigated regions where the isolates from the UK are different from isolates from mainland Europe and North America, and showed that there are likely phenotypic differences between isolates from Croatian sea beet and the English sea beet isolates. We illustrate a region where the isolates from Croatia contained a premature stop codon in a gene involved in the production of an aflatoxin in high frequency, while it was present at a low frequency the isolates from English sea beet. Thus, we show that while *C. beticola* may not show strong signatures of host specialisation yet, there are some differences between isolates from different locations indicating the potential for future population divergence.

In **Chapter 3**, we compare and contrast *C. beticola* to four other *Cercospora* species to elucidate differences and similarities in genome content and synteny within the genus. We show that *C. beticola* has a higher number of genes encoding proteins that are involved host-pathogen interaction. We also note that the other *Cercospora* species that has a broad host range included in this, *C. cf. flagellaris*, has a similar repertoire of genes. We also show that these two species

share substantial synteny. We postulate that they most recent common ancestor of these two species likely had a plastic genome that underwent several translocation events.

Taken together, we show that the *Cercospora* genus is shaped by its interactions with its environment and the various hosts. We show that *C. beticola* has not yet shown strong association with either host or location.

Zusammenfassung

Pflanzenpathogene Pilze stellen eine ernsthafte Bedrohung für die weltweite Lebensmittelsicherheit dar. Infektionen von Nutzpflanzen mit phytopathogenen Pilzen können zu erheblichen Ertragseinbußen führen. Die phytopathogene Pilze entwickelten sich zusammen mit ihren Wirten während der Geschichte der Domestizierung von Nutzpflanzen. Während einige phytopathogene Pilze moderner Nutzpflanzen die Fähigkeit behalten haben, die wilden Verwandten dieser Nutzpflanzen zu infizieren, haben viele aufgrund des evolutionären Wettrüstens eine Wirtsspezifität entwickelt. Diese Koevolution zwischen Pflanzen und ihren Pathogenen erstreckt sich über viele Generationen. Daher haben wir das Pathosystem des Pilzerregers *Cercospora beticola* und seiner Wirte *Beta vulgaris* ssp. (domestizierte Rübe) und *B. maritima* (Meerrübe) verwendet. Dieses Pathosystem ist besonders, da die Zuckerrübe eine vergleichsweise kurze Domestikationsgeschichte von etwa 300 Jahren aufweist, verglichen mit der mehrere tausend Jahre langen Geschichte anderer moderner Nutzpflanzen. Die Untersuchung des Einflusses der Domestizierung von Nutzpflanzen auf die Pilzevolution in einem so kurzen Zeitrahmen kann einen Einblick in die frühen Prozesse geben, die der Evolution der Wirtsspezifität zugrunde liegen.

Die Verfügbarkeit von Daten der Ganzgenomsequenzierung gesamter phytopathogener Pilzpopulationen hat detaillierte Analysen der genomischen Variation innerhalb und zwischen Feldpopulationen ermöglicht. Mit Hilfe dieser genomischen Populationsdaten sind wir in der Lage die Populationsstruktur phytopathogener Pilze zu erkennen, sich zwischen Isolaten stark unterscheidende Regionen zu identifizieren und den evolutionären Verlauf von Krankheitsepidemien vorherzusagen. Der primäre Fokus dieser Arbeit lag darauf, die Populationsgenomik des Pilzes *Cercospora beticola* darzustellen und den Einfluss der Wirtsdomestikation auf die rezente Evolution und Populationsstruktur dieses Pilzes zu bestimmen.

In **Kapitel 1** stellten wir uns der herausfordernden Aufgabe die populationsgenomischen Daten von Arten mit struktureller Variation, wie sie bei vielen pathogenen Pilzen der Fall ist, zu assemblieren und zu analysieren. Wir verglichen zwei in der Populationsgenomik verwendete Methoden zum

Detektieren von Varianten („variant calling“). Wir zeigen, dass die am häufigsten verwendete Methode des „variant calling“, die Referenz-Mapping-basierten Ansätze, sowie neuere Multiple-Genom-Alignment-basierte Methoden bei hohen Sequenzierungstiefen in Arten mit variablen Mengen an repetitiven Elementen identische Leistungen erbringen. Wir fanden jedoch heraus, dass Referenz-Mapping-basierte Ansätze auch bei mittleren und hohen Sequenzierungstiefen zuverlässige Ergebnisse hervorbringen, unabhängig vom repetitiven Anteil des Genoms. Wir bieten daher die Möglichkeit zur Verwendung Multiple-Genom-Alignment-basierter Methoden in populationsgenomischen Studien von Pilzen unter der Berücksichtigung von Genom-Assemblierungen, die mit Sequenzierplattformen der dritten Generation erzeugt wurden, sowie Genom-Assemblierungen, die durch qualitativ hochwertige Assemblierungen des Next-Generation Sequencing erzeugt wurden.

In **Kapitel 2** analysierten wir die populationsgenetische Struktur von *C. beticola*, um die genetische Variation in Populationen von domestizierten und wilden Rübenarten zu vergleichen. Insbesondere nutzten wir die Analyse-Tools der Populationsgenomik, um aufklären, ob *C. beticola*-Isolate von wilden und domestizierten Wirten starke Signale der Wirtsspezialisierung aufweisen. Die Zuckerrübe ist eine vergleichsweise neue Kulturpflanze, die einen Einblick in die frühen Spezialisierungsprozesse von Pathogenen domestizierter Pflanzen gibt. Wir sammelten Isolate von wilden und domestizierten Rüben aus Europa und Nordamerika und zeigen, dass es keine eindeutigen Populationen von *C. beticola*-Isolaten gibt, die entweder wilde oder domestizierte Rüben infizieren. Wir zeigen, dass es sich bei den *C. beticola*-Isolaten voraussichtlich um eine globale Population handelt, mit einer erheblichen Vermischung zwischen Individuen aus verschiedensten Wirten und Standorten. Während es eine Vermischung zwischen Individuen aller untersuchten Standorte gab, unterschieden sich Isolate aus Meeresrüben in Großbritannien stärker von den Isolaten anderer Standorte, was auf Barrieren für den Genfluss und unterschiedliche Populationsgeschichten der Meeresrüben-Isolate schließen lässt. Wir untersuchten Regionen, in denen sich die Isolate aus Großbritannien von Isolaten vom europäischen Festland und aus Nordamerika unterscheiden, und zeigen, dass es vermeintliche

phänotypische Unterschiede zwischen Isolaten aus kroatischen Meeresrüben und den englischen Meeresrübenisolaten gibt. Wir veranschaulichen eine Region, in der die Isolate aus Kroatien in hoher Frequenz ein vorzeitiges Stopcodon in einem an der Produktion eines Aflatoxins beteiligten Gens enthalten, während es in den Isolaten aus englischen Meerrüben in niedriger Frequenz vorhanden ist. Wir zeigen daher, dass *C. beticola* bisher zwar keine starken Anzeichen einer Wirtsspezialisierung aufweist, dass es aber einige Unterschiede zwischen Isolaten von verschiedenen Standorten gibt, die auf das Potenzial einer zukünftigen Populationsdivergenz hinweisen.

In **Kapitel 3** vergleichen wir *C. beticola* mit vier anderen *Cercospora*-Arten, um Unterschiede und Gemeinsamkeiten im Genominhalt und in der Syntänie innerhalb dieser Gattung aufzuklären. Wir zeigen, dass *C. beticola* eine höhere Anzahl von protein-kodierenden Genen besitzt, die an der Wirt-Pathogen-Interaktion beteiligt sind. Wir stellen fest, dass eine weitere *Cercospora*-Art mit ein breites Wirtsspektrum, *C. cf. flagellaris*, ein ähnliches Repertoire an Genen aufweist. Wir zeigen zudem, dass diese beiden Arten eine erhebliche Syntänie aufweisen. Wir postulieren, dass der jüngste gemeinsame Vorfahre dieser beiden Arten wahrscheinlich ein plastisches Genom hatte, das mehrere Translokationsereignisse durchlief.

Zusammengenommen zeigen wir, dass die Gattung *Cercospora* durch ihre Interaktionen mit ihrer Umwelt und ihren unterschiedlichen Wirten geprägt ist. Wir zeigen, dass *C. beticola* bisher keine starke Assoziation mit ihren Wirten oder den verschiedenen Standort aufweist.

General

Introduction

General Introduction

Introduction

Humans have always been relying on plants for nourishment. However, plant domestication did not occur until 9500 BC with the birth of agriculturally based societies (Allaby et al., 2008). It is widely believed that localised climate change following the last ice age resulted in the transition from hunter-gatherer groups to settled villages during the Neolithic revolution (Olsson and Hibbs, 2005). Due to long dry periods, the villages were forced to stockpile seeds and tubers to ensure that food was available throughout the year. Initially only animals were purposefully domesticated with plants following about 2000 years later around 9500 BC. Concentrated domestication efforts of plants was not required as some of the first grain species were domesticated naturally through continued cultivation during the Epi-Paleolithic period (Hillman et al., 2001). Rye was the first plant to be domesticated by humans in the Epi-Paleolithic period (Hillman et al., 2001). Some of the other plant species that were domesticated later by humans, termed the Neolithic founder crops, were emmer wheat, einkorn wheat, barley, lentil, pea, chickpea, and bitter vetch (Zohary et al., 2012).

Initially it was thought that plants were rapidly domesticated rather than during prolonged selection. Phylogenetic analyses of the eight major ancient crops have contributed significantly to our understanding of crop domestication. If there was only one domestication event of each crop, the accessions would all be monophyletic. However, it has been found that several crop species have polyphyletic origins (Londo et al., 2006; Molina-Cano et al., 2005; Zohary, 1999). It has also been supported by the staggered, rather than simultaneous, appearance of agriculturally important traits (Fuller, 2007). For example, in domesticated grasses, traits for grain size and shape were selected for before non-shattering ears or panicles (Fuller, 2007).

The term domestication syndrome was coined by Hammer (Hammer, 1984) to describe the accumulation of traits favoured by concentrated domestication efforts. The genetic consequences of the domestication syndrome have been studied extensively, and with the increased application of systems biology, the true effects of are being discovered (Martínez-Ainsworth and Tenaillon,

2016; Moyers et al., 2018). Many of the traits selected for have pleiotropic effects. One of major efforts in plant breeding is to reduce the plant's susceptibility to various biotic and abiotic factors that result in yield loss. These factors include drought resistance, nutrient uptake and utilisation, and pathogen resistance as few examples.

Plant disease is a vital biotic component considered by breeding efforts due to the threat pathogens pose to global food security and safety (Fisher et al., 2012). Fungal plant pathogens are known to cause severe crop loses (Fones et al., 2020; Godfray et al., 2016). Global climate change and increased mono-culture based farming, and globalised trading pose risks to future fungal pandemics (Almeida et al., 2019; Fones et al., 2020). As such, breeding efforts towards fungal resistant and tolerant crop varieties have been underway for all major crop types (Melchers and Stuiver, 2000; Zuccaro and Langen, 2020). In many crop species, including wheat and tomato, disease resistance, as well as other favourable traits are bred into crops from wild relatives (Hajjar and Hodgkin, 2007; Harlan, 1976; Migicovsky and Myles, 2017; Pimentel et al., 1997).

Below, I will summarise our current knowledge on the effect of plant domestication on pathogen evolution. Moreover, I will provide an overview of molecular plant-pathogen interactions which are crucial in driving host-pathogen co-evolution. Finally, I will introduce the model system studied in this research, the fungus *Cercospora beticola* and its host plant species of *Beta*.

Domestication Processes Affecting Fungal Pathogen Evolution

Many modern crops were domesticated around 7000-10000 years ago (Balter, 2007). Pathogens have been associated with crops since domestication occurred (Diamond and Guns, 1997). The breeding of new crop species has resulted in the evolution of novel plant pathogens as well as significant changes in the populations that infected the wild ancestor populations (reviewed by Stukenbrock and McDonald, 2008).There are many differences between wild and domesticated plant populations that have the potential to drive evolution of pathogens of domesticated plants including, among others the use of pesticides, lack genetic variation, introduction of R-genes, and crop rotations in agricultural systems.

The agricultural landscape is very different from that of the natural ecosystems. The density of monoclonal plants is much higher in agricultural systems. In natural ecosystems, there is a considerably larger heterogeneity in ecological and genetic conditions (Figure 1). It is hypothesised that the virulence of a pathogen is an interplay between the pathogen's ability to colonise and reproduce within a host and transmission between hosts (Anderson and May, 1982; May et al., 1983; Stukenbrock and McDonald, 2008). If a pathogen is excessively virulent and kills its host prior to successful reproduction, the spread of the disease is unlikely to occur. However, if a pathogen is less virulent, the chances of successful reproduction and infection of new plants are increased. Due to the compact nature of agricultural fields, transmission of inoculum between plants is made significantly easier than it would be in wild ecosystems. There is also an abundance of genetically identical hosts that removes the selective pressure against evolving less virulent phenotypes. This, in turn, allows for the evolution of more virulent pathogens in agricultural ecosystems than in natural ecosystems (Thrall and Burdon, 1999).

Figure 1: An illustration of the environment that the domesticated beet, sugar beet (A) and the wild beet, sea beet (B) are encountered in. **1A**: Sugar beet are grown closely to one another with no genetic diversity between plants. Soils are supplemented with nutrients, pesticides are applied, and fields are irrigated regularly. Photograph was taken at an experimental sugar beet field from the British Beet Research Organisation (BBRO) in Norfolk, UK in September 2018.

1B: Sea beet are interspersed between other naturally occurring plant species. The soils are usually poor in nutrients, no unnatural irrigation occurs, and no disease control measures are taken. The photograph was taken in Orford, UK in September 2018

In addition to host density and distribution, plants grown in agricultural settings are also treated by artificial means to support their growth. These treatments include being supplemented with nutrients to maximise their output (Johnston et al., 2000). In many instances, plants grown on nutrient poor soil are more susceptible to disease than plants that are grown in soils with a suitable nutrient profile (Spann and Schumann, 2010). While a balanced nutrient profile in the soil is beneficial for plant growth, other soil-borne microorganisms can also make use of the available nutrients to proliferate, and potentially metabolise the nutrients before they can be used by the plants (Rovira, 1965). Therefore, preventative measures, such as fungicide application, are important in agriculture (Poole and Arnaudin, 2014).

The use of pesticides drives different evolutionary trajectories between pathogens in domesticated and wild host populations (Russell, 2005; Stukenbrock and McDonald, 2008). The use of pesticides provides a strong selective pressure on fungi infecting crops (Hahn, 2014). The evolution of fungicide resistance not only threatens global food security (Lucas et al., 2015). It also threatens the biodiversity of local ecosystems, and should be used with caution (McMahon et al., 2012). There are several different classes of fungicides employed in agriculture, namely benzimidazoles, demethylation inhibitors, Qo respiration inhibitors, and dicarboximides (Ma and Michailides, 2005). There are four primary modes of fungicide resistance that have evolved by fungi: (1) altering the target site of the fungicide so that binding is reduced, (2) expression of an enzyme that binds to the fungicide's target, (3) over-expression of the target of the fungicide to ensure that some products remain active, and (4) a reduction in the uptake or an active extracellular export of the fungicide (Fluit et al., 2001; Gisi et al., 2000; Gullino et al., 2000; McGrath, 2001). There are likely other undiscovered mechanisms employed by fungi to overcome the use of fungicide, and the ability can evolve rapidly (Ma and Michailides, 2005). As such, it is vital not to rely solely on fungicide use to control fungal diseases. One of the most important

contributing factors to fungal plant disease management is the use of resistance breeding (e.g. Melchers and Stuiver, 2000; Secor and Gudmestad, 1999; Singh and Reddy, 1991).

The responses to the various different environmental pressures are mediated through three primary mechanisms that act similarly under the various pressures. Fungicide resistance can arise in various ways (reviewed by Hawkins et al., 2019). The response to environmental pressures, such as fungicide resistance can arise from *de novo* mutations, standing variation in the population, and interspecific transfer of resistance genes. New mutations that give rise to fungicide resistance will increase in frequency at sites where fungicide is applied. At population level, this leaves a signature in the genome represented by a selective sweep (Messer and Petrov, 2013). Standing variation can contribute differently to fungicide resistance, depending on the frequency that the trait is present in the population (Barrett and Schluter, 2008). If the trait is present at a low frequency, selective pressures can contribute to the mutation becoming fixed in the population. If the trait is already present at a high frequency prior to the environmental change, the population can be considered to be intrinsically resistant to the environmental change (Lucas et al., 2015). Finally, fungicide resistance may be acquired through interspecific gene transfer whereby distinct species of pathogens hybridise. Hereby, one species carrying a fungicide resistance allele will transfer its resistance to another species. Genome studies have, in the recent years, documented that interspecific hybridisation occurs more frequently than previously though, and may be a relevant mechanism in the evolution of new fungicide resistance (Lucas et al., 2015; Stern, 2013).

The different evolutionary pressures on pathogens of wild and domesticated plants can result in gene flow into wild populations from populations on domesticated plants. In instances where populations are undergoing parallel evolution, alleles can flow freely between populations until a reproductive barrier between the populations evolves. This allows for changes in allele frequency in the wild population driven by diversity found in pathogens of domesticated plants that could be detrimental to wild plants. In many cases, host specialisation becomes an important reproductive barrier between populations on wild and domesticated plants, or when resistant cultivars are planted that limit a fungus from completing its life cycle (Gavrilets, 2004; Giraud et al., 2010). As

such, cross-infection between hosts cannot occur freely, and pathogens cannot undergo sexual reproduction on their non-native host (Tellier et al., 2010). This limits gene flow between the populations, and often increases population segregation. As this specialisation occurs, the plant's ability to resist or tolerate the diseases caused by the different populations changes (Mercier et al., 2019).

The distinction between disease resistance and disease tolerance is important to note in discussions regarding plant-pathogen interactions. Disease tolerance is a measure of the plant's ability to grow and yield a crop despite infection, while disease resistance pertains to the ability of the plant to resist infection by a pathogen by preventing the infection, establishment, and reproduction of the pathogen (Graham and Webb, 1991). The outcome of which is determined by the molecular interactions between the plant and the infecting pathogen.

Evolutionary Theory on the Effect of Domestication on Fungal Pathogens

There are two primary categories of theories describing the evolution of a plant-pathogen system. One category pertains to co-evolutionary dynamics, and the other to environmental factors that drive evolution. Within the category relating to co-evolutionary dynamics, there are two primary hypotheses. The first is the Red Queen hypothesis (Van Valen, 1977). This hypothesis proposes that species must constantly evolve and reproduce in the face of opponents, and ties in with the arms race discussed previously (Clay and Kover, 1996). This hypothesis is supplemented by the Black Queen hypothesis that argues that communities of microbes are able to selectively lose genetic content, assuming that other members of the community are able to supplement the function of the lost content (Morris et al., 2012). In the category relating to environmental factors, the most widely discussed theory is the Court Jester hypothesis (Barnosky, 2001). This hypothesis states abiotic factors drive evolution, rather than interactions between individuals in a system. The three hypotheses have been seen to contribute in the interaction of plants and their fungal pathogens, and the resulting signatures can be detected within genomes.

Mutations occur naturally at different rates within individual genomes, and the fates of these mutations depends on several factors. The fitness effect of a mutation influences the distribution of mutations (Eyre-Walker and Keightley, 2007). If a mutation is deleterious, but not lethal, or simply neutral, the chances of the mutation spreading through the population and becoming fixed (i.e. shared by all individuals) are lower than if the mutation has a fitness benefit (reviewed by Loewe and Hill, 2010). If the effects of a mutation are strongly deleterious, selection against it acts effectively (Keightley and Eyre-Walker, 2010; Trindade et al., 2010). If mutations are only slightly deleterious, these can accumulate within genomes through random genetic drift and non-random mating (Glémin, 2003; Kondrashov, 1995; Ohta, 1973). Mutations with a strong fitness benefit are rarer, but contribute heavily to evolution (Eyre-Walker, 2006). Mutations can become fixed by either random genetic drift, or selection.

Selective sweeps are a process by which a mutation becomes fixed in a population by means of positive selection (reviewed by Stephan, 2019). Through selective sweeps, the mutation landscape of the region may lack genetic variation due to the regions hitchhiking along with the beneficial mutation (Smith and Haigh, 1974). There are three types of sweeps: i) a hard selective sweep where the beneficial mutation rapidly becomes fixed in a population (Smith and Haigh, 1974), ii) a soft sweep from standing genetic variation driven by external factors such as environmental pressures (Hermisson and Pennings, 2005), and iii) a multiple origin soft sweep where mutations are common in a population and the same mutation occurs against different genomic backgrounds and no hitchhiking occurs (Pennings and Hermisson, 2006). Selective sweeps can be detected by large haplotypes being present in a population that result in reduced diversity at linked loci (reviewed by Booker et al., 2017).

The use of population genomics is fundamental in the identification of the genomic regions underlying such phytopathogenic phenotypes. One of the phenotypes studied is the genomic basis of host specialisation, particularly within a single fungal species that has a wide host range. In the case of *Botrytis cinerea*, several different populations exist which infect more than 1400 dicotyledonous and monocotyledonous plant species (Mbengue et al., 2016; Williamson et al.,

2007; 2016). The agriculturally important crops *B. cinerea* is known to infect include grapevine and tomato. Since isolates obtained from tomato and grapevine are known to cause more disease on their respective hosts, this system has been used extensively to study host specialisation (Mercier et al., 2019, 2020). Microsatellite markers indicate population subdivision between isolates from various hosts in France, and suggests that populations include a collection of generalists with some specialist individuals (Mercier et al., 2019). Populations of *B. cinerea* also showed differentiation between individuals isolated from greenhouses and from outdoor hosts (Walker et al., 2015). Population genomics approaches have been used to locate the genomic regions in the various populations to determine which genes are under some degree of positive selection that may be important for host specialisation (Mercier et al., 2020). Several PCWDEs are shown to be under different selective pressures in the various populations, suggesting that these are important for host specialisation. In addition, two CAZymes that are known virulence factors in other species are unique to populations from grapevine (Beauvais et al., 2013; King et al., 2017). These CAZymes are known to contribute to host recognition and fungal morphology, and thus likely contribute to *B. cinerea*'s ability to infect grapevine more effectively. In the populations that infect tomato, a phytotoxin that is present in all *B. cinerea* populations was found to undergo strong positive selection (Schouten et al., 2008). Additionally, regions containing secondary metabolite (SM) clusters were also shown to be part of a selective sweep. SMs are known to impact host range (Thynne et al., 2019). Therefore, this cluster may also be a contributing factor to host specialisation.

Taken together, the interaction between plants and their pathogens drive genome evolution under various scenarios. In order to overcome plant defenses, the fungal genome must constantly adapt to the environment and changes in the hosts.

Plant-Pathogen Interactions During Infection

Plants possess an immune system allowing them to recognise and defend against pathogens. In order to evade the plant's immune system, fungal pathogens must either avoid detection, or manipulate the immune response (Boyd et al., 2013). Plant defense responses operate on two

levels distinguished by the host plant detecting the phytopathogen, mounting a defense against the pathogen, and the pathogen managing the response (Jones and Dangl, 2006).

Upon infection, plants detect pathogen associated molecular patterns (PAMPs), resulting in a PAMP triggered immune (PTI) response within the plant (Jones and Dangl, 2006). PAMPs are often integrated components of pathogen cell walls or cell components. Plants have evolved the ability to detect PAMPs due to the conserved nature of these molecules. In fungi, chitin is one of the most commonly recognised PAMPs (Felix et al., 1993). PAMPs are recognised by pattern recognition receptors (PRRs) localised on the plant's cell surface, and trigger the PTI response (Zipfel, 2014). PRRs trigger a cascade of internal responses, one of the first in the cascade is an oxidative burst aimed at killing the invading pathogens (Torres et al., 2006). Later PTI responses include callose deposition, and the production of phytohormones including ethylene and salicylic acid (Zipfel and Robatzek, 2010).

Following the PTI response, pathogens must manipulate the cellular space to ensure successful establishment in the host (Boyd et al., 2013). To achieve this, pathogens must suppress the host immune response, and manipulate the environment to allow for successful establishment and propagation. This is often achieved by the expression of effectors. Effectors can fulfill a multitude of functions, and often defines host ranges (Kim et al., 2016). Effectors are small, cystein-rich peptides that interact with proteins or cell wall components of plants (Stergiopoulos and de Wit, 2009). The response to effectors is known as effector triggered immunity (ETI) (Jones and Dangl, 2006). Phytopathogenic fungal genomes can contain hundreds of different effector genes (Lo Presti et al., 2015). Some of effectors are small secreted proteins (SSPs) that are involved in host colonisation (Rep, 2005). A number of effectors have been functionally characterised and revealed a multitude of ways to manipulate host defenses (e.g. Kleemann et al., 2012; Lo Presti et al., 2015; Stergiopoulos and de Wit, 2009). However, for the vast majority of effector proteins, the functional relevance is still unknown.

In addition to the arsenal of effectors, phytopathogens also make use of carbohydrate active enzymes (CAZymes) during the infection process (Zhao et al., 2013). CAZymes are enzymes

targeting only carbohydrates. Plant cell walls are rich in carbohydrates, therefore the use of these enzymes to break down cell walls for cell invasion and nutrient acquisition is vital (Karr and Albersheim, 1970). There are many different classes of CAZymes, grouped based on their biological function and amino acid similarity (Henrissat, 1991; Henrissat and Bairoch, 1993; Lombard et al., 2010). The genomic composition of CAZyme content differs between different fungal phytopathogens, depending on the plants they infect (Zhao et al., 2013). Pathogens that rely on “brute force” to overcome the plant immune system tend to have a higher number of CAZymes in their genomes to successfully destroy the plant’s cell wall (Bolton et al., 2006).

In response to the pathogen expressed effectors plants launch the ETI response. The response is triggered by the expression of resistance (R) genes within the plant. The response is commonly mediated by leucine rich repeat (LRR) proteins. This R/Avr recognition pattern is often a gene-for-gene resistance system (Bergelson et al., 2001). Within-species polymorphisms between R genes are not uncommon, and are commonly non-synonymous substitions located in the exposed residues of the LRR protein (e.g. McDowell et al., 1998; Meyers et al., 1998; Parniske et al., 1997). R-genes are also present in various copy numbers within a species (reviewed by Dolatabadian et al., 2017).

The evolutionary arms race that plants and their pathogens are engaged in result in various evolutionary scenarios that influence fungal genome evolution. Phytopathogens often evolve new mechanisms to overcome the plant defense responses (Knogge, 1996). Many of these mechanisms rely on genomic variation.

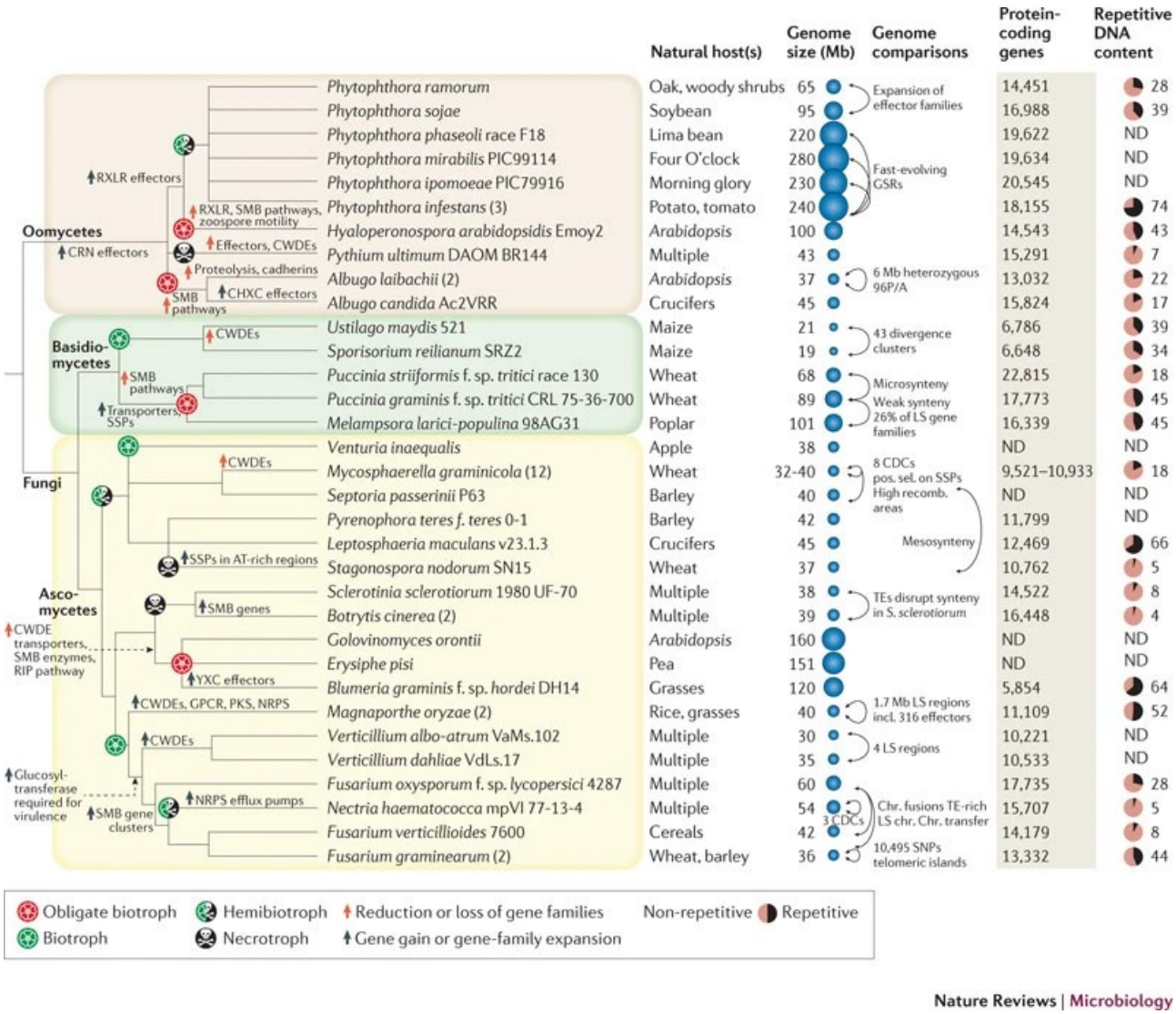
Genome Evolution in Fungal Plant Pathogens

Fungal genomes, like most genomes, are influenced by small and large scale mutational events (reviewed by Priest et al., 2020). Variable mutation rates between species as well as within genomes influences the rate of evolution. These mutational events can occur as single nucleotide polymorphisms (SNPs), and structural variants. Structural variants that include large rearrangements such as inversions, insertions, deletions, and translocations, chromosome gains

and/or losses, large-scale HGT, and polyploidy. These rearrangements can be driven by transposable elements (TEs), meiotic recombination, and double stranded break repairs during mitosis (reviewed by Priest et al., 2020).

Genome size differs drastically among fungi and oomycetes (reviewed by Raffaele and Kamoun, 2012) (Figure 2). The species considered by Raffaelle and Kamoun showed a 15-fold difference between the smallest genome used, that of *Ustilago maydis*, with a size 21 Mb and the largest 280 Mb in *Phytophthora infestans* (Haas et al., 2009; Raffaele et al., 2010). Since the review was published in 2012, several other similar comparative studies have been performed, noting similar variation in genome size, gene content, and repetitive content (e.g. Aylward et al., 2017; Mohanta and Bae, 2015; Möller and Stukenbrock, 2017). Gene gains and losses, accessory chromosomes, and repetitive content variation all drive fungal evolution in various manners.

One contributing factor to genome variation in fungi are accessory compartments, in some cases even accessory chromosomes. Accessory chromosomes occur in a variety of plant pathogenic fungi, but they are not a trait unique to fungi, and also occur in various plant and insect species (e.g. Alfenito and Birchler, 1993; Camacho et al., 1997; John and Lewis, 1966; Möller and Stukenbrock, 2017; Raffaele et al., 2010; Wilson, 1907). These chromosomes have been detected using pulsed field gel electrophoresis, and more recently with next and third generation sequencing (NGS) technologies (Croll and McDonald, 2012; van Dam et al., 2017; Weiland and Koch, 2004). With advances in genome sequencing, the possibility to identify and characterise accessory chromosomes *in silico* has become easier. The gene content of these regions is quite variable between species (Croll and McDonald, 2012). In some instances, these chromosomes contain genes that are important for successful infection of the host plant (Coleman et al., 2009; Ma et al., 2010; Miao et al., 1991). In other species, the accessory chromosomes do not contribute to successful host infection, and possibly carry negative fitness consequences (Houben, 2017). Inheritance of these chromosomes sometimes follows Mendelian patterns, while in other species they do not (Habig and Stukenbrock, 2020; Möller et al., 2018).



Nature Reviews | Microbiology

Figure 2: The review by Raffaele and Kamoun (2012) illustrates the differences between various fungi and oomycetes phytopathogens. The review shows differences in genome size, number of protein coding genes, repetitive genome content, as well as gene family expansions. The following is the key used in the review: CDC, conditionally dispensable chromosome; chr., chromosome; CNV, copy number variation; CRN, Crinkler; CWDE, cell wall-degrading enzyme; GPCR, G protein-coupled receptor; GSR, gene-sparse region; incl., including; LS, lineage-specific; ND, not determined; NRPS, non-ribosomal peptide synthetase; P/A, presence/absence; PKS, polyketide synthase; pos. sel., positive selection; rec., recombination; RIP, repeat-induced point; SMB, secondary metabolites biosynthesis; SNPs, single-nucleotide polymorphisms; SSP, small secreted protein; TE,transposable element.

TEs are important drivers in genome evolution (Kazazian, 2004). These elements are often considered to be parasitic, or selfish genetic elements that can lead to an increase in genome size (Möller and Stukenbrock, 2017; Wendel et al., 2018). TEs exist as retrotransposons (class I) that transpose themselves with an RNA transcript intermediate, while DNA transposons (class II) that are transposed without the RNA intermediate (Wicker et al., 2007). Genomes contain a varying amount of TE's from relatively few to up to 80%. The insertion of TEs can be detrimental to the genome. TE insertion can disrupt genes inherited together, alter gene expression, and promote recombination (Kent et al., 2017; Laricchia et al., 2017; Werren, 2011). Class I TEs can result in novel or pseudogenes once the mRNA intermediates are incorporated into the genome, and can influence the phenotype of the organism (Vinckenbosch et al., 2006). An important example showed that a long terminal repeat relic of a retrotransposon in a major facilitator gene promoter resulted in fungicide resistance in *Zymoseptoria tritici* (Omrane et al., 2018). Most TEs remain silent or inactive, and evolve in a neutral manner, but some may contribute to rapid genome evolution and promote adaptation (Arkhipova, 2018). The location of TEs is often spread evenly across the genome in a non-random manner (Muszewska et al., 2019). Compact genomes contain more TE remnants in genes than in genomes that have more non-genic space. Fungal pathogens also have more TEs inserted in genes than in species that are not pathogenic. SSPs are, however, generally not strongly associated with TE neighbourhoods (Muszewska et al., 2019).

The invasion of TEs is controlled by several genome defense mechanisms. These mechanisms include, among others, DNA methylation, RNA interference (RNAi), and a fungus specific mechanism termed "repeat-induced point mutations (RIP)" (John Clutterbuck, 2011; Torres-Martínez and Ruiz-Vázquez, 2017; Zemach et al., 2010). Genome defense mechanisms are crucial to maintain the integrity of the genome, as shown when these mechanisms are inactivated (e.g. Wang et al., 2020).

RNAi is a eukaryotic regulatory mechanism mediated by small non-coding RNA molecules of 20-30 nucleotides (reviewed by Dang et al., 2011). By this mechanism, incorrectly transcribed RNA is recongised by the RNA complex, converted to dsRNA, processed, then packaged into the

Argonaute proteins (Nakayashiki et al., 2006). Within the broad scope of this process, fungi show a great deal of variation in the proteins (Fulci and Macino, 2007; Li et al., 2010). In *Neurospora crassa*, RNAi is shown to be important in the post-transcriptional silencing of repetitive or homologous sequences in the vegetative or asexual state, while in the non-vegetative or sexual stage, RIP and meiotic silencing are responsible for this (Romano and Macino, 1992; Selker, 1990; Shiu et al., 2001).

DNA methylation is a genome defense mechanism found in all domains of life (Zilberman, 2008). However, some species have lost the function, including yeast (Capuano et al., 2014). TEs and other repeats are consistently methylated in species that possess the ability to methylate DNA, indicating that methylation is important for the regulation of TEs as well as gene regulation (Goll and Bestor, 2005; Henderson and Jacobsen, 2007). Methyl groups are added to adenine and cytosine nucleotides by various enzymes, including mehyltransferases DIM2 and DMNT1 that recognise CG dinucleotides and methylate the cytosine, causing the DNA strands to coil more tightly together (Jeon et al., 2015; Jullien et al., 2012; Kouzminova and Selker, 2001; Tamaru and Selker, 2001). DNA methylation silences target regions, including those that were affected by RIP (Honda et al., 2010; Lewis et al., 2010; Selker et al., 2003). DNA methylation is also rarely encountered in genes, and most in transcription activation sites, contributing to the theory that DNA methylation is a genome defense mechanism.

RIP is a homology based silencing mechanism employed in some ascomycete fungi to regulate selfish genetic elements (Selker, 1990; Selker et al., 1987). RIP is present in many, but not all fungi at different levels of activity and has only been experimentally proven in ascomycetes (Van Wyk et al., 2020). RIP targets duplicated regions larger than ~400 bp with sequence identity of higher than 80%, and introduces G:C to T:A mutations (Cambareri et al., 1991; Watters et al., 1999). Following RIP mutation, up to 30% of the C:G pairs can be mutated. RIP signatures are more prominent in the accessory compartment of the genome due to its association with TEs and other repeated sequences (van Wyk et al., 2019, 2020). RIP is present in a variety of ascomycete fungi, but many species have lost the ability to induce RIP (Horns et al., 2012; John Clutterbuck, 2011; Selker and

Stevens, 1985; Selker et al., 1987; Van Wyk et al., 2020). RIP is restricted to fungi that are actively recombining as the process is confined to the cell cycle that immediately precedes meiosis (Selker, 1990; Selker and Garrett, 1988).

Fungal genomes contain a variety of mechanisms to ensure that genome integrity remains intact. The level of redundancy seen at the various levels shows the evolutionary pressure that selfish genetic elements place on the genome. While these selfish genetic elements come with a fitness cost, they contribute to the ability of fungal pathogens to evolve rapidly in response to changed environmental conditions. Genomes are clustered into two primary compartments. The first usually contains housekeeping genes with fewer TEs, and evolves more slowly than the other compartment that has more TEs along with genes that are important for plant-pathogen interactions (Dong et al., 2015).

Due to the rapidly evolving nature of genome compartments that are enriched with TEs, new genes can be more readily gained (Dong et al., 2015). A primary genes that are enriched in such compartments are effectors. The origin and location of effectors is tightly linked with the location of TEs in gene sparse regions (Castanera et al., 2016; Dutheil et al., 2016; Faino et al., 2016; Rouxel et al., 2011; Yoshida et al., 2016). Effectors can evolve from pre-existing genes, or *de novo* mutations. Effectors originate from gene duplications more often than genes that are not effectors. Effectors often lack homologous copies in closely related species, and likely evolve from orphan genes (reviewed by Plissonneau et al., 2017). Orphan genes are genes that are exclusive to a species or a lineage, therefore it is unsurprising that these are a source of pathogen specific effectors (Tautz and Domazet-Lošo, 2011) (Figure 3). The evolution of orphan genes can occur through two different scenarios, but often involve duplication of already existing genes. New gene copies may mutate due to relaxed evolutionary pressure, and may progressively lose homology to the original gene. Eventually new functions may evolve. *De novo* evolution of gene sequences can occur by mutations in transcription factor binding sites combined with mutations that generate an open reading frame (McLysaght and Guerzoni, 2015). The second scenario is less likely than the first due to the high selective pressure on genes involved in virulence. *De novo* genes tend to be

expressed at a consistently low level throughout the cell cycle (Carvunis et al., 2012). The transcription of these genes is also tightly controlled during the various stages of plant infection (Haueisen et al., 2019; Kleemann et al., 2012; Palma-Guerrero et al., 2016; Soyer et al., 2014). This control is also important during the shifts of lifestyles, such as biotrophic and necrotrophic infection stages (Haueisen et al., 2019).

The product of effector genes results in proteins that are often secreted, and interact directly with the pathogen's host. The effects of effectors include altering cell metabolism and hormone homeostasis, disabling the host's ability to detect the pathogen, as well as lead to necrosis (Cook et al., 2015). Fungi contain hundreds of different effectors that perform these functions (Lo Presti et al., 2015). While many effectors are lineage-specific, core effectors that are present in most fungi exist (Akcapinar et al., 2015; Jonge et al., 2012; Sánchez-Vallet et al., 2015). Effectors that are recognised by the plant are quickly lost (Plissonneau et al., 2017). As a result, core effectors may be mistaken as orphans in some species.

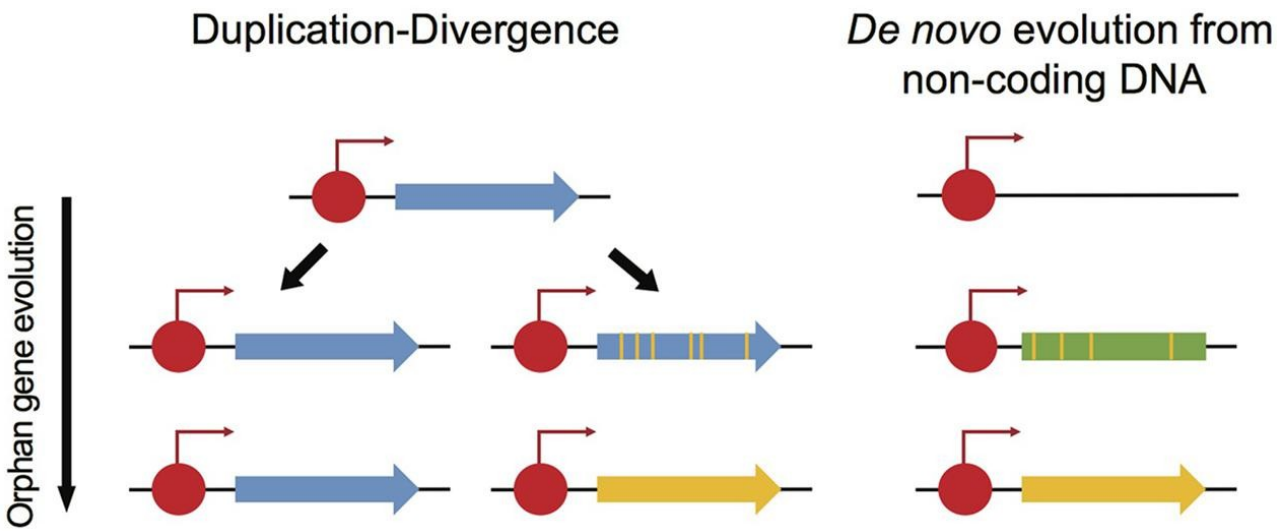


Figure 3: The two evolutionary scenarios of gene birth modified from (Plissonneau et al., 2017). In the first scenario, genes are duplicated and one copy diversifies until little to no

homology to the original gene remains while the original copy maintains its function. In the second scenario, a transcription factor binding site exists within the genome, and the downstream regions mutates to become an open reading frame that could become a viable product when expressed.

Description of Model System

The evolutionary arms race between plants and their fungal pathogens is an interplay between environmental factors, plant evolution, and the fungal ability to adapt and overcome the changes. Several well studied model systems exist, but there are many other evolutionary scenarios that have not been studied yet. Here, we use the fungal pathogen, *Cercospora beticola* of domesticated and wild beet to determine whether a view of early host speciation can be obtained. This pathosystem is of interest due to the relatively modern domestication of sugar beet compared to other extensively studied crop species. Additionally, the relationship of *C. beticola* species on wild beet and sugar beet remains unknown. Following, a description of the domestication process of beet, as well as an overview of what is known about *C. beticola*.

Beta species

Currently, 20% of the world's total sugar is extracted from sugar beet, *Beta vulgaris*, (International Sugar Organisation; https://www.isosugar.org/sugarsector/sugar, FOA 2009, http://www.eastagri.org/publications/pub_docs/4_Sugar_web.pdf). Due to climate change and resulting water shortages, it is projected that the amount of sugar produced by sugar beet, particularly in Northern Europe will increase dramatically (Jones et al., 2003). As with all plants, sugar beet, and its wild relative, sea beet, are also threatened by biotic and abiotic factors (Biancardi et al., 2011). Understanding the evolution of the pathogen and its association with its hosts in relation to how other pathogens interact and evolve with their hosts may allow predictions of future evolutionary trajectories and disease emergence.

The evolutionary history, as well as the domestication history of beet is well studied. The genus of *Beta* is a member of the Betoideae sub-family that diverged from the ancestral family of

Chenopodiaceae 38 to 27 million years ago based on molecular dating (Hohmann et al., 2006). Various beet species have been associated with humans for thousands of years (Biancardi et al., 2011; Olmo and Simmonds, 1976). The domestication of beet occurred around 8000 B.C after the knowledge of plant domestication traveled with ships from the fertile crescent to the Mediterranean. There, wild beets were domesticated in high altitudes in cool growing climates. Following domestication, a number of varieties arose, and to this day, a high level of diversity of beet species exists in the Mediterranean. Mitochondrial DNA analyses have shown that there was a unique ancestor to all cultivated beets that is no longer in existence (Santoni and Bervillé, 1992). Even after domestication of beet, wild beet are still commonly collected for food.

The distribution of beet remnants throughout Europe aids in dating the domestication and spread of the crop. In regions that are far from the native distribution range of wild beet, it is assumed that beet was introduced with agriculture. The anatomical similarities between the wild and various cultivated beets makes it very difficult, almost impossible to distinguish the wild from the domesticated beets. As such, traces of pollen, seeds, and roots within excavation sites have been used to date the distribution of beet throughout mainland Europe (Kubiak-Martens, 1999, 2002; Robinson and Harild, 2002; Voigt et al., 2008).

The closest extant wild relative to the original wild beet that beets were domesticated from is sea beet, or wild beet (Figure 4). Sea beet is known to be able withstand very stressful conditions including high soil salinity, drought, and nutrient deficiencies (Shaw et al., 2002). Sea beet can therefore out-compete neighbouring plants in harsh conditions (Biancardi and De Biaggi, 1979). In comparison, sugar beet is grown under comparably milder conditions. Unlike sea beet, sugar beet requires soils with high nutrient content, and the ability to store a lot of moisture (Rolph, 1917).

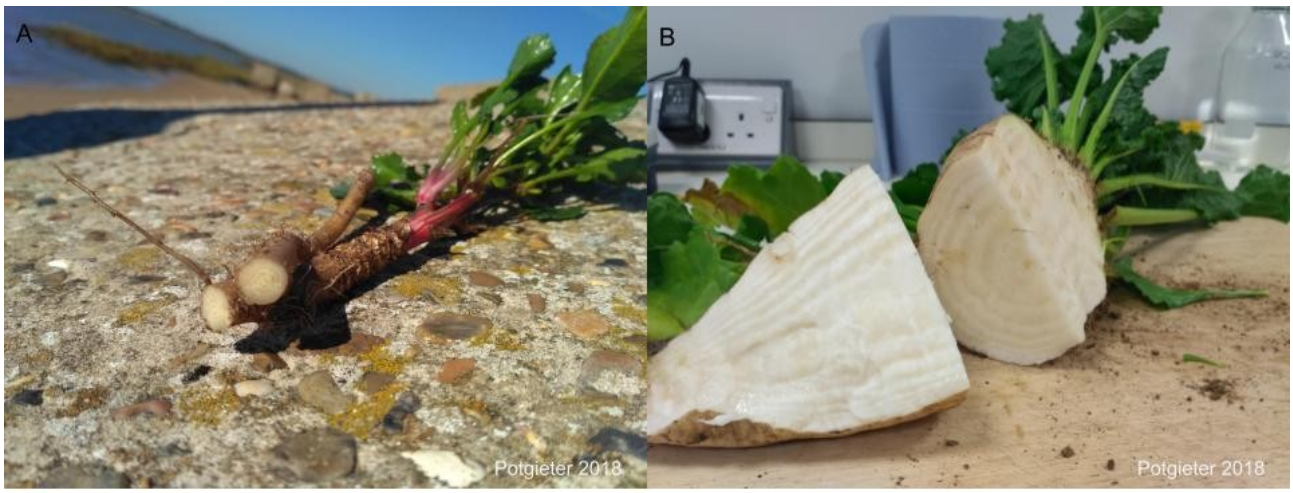


Figure 4: The roots of sea beet (**A**) and sugar beet (**B**). As both species are very closely related, the structure of the root remains consistent. However, sea beet roots contain significantly less sucrose than that of sugar beet, and are generally smaller.

The different growth conditions have an effect on microbes associated with the plants. Seeds of sugar beet can transmit microbes between generations (Dent et al., 2004). Comparative microbiome studies have shown that sea and sugar beet share a core microbiome associated with the genus, with some differences at higher taxonomic levels (Zachow et al., 2014). It has been proposed that these differences contribute to the survival of the plant, however this hypothesis remains to be tested. The microbiome of sugar beet was shown to consist of species with higher antagonistic potential against pathogens than sea beet (Zachow et al., 2014). This may be due to the locations that sea and sugar beet grow in as the harsher conditions that sea beet usually grow in have lower pathogen pressure (Biancardi et al., 2011). Plants grown in an agricultural milieu also lack the protection that the genetic diversity provided by natural ecosystems (Latz et al., 2012). The microbes associated with sea beet were, however, able to withstand more drastic conditions such as high salt concentrations and drought, similar to what sea beet can endure (Biancardi and De Biaggi, 1979; Shaw et al., 2002; Zachow et al., 2014).

Despite their differences, introgression between domesticated, wild, and weedy *Beta* plants has become a textbook case to illustrate gene flow between domesticated and wild plants (Arnaud et al., 2010; Bartsch, 2010; Ellstrand et al., 2013; Van Geyt et al., 1990). In sea beet populations that

grow near domesticated beet, alleles that are common in domesticated beet but rare in wild beet, were common (Bartsch et al., 1999). Although there was hybridisation between wild and domesticated populations, the genetic diversity of the wild beet population was not affected. There are no reproductive barriers between sea and sugar beet facilitating spontaneous hybridisation events (Abe and Shimamoto, 1989; Bartsch, 2010). This hybridisation is sometimes problematic when it comes to controlling bolting traits (Arnaud et al., 2010). Bolting is controlled by the B locus that is recessive (*bb*) in sugar beet. Breeders have been selecting against plants carrying the dominant allele (*B*) to maximise the vegetative growth stage of beets, ensuring increased yield (Abegg, 1936). Beets carrying the dominant *B*-allele reproduce without vernalisation within their first year of growth when exposed to long day conditions, while beets carrying the recessive *b*-allele require vernalisation to reproduce before being exposed to long day conditions (Abe et al., 1997; Bell and Bauer, 1942; Owen, 1954; Owen et al., 1940). Sea beets contain high levels of variation at the *B*-locus compared to sugar beet populations (Dohm et al., 2014).

Comparative genomic analyses between sugar and sea beet genomes shows that sugar beet accessions have regions in the genome without diversity, so-called “variation deserts” (Dohm et al., 2014). One of the genes contained in one of these regions is the *B* locus. Once beets have bolted, their roots become woody, potentially damaging farming equipment, and their sucrose yield is reduced. Seedlings of hybrids are impossible to distinguish from non-hyrids, and can easily cross with one another into introgressed wild weed populations that survive well in sugar beet fields (Arnaud et al., 2010; Bartsch, 2010; Ellstrand et al., 2013).

Introgression can also occur from the domesticated species into wild beets. Instances of gene flow from sugar beet to sea beet have been reported in Denmark and Southern Europe (Andersen et al., 2005; Arnaud et al., 2009). Most sugar beet propagation takes place in Northern Italy and Southern France due to favourable climate conditions for sugar beet maturation and harvest. More recently, also Denmark has become a significant contributor to sugar beet seeds (Andersen et al., 2005; Bornscheuer et al., 1993). In Southern California, the wild beet species *B. metacarpa* may have evolved from feralised populations of *B. vulgaris* ssp. v*ulgaris* and hybridised with *B. maritima*

species (Bartsch and Ellstrand, 1999). These instances are important to note as the receiving gene pool may undergo evolutionary changes due to the gene flow, and potentially undergo extinction if the population is too small (Levin et al., 1996). In addition to the biodiversity concerns, the use of genetically modified organisms may result in unnatural alleles being introduced into wild populations (Bartsch et al., 2003).

While unintentional hybridisation between sea and sugar beet is potentially problematic, guided crosses have many positive aspects for agriculture. Sea beet is often a source for the introduction of resistance to diseases in domesticated beet (e.g. Skaracis and Biancardi, 2000; Stevanato et al., 2001; Yu et al., 1999). While resistance can be bred into sugar beet, the inheritance of these traits between generations of sugar beet is relatively low (Smith and Ruppel, 1974). The genetic differences and similarities between domesticated and wild beet allow for comparisons of pathogens of the various populations to establish whether these drive pathogen evolution, or not.

The Ascomycete Genus *Cercospora*

The *Cercospora* genus contains 650 species all of which are associated with plants (Groenewald et al., 2013). Most species are causal agents of leaf spot diseases, but some are associated with lesions on fruit, seeds, and flowers (e.g. Silva and Pereira, 2008). The phylogeny of the *Cercospora* genus has been difficult to establish (Goodwin et al., 2001; Groenewald et al., 2013). Phylogenetic studies have used conserved marker regions, such as the internal transcribed spacer (ITS) region, β-tubulin, and actin. However, these regions do not contain sufficient informative sites per gene to construct a reliable phylogeny, and combinations of multiple genes have been employed to reconstruct phylogenetic relationships (Groenewald et al., 2013).

Cercosporin is a photoactivated toxin produced by a many *Cercospora* species (Daub, 1982; Daub and Ehrenshaft, 2000). Due to the widespread nature of the toxin in the genus, it was postulated that all *Cercospora* species produce cercosporin. However, as more Cercospora species were described, this was found to be inaccurate (Assante et al., 1977; Goodwin et al., 2001; Groenewald et al., 2013). Cercosporin is a perylenequinone toxin that, once activated by light, interact with

oxygen to produce activated oxygen species such as singlet oxygen that destabilises cell membranes (Daub and Briggs, 1983; Daub and Chung, 2007; Valenzeno and Pooler, 1987). The destabilisation happens when lipids in the cell membrane are peroxidated, leaving the membrane porous, and eventually results in cell death. Cercosporin has been shown to play an important, and often vital role in the successful infection of plants by various Cercospora species (Daub and Ehrenshaft, 2000). The deletion of the cercosporin gene cluster results in significantly reduced disease suggesting that in some species it is not critical for infection, merely one of several virulence factors (Choquer et al., 2005). Cercosporin production is inhibited beyond 30^{o}C, and is additionally influenced by environmental factors such as pH, carbon:nitrogen ratio, as well as the source of the carbon and nitrogen elements (Daub and Ehrenshaft, 2000).

Cercosporin is a universal toxin, but *Cercospora* species that produce the toxin are themselves immune to its effects (Daub et al., 2005). Other fungi that also produce perylenequinones are also resistant to cercosporin, as well as some bacteria including *Xanthomonas campestris* pathovar *zinniae* (Chen et al., 2007; de Jonge et al., 2018; Taylor et al., 2006; Upchurch et al., 2002). Resistance is conferred by an oxidoreductase (Taylor et al., 2006). Cercosporin is accumulated by the protein expressed by the cercosporin facilitator protein (*CFP*) gene (Callahan et al., 1999). CFP has high similarity to other known membrane transporters of the major facilitator superfamily (MFS). These proteins are presumed to be important in the export of cercosporin from the mycelium. The deletion of *CFP* gene also results in heightened sensitivity towards cercosporin in Cercospora species, and introduction of the gene introduces resistance in other species (Upchurch et al., 2002). The vitamin B6 pathway has also shown to confer resistance with an antioxidant activity (Denslow et al., 2005; Ehrenshaft et al., 1999).

The Molecular Interaction of *Cercospora beticola* and its *Beta* Hosts

As with many *Cercospora* species, *C. beticola* has a wide host range (Chupp, 1953; Groenewald et al., 2006a). *C. beticola* has been described to form lesions on a variety of domesticated beet species including *B. vulgaris* subspecies such as sugar beet, table beet, fodder beet, and Swiss chard (Pool and McKay, 1916; Vestal, 1933). Importantly, *Beta maritima*, commonly known as sea

beet, is also commonly afflicted with CLS. Agriculturally important plants outside of the *Beta* genus are also affected by *C. beticola*, including safflower, lettuce, and spinach (Hotegni Houessou et al., 2011; Lartey et al., 2005; Vestal, 1933). Several species in the genera *Acanthus, Amaranthus, Apium, Atriplex, Chenopodium, Chrysanthemum, Cycloloma, Goniolimon, Limonium, Malva*, and *Plantago* were also shown to be susceptible to *C. beticola* (Bobev et al., 2009; Groenewald et al., 2006a; Jacobsen and Franc, 2009; Rooney-Latham et al., 2010). Cross infection studies of *C. beticola* on a variety of plant hosts confirmed the ability of the fungus to infect the broad host range, but with varying degrees of success (Knight et al., 2019a). It is, however, sonsidered that *C. beticola* can infect some of its alternate hosts following insect damage to the leaf surface, or in combination with other lesion forming fungi (Knight et al., 2019a).

The *C. beticola* inoculum that is responsible for sugar beet infection can take various forms (Figure 5). Stromata overwinter in the soil on leaf debris after a growing season, or the fungus grows as vegetative saprophytic mycelia (Weiland and Koch, 2004). *C. beticola* can survive in this saprophytic manner between 10 and 22 months (Khan et al., 2008). Once conditions are favourable, conidiation results in spores that are transferred to leaves by water splash or wind (Lawrence and Meredith, 1970; Meredith, 1967; Pool and McKay, 1916). Healthy plants can then be infected by wind-borne conidia, and stromata from other plants (Khan et al., 2008; Skaracis et al., 2010; Tedford et al., 2018). Recently, it has been shown that *C. beticola* can also be detected in seeds produced from infected plants (Knight and Pethybridge, 2020). In an agricultural setting, inoculum is also spread across fields by human activity that includes moving diseased plant material and infected machinery (Knight et al., 2018, 2019b). The introduction of *C. beticola* inoculum from external sources has been responsible for CLS epidemics (Vaghefi et al., 2017a).

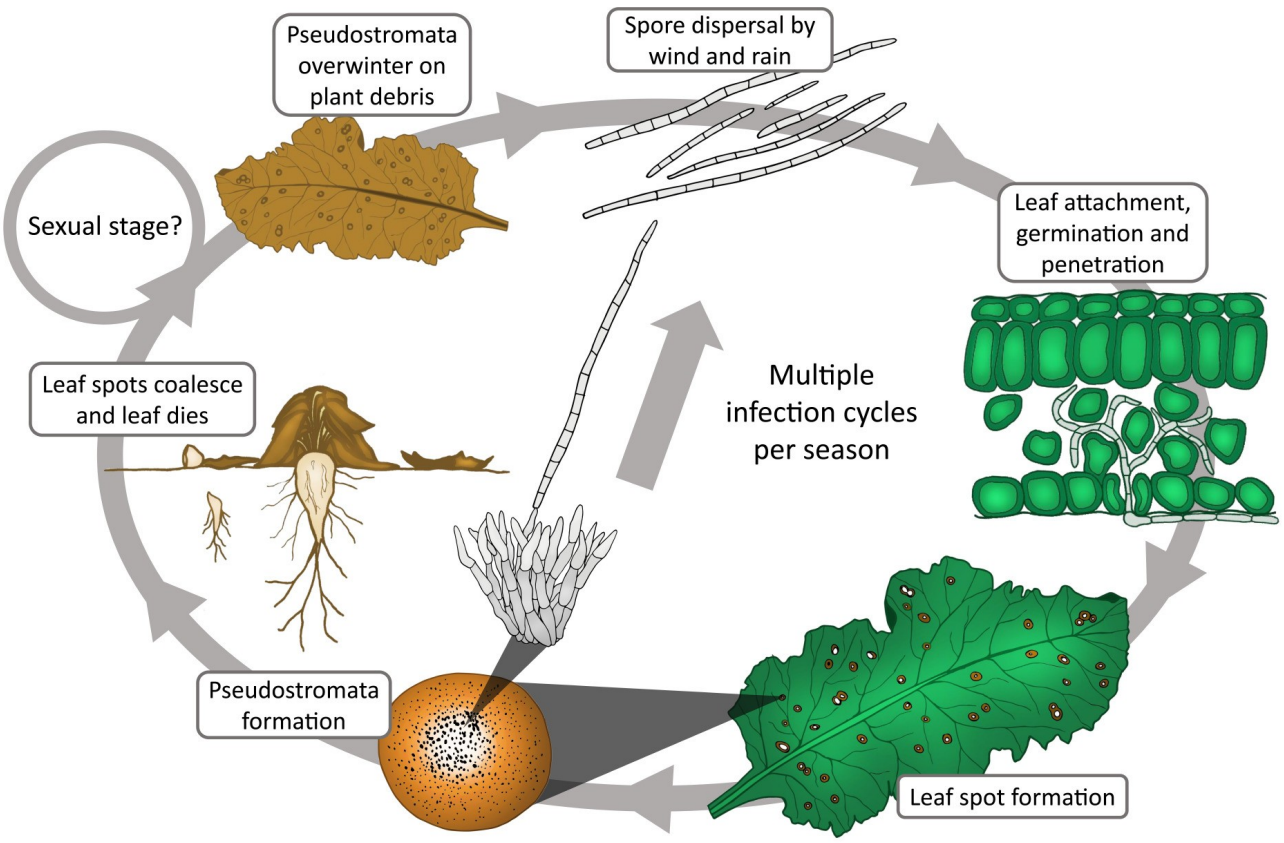


Figure 5: The disease cycle of *C. beticola* on sugar beet (figure obtained from Rangel et al. (2020)). Infection begins by dispersed conidia that penetrate the leaf surface through the stomata. This is followed by intercellular hyphal growth. Seven days after infection leaf spots start forming. Pseudostromata develop within the lesions, and produce spores via asexualbreproduction. These pseudostromata can result in multiple infections per growing season, as well as overwinter in the soil to serve as inoculum for the next growing season. The sexual stage of *C. beticola* has never been observed in the field

Management of CLS is vital for successful sugar beet cultivation. Strategies include the use of crop rotation, fungicides, and resistance breeding. To reduce the risk of inoculum from crop remains, a three year rotation period is often used in sugar beet fields (Pethybridge et al., 2018). Physical distance between fields may also be advantageous to reduce the effects of CLS dispersal among fields.

The lifecycle of *C. beticola* predisposes it to be an efficient pathogen. Isolates are able to undergo multiple asexual cycles within a single growing season. In populations in some locations, the two different homothallic individuals exist within single populations, in other of heterothallic populations the various mating type (*MAT*) genes are present in almost equal ratios, while in other populations these ratios deviate from the expected 1:1 ratio (Bakhshi et al., 2011; Bolton et al., 2012; Groenewald et al., 2006b, 2008). However, the ancestral state of the mating system in *Cercospora* was likely heterothallic due to the fragments of the *MAT* locus interspersed randomly across the genome (Bolton et al., 2014). While no sexual cycle has been observed in *C. beticola*, possibly due to the loss of the ability to form a teliomorph, population genomics studies have suggested the widespread occurrence of sexual recombination (Goodwin et al., 2001; Groenewald et al., 2006b, 2008; Knight et al., 2018, 2019b).

Population studies on *C. beticola* have been conducted using several population genetics and genomics approaches. Population genetics approaches that have been used prior to the widespread use of whole-genome sequencing, amplified fragment length polymorphisms (AFLPs), microsatellites, and SNPs in single genes were used (Groenewald et al., 2007; Turgay et al., 2010; Vaghefi et al., 2017b). The microsatellites developed for population genetic studies have since been widely applied for clone correction, a step necessary for several population genetic analyses. Earlier population genetics studies were focused on providing evidence for migration, sexual recombination, and the mutational landscapes (Rangel et al., 2020). These studies showed high genetic diversity that coincided with the high phenotypic diversity among isolates (Moretti et al., 2004; Ruppel, 1972). More recent studies based on population genomic sequencing have confirmed the gene flow between *C. beticola* isolates form Europe and North America, indicating that there is a long dispersal range for *C. beticola* (Vaghefi et al., 2017b). *C. beticola* populations in North America were determined to be more diverse than those in Europe. This leaves much to be discovered in terms of population genomics in terms of global population dynamics, direction of gene flow, and whether hosts exert different evolutionary pressures on the fungus.

Beta vulgaris and *Cercospora beticola* as a Model System

This model system is interesting out of both a biological and a historical point of view. As sugar beet is a relatively modern crop compared to other domesticated beets, and other widely used crops, studying the host-pathogen interactions between sea beet, sugar beet, and *C. beticola* would be of great interest. Due to the recent domestication of sugar beet, many of the evolutionary processes that have occurred in more ancient agrosystems, may not have occurred in this system. It also allows for the unique opportunity to track the recent evolution of a pathogen associated with wild and domesticated hosts. There have been several studies on other domesticated plants that have a longer domestication history than sugar beet, their wild relatives, and their associated fungal pathogens. Much of this research is driven by the need to understand the mechanisms that confer disease tolerance or resistance in wild populations. Understanding how host domestication drives evolution allows concentrated breeding efforts to allow for resistant plants, as well as potential fungicide development to inhibit the growth of pathogens. Further, this systems allows for the characterisation of gene flow between populations in managed and wild ecosystems.

Scope of the Thesis

Chapter 1

On variant discovery in genomes of fungal plant pathogens

Fungal pathogens often exhibit a large extent of structural variation in their genomes. This presents a challenge to the analyses of genome-wide polymorphisms, including structural variants. There are two primary approaches to calling variants, namely reference-based mapping and multiple genome alignment-based methods. Currently, reference-based mapping approaches are most frequently used in fungal population genomics. Multiple genome alignment approaches can be used to determine structural variants as well as SNPs without using additional tools, and may be a viable replacement for reference-based mapping approaches. Therefore, the main objectives of this study are:

- Determined the difference in accuracy between the reference-based mapping approach and the multiple genome alignment approach
- Quantify the influence genomic repeat content has on the reliability of variant calling methods
- Establish the effect read depth has on the reliability of variant calling methods
- Recover structural variants using the multiple genome alignment approach

Chapter 2

Population genomics of *Cercospora beticola* suggests that recent host domestication has not influenced genome evolution

Cercospora beticola is the causal agent of Cercospora Leaf Spot in domesticated and wild beet in Europe and North America. Sugar beet is a very recently domesticated crop, and therefore its pathogens may not have had enough time to undergo host specialisation. However, there was

more differentiation between isolates from sea beet in the UK and the remaining isolates from mainland Europe and North America from various domesticated beet and sea beet. This study aimed to:

- Asses the genetic diversity of *C. beticola* isolates found on sea beet in the UK
- Determine if location or host domestication influences population structure
- Identify regions that differentiate *C. beticola* from different hosts and locations
- Describe the biological relevance of the differentiated regions

Chapter 3

Genome Content of the *Cercospora* Genus Predicts Host Range

In chapter 2 we showed that the recent host domestication of *Cercospora beticola* has not yet driven clear populations to form between isolates from wild and domesticated hosts. The *Cercospora* genus contains several species that are host specific as well as many species that have a wider host range. The carbohydrate active enzyme (CAZyme) and effector repertoires of these specialist and generalist *Cercospora* species have not been compared. Additionally, the *Cercospora* genus has proven to be challenging to establish a reliable phylogeny. To investigate this, we aimed to:

- Establish a species tree for the *Cercospora* genus
- Characterise CAZyme and effector repertoires of included *Cercospora* species
- Compare and constrast the CAZyme and effector repertoires of two generalist *Cercsopora* species to those of specialist species
- Describe the genome-wide synteny between the two generalist species


Chapter 1

On variant discovery in genomes of fungal plant pathogens

Chapter 1: On variant discovery in genomes of fungal plant pathogens

Lizel Potgieter[1,2], Alice Feurtey[1], Julien Y. Dutheil[3*], Eva H. Stukenbrock[1,2*]

[1]Environmental Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany, [2]Environmental Genomics, Christian-Albrechts University of Kiel, 24118 Kiel, Germany, [3]Molecular Systems Evolution Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany

* Corresponding authors

Julien Y. Dutheil, Research group Molecular Systems Evolution, Max Planck Institute for Evolutionary Biology, August-Thienemann-Str. 2, 24306 Plön, Germany, Email: dutheil@evolbio.mpg.de, phone: +49 (0) 4522 763 298

Eva H. Stukenbrock, Environmental Genomics, CAU Kiel, Am Botanischen Garten 1-11, 24118 Kiel, Germany, Email: estukenbrock@bot.uni-kiel.de, phone: +49 (0) 431 880 6368

Keywords

Population genomics, fungal pathogens, next-generation sequencing, genome alignment, variant calling, genome assembly

Abstract

Comparative genome analyses of eukaryotic pathogens including fungi and oomycetes have revealed extensive variability in genome composition and structure. The genomes of individuals from the same population can exhibit different numbers of chromosomes and different organization of chromosomal segments, defining so-called accessory compartments that have been shown to be crucial to pathogenicity in plant-infecting fungi. This high level of structural variation confers a methodological challenge for population genomic analyses. Variant discovery from population sequencing data is typically achieved using established pipelines based on the mapping of short reads to a reference genome. These pipelines have been developed, and extensively used, for eukaryote genomes of both plants and animals, to retrieve single nucleotide polymorphisms and short insertions and deletions. However, they do not permit the inference of large-scale genomic structural variation, as this task typically requires the alignment of complete genome sequences. Here, we compare traditional variant discovery approaches to a pipeline based on de novo genome assembly of short read data followed by whole genome alignment, using simulated data sets with properties mimicking that of fungal pathogen genomes. We show that the latter approach exhibits levels of performance comparable to that of read-mapping based methodologies, when used on sequence data with sufficient coverage. We argue that this approach further allows additional types of genomic diversity to be explored, in particular as longread third-generation sequencing technologies are becoming increasingly available to generate population genomic data.

Introduction

Comparative genome studies of fungal and oomycete pathogens have revealed highly variable genome architecture and content [reviewed by Raffaele and Kamoun, 2012; Möller and Stukenbrock, 2017]. The genome size and ploidy level of pathogenic fungi and oomycetes can vary significantly between individuals of the same species. Differences can be attributed to the dynamics of transposable elements, chromosome instability, and genome compartmentalization (Möller and Stukenbrock, 2017). Fungal genomes are known to contain accessory compartments that are thought to be relevant for rapid evolution of phytopathogens [reviewed by Croll and McDonald, 2012; Möller and Stukenbrock, 2017]. Typically, these compartments contain a lower density of genes than the core genome, and have a higher content of repetitive elements (Coleman et al., 2009; Ma et al., 2010). Rapidly evolving genome compartments were shown, in some species, to encode virulence determinants (e.g. Does et al., 2016). However, in spite of their functional importance, it is challenging to analyze genetic variation in these regions due to the high extent of structural variability of the genomic sequences.

Population genomic datasets based on next generation sequencing (NGS) can be used to recover genomic variants such as single nucleotide polymorphisms (SNPs), insertions and deletions (indels), and structural variants (SVs). The latter category includes translocations, inversions, duplications, either tandem or interspersed, deletions, and novel sequence insertions (Alkan et al., 2011). Two different frameworks are traditionally used for the detection of variants (Mahmoud et al., 2019). Firstly, a reference-based approach, whereby short read data generated from NGS is mapped on a reference genome, is used to recover SNPs and short indel variants (Horner et al., 2010; El-Metwally et al., 2013). Secondly, from whole genome alignments based on de novo assembled genomes. The recovery of small structural variants from short read mapping makes use of mapping distance and orientation information of the reads, as well as read depth and pair-end discordance (Chen et al., 2009; Rausch et al., 2012; Layer et al., 2014). State-ofthe-art methods further use a local assembly of the identified inserted material (e.g. (McKenna et al., 2010; Rimmer

et al., 2014). Conversely, recovery of large-scale structural variants is typically achieved by first assembling individual genomes, which are then combined into a whole genome alignment (WGA) (Tian et al., 2018). The WGA enables the accurate location of large indels (typically larger than 3 kb) (Nattestad and Schatz, 2016; Tian et al., 2018).

While typically used to compare distinct species, if applied at the population level, WGAs potentially provide a crucial resource to conduct population genomic analyses in species with a significant proportion of structural variation since they can, in principle, capture both large and small variants (Faino et al., 2016). However, methods for calling variants in populations from WGAs are currently limited and the available approaches have not been benchmarked with fungal genome data. In this study, we take the first step to compare variant discovery approaches for population genomic analyses of fungal pathogen genomes. We assess the accuracy of a pipeline based on de novo genome assembly followed by whole genome alignment (referred to as dnWGA, Figure 1) to simultaneously recover single nucleotide polymorphisms (SNPs) and large structural variants known (Wu et al., 2017). The resulting called positions can be subsequently classified into one of four categories: (1) correctly identified variant positions [true positives (TP)], (2) correctly identified non-variant positions [true negatives (TN)], (3) variants incorrectly called in non-variant positions [false positives (FP)], and (4) variable positions that were not identified by the calling method [false negatives (FN)] (Rosner, 2006). The proportion of variants falling in each of these categories allows to compute several measures of performance (Goutte and Gaussier, 2005). Hereby “precision” is defined as the proportions of correctly inferred positives (TP/(TP + FP)), while the “recall” measure denotes the proportion of variable positions that were recovered (TP/(TP + FN)). Like many classification procedures, most variant calling methods are subject to a trade-off between precision (the higher the precision value, the more confident we can be in the prediction), and recall (the higher the recall value, the more exhaustive the variant discovery is). The performance of a given method along this trade-off can be captured by the F1 score, defined as the harmonic mean of the precision and recall values:

$$F1 = 2 * (recall * precision)/(recall + precision)$$

The F1 score is, therefore, a global measure of the reliability of the variant discovery method (Goutte and Gaussier, 2005). Several studies have demonstrated that the data used for benchmarking of variant callers is critical (e.g. Hwang et al., 2015; Sandmann et al., 2017; Wu et al., 2017; Bian et al., 2018). Notably, human population genomic data have been considered producing well-defined benchmarking tools, including the “Genome in a Bottle” project that has published a set of high-confidence variants for a reference genome (see1 : Hwang et al., 2015). Since fungal pathogen genomes differ from human genomes in many aspects, we here aimed to compare variant calling approaches on data sets specifically mimicking the characteristics of fungal pathogen genomes, including accessory genome compartments and high nucleotide diversity.

Method Overview

To compare the performance of dnWGA and referencebased mapping for variant calling, we generated population genomic data sets from chromosomal sequences of two fungal plant pathogens, Cercospora beticola and Zymoseptoria tritici using simulations (see Supplementary Methods for a detailed description of methods and materials). We selected these two different species with distinct repeat content, since repeats are known to hamper the variant calling process. While the C. beticola chromosome was virtually deprived of repeats (0.2% of 5.8 Mb) (de Jonge et al., 2018), a comparatively high proportion [11% of 6.2 Mb (Grandaubert et al., 2015)] is annotated in the chromosome of Z. tritici. We employed the chromosomes to simulate a population genomic data set that resembled empirical population genomic data. The genetic diversity of the simulated populations, measured by Watterson's theta, was 0.0077 and 0.0073 for C. beticola and Z. tritici, respectively (Watterson, 1975). The simulated population data sets comprised SNPs, indels, and accessory genome segments at known positions allowing us to evaluate the variant recovery (Figure 1).

We simulated NGS reads from the simulated genomes with both low (25X) and high (100X) sequencing coverage. To compare the efficacy of SNP discovery methods on each of the four data sets (two species, and two depths of coverage), we computed the recall, precision, and F1 scores in each case. We further assess whether the large SV were properly recovered by the dnWGAs. Details on the data generation and analyses are provided in the Supplementary Text.

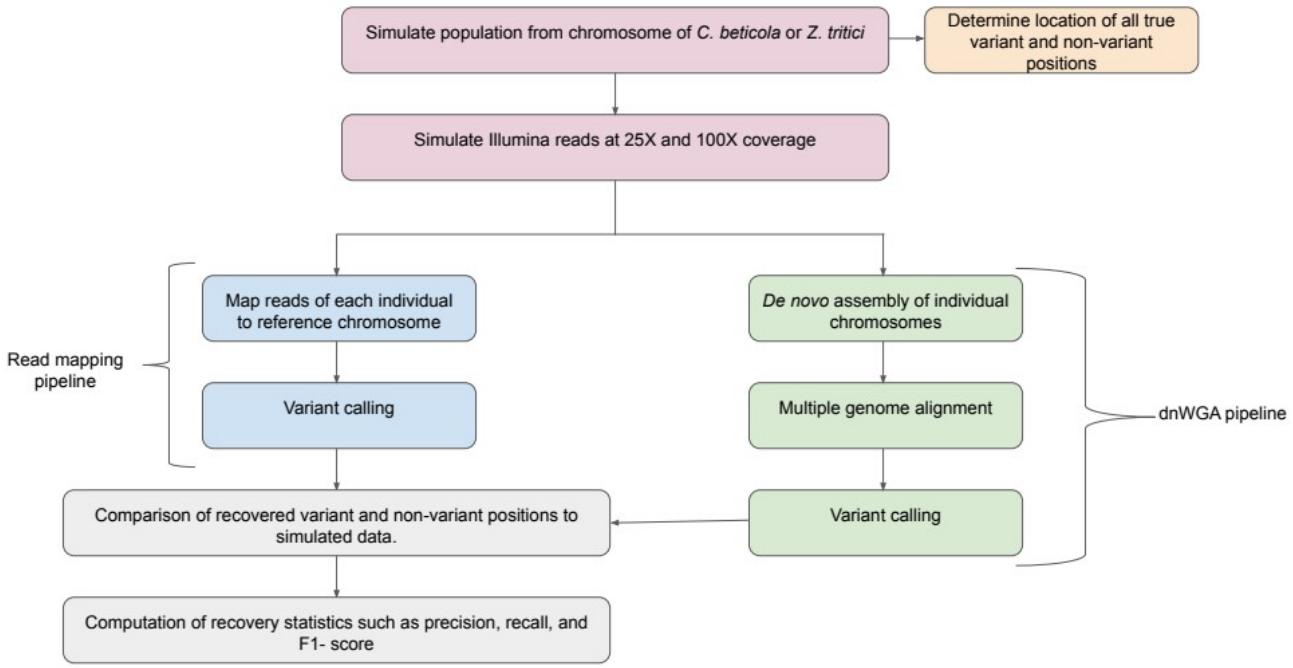


Figure 1: Overview of the pipelines used for the compared approaches. Samples of 20 individuals were simulated from *C. beticola* and *Z. tritici* chromosomes. The simulated chromosome samples were used to establish a set of true variant and non-variant positions. Reads were simulated from the simulated chromosomes at both low (25X) and high (100X) read depth. The reads were then processed by the read mapping pipeline and the dnMGA pipeline. The variant and non-variant positions recovered by each approach at each read depth were then compared to the known introduced variants, and the recovery statistics computed.

Results and Discussion

While WGAs are used to infer structural variation, how well they can recover single-nucleotide variation has not been systematically tested in fungi. We first set out to compare the performance of SNP recovery of dnWGA and reference-based approaches. We specifically ask how the sequencing depth and repeat content of the genomes affect the relative performance of the two methods: the F1 score of the reference-based approach was found to be higher than 99.7% for both Z. tritici and C. beticola, at low (25X) and high (100X) coverage. The F1 score of the dnWGA approach, however, depends on the sequence depth and the species (Figure 2A). When using high coverage data, the F1 score in both species reaches 99.9%. When low coverage sequencing was used, however, the F1 score was found to be similarly high (99.9%) for C. beticola, but only 43% for the Z. tritici data set. This effect is essentially due to the precision getting as low as 28%, while the recall value remains comparatively high (93%), suggesting that the false positive rate is high for the repeat rich chromosome with low sequencing coverage (Supplementary Table S1).

We further investigated the drop of performance at low coverage of the dnWGA approach in the repeat-rich Z. tritici data set by comparing the genome assemblies. N50 was equal to 219 kb with the 100X data set, but only 12 kb when using a 25X read depth (Supplementary Tables S2, S3). In comparison, the de novo assemblies of the C. beticola chromosomes showed a comparable N50 of 1.8 Mb and 1.7 Mb at 100X and 25X, respectively (Supplementary Tables S4, S5) underlining the impact of high repeat content in Z. tritici on chromosome assemblies. We used Quast (Gurevich et al., 2013) to further quantify the accuracy of the assembled genomes and identify misassemblies, defined as regions of the de novo assemblies that did not align to the original chromosome at the correct positions. In the repeat poor C. beticola data set, the number of misassemblies remained comparable for both sequencing depths. For the Z. tritici data set, however, we find four times more misassemblies in the 25X than in the 100X data. Therefore, we conclude that the low performance of the dnWGA procedure at low sequencing coverage is essentially due to failure of de novo assembling the chromosome sequences in the presence of a higher repeat content.

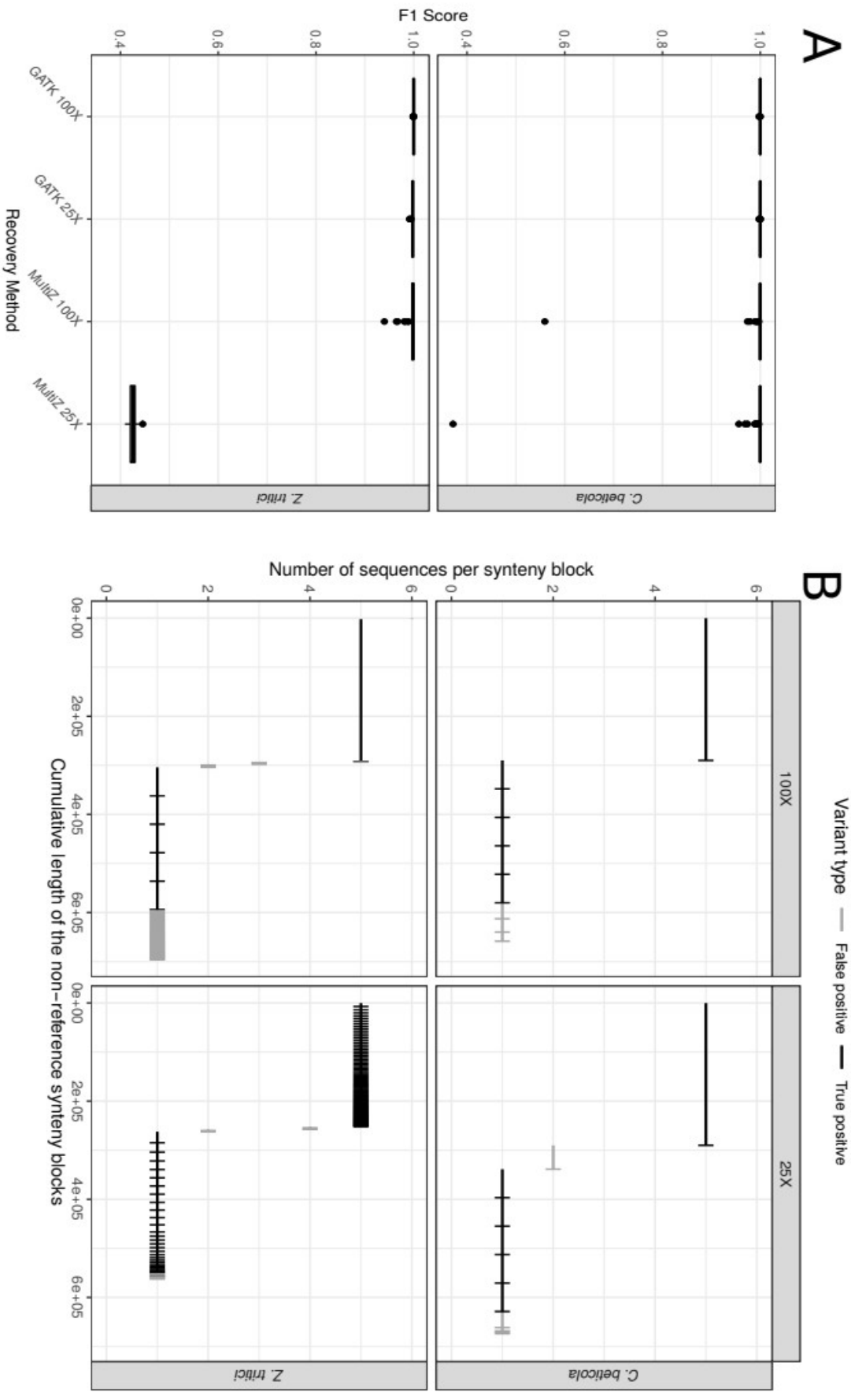

A
F1 Score
1.0
0.8
0.6
0.4
C. beticola
Z. tritici
GATK 100X
GATK 25X
MultiZ 100X
MultiZ 25X
Recovery Method
B
Variant type
False positive
True positive
100X
25X
Number of sequences per synteny block
0
2
4
6
0e+00
2e+05
4e+05
6e+05
Cumulative length of the non−reference synteny blocks

Figure 2: Comparison of the efficacy of variant calling pipelines. **A:** distribution of the F1 scores of SNP recovery methods for *C. beticola* and *Z. tritici* at 25X and 100X sequencing depths. **B:** Recovery of non-reference sequences (accessory regions) in the multiple-genome alignments (MGA) for both species and at both sequencing depths. Six accessory regions have been introduced in the simulated chromosomes: one long region (290 kb) in five individuals, and five smaller regions (58 kb) in one individual. Each segment corresponds to a synteny block in the MGA. Segments have been ordered by decreasing frequency in the sample (y-axis). The x-axis represents the resulting cumulative sum of segment lengths. Each segment is classified as true (black) or false (grey) positive, according to whether it corresponds to a simulated insertion or not. A false positive is a sequence detected as non-reference in the final MGA but which was not inserted during the simulation process (i.e. an artefact of either assembly or alignment) or a sequence inserted during the simulations but recovered in only some of the genomes and/or at an incorrect position.

We then investigated whether the dnWGA approach could recover the simulated accessory regions. Such regions should appear in the WGA as synteny blocks that do not contain the reference sequence. We extracted such synteny blocks from the WGA and compared their size and sequence to the known introduced regions to identify false and true positives. In the C. beticola alignment, the accessory regions were recovered entirely as single regions and in all the chromosomes they were introduced into (Figure 2B). The accessory regions introduced in the Z. tritici chromosomes could be recovered with a similar level of quality at a depth of 100X. Conversely, at 25X, all recovered insertions were fragmented, but 542 out of the 580 kb inserted (93%) were recovered (Figure 2B). False positive regions (non-reference DNA fragments that did not match with the introduced sequences in all WGA) were also detected in all data sets, with a total size ranging from 15 kb to 116 kb per WGA. These regions were found to be comparatively small, and more abundant in the repeat-rich Z. tritici data set. In summary, we find that dnWGA allows the recovery of accessory regions in population genomic datasets. For genomes with a low

frequency of repeats high performance is achieved even with low coverage data, while high coverage data is required in the presence of repeats.

Perspective

The genomes of eukaryote pathogens including fungi and oomycetes can comprise extensive structural variation such as accessory regions, not found in reference genomes. So far, methods to analyze genetic variation in populations of individuals with different genome content and structure are sparse. Whole genome alignment of de novo assembled genomes permits the joint analysis of genetic variation ranging from single nucleotide substitutions to large structural variation. We here show that SNPs can be called from WGAs with a precision similar to that of mapping-based approaches when sufficient sequencing coverage is achieved. We note that with our benchmark based on fungal data, the performance of the dnWGA approach was higher than what was observed in previous comparisons performed on human datasets, where the precision and recall were 87 and 50% at 20X, and 93 and 56% at 50X (Wu et al., 2017). Moreover, the dnWGA approach also allowed us to recover accessory chromosome fragments, genomic features that were shown to occur frequently in fungal genomes. The computational framework based on de novo assembled genomes, therefore, also potentially allows for the analyses of genome segments encoding orphan genes, the comparison of highly dynamic genome compartments, the detection of accessory chromosomes, and the study of repeat dynamics within a population.

In genomes with high frequencies of SVs and accessory regions, the use of dnWGA allows for reference-based mapping to be skipped entirely for variant discovery. However, current methods based on WGA are computationally more demanding than reference-based mapping approaches. As assembly algorithms are improving in quality and efficiency, fostered by the development of long-read sequencing technologies, whole genome alignment constitutes the next methodological challenge. Current state-of-the-art methods are designed for interspecific comparisons and relatively small sample sizes (typically less than 100 genomes). As such, they are not sized to cope with population genomic data sets, for which mechanisms such as recombination can no longer be ignored and prohibits the use of a single guide tree when aligning multiple genomes. The average higher similarity of genomes from a single species, however, should permit more efficient

alignment algorithms. A new generation of genome aligners is, therefore, needed to exploit the full potential of long-read sequencing technologies to characterize genome variation in populations.

Acknowledgements

This work was supported by the International Max Planck Research School (IMPRS) and the German Research Foundation (DFG) SPP1819. Research of EHS is furthermore supported by the Canadian Institute of Advanced Research (CIFAR).

References

Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery and genotyping. Nat. Rev. Genet. *12*, 363–376.

Bian, X., Zhu, B., Wang, M., Hu, Y., Chen, Q., Nguyen, C., Hicks, B., and Meerzaman, D. (2018). Comparing the performance of selected variant callers using synthetic data and genome segmentation. BMC Bioinformatics *19*.

Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., et al. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods *6*, 677–681.

Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., Schmutz, J., Taga, M., White, G.J., Zhou, S., et al. (2009). The Genome of *Nectria haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion. PLOS Genet. *5*, e1000618.

Croll, D., and McDonald, B.A. (2012). The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. PLOS Pathog. *8*, e1002608.

Does, H.C. van der, Fokkens, L., Yang, A., Schmidt, S.M., Langereis, L., Lukasiewicz, J.M., Hughes, T.R., and Rep, M. (2016). Transcription Factors Encoded on Core and Accessory Chromosomes of *Fusarium oxysporum* Induce Expression of Effector Genes. PLOS Genet. *12*, e1006401.

El-Metwally, S., Hamza, T., Zakaria, M., and Helmy, M. (2013). Next-Generation Sequence Assembly: Four Stages of Data Processing and Computational Challenges. PLOS Comput. Biol. *9*, e1003345.

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

Goutte, C., and Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In Advances in Information Retrieval, D.E. Losada, and J.M. Fernández-Luna, eds. (Springer Berlin Heidelberg), pp. 345–359.

Grandaubert, J., Bhattacharyya, A., and Stukenbrock, E.H. (2015). RNA-seq-Based Gene Annotation and Comparative Genomics of Four Fungal Grass Pathogens in the Genus *Zymoseptoria* Identify Novel Orphan Genes and Species-Specific Invasions of Transposable Elements. G3 Genes Genomes Genet. *5*, 1323–1333.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.

Horner, D.S., Pavesi, G., Castrignanò, T., De Meo, P.D., Liuni, S., Sammeth, M., Picardi, E., and Pesole, G. (2010). Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. Brief. Bioinform. *11*, 181–197.

Hwang, S., Kim, E., Lee, I., and Marcotte, E.M. (2015). Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci. Rep. *5*, 17875.

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., et al. (2018). Gene cluster conservation provides insight into

cercosporin biosynthesis and extends production to the genus *Colletotrichum*. Proc. Natl. Acad. Sci. 201712798.

Krusche, P., Trigg, L., Boutros, P.C., Mason, C.E., Vega, F.M.D.L., Moore, B.L., Gonzalez-Porta, M., Eberle, M.A., Tezak, Z., Lababidi, S., et al. (2019). Best practices for benchmarking germline small-variant calls in human genomes. Nat. Biotechnol. *37*, 555–560.

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. *15*, R84.

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

Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., and Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it. Genome Biol. *20*, 246.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

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

Nattestad, M., and Schatz, M.C. (2016). Assemblytics: a web analytics tool for the detection of variants from an assembly. Bioinformatics *32*, 3021–3023.

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

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., Wilkie, A.O.M., McVean, G., and Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat. Genet. *46*, 912–918.

Rosner, B.A. (2006). Fundamentals of biostatistics (Thompson-Brooks/Cole, Belmont, CA). Fundam. Biostat. 6th Ed Thomson BrooksCole Belmont CA.

Sandmann, S., de Graaf, A.O., Karimi, M., van der Reijden, B.A., Hellström-Lindberg, E., Jansen, J.H., and Dugas, M. (2017). Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. Sci. Rep. *7*, 43169.

Tian, S., Yan, H., Klee, E.W., Kalmbach, M., and Slager, S.L. (2018). Comparative analysis of *de novo* assemblers for variation discovery in personal genomes. Brief. Bioinform. *19*, 893–904.

Watterson, G.A. (1975). On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. *7*, 256–276.

Wu, L., Yavas, G., Hong, H., Tong, W., and Xiao, W. (2017). Direct comparison of performance of single nucleotide variant calling in human genome with alignment-based and assembly-based approaches. Sci. Rep. *7*.

Supplementary Materials for Chapter 1

Supplementary Tables

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb. 2020.00626/full#supplementary-material. Here, only Table S1 will be shown.

All custom scripts cited in this text are available at https://gitlab.gwdg.de/alice.feurtey/variant-discovery-methods

All data used in this text are available at the DOI: 10.5281/zenodo.3696563

Table S1: Recovery statistics of known variant and non-variant sites in the simulated *C. beticola* and *Z. tritici* populations from the reference based mapping and MGA approaches at 25X and 100X read depth

Species	**Variant Recovery Method**	**Read Depth**	**Precision**	**Recall**	**F_1 score**
C. beticola	GATK	25X	0.999	0.999	0.999
		100X	0.999	0.999	0.999
	MultiZ	25X	0.999	0.999	0.999
		100X	0.999	0.999	0.999
Z. tritici	GATK	25X	0.999	0.995	0.997
		100X	0.999	0.999	0.999
	MultiZ	25X	0.276	0.993	0.425
		100X	0.991	0.999	0.995

List of Supplementary Tables Available Online

Table S2: QUAST report for genome assemblies of *C. beticola* isolates at 25X read depth

Table S3: QUAST report for genome assemblies of *C. beticola* isolates at 100X read depth

Table S4: QUAST report for genome assemblies of *Z. tritici* isolates at 25X read depth

Table S5: QUAST report for genome assemblies of *Z. tritici* isolates at 100X read depth

Population Simulation

In order to simulate data with different characteristics (such as sequence composition or repeat content) similar to that of fungal pathogens, we based our simulations on two Dothideomycete species for which a complete genome sequence was available: *Cercospora beticola,* the causative agent of cercospora leaf spot on sugar beet as well as a variety of uncultivated plants (de Jonge et al., 2018), and *Zymoseptoria tritici*, the causal agent of the wheat disease septoria leaf blotch (Goodwin et al., 2011).

The first simulations were based on chromosome 1 of the *C. beticola* (accession number: PRJNA270309, chromosome: CM008499.1). The chromosome sequence was extracted and all ambiguous characters were removed. The final sequence used here includes 5.8 Mb and had an average GC content of 50.3%. The second simulations were based on the first chromosome of *Z. tritici* (accession number: PRJNA19047, chromosome: CM001196.1). The sequence used was 6.1 Mb in length and had an average GC content of 53.06%. The repeat content of these chromosomes is different: 0.2% for *C. beticola*, and 11% in *Z. tritici*, allowing us to take into account the impact of repeats on the variant recovery.

From the two chromosome sequences, we simulated a population of 20 individuals with ALF (Dalquen et al., 2012). ALF allows simulating full genome sequence from an ancestral sequence under a given evolutionary model along a tree. We made use of the native birth-death model from ALF to generate a random tree, and applied an HKY substitution model (Hasegawa et al., 1985). A total tree height of 5 and 0.3 were set for the *C. beticola* and *Z. tritici* simulations respectively, to ensure that resulting samples have a Watterson's theta value of 10^{-3} to resemble empirical data sets (Stukenbrock and Dutheil, 2018).

We used MafFilter to determine the variant positions from the multiple sequence alignment produced by ALF (Dutheil et al., 2014). We used simuG (Yue and Liti, 2018) to simulate indels according to the genealogies simulated by ALF and a custom python script to insert them into the simulated genomes. In one branch of the tree an accessory region that was 5% of the total genome size was inserted, and 5 orphan regions each 1% the size of the total chromosome were

added in a single individual. From these simulated genomes Illumina reads of 25X and 100X coverage were generated with ART (Huang et al., 2012). The simulated genomic reads were used in two separate pipelines (Figure 1).

Short Variant Calling

In the following, we define the reference chromosome as the ancestral sequence from which the synthetic population was simulated. In the reference-based mapping pipeline, the short reads were mapped to the reference chromosome using BWA-mem (Li and Durbin, 2009). The alignment files were sorted, indexed, and duplicates were removed with the SAMtools (Li et al., 2009), and further processed with the Picard Tools (http://broadinstitute.github.io/picard/). Short variant discovery was done with GATK v4.1.0.0 (Auwera et al., 2013; DePristo et al., 2011; McKenna et al., 2010). HaplotypeCaller was used to produce gVCF files. These were combined with CombineGVCFs, and the resulting SNP coordinates and frequencies were written in a VCF file with GenotypeGVCFs, following the GATK Best Practices Workflow (https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-).

In the dnWGA pipeline, the short reads were *de novo* assembled with SPAdes (Bankevich et al., 2012). The quality of each assembly was determined by QUAST (Gurevich et al., 2013). Whole genome alignment of the *de novo* assembled genomes with the reference chromosome by MultiZ (Blanchette et al., 2004) was run in parallel with GNU parallel (Tange, 2011). The alignment was projected on the reference sequence with maf_project (Blanchette et al., 2004), and SNPs were called using MafFilter (Dutheil et al., 2014).

The variants called by the two pipelines were stored as VCF files and filtered to remove all missing data and indels with the VCFtools (Danecek et al., 2011). We compared the filtered variants using a custom R script. We computed the precision, recall, and F1 score of each method (Goutte and Gaussier, 2005) by comparing the variants discovered by each pipeline to the true set of variants from the simulations.

Recovery of Accessory Regions

The multiple genome alignments from MultiZ were used to call accessory regions, which are not part of the reference genome. The alignments were filtered with MafFilter to remove all blocks smaller than 1 kb. A custom python script was used to determine the identity of scaffolds contained in alignment blocks. Another custom R script was used to extract the coordinates of the synteny blocks that did not contain any reference sequence. The corresponding regions were extracted from the assemblies with the bedtools getfasta function (Quinlan and Hall, 2010). These regions were then compared to the simulated genomes with BLASTn (Altschul et al., 1990) to determine the location of the unaligned regions. These were compared to the known locations of the inserted accessory regions in order to identify true and false positives in the recovery of accessory regions, with a true positive being defined as a block in which the correct number of sequences are recovered and matching the correct coordinates.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Auwera, G.A.V. der, Carneiro, M.O., Hartl, C., Poplin, R., Angel, G. del, Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr. Protoc. Bioinforma. *43*, 11.10.1-11.10.33.

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., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. *19*, 455–477.

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. Genome Res. *14*, 708–715.

Dalquen, D.A., Anisimova, M., Gonnet, G.H., and Dessimoz, C. (2012). ALF—A Simulation Framework for Genome Evolution. Mol. Biol. Evol. *29*, 1115–1123.

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., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Dutheil, J.Y., Gaillard, S., and Stukenbrock, E.H. (2014). MafFilter: a highly flexible and extensible multiple genome alignment files processor. BMC Genomics *15*, 53.

Goodwin, S.B., M'Barek, S.B., Dhillon, B., Wittenberg, A.H.J., Crane, C.F., Hane, J.K., Foster, A.J., Lee, T.A.J.V. der, Grimwood, J., Aerts, A., et al. (2011). Finished Genome of the Fungal Wheat

Pathogen *Mycosphaerella graminicola* Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. PLOS Genet. *7*, e1002070.

Goutte, C., and Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In Advances in Information Retrieval, D.E. Losada, and J.M. Fernández-Luna, eds. (Springer Berlin Heidelberg), pp. 345–359.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. *22*, 160–174.

Huang, W., Li, L., Myers, J.R., and Marth, G.T. (2012). ART: a next-generation sequencing read simulator. Bioinformatics *28*, 593–594.

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., et al. (2018). Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus Colletotrichum. Proc. Natl. Acad. Sci. 201712798.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760.

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

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

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

Stukenbrock, E.H., and Dutheil, J.Y. (2018). Fine-Scale Recombination Maps of Fungal Plant Pathogens Reveal Dynamic Recombination Landscapes and Intragenic Hotspots. Genetics *208*, 1209–1229.

Tange, O. (2011). Gnu parallel-the command-line power tool.; login: The USENIX Magazine, 36 (1): 42–47 (Feb).

Yue, J.-X., and Liti, G. (2018). simuG: a general-purpose genome simulator. BioRxiv 491498.


Chapter 2:

Population genomics of *Cercospora beticola* suggests that recent host domestication has not influenced genome evolution

Chapter 2: Population genomics of *Cercospora beticola* suggests that recent host domestication has not influenced genome evolution

Lizel Potgieter[1,2], Alice Feurtey[1], Kim Hufnagel[2], Doreen Landermann[1], Mark McMullan[3], Melvin D. Bolton[4,5], Eva H. Stukenbrock[1,2]

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

[2]Environmental Genomics, Christian-Albrechts University of Kiel, Kiel, Germany

[3]Department of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

[4]Edward T. Schafer Agricultural Research Center, USDA Agricultural Research Service, Fargo, North Dakota, USA

[5]Department of Plant Pathology, North Dakota State University, Fargo, North Dakota, USA

Abstract

Cercospora beticola is the fungal pathogen of wild and domesticated beets, causing Cercospora Leaf Spot (CLS) in all species. Sea beet is the extant wild ancestor of all domesticated beets. Sugar beet is a relatively modern crop, and has only been domesticated ~300 years ago. Here we have employed population genomics approaches to determine whether recent host domestication has influenced the genome evolution of *C. beticola*. We included *C. beticola* isolates from wild and domesticated beet from Europe and North America to study signatures of population dynamics, admixture, and population differentiation. We have found that *C. beticola* isolates show a high levels of admixture, and that these exist in a global population that experiences substantial gene flow. We have not found that clusters of *C. beticola* exist that are strongly associated with either host or location. However, *C. beticola* isolates from the UK showed slightly different patterns of admixture, and showed some differentiation in the phylogenetic network. Therefore, we aimed to determine what differentiated these accessions from those from the remaining sampling locations. We found regions within the genomes of isolates between the UK and the remaining sampling locations that indicate some differentiation between individuals. These regions contain protein-coding genes that are likely important in the fungus's biology, and may be regions involved in future evolution and host specialistion.

Introduction

Sugar beet is becoming an increasingly important agricultural crop (Jones et al., 2003). In the 300 years following its domestication, sugar beet production has been expanded and the sector currently contributes 14-20% of the global sugar production, with Germany as one of the leading producers (https://www.isosugar.org/sugarsector/sugar). The contribution of sugar beet to global sugar production is strongly debated with some projections suggesting that it will increase due to climate change, while others suggest that the ratio of sugar refined from sugar beet will remain relatively constant (Jones et al., 2003; Michèle, 2018). Both crops are also used in the production of biofuels. Sugar beet has several advantages over sugarcane that is traditionally used to refine sugar. Sugar beet can be grown in temperate conditions and requires less water to be cultivated successfully than sugar cane (Biancardi et al., 2011). The refining of sugar beet is also not as time sensitive as that of sugarcane Due to the important economic and agricultural roles that sugar beet fulfills, understanding factors that lead to crop loss are becoming more important.

Sugar beet was domesticated from fodder beet, *Beta vulgaris* subsp. *vulgaris L.,* ~300 years ago, and therefore a crop that is significantly younger than most other crops currently used (Hanelt et al., 2001). Mitochondrial phylogenetic analyses have shown that the sea beet is the extant species most closely related to the wild ancestor of beets (Santoni and Bervillé, 1992). All beets, including beetroot, table beet, sugar beet, and sea beet are all subspecies of *Beta vulgaris*, and are able to cross with ease (Arnaud et al., 2010; Bartsch, 2010; Ellstrand et al., 2013; Van Geyt et al., 1990). As sea beet is often used to breed resistance traits into sugar beet populations, surveying the diseases present in wild populations might shed light onto the future of infections in the domesticated populations (Van Geyt et al., 1990). Wild plant populations can potentially act as reservoirs that can lead infection for nearby domesticated crop species (Power and Mitchell, 2004). Furthermore, host domestication has also been shown to be an important factor driving pathogen evolution in some, but not all, crops (Gladieux et al., 2010; Munkacsi et al., 2007; Shapiro et al., 2012, 2018). Some fungi with broad host ranges are able to infect both wild and domesticated hosts, that may result in gene flow between isolates from the different hosts (Feurtey et al., 2020).

The genomes of several sea and sugar beet accessions have been sequenced (Dohm et al., 2014; Rodríguez Del Río et al., 2019). The 567 Mbp sugar beet genome was predicted to contain 27421 protein-coding genes, and a repetitive content of 42.3%. The sea beet genome was slightly larger and was 590 Mbp long, and predicted to contain 16102 protein-coding genes. Previous studies showed high genotypic divergence within the *Beta* genus, and was further confirmed when comparing whole genome sequences of various sugar beet accessions (Dohm et al., 2012, 2014; Schneider et al., 2007). Regions of low diversity, so-called “variation deserts” were identified among sugar beet accessions, and are a product of selective breeding. These regions contain the bolting locus that is of agricultural importance. These variation deserts were not present in sea beet genomes.

Cercospora beticola is a fungal pathogen that causes Cercospora Leaf Spot (CLS) on domesticated and wild beet as well as a variety of other weeds including white goosefoot and bitter dock (Knight et al., 2019a; Pool and McKay, 1916; Vestal, 1933). CLS causes leaf spots on the adaxial leaf surface, reducing the photosynthetic potential of the plant leading to smaller roots that contain less sucrose (Weiland and Koch, 2004). Infection with *C. beticola* can also lead to more rapid root spoiling prior to refining as well as increased molasses (Skaracis et al., 2010). All of these consequences result in economic losses.

Understanding the effect of host domestication as well as geographical isolation on the evolution of phytopathogens (e.g. De Gracia et al., 2015; Giraud et al., 2008, 2010). Domestication of host plants often drives divergent evolution of fungal plant pathogens of either domesticated and/or wild pathogens due to the evolutionary arms race between the host and the fungus (Anderson et al., 2010; Frenkel et al., 2010; Xhaard et al., 2011). In addition to the plant-pathogen interaction influences on evolution, agro-ecosystems are also markedly different from wild ecosystems (Stukenbrock and McDonald, 2008). Some of these differences include higher genetic diversity in wild ecosystems, higher frequency of monoclonal plants in agro-ecosystems, and the human-mediated management of pests and diseases. The abundance of genetically identical hosts in

agro-ecosystens removes the slective pressure of evolving less virulent phenotypes, resulting in the evolution of more virulent pathogens in agricultural ecosystems (Thrall and Burdon, 1999).

There is substantial variation among fungal plant pathogens isolated from agro-ecosystems. This variation is influenced by the crop as well as the geographic location of the crop. In the plant pathogen *Botrytis cinerea*, accessions isolated from different hosts showed reduced ability to infect hosts other than the one they were isolated from due to selective sweeps (Mercier et al., 2019). In *Sclerotinia sclerotiorum*, there are two main geographic clusters, and a selective sweep in one population that was close to a gene that may affect nutrient uptake in the host (Derbyshire et al., 2019). In *Fusarium graminearum*, a gene cluster that produces a novel mycotoxin, and other genes related to host-pathogen interactions, were divergent in North American isolates (Kelly and Ward, 2018). It is proposed that the differentiated traits reflect the evolutionary trajectories followed by *F. graminerum* populations during their host shift from an endophytic lifestyle on wild grass to a pathogenic lifestyle when infecting domesticated cereals (Kelly and Ward, 2018). *F. graminearum* isolates in Germany have been shown to form part of a recombining meta-population (Talas and McDonald, 2015). However, isolates from North America exist as multiple populations that have recently evolved to infect the introduced domesticated cereals (Lofgren et al., 2018).

The comparison between the same fungal plant pathogen that infects both the wild ancestor of a modern crop and the crop itself shows the effect that host domestication has on the domestication of its pathogens. In domesticated crops, *F. graminearum* causes Fusarium head blight, but infect native grasses in an asymptomatic manner with little to no accumulation of mycotoxins (Lofgren et al., 2018). This may be due to the long evolutionary history the fungus shares with the wild hosts, resulting in distinct biochemical interactions. In *Venturia inaequalis*, the transition from wild to agricultural apple trees resulted in larger spores and increased sporulation capacity (De Gracia et al., 2015). These few examples are several of many detailing plant-pathogen interactions, and the impact host domestication has on fungal evolution.

Prior population studies of *C. beticola* populations have illustrated global genetic diversity of isolates infecting domesticated beet. These studies have mostly been carried out on various

populations using microsatellite markers (Groenewald et al., 2007; Knight et al., 2018, 2019b; Vaghefi et al., 2017a). In some populations recent expansions were detected suggesting that population structure can be influenced by local demographic events (Knight et al., 2019b). Long range dispersal of *C. beticola* has also been shown, with population clusters in the USA arising from a Eurasian population, as well as another unknown location (Knight et al., 2019b). *C. beticola* has been detected in seeds of sugar beet, and sugar beet seeds used in the USA are often produced in Australia, Europe, South America, and South Africa (Noel Knight, personal communication Knight and Pethybridge, 2020). While the viability of the *C. beticola* has not been confirmed to be a source of inoculum for infections, seed dispersal may contribute to the spread of Eurasian isolates to North America, as has been the case with several other fungal species (reviewed by Fisher et al., 2012). These studies have been restricted to *C. beticola* accessions isolated from sugar beet, with little known about the *C. beticola* isolates from sea beet.

We hypothesise that there will be weak signals of host specialisation as the host system is very modern compared to other domesticated plants, and the wild and domesticated hosts readily hybridise. We further hypothesise that the high levels of admixture seen in other *C. beticola* populations will also be present in the isolates from sea beet. While we hypothesise that there will be little differentiation between populations, we anticipate that some genomic regions may show the hallmarks of early differentiation. This study aims to compare populations of *C. beticola* from wild and domesticated beet using whole genome sequences. Since there have been no comparisons between isolates from wild and domesticated hosts in this system, questions regarding the effect host domestication and location have on pathogen evolution. Therefore, we aim to 1) determine the genetic diversity of *C. beticola* isolates on sea beet on the East coast of the UK, 2) establish whether host and/or location influences population structure, 3) identify what the regions that differentiate the isolates from the UK from the *C. beticola* isolates from the rest of the sampling sites, and 4) predict the biological implications of the differentiated regions. Our analyses will allow us to gain novel insights into the recent evolution of this important sugar beet pathogen.

Understanding the population dynamics of pathogens on wild and domesticated plants is crucial to develop novel crop protection strategies.

Methods and Materials

Sample Collection and Clone Correction

In order to compare genetic diversity of *C. beticola* on wild and domesticated plants, we conducted a field collection of sea beet isolates. Sea beet plants infected with *C. beticola* were sampled on the East Coast of the United Kingdom in September 2018. Three locations, Southwold, Orford, and Bawdsey Quay, were selected listed from North to South. Up to 10 infected leaves were collected per site, and between 7 and 10 sites were selected for each site, depending on the prevalence of sea beet infected with *C. beticola*. There was at least 10 m between each sampled plant. For each leaf with more than one leaf spot, two spots were chosen at random for spore isolation.

Spores from leaf spots were isolated by dislodging these with 10 µl of water containing ampicillin (0.1 mg/l) and transferred to water agar plates. The drop containing the spores was spread across the surface of the agar with an additional 50 µl of water and a sterile hockey stick. The plates were incubated in the dark at room temperature for up to 14 days, and inspected daily for germinated spores. Most plates contained more than one germinated spore after 5 days with enough growth to identify *C. beticola*. Two germinated spores with hyphae spatially separated from contaminants and other germinated spores were cut from the water agar plates. These squares were transferred to tomato juice agar plates containing streptomycin (0.1 mg/l) that were made with the house brand tomato juice from the local supermarket, Rewe (Supplementary Methods). Isolates were incubated in the dark at room temperature, and underwent several rounds of pure culturing until each isolate was free from contaminants.

The resulting pure cultures were used to extract DNA from, and placed in long-term storage. DNA was extracted following the CTAB protocol (Clarke, 2009). *C. beticola* isolates were cultivated in liquid tomato juice based media in the dark at room temperature for 5 days, while being gently shaken at 30 rpm. Hyphae were removed from the media, and ground with a mortar and pestle with liquid nitrogen and transferred to 2 ml Eppendorf tubes. 1.2 ml of extraction buffer was added to each sample, and shaken with a vortex briefly (Supplementary Methods). The samples were

incubated at 65°C for 1 hour, and shaken every 15 minutes. Samples were spun in a centrifuge for 10 minutes at 13500 g. The supernatant was transferred to a fresh tube, and 800 µl phenol/chloroform was added. Samples were mixed on a rotating platform for 20 minutes at room temperature, and spun in a centrifuge at 13500 g for 10 minutes. The aqueous phase was washed again with phenol/chloroform and spun. The aqueous phase was transferred to 800 µl ice cold isopropanol and incubated at -20°C. Samples were spun in a centrifuge for 10 minutes at 13500 g, and the supernatant was discarded. The pellet was suspended in 250 µl TE buffer, and 25 µl RNase was added and samples were incubated at 37°C for 30 minutes. 25 µl NaOAc and 600 µl ice cold 100% EtOH were added to the samples, and incubated overnight at -20°C. Samples were spun in a centrifuge at 13500 g for 10 minutes, and the supernatant was discarded. The pellet was washed twice with 70% EtOH, and air dried. The pellet was suspended in 100 µl TE buffer. The DNA concentrations were determined by fluorometry and quality was assessed by gel electrophoresis.

To optimise conditions and primers for clone correction, a small subset of the isolates were used to test the primers used for *C. beticola* clone correction (Vaghefi et al., 2017a). Unlabeled primers were used to test the successful amplification of all target regions by the polymerase chain reaction (PCR) (Supplementary Table 1). The microsatellite markers were also used to confirm species identity. The markers only successfully amplify *C. beticola*, and commonly used genes such as ITS are not useful for *C. beticola* due to the lack of informative sites within the genus (Vaghefi et al., 2017a).

A PCR amplification of each target region was performed using the same reaction mix and cycles, but with different annealing temperatures (Table 1). A subset of 11 *C. beticola* isolates were randomly chosen to optimise the amplification reaction with fluorescent primers. DNA concentration, number of amplification cycles, and primer concentration were important to ensure that the peaks were not too high. This was also done to establish which markers were informative for the isolates. From the original 12 primer pairs, 6 pairs were informative for these populations, and these were further used to genotype the remaining isolates (Table 1). The allele sizes of the

indivduals was with determined with GeneMapper 5 (Applied Biosystems). Haplotype Analysis v 1.05 was employed to determine haplotypes and compute genetic diversity (Eliades and Eliades, 2009).

Table 1: Microsatellite primers used for *C. beticola* clone correction and species confirmation. Included are assigned primer names, accession, primer sequences, and dyes from previous research, as well as the optimised multiplexes and annealing temperatures optimised for this study.

Primer name	GenBank Accession Number	Sequence	Multiplex	Annealing Temperature	Dye	Reference
SSRCb1		F: TGCGATCTGGGCATAAATATC R:AGATTTGCATTTGCCCACAC	2	55ºC	Hex	(Groenewald et al., 2007)
SSRCb3		F: ATAGAGTCAAACCAAGCCAAG R: CCCGTTATAGCGCCCTTAG	2	55ºC	Fam	(Groenewald et al., 2007)
SSRCb21	KX452351	F: GACTTTGGCATTCGAGAAGATGG R: CCACTAAACGTATCTCTTTGCTGT	2	55ºC	Fam	(Vaghefi et al., 2017a)
SSRCb22	KX452352	F: GCCACTTCATTACCACCTTGAAT R: TGAGCTGATGTGAAAGGTAGAGG	1	58ºC	Fam	(Vaghefi et al., 2017a)
SSRCb25	KX452355	F: GACGAGCATTCCATTGAGAAGTC R: TCGTCGTTTTGGTCCTCTTCTTC	1	58ºC	Hex	(Vaghefi et al., 2017a)
SSRCb27	KX452357	F: CGTCAAAGCAGTCCCTCGAT R: AATTGAACAAGCGCCCAACC	1	58ºC	Fam	(Vaghefi et al., 2017a)

Genome Sequencing

To analyse genome-wide variation, we generated a population genomic dataset of the *C. beticola* isolates. In addition to the populations from the UK that were collected, Melvin Bolton (USDA, Fargo, USA) contributed DNA samples of *C. betiola* isolated from sea beet plants in two locations in Croatia. We also included five *C. beticola* isolates from Germany, from the Institut für Zückerrübenforschung (IFZ, Göttingen, Germany). These isolates were sequenced at the Max Planck Genome Center in Cologne along with the isolates from the UK. Illumina sequencing was performed using the HiSeq 3000 platform to render paired end reads of 150 bp. An average of 20X read coverage was requested.

Quality Analysis and Read Processing

To determine high quality variants, read processing and read mapping was performed. The NGS data received was inspected for primer contamination and overall sequence quality with FastQC v0.11.6 (Andrews, 2010). Number of reads were calculated with GNU parellel and a custom script (count.reads.sh) (Tange, 2011). All scripts are available in the supplementary material as well as online (https://github.com/lpotgieter/phd.scripts). The reads were trimmed to remove regions of poor quality with FASTQ quality trimmer, a function of FASTX Toolkit v0.0.14 (Gordon and Hannon, 2010), and read pairs were extracted with BBmap (https://sourceforge.net/projects/bbmap/). Reads were mapped to the 10 fully assembled chromosomes of the *C. beticola* reference genome (GCA_002742065.1, all chromosomes over 1 Mb in size (de Jonge et al., 2018)) with bwa-mem (Li and Durbin, 2009). Aligned reads were processed with SAMtools to sort and index reads, as well to remove PCR duplicates (Li et al., 2009). Where isolates were resequenced, SAMtools was also used to merge the reads from both runs. Read groups were added with Picard (Horner et al., 2010). Variant sites were determined by GATK v4.1.0.0 (Auwera et al., 2013; DePristo et al., 2011; Poplin et al., 2018). First, HaplotypeCaller was used to produced genomic variant call format (GVCF) files for each sample. Second, CombineGVCFs was used to merge the GVCFs from each sample into a combined GVCF file. Finally, GenotypeGVCF was used to produce a combined VCF that was used in the subsequent analyses.

Analyses of Population Structure

To establish population structure, variants were filtered to remove low quality variants, as well as linked sites. Before the variant file was used in downstream analyses, poor quality positions were filtered, and basic statistics about the quality of the variants computed. Variants were filtered with VCFtools (minimum quality per site 15; minimum mean depth per site 3, maximum mean depth per site 100) to remove positions that had low quality as well as disproportionate sequencing depth as well as insertions and deletions (indels) (Danecek et al., 2011). These parameters were selected based on the average read depth across the genome for all individuals deduced from the alignment files. A minimum mean depth per site of 3 allowed for rare variants that are likely true

variants. A maximum mean depth per site of 100 was 3 times the average sequencing depth, and would remove excessive coverage of repetitive regions such as transposable elements. The stats function of BCFtools was used to calculate the mean read depth, the number of variant sites, and the amount of missing data for each individual (Li et al., 2009). Population statistics including pi and Tajima's D were also calculated with VCFtools. To determine whether Tajima's D values differ significantly between geographic locations and hosts, an ANOVA was performed (anova.of.tajd.R). All plots were made using a custom R script (cercospora.plots.R).

Biallelic sites across the genome were extracted, converted to the nexus format using PGDSpider2 (Lischer and Excoffier, 2012), and used to construct a reticulation network with SplitsTree v4 (Huson, 1998). The biallelic sites across the genome were used to perform a principle component analysis (PCA) with PLINK v1.9 (Purcell et al., 2007).

To determine the distance at which SNPs were linked, the decay of linkage disequilibrium (LD) was calculated. Due to computational restraints, the mimimum number of SNPs on a single chromosome across all populations was determined and used to subsample the SNPs on each chromosome. PLINK was employed to calculate a windowed R^2 for each chromosome based on the size of each chromosome. A custom R script (ld.R) was used to calculate the decay across each chromosome for each sampling site. The distance at which the maximum R^2 value halved, was used to thin the genome wide SNPs for subsequent population structure determination.

To quantify the population structure, sNMF as a part of the LEA v2.0 R package was used (Frichot and François, 2015). Genome wide SNPs that were thinned to 5 kb were converted to a genotype format with a custom script (vcf2geno.sh). A custom R script (snmf.R) was used to compute K=2..20 over 10 iterations.

Analyses of Population Differentiation

Individuals from the UK clustered differently from the other *C. beticola* isolates, and further analyses were performed to explore these differences. From the reticulation network, PCA, and ancestry analyses, it was not clear whether the differences between the isolates from the various

sites was due to the host they were isolated from or the location they were isolated from. To quantify the differences between the isolates collected from different sites, the fixation index (F_{ST}) was calculated in a pairwise manner between all sampling sites with VCFtools in 50kb windows. The regions 50 kb up- and downstream of the global maxima of F_{ST} were further used. The genes published annotation was used to query the NCBI database with a blastn to determine homologous genes in other species as well as their function with Blast2GO v5.1 (Conesa et al., 2005). The products of these genes were also queried to determine whether they contained signal peptides with SignalP (Petersen et al., 2011). EffectorP 2.0 was employed to evaluate whether the genes were potentially effectors (Sperschneider et al., 2016). To detect fine-scale selection, Tajima's D and F_{ST} were calculated for each gene with VCFtools. The functional effect of SNPs within the F_{ST} outlier regions was determined by SnpEff v4.3t (Cingolani et al., 2012).

As the reference genome was hard masked, several ambiguous regions were located. To determine whether these ambiguous regions could be recovered in other genomes, a small subset of genomes were assembled in a *de novo* manner with SPAdes v3.11.1 (Bankevich et al., 2012). The regions flanking the ambiguous regions were used to query the *de novo* assemblies with a local blastn to determine the homologous contigs (Altschul et al., 1990). To align the identified contigs with the reference genome, reference chromosome and the identified contigs were aligned with nucmer, a part of mummer v3.9.4 (Delcher et al., 2003). The show-coords function was then used to determine the coordinates of the contigs from the assemblies.

Results

Sample Collection and Clone Correction

To characterise genetic variation of *C. beticola* on sea beet, we collected isolates from three field sites in the UK. The three sites on the East Coast of the UK had different distributions of sea beet, and had varying environments. As such, different numbers of spores germinated from each site, and different numbers of isolates were used for clone correction from each site. Overall, the success rate of obtaining pure cultures from sea beet was low. The antibiotics used eradicated the vast majority of bacterial contaminants, but the fungal contaminants were more difficult to control. Often this contamination was only found after revival of the isolates from -80^{o}C conditions since *C. beticola* takes a longer time to start growing than many other fungi. The final number reported in Table 2 consisted only of pure cultures that were able to withstand freezing and thawing cycles.

Microsatellite markers were used to identify potential clones between the isolates collected. The preliminary test of the primers on the isolates collected from the UK showed that six out of the original 12 primer pairs provided sufficient resolution to identify clones. Several primer pairs also had a high error rate when used to amplify the target regions (Supplementary Table 1). Since there was no ambiguity regarding the clonal nature of isolates using only the six primers, further optimisation was not required. The primers used identified between 4 and 16 different alleles for each region respectively (Table 3).

Table 2: Description of the sampling sites, germination, and pure culturing success rate of *C. beticola* isolates collected from the UK.

Location	**Description of Site**	**Number of Plots Sampled**	**Number of Pure Isolates**	**Number of Isolates Included for Whole Genome Sequencing**
Southwold	Fishing harbour High human activity around plants Moderate distribution of sea beet plants Sea beets of a moderate size	13	29	18
Orford	Grassland surrounding military base Human activity contained to paths Across the path from a healthy sugar beet field High prevalence of sea beet plants Large sea beet plants	10	27	14
Bawdsey Quay	Quay High human activity Harsh winds Low distribution of sea beet plants Small sea beet plants	8	12	10

Table 3: Number of haplotypes identified by each microsatellite primer pair, as well as the allele sizes

Primer Name	Number of Haplotypes	Allele Sizes (bp)
SSRCb1	4	221, 223, 225, 237
SSRCb3	10	245, 262, 264, 271, 273, 275, 279, 284, 291, 345
SSRCb21	5	165, 167, 173, 178, 180
SSRCb22	4	185, 188, 191, 199
SSRCb25	7	231, 235, 255, 258, 261, 264
SSRCb27	16	368, 373, 376, 379, 382, 398, 404, 406, 410, 418, 425, 427, 433, 436, 439, 442

The microsatellite markers was used to compare the genetic diversity between the three sampling sites, as well as selecting unique isolates (Table 4). The number and frequency of shared and unique haplotypes was used to calculate haplotypic richness, genetic diversity, and mean genetic diversity between individuals within each sampling site. These measures showed that isolates collected from Southwold harboured the highest amount of genetic diversity (Table 4). The populations in Orford and Bawdsey Quay contained similar levels of genetic diversity, and were less diverse than the population collected from Southwold. Although the population from Orford contained more than double the amount of individuals than the population from Bawdsey Quay, both the measure for genetic diversity and diversity between individuals was comparable. This diversity is reflected in the distribution of haplotypes among the sites (Figure 1). From these isolates, the unique isolates were selected for whole genome sequencing to represent the genetic diversity present in the populations of *C. beticola* from the sampling sites (Supplementary Table 2).

Table 4: Population statistics from microsatellite markers from each sampling site in the UK

Sampling Site	Sample size	Number of Haplotypes Detected	Number of Private Haplotypes	Effective Number of Haplotypes (Ne)	Haplotypic Richness (Rh)	Genetic Diversity (He)	Mean Genetic Diversity Between Individuals (D^2 sh)
Southwold	29	20	18	15.868	9.23	0.97	17693.702
Orford	27	13	12	6.451	6.758	0.877	5470.108
Bawdsey Quay	12	7	6	4.8	6	0.864	5376.29

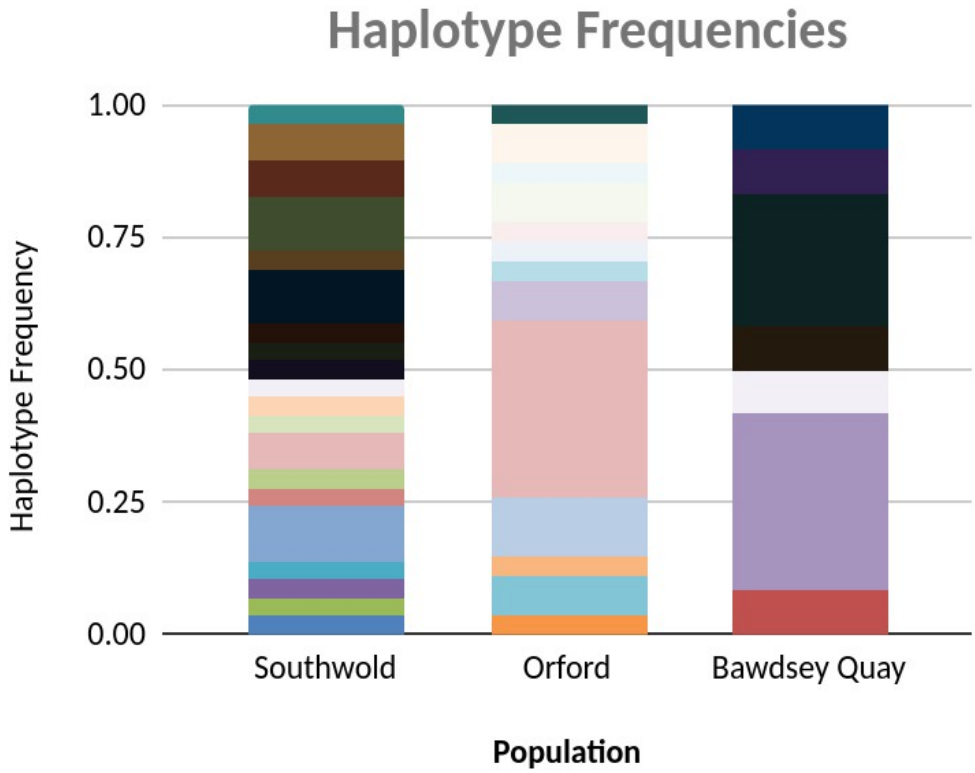


Figure 1: Haplotype frequencies at each of the 3 sampling sites on the East coast of the UK. Haplotypes were determined by six unique microsatellite markers. Each colour represented a different haplotype. In total, 38 haplotypes were identified among all isolates.

Genome Sequencing, Quality Analysis and Read Processing

The unique *C. beticola* isolates from the UK as well as isolates provided by M. Bolton were sequenced by the Max Planck Genome Center in Cologne, Germany (Supplementary Table 2). For most individuals, the requested number of reads were delivered after the first run. For those that had a lower than requested coverage, a second run was performed. These were merged later in the read mapping pipeline. Prior to filtering, isolates contained between 3.4 million and 7.6 million forward and an equal number of reverse reads. The reads mapped more than 69% of the reference genome for all but four isolates (Supplementary Tables 3 and 4).

Detection of Population Structure

We combined the data from our own collection with the genome data provided by M. Bolton. Hereby, we have a collection of *C. beticola* representing diversity of the pathogen in Europe and North America on wild and cultivated accessions (Table 5). The distribution of isolates from sea beet, sugar beet, and table beet was spread across Europe and North America. Isolates from sugar beet were collected from North Dakota and Italy, isolates from sea beet from the UK and Croatia, and a single population of isolates from table beet was collected from New York. To ensure that the data included in the subsequent analyses was of good quality, each site was filtered to remove sites that were potentially false positives. All individuals had an average coverage of at least 10X per site (Supplementary Table 4, Supplementary Figures 1 and 2).

Table 5: Number and geographic origin of *C. beticola* genomes, as well as the host from which they were isolated, included in this study

Location	**Host**	**Number of Genomes**
Southwold, UK	Sea beet	18
Orford, UK	Sea beet	14
Bawdsey Quay, UK	Sea beet	10
Antenal, Croatia	Sea beet	9
Jadransko, Croatia	Sea beet	15
Various, Italy	Sugar beet	17
Various, Germany	Sugar beet	5
Fargo, ND, USA	Sugar beet	88
Foxhome, ND, USA	Sugar beet	17
New York, NY, USA	Table beet	24

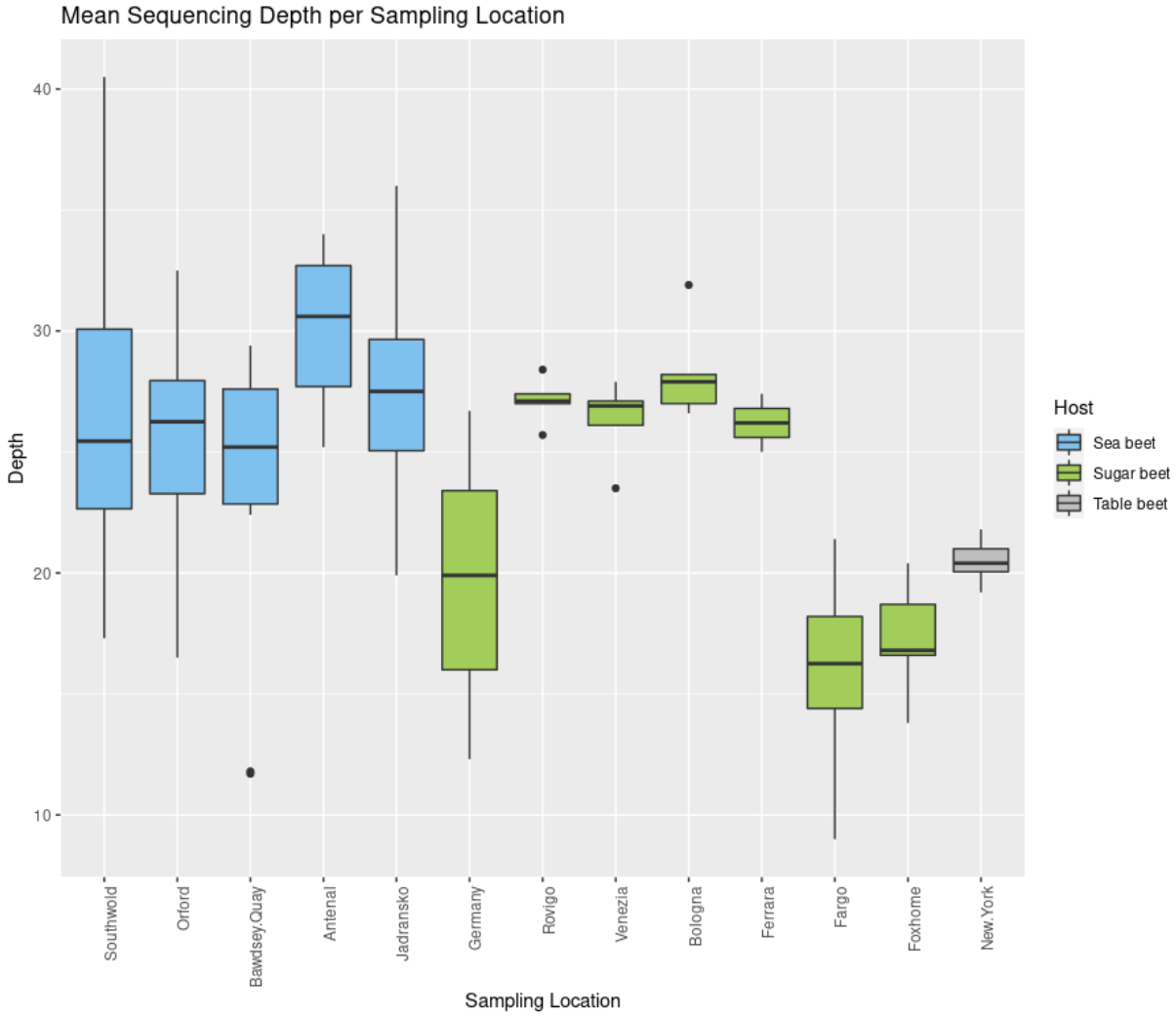


Figure 3: Box plot of the mean read depth per individual of all individuals at each sampling site. There was variation between the sample size between the number of isolates from the various locations, namely Southwold, UK (n = 18), Orford, UK (n = 14), Bawdsey Quay, UK (n = 10), Antenal, Croatia (n = 9), Jadransko, Croatia (n = 15), Germany (n = 5), Rivigo, Italy (n = 5), Venezia, Italy (n = 5), Bologna, Italy (n = 5), Ferrara, Italy (n = 2), Fargo, USA (n = 88), Foxhome, USA (n = 17), New York, USA (n = 24).

To confirm that the majority of the genome was included, the fraction of missing sites per individual was calculated. No isolate had more than 8% of the sites missing (Supplementary Figure 3). From the combination of read depth per site, average quality per individual, and fraction of missing sites,

the quality of the data used for the remainder of the analyses was considered to be of sufficient quality for downstream analyses.

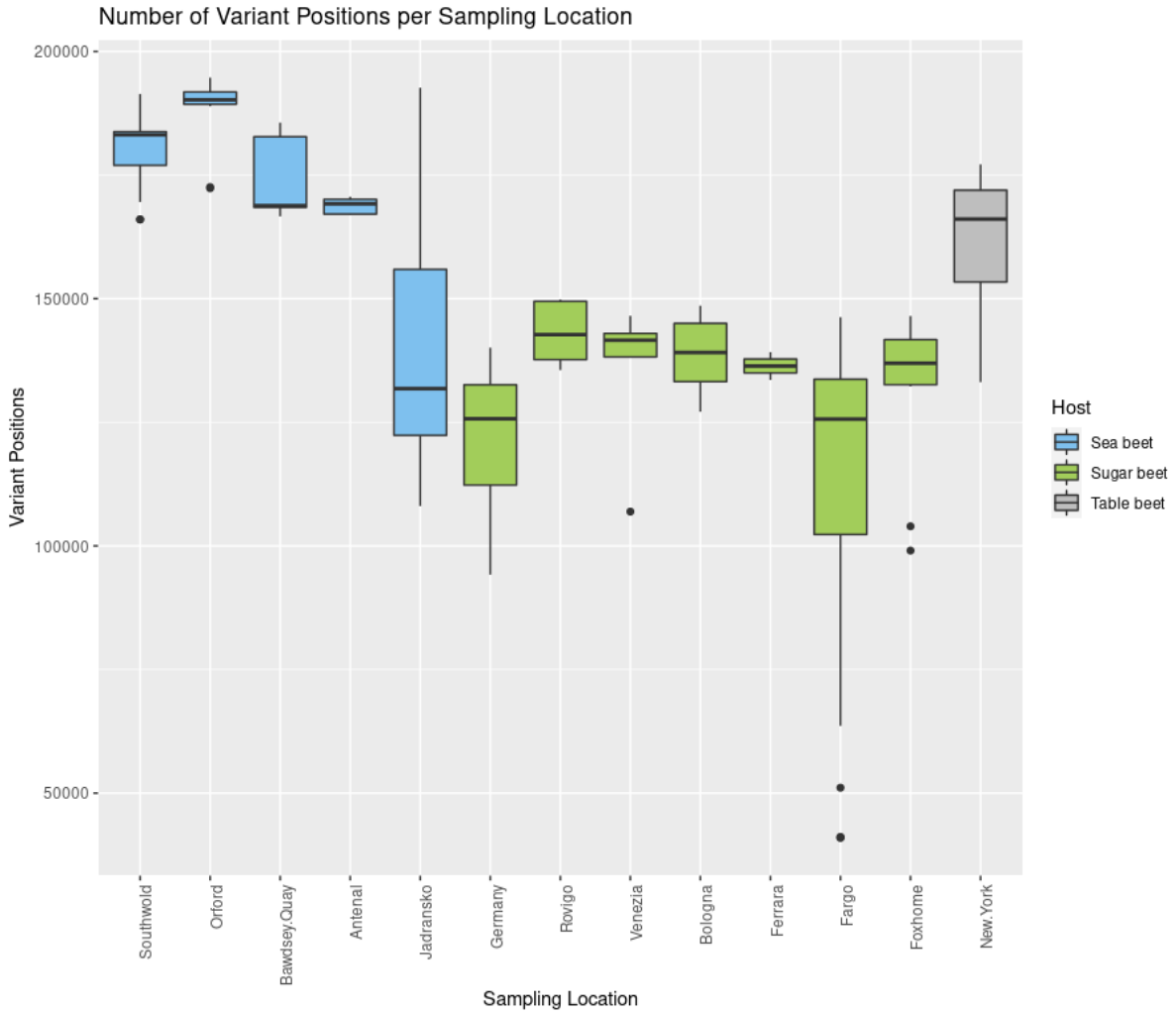


Figure 4: Box plot illustrating the number of variant positions from isolates from each sampling site. There was variation between the sample size between the number of isolates from the various locations, namely Southwold, UK (n = 18), Orford, UK (n = 14), Bawdsey Quay, UK (n = 10), Antenal, Croatia (n = 9), Jadransko, Croatia (n = 15), Germany (n = 5), Rivigo, Italy (n = 5), Venezia, Italy (n = 5), Bologna, Italy (n = 5), Ferrara, Italy (n = 2), Fargo, USA (n = 88), Foxhome, USA (n = 17), New York, USA (n = 24).

The comparison of individuals collected from the different hosts and locations was conducted on an individual as well as on a population scale. On an individual scale, the number of SNPs were counted. The individuals collected from sea beet in the UK contained more SNPs relative to the reference genome than isolates from Croatian sea beet, and isolates collected from domesticated beets. Between the individuals collected from domesticated beet, isolates from North Dakota contained the lowest number of SNPs relative to the reference genome. This was likely due to the origin of the reference genome. The reference genome was collected from sugar beet found in the USA (de Jonge et al., 2018).

Due to the distribution of SNPs per individual, population statistics were determined for each location (Table 6). The number of SNPs per location varied between 265850 and 514434, with a combined total of 774662 SNPs across the whole genome (Figure 4, Supplementary Figure 4, Supplementary Table 5). The average pi across the genome ranged between 0.00322 and 0.00424 (Supplementary Figure 5). The average Tajima's D measure was also not highly divergent from 0, indicating that the vast majority of the sites within each population were not represented by an excess or an absence of rare alleles (Figure 5).

Table 6: Population statistics of the *C. beticola* populations, including the number of SNPs, average pi and Tajima's D across the genome in 50 kb windows.

Population	**Host**	**Number of Individuals**	**Number of SNPs**	**Average Pi**	**Average Tajima's D**
All Combined	Combined	217	774662	0.00429	0.016
Croatia	Sea beet	24	514434	0.00424	-0.007
UK	Sea beet	42	494315	0.00403	0.781
Germany	Sugar beet	5	265850	0.00364	-0.055
Italy	Sugar beet	17	368872	0.00325	0.134
North Dakota	Sugar beet	105	508751	0.00322	0.24
New York	Table beet	24	424795	0.00361	0.301

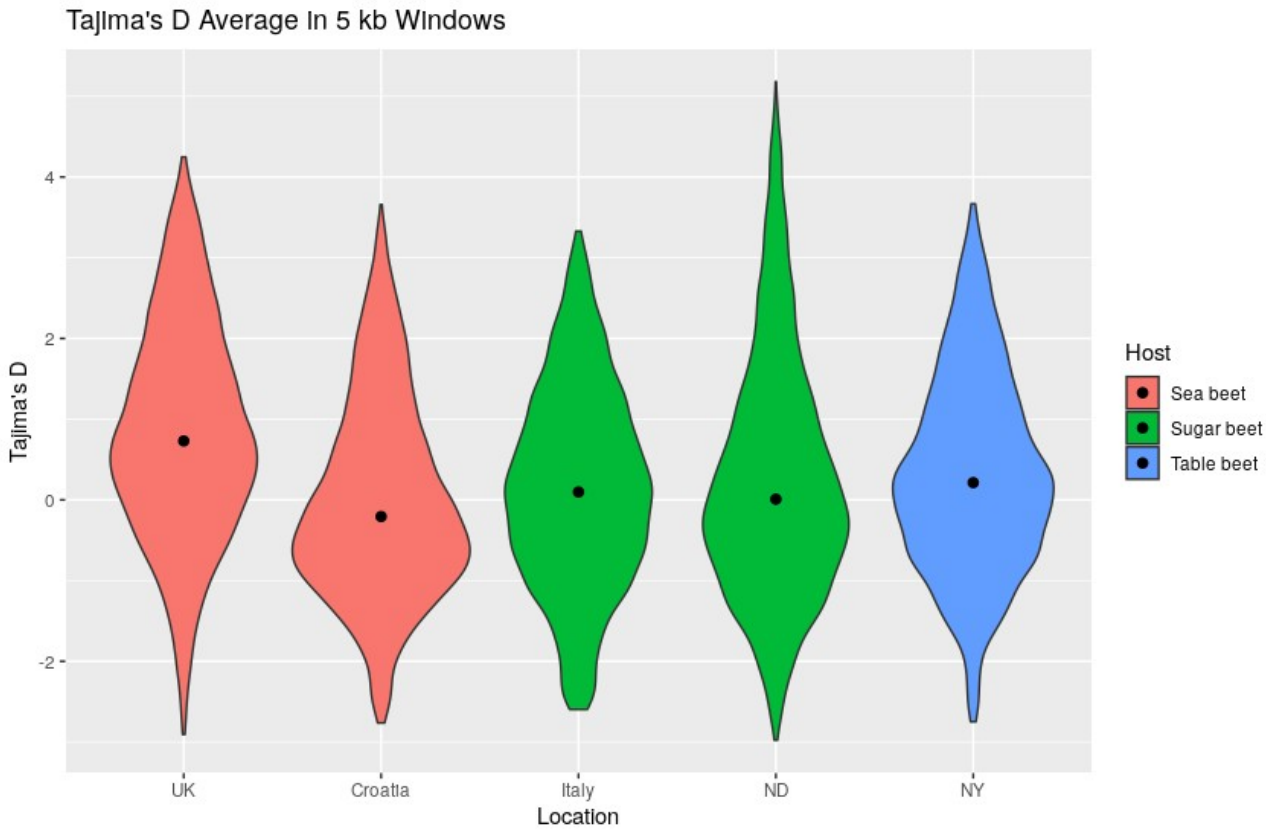


Figure 5: Violin plots of the Tajima's D distribution across the genome for *C. beticola* in 50 kb windows from the various sampling sites.

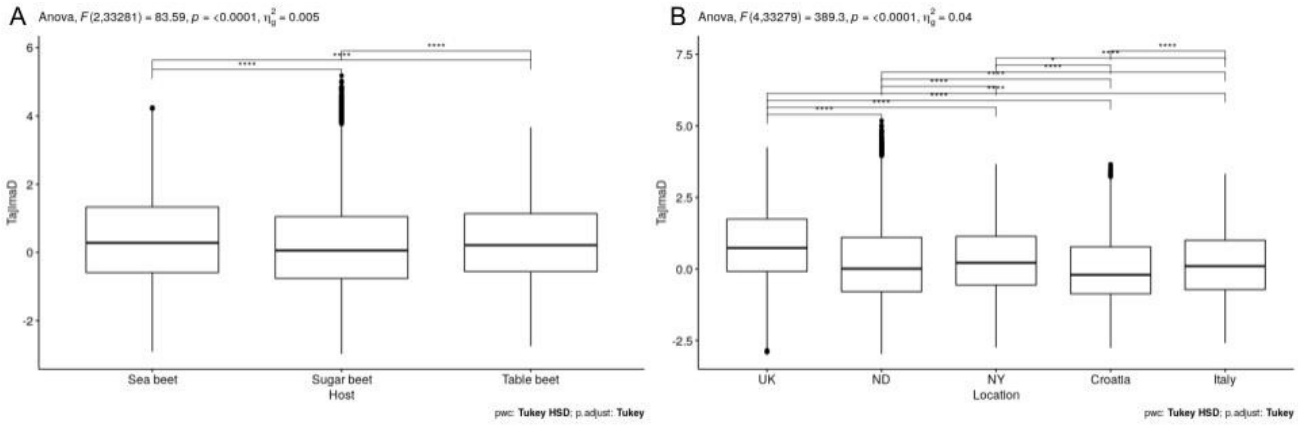


Figure 6: The variance between the Tajima's D distribution in 50 kb windows along the genome was significantly different among all locations and all hosts.

Significance values were indicated as follows: p≈0 '***', p<0.001 '**', p<0.01, '*' p<0.05

To compare how closely individuals were related to each other purely based on sequence similarity, an identity by state matrix was constructed. The range of identities of biallelic SNPs

between individuals ranged from 0.72 to 0.99, with median decent by state of 79%. To establish the phylogenetic relationships individuals have to one another, a reticulation network was computed (Figure 7). The network showed that the individuals were all closely related, and did not separate into clear populations. The isolates from the UK clustered separately from the isolates from the other locations in two groups. The isolates from the UK also showed some genetic exchange between the three locations in one of the clusters, and less in the other. The cluster that showed less genetic exchange between the individuals was more closely related to isolates collected from domesticated beet. One of the sites from Croatia, Antenal, also clustered separately from the isolates collected from the UK and USA, and shared ancestry with some of the isolates from Jadransko. The other isolates from Jandransko were more closely related to isolates from domesticated beet from domesticated beet. The *C. beticola* isolates from domesticated beet in Europe and USA were not separated by location or host. The reticulation network also showed that there was genetic exchange between all individuals, as seen with the reticulation at the center of the network. In some clusters, there have been more reticulation events confined to individuals more closely related to one another.

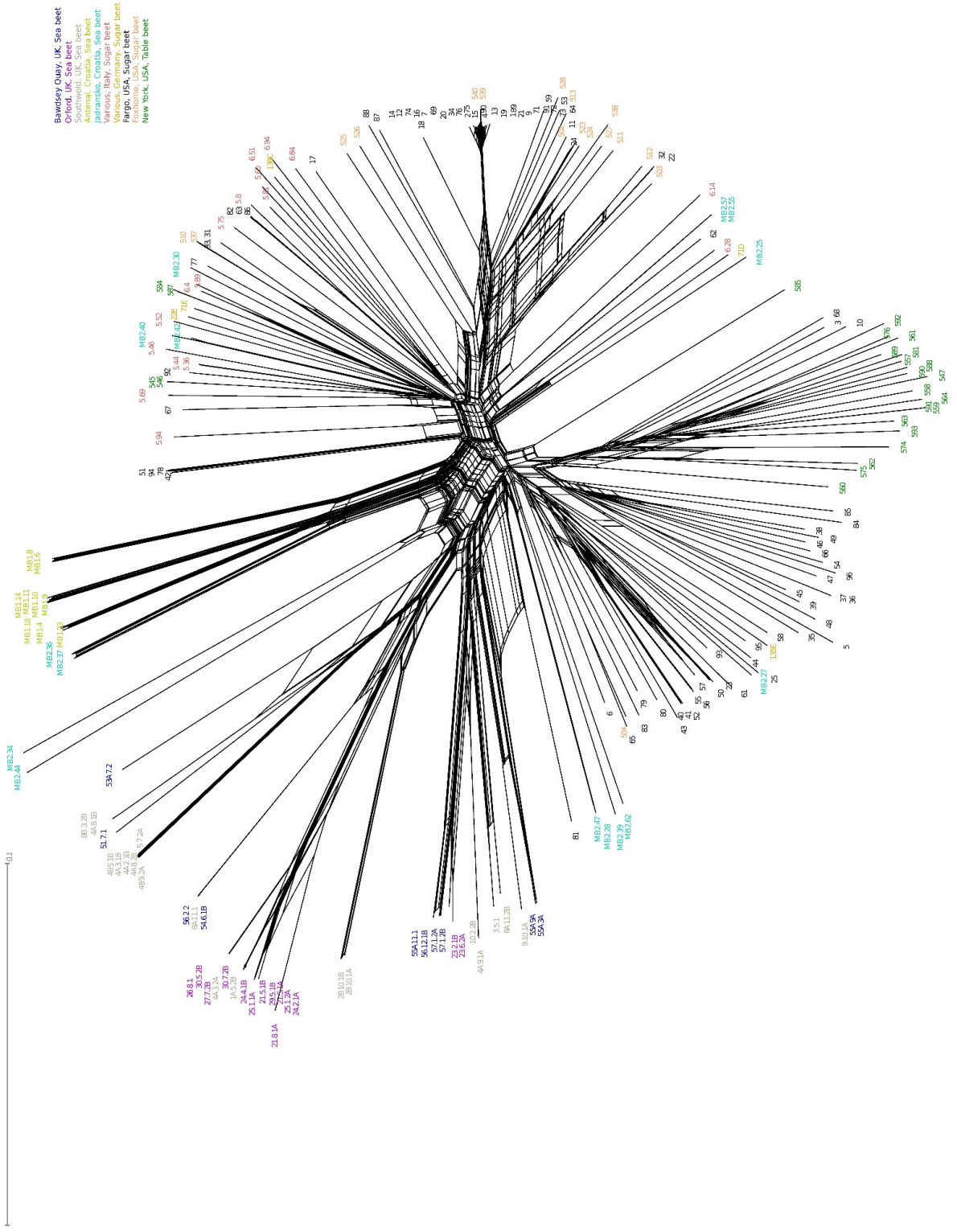

Bawdsey Quay, UK, Sea beet
Orford, UK, Sea beet
Southwold, UK, Sea beet
Anterak, Croatia, Sea beet
Rovinj, Croatia, Sea beet
Various, Italy, Sugar beet
Various, Germany, Sugar beet
Fargo, USA, Sugar beet
Foxhome, USA, Sugar beet
New York, USA, Table beet
0.1

Figure 7: The reticulation network of the *C. beticola* isolates included. The isolates from the UK and Anetnal were the only isolates with a distinct separation from the other isolates among themselves. The remaining isolates did not show clear separation, and were interspersed between each other in the network. Individual labels were highlighted based on sampling location as indicated in the key. *C. beticola* isolates from the UK clustered by themselves while the other isolates were more closely related to each other, and did not cluster by location. The reticulation events indicated in the center of the network showed that there was genetic exchange between isolates.

To further examine the relationship between isolates, a PCA was performed (Figure 8). The first component component explained 38% of the variation, and the second, 25%. As with the reticulation network, some of the isolates from the UK clustered separately from the other isolates while others were more closely related.

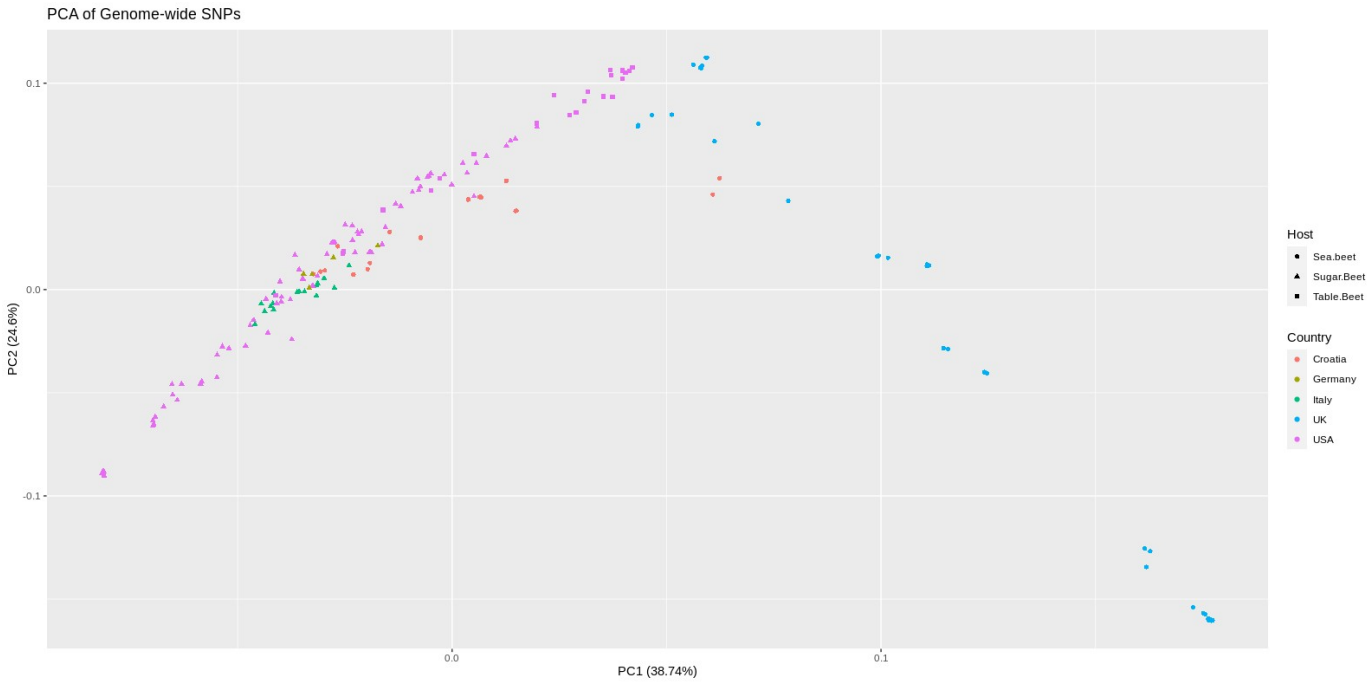


Figure 8: The principle component analyses of the *C. beticola* isolates included. The first component explained 38.74% of the variation observed among the isolates, and the second 24.6%. In agreement with the reticulation network, the isolates from the UK clustered

separately from the remaining isolates. Isolates collected from sea beet are indicated with circles, those from sugar beet with triangles, and the isolates from table beet were squares. The country that the isolates were collected from were shown with different colours.

The reticulation network, and the PCA showed no strong population separation based on either host or location, but that there were some differences between the various locations in North America and mainland Europe. To quantify the differences between the population ancestry between the individuals, the unlinked SNPs were extracted. Linkage distance was used to determine the distance at which SNPs were considered to be no longer linked.

Linkage distance instead of R^2 was used for filtering because none of the sampling locations had comparable R^2 profiles, but most chromosomes had a similar distance at which R^2 decayed to half of its maximum (Figure 9). Isolates collected from Antenal and Orford showed similar patterns of LD, with almost no reduction in the first 100 kb of most chromosomes, but started decreasing at greater distances with minima similar to that of the other sampling sites for each chromosome (Supplementary Figure 6, Supplementary Table 6). The R^2 decayed to half of its maximum at between 5 and 10 kb for each sampling site. As such, a global thinning distance of 5 kb was selected to ensure that SNPs used for the population ancestry were unlinked.

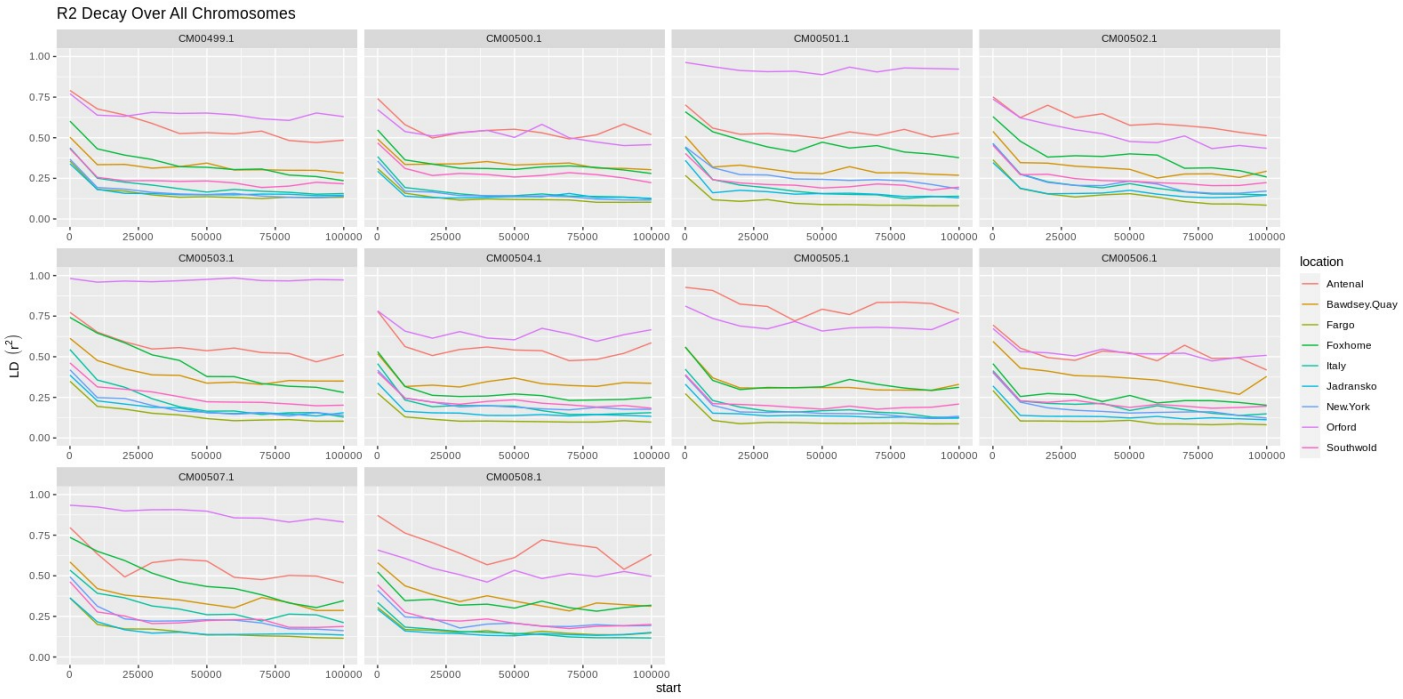


Figure 9: R^2 decay over the first 100 kb of each chromosome separated by sampling site. The R^2 value was not comparable between the populations. However, the R^2 value reduced to half of its maximum in all populations on each chromosome at 8 kb.

Following the thinning of the sites, the signals of admixture among the individuals was determined. Thinning to an 5 kb distance reduced the number of SNPs to 6184 variant positions across the entire genome. The cross entropy estimate did not reach a constant minimum, and was not informative for indicating an ideal number of populations present. The cross entropy estimation started decreasing at K = 2, and thus, K = 2-5 were shown since these were the only scenarios with likely biological explanations (Figures 10 and 11, Supplementary Figure 7). For each K, only the most supported iteration was shown.

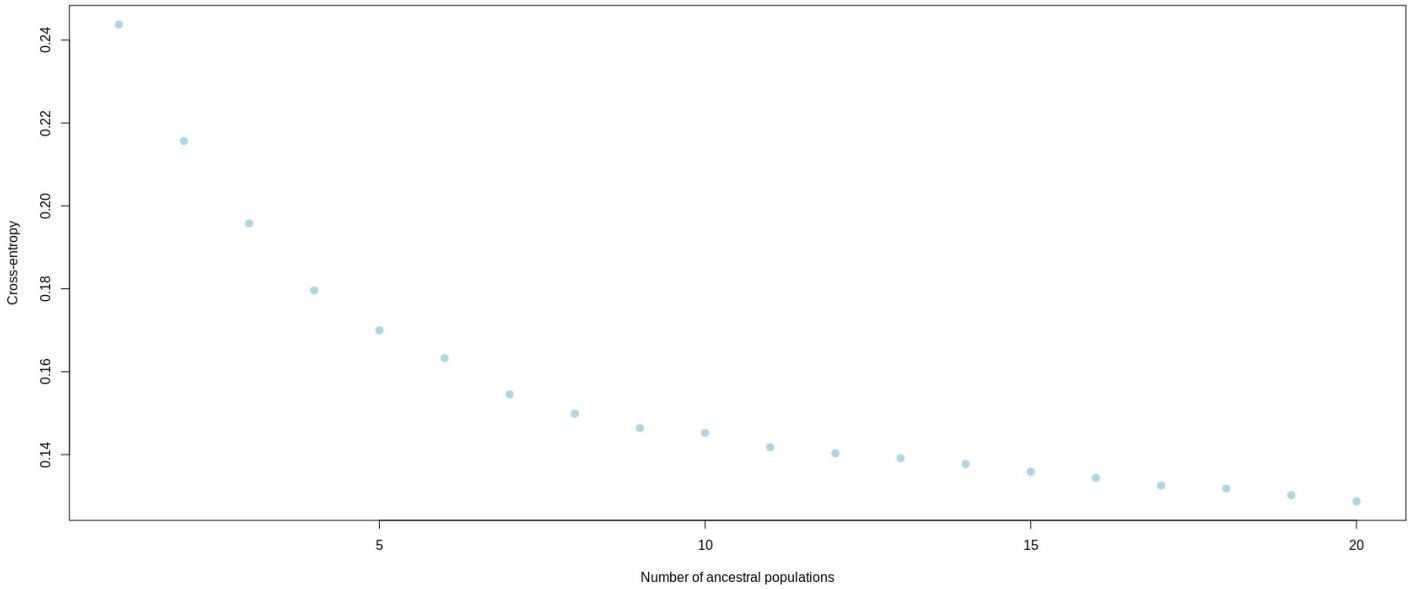


Figure 10: Cross entropy value for the different values of K after thinning the genome wide SNPs at 5 kb distances. At K=2 a decrease in cross entropy begins, but a value close to 0 was never achieved.

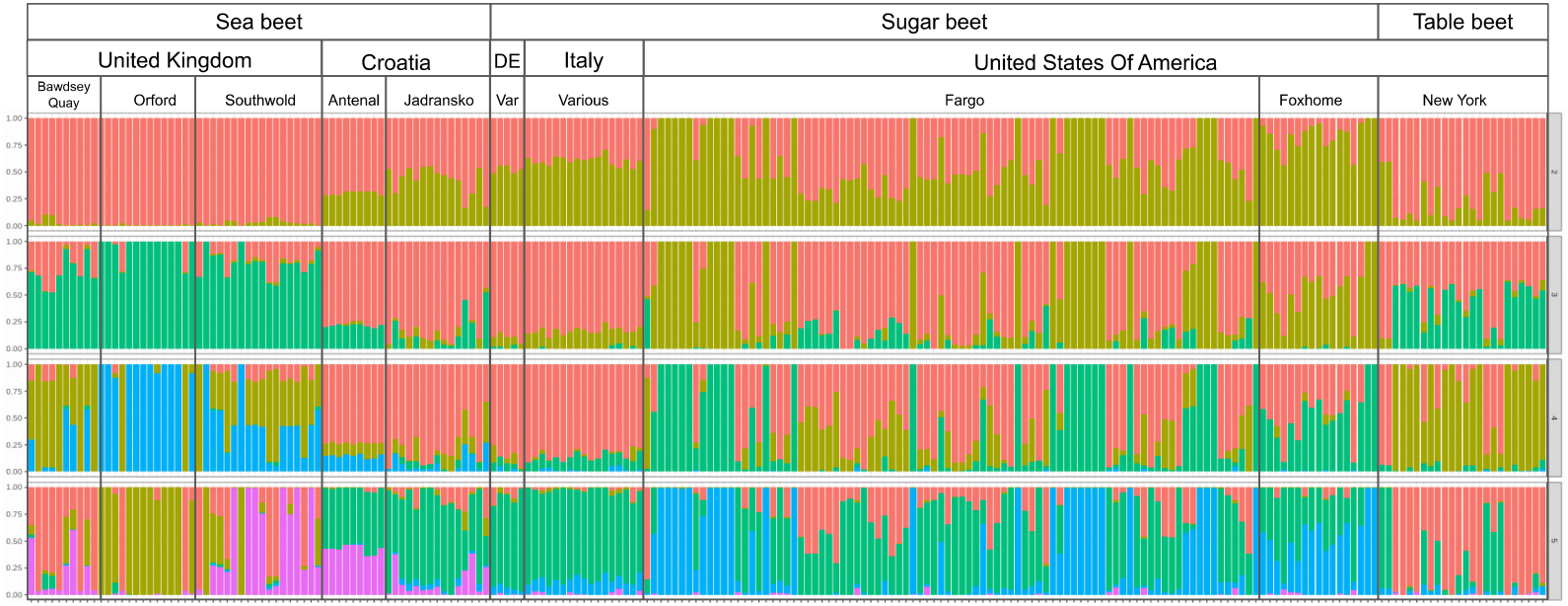

Sea beet
Sugar beet
Table beet
United Kingdom
Croatia
DE
Italy
United States Of America
Bawdsey Quay
Orford
Southwold
Antenal
Jadransko
Var
Various
Fargo
Foxhome
New York
1.00
0.75
0.50
0.25
0.00
2
3
4
5

Figure 11: Population structure of the *C. beticola* isolates. SNPs were thinned to 5 kb distance of one another. 6184 SNPs were used to perform the admixture analyses. K = 2-5 were selected as the most likely to explain the biological implications of the data. The isolates from the UK showed higher levels of admixture than isolates from mainland Europe and the USA.

Regions Differentiation the *C. beticola* Isolates from the UK from the Isolates from Mainland Europe and North America

To determine what distinguishes the UK population from the other four sampling sites, regions that were more differentiation were determined. The pairwise F_{ST} among all sampling locations showed that, on average, the isolates from the UK were more differentiated from the isolates from the other sampling sampling sites (Table 7). The pairwise F_{ST} between the UK population and the isolates from the other four broad sampling locations, Croatia, Italy, North Dakota, and New York, showed that most of the genome was not greatly differentiated. There were some regions that were higher than that of the surrounding regions (Figure 12).

Table 7: Weighted pairwise F_{ST} between *C. beticola* isolates from different sampling locations in 50 kb windows

Locations	**UK**	**Croatia**	**Italy**	**North Dakota**	**New York**
UK	0	0.26	0.32	0.33	0.24
Croatia		0	0.14	0.18	0.21
Italy			0	0.14	0.24
North Dakota				0	0.23
New York					0

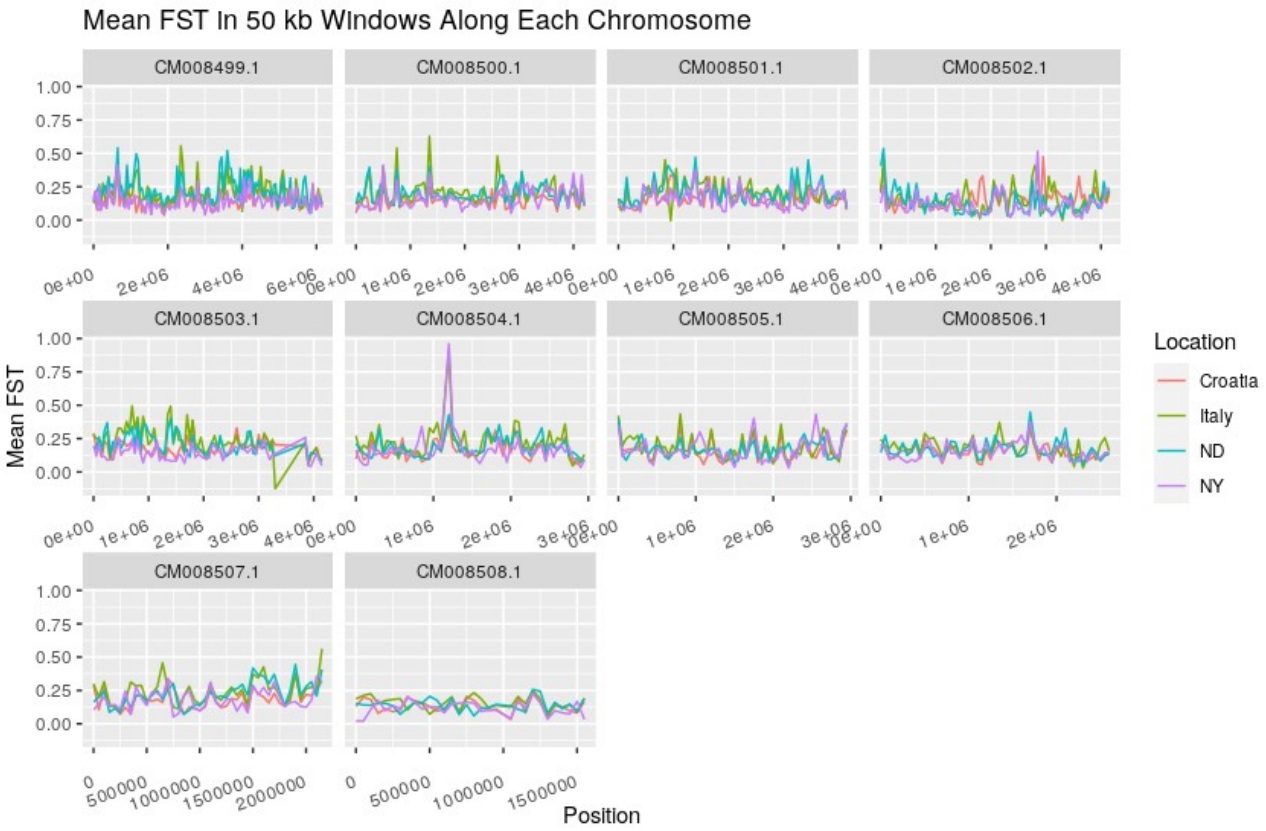


Figure 12: Pairwise F_{ST} in 50 kb windows between the isolates from the UK and the four other sampling sites, Croatia, Italy, North Dakota, and New York, per chromosome.

The average F_{ST} between the sampling locations was comparable (Table 8). The mean F_{ST} between the various sites and the UK ranged between 0.147 and 0.201. The F_{ST} between the isolates from sea beet in the UK showed that these isolates were the least differentiated, and that the UK population was the most differentiated from the *C. beticola* isolates from sugar beet found in Italy. While the mean F_{ST} between locations was comparable, there were regions that showed a higher fixation index than the mean in each sampling location (Table 9). Here, only the global maxima of the pairwise F_{ST} indexes were considered for further characterisation.

Table 8: Mean and median F_{ST} between *C. beticola* isolates from different sampling locations relative to isolates from the UK along the genome in 50 kb windows

Location	Host	Mean F_{ST}		Weighted F_{ST}	
		Mean	Median	Mean	Median
Croatia	Sea beet	0.147	0.158	0.245	0.235
Italy	Sugar beet	0.201	0.187	0.304	0.289
North Dakota	Sugar beet	0.191	0.178	0.310	0.296
New York	Table beet	0.150	0.137	0.222	0.202

Table 9: Regions identified from the pairwise F_{ST} analysis along the genome in 50 kb windows. Values in bold typeface indicate F_{ST} outliers

Chromosome	Contig coordinate of the start of the outlier region	Mean F_{ST} per Sampling Location			
		Croatia Sea beet	Italy Sugar beet	North Dakota Sugar beet	New York Table beet
CM008499.1	650001	0.357	0.420	**0.637**	0.410
CM008502.1	2950001	**0.635**	0.054	0.186	0.124
CM008504.1	1200001	0.382	**0.974**	0.428	**0.980**

Characterising the Region on Conting CM008504.1 Located at 1.2 Mb

The region with a high degree of differentiation was detected with large windows of 50 kb. The region preceding the window with the global maximum between the isolates from Italy and New York, and the isolates from the UK, consisted of solely ambiguous characters. As such, no genes could be annotated within this region. As an example, the *de novo* assembly of isolate 10.2.2B to bridge the ambiguous region will be shown. Additional assemblies were also used, but the region could not be characterised.

Characterising the F_{ST} Outlier Region on Contigs CM008499.1 and CM008502.1

The regions of the chromosomes surrounding the target region was well assembled and did not contain many ambiguous characters. The gene annotations of ten genes upstream and downstream of the start of the outlier regions were extracted and compared to the NCBI database, and the best characterised match was extracted (Supplementary Tables 7 and 8). The genes contained in the regions 50 kb and and downstream were varied, and had different F_{ST} when the genes from the various sampling locations were compared to the UK (Supplementary Tables 9 and 10, Figures 13 and 14). Most genes were not effectors, and did not contain signal peptides (Table 10).

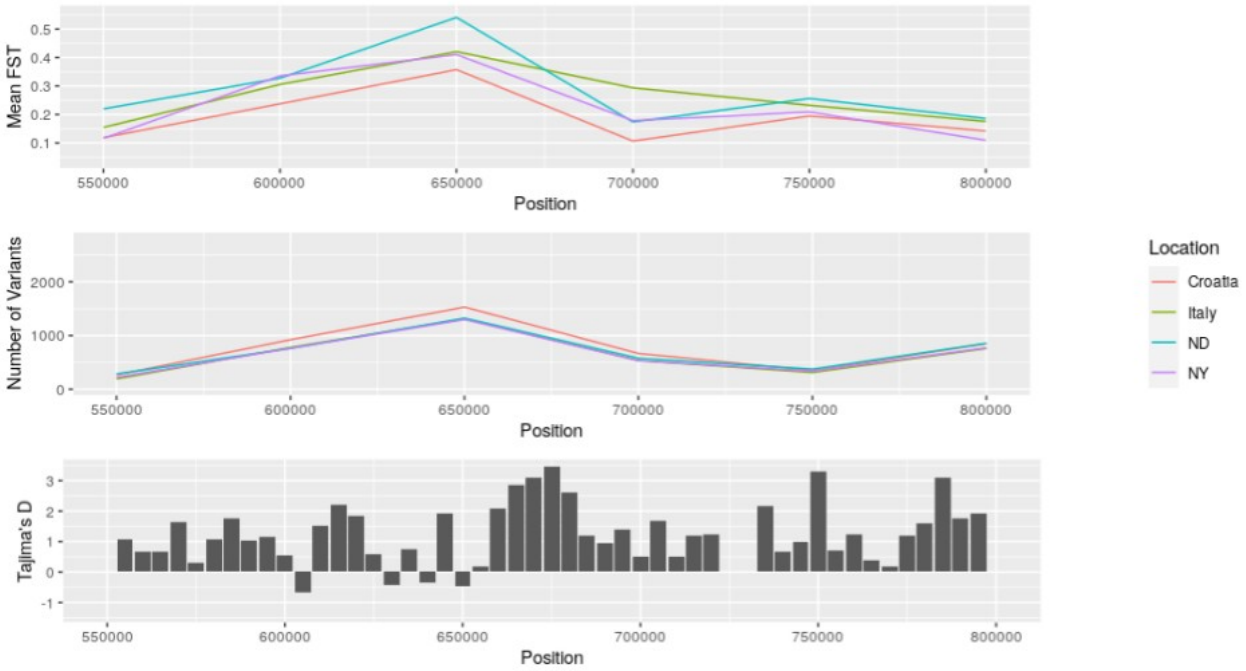


Figure 13: The pairwise F_{ST} between the *C. beticola* isolates from the UK and the other locations around the 650 kb region of the CM008499.1 chromosome, the number of variants in the region, as well as the Tajima's D distribution for the corresponding region. There was an increase in Tajima's D directly after the region that showed a higher F_{ST} than the surrounding regions.

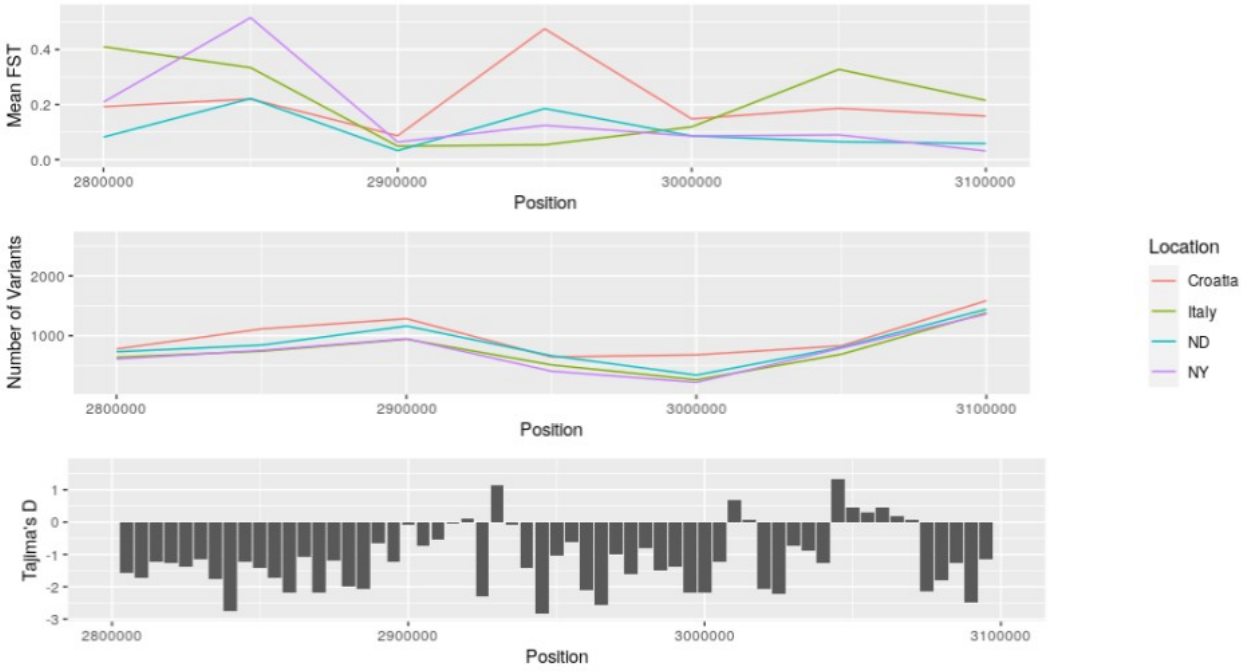


Figure 14: The pairwise F_{ST} between the *C. beticola* isolates from the UK and the other locations around the 2.95 Mb region of the CM008502.1 chromosome, the number of variants in the region, as well as the Tajima's D distribution for the corresponding region. There was a decrease in Tajima's D directly before the region that showed a higher F_{ST} than the surrounding region.

Table 10: Notable gene exceptions that are likely involved in the interaction of *C. beticola* with its hosts

Chromosome	mRNA ID	Signal peptide	Effector	Description	F_{ST}	Tajima's D
CM008499.1	XM_023592896.1	No	No	HC-toxin efflux carrier	Croatia: 0.24 Italy: 0.47 ND: 0.29 NY: 0.48	1.83
	XM_023592900.1	Yes	No	No similarities	Croatia: 0.04 Italy: 0.008 ND: 0.28 NY: 0.04	2.42
	XM_023592906.1	No	No	Tubulin gamma chain	Croatia: 0.39 Italy: 0.43 ND: 0.66 NY: 0.46	-1.05
	XM_023592912.1	No	No	Zinc finger protein	Croatia: 0.57 Italy: 0.15 ND: 0.35 NY: 0.57	0.03
	XM_023592913.1	No	No	No similarities	Croatia: 0.38 Italy: 0.5 ND: 0.51 NY: 0.21	2.8
	XM_023592916.1	No	No	Vegetative incompatibility protein HET-E-1	Croatia: 0.56 Italy: 0.53 ND: 0.76 NY: 0.56	0.2
	XM_023592924.1	Yes	Yes	60S ribosomal protein L14-B	Croatia: 0.21 Italy: 0.45 ND: 0.26 NY: 0.28	0.16
	XM_023592926.1	No	No	Phosphatidylserine decarboxylase proenzyme 3	Croatia: 0.11 Italy: 0.61 ND: 0.13 NY: 0.2	4.52
CM008502.1	XM_023597588.1	No	No	Ethanolamine-phosphate cytidylyltransferase	Croatia: 0.29 Italy: 0.37 ND: 0.02 NY: 0.41	0.22
	XM_023597609.1	Yes	No	Alpha-L-fucosidase	Croatia: 0.69 Italy: 0.003 ND: 0.4 NY: 0.17	4.36
	XM_023597610.1	No	No	O-methylsterigmatocystin oxidoreductase	Croatia: 0.75 Italy: 0.04 ND: 0.35 NY: 0.14	2.9
	XM_023597613.1	No	Yes	Putative oxidoreductase	Croatia: 0.75 Italy: 0.14 ND: 0.17 NY: 0.14	2.52

The region spanning XM_023597609.1 to XM_023597613.1 was further analysed as several genes next to each other showed a high level of differentiation between the isolates from UK and Croatia. Within the genome, this region was 11.6 kb long. This corresponded closely with the R^2 determined, and indicated that SNPs within this region were linked in the isolates from Croatia. The XM_023597611.1 did not contain any variants and it is likely evolutionary conserved as it is a mediator of RNA polymerase, a conserved function. Two genes within this region were predicted to encode oxidoreductases. In XM_023597613.1, the mutations were only synonymous mutations (Supplementary Table 11). Within the XM_023597610.1 mRNA, isolates from the different locations harboured different frequencies of early stop codons at the first SNP (Supplementary Table 11). Isolates from the UK had the lowest frequency of the premature stop codon, and it was present at a frequency of 0.05. Similarly, in isolates from Italy the premature stop codon was also present in 6% of isolates. 25% of isolates from New York, and 51% of isolates from North Dakota contained this mutation. In isolates from Croatia, this premature stop codon was present in 79% of individuals. The product of XM_023597613.1 was predicted to be a O-methylsterigmatocystin oxidoreductase that was homologous to proteins that play important roles in plant-pathogen interactions in other phytopathogens. The individuals that did not contain the premature stop codon clustered with the isolates from North America and Italy in the reticulation network, that indicated that this mutation may be beneficial for the survival and successful infection of *C. beticola* in domesticated beet. The individuals that do not contain this premature stop codon, may be indicators that there are some genetic barriers to the ability of all *C. beticola* isolates to be transferred between wild and domesticated hosts.

In summary, our detailed inspection of outlier regions revealed a set of candidate genes, some of which may play a role in plant-pathogen interactions. The presence of the premature stop codon in XM_023597613.1 showed that there may be differences between isolates collected from wild and domesticated beets, and that these variants may be important drivers of future host specialisation.

Discussion

This study aimed to determine the effect of host domestication on the evolution of a fungal pathogen that was associated with both hosts. To quantify this effect, a variety of measures were considered, including genetic diversity, linkage decay, population structure, and F_{ST}. Regardless of which measures were included, a clear distinction between whether host or location influenced the evolution of the fungus could not be confirmed. Overall, there was little support for strong population structure or genetic diversity based on host or location. What was well supported in all analyses of genetic distance was that the individuals from the UK were different from the isolates from wild and domesticated hosts from mainland Europe and the USA. Here, we will discuss findings from this system in the broader scope of existing literature, and evaluate whether this system can be used as an insight into the early stages of pathogen specialization due to host domestication.

Genetic Diversity of *C. beticola* from Sea Beet in the UK

The isolates from the UK were isolated from three distinct sampling locations. The locations differed in the level of human impact on the environment that has resulted in a change in the plants surrounding the sea beet plants. The first site, Southwold, was a small commercial harbour with sea beet plants growing along the edges of the walkways with little competition from other plants. The second site, Orford, was next to a naval base, and next to a sugar beet field. There, the sea beet plants grew among tall grasses, and the density was much higher than at the other two sites. Interestingly, the sugar beet field next to the infected sea beet plants did not show signs of infection along the edges of the field aside from a single plant that had two CLS lesions. The third site, Bawdsey Quay, was a quay with fewer sea beet plants than the other two sites. Wind is an important dispersal mechanism for *C. beticola* spores, and alongside human activity, may have contributed to the dispersal of *C. beticola* among plants (Lawrence and Meredith, 1970; Weiland and Koch, 2004).

The three sites showed different degrees of genetic diversity. Isolates from Southwold were the most diverse, and had the higher number of effective haplotypes. Each site had several clones isolated from different plants that showed that short range dispersal occurs within each site. Each site had a high percentage of private haplotypes that showed that there is not much dispersal between the three sites. Southwold shared a single haplotype with Bawdsey Quay, and a different haplotype with Orford. These isolates may have been transported between sites by human activity. The haplotype that is shared by Orford and Southwold occurred on four separate plants in Orford and on a single plant in Southwold. It was likely that the isolate was transferred from Orford to Southwold. The shared haplotype was found on a single plant in both sites, and no direction of infection can be deduced from this. The lower levels of genetic diversity between individuals at Bawdsey Quay could be attributed to the size and abundance of sea beet at the site. The harsher weather conditions might also contribute to the selection of individuals that are able to establish infect the plant more quickly before the wind blows the spores away. Isolates from Southwold and Orford were not challenged by the environment as much, which could have resulted in the sites accumulating isolates with a range of virulence as they might undergo less pressure to infect plants quickly. An abundance of sea beet plants may also reduce the pressure on the isolates to infect plants more quickly. Future cross infection studies of these isolates on sea and sugar beet would shed light on whether this is indeed the strategy used by these isolates. Additionally, *in planta* infection assays would also be important to determine whether these isolates are able to infect sugar beet to explain why the sugar beet field was not infected by *C. beticola*.

The isolates collected and used for clone correction were sampled using a different strategy from previous studies of *C. beticola*. One of the most recent *C. beticola* studies using microsatellite markers used in this study sampled few plants very intensively and comprised of 649 individuals from various locations in the USA compared to the 68 isolates used here (Vaghefi et al., 2017b). Isolates were clustered based on the sampling site, and Nei's index of gene diversity (H_e) ranged between 0.29 and 0.57, drastically lower than that of isolates included in this study. This difference may be due to differences in sampling strategies and sample size. An explorative study of *C.*

beticola isolated from table beet in New York also showed an H_e ranging from 0.045 to 0.566 (Knight et al., 2018). A more in-depth survey of *C. beticola* across the USA and Europe showed H_e ranging from 0.05 to 0.64 (Knight et al., 2019b). In these studies, the sample sizes from some different fields were comparable to the ones used in this study. Therefore, the *C. beticola* populations from the sites in the UK were shown to be more diverse than the isolates from table beet and sugar beet.

Fungal isolates from wild and agricultural hosts show differences in genetic diversity. In *S. sclerotiorum* the inverse of what was seen in *C. beticola* was observed (Kohn, 1995). Isolates from Canadian canola and Norwegian *Ranunculus ficaria*, fungal diversity was higher in the isolates from agricultural fields. Due to the distribution of DNA fingerprints, it was difficult to distinguish clonality from inbreeding in isolates from wild populations. In most cases, however, pathogens of wild species harbour more diversity than pathogens of domesticated crops (reviewed Burdon and Thrall, 2008). Wild plant populations are more diverse than crops, and therefore drive the maintenance in the diversity within their associated pathogens.

Does Location of Host Domestication Have a Clear Influence on Population Structure?

To assess the presence or absence of population structure, a reticulation network, a PCA, and an admixture analysis were performed. There is a difference between the analyses that can be performed downstream differ on data sets that contain individuals that form clear, and well supported populations compared to those that do not have clear structure or high levels of differentiation. Among the *C. beticola* isolates used, there was no clear distinction clustering of individuals based on either host or location. Instead, there was a gradient in most cases, with some individuals from that were very closely related even though they were isolated from different hosts or continents, while others were less closely related to isolates that share a host or a location with them.

In the phylogentic analyses, the isolates from the UK were the most distantly related to the isolates from mainland Europe and the USA. Within the UK population, there was variation between the

individuals regarding how closely they were related to the remaining isolates. The first and second components of the PCA illustrated this well. Regardless of the host they were collected from, the isolates from mainland Europe and the USA clustered more closely together. Although it was not definitely separated from the main cluster, a second cluster containing the individuals from the UK as well as two individuals from Croatia. Unlike many fungal pathogens that have been isolated on different hosts, the reticulation network did not show strong differentiation among the isolates.

The LD differences between the various sampling locations was indicative of sexual reproduction occurring within the species. LD has been shown to be closely correlated to the frequency of sexual reproduction a fungus undergoes (Nieuwenhuis and James, 2016). In species that undergo very frequent or obligatory sexual reproduction, LD half-decay is very short as in *Schizophyllum commune* where this distance is merely 110 bp (Baranova et al., 2015; Nieuwenhuis and James, 2016). In species that don't undergo sexual reproduction and are known to be highly clonal, the distance at which LD reaches its half-decay value exceed 100 kb. In the case of *Candida albicans*, the distance of LD half decay is 162 kb (Hirakawa et al., 2015; Nieuwenhuis and James, 2016). The LD half-decay of *C. beticola* was comparable to that of other fungi that have mixed modes of reproduction, i.e. where both sexual and clonal reproduction occurs. In *Saccharomyces cerevisiae* the LD half-decay is 2.3 kb while in *Saccharomyces pombe* it is closer to 20 kb (Bergström et al., 2014; Nieuwenhuis and James, 2016). It has also been shown that LD half-decay can vary among populations of the same species (Derbyshire et al., 2019). As LD half-decay is closely associated with reproduction strategy, the presence of LD decay in *C. beticola* strongly suggests that sexual reproduction occurs at different rates within, and potentially between isolates from the different locations. The reticulation network further confirmed that there was genetic exchange among individuals (Huson and Bryant, 2006). The differences in the LD decay among chromosomes could be indicative of different recombination rates among chromosomes. Future studies to detail the recombination landscape of *C. beticola* would be required to confirm this.

To further confirm that the *C. beticola* isolates did not cluster in distinct populations, admixture analyses were used. The population structure analysis showed a high level of admixture among *C.*

beticola isolates included in this study, and that there were different patterns of admixture based on both host and geography. Fungal plant pathogens have been known to cluster in populations based on host, as well as geographic origin. In a study considering whole genome SNPs of *Sclerotinia sclerotiorum* isolates from different hosts, the corresponding reticulation network showed two distinct clusters as well as several genotypic outliers (Derbyshire et al., 2019). These populations were further supported by an admixture analyses that showed that the clusters were indeed supported as distinct populations. In other fungal phytopathogens such as *Botrytis cinerea*, the lack of structure associated with geography has been attributed to frequent migration and/or large populations (Walker et al., 2015). However, *B. cinerea* populations are often subdivided based on the host they were isolated from (Mercier et al., 2020). In the case of *C. beticola*, it could be shown that the host that the fungus was isolated from did not influence the population structure.

The lack of strong population structure among *C. beticola* isolates is unsurprising. Fungal pathogens of many of the major crops have had several thousand years to establish a co-evolutionary dynamic (Lo Presti et al., 2015; Stukenbrock et al., 2007). Due to the co-evolutionary dynamic, differences in evolutionary pressures in wild ecosystems and agro-ecosystems drive different evolutionary trajectories in fungal pathogens (Kohn, 1995; Stukenbrock and McDonald, 2008). Fungal populations contain substantial genetic variation that allows for the rapid evolution of advantageous phenotypes (Barrett and Schluter, 2008). This allows phytopathogens to rapidly overcome host resistance, as well as evolve to adapt to new environments (Brown, 2015; Rouxel et al., 2003). While co-evolution of fungi with the host occurs within wild plant populations, the selective pressures are not always as strong as those presented by agro-ecosystems (Burdon and Thrall, 2009; Möller and Stukenbrock, 2017).

The comparison of *C. beticola* to a more ancient crop pathogen shows the effect of a longer co-evolution relationship. *Zymoseptoria tritici* is the causal agent of *Septoria tritici* blotch (STB) on wheat (reviewed by McDonald et al., 2015). It has been hypothesised that *Z. tritici* evolved from an ancestral species that colonised grasses ~11000 years ago and speciation occurred alongside wheat domestication (Stukenbrock et al., 2007). *Z. tritici* populations have high genetic diversity, as

well as a high rate of adaptive substitutions (Grandaubert et al., 2019; Linde et al., 2002; Mekonnen et al., 2020). The pangenome of *Z. tritici* shows an equal plasticity (Plissonneau et al., 2018). Isolates from different geographic regions also cluster in distinct populations (Linde et al., 2002; Vagndorf et al., 2018). Given the large evolutionary scale that has given rise to the variation in *Z. tritici*, it is expected that a modern crop, like sugar beet, and its pathogens may not show the same variation, structure, and differentiation as seen in an ancient species. To date, global population genetics surveys of other pathogens that share wild and domesticated beets as host has not been done.

The high degree of gene seen between isolates from mainland Europe and North America may be a cause of concern for plant breeders. In the pathosystem of *Puccinia graminis* and wild and domesticated oats, gene flow from pathogens from wild oats to those on domesticated oat resulted in the evolution increased virulence of *P. graminis* isolates on domesticated oat (Burdon et al., 1992; Oates et al., 1983). The fungal pathogen *Magnaporthe oryzae* causes several blast diseases on rice and various grasses. Gene flow among between multiple lineages of this fungus from different hosts occurs, and detection thereof has indicated the need for surveillance against emerging plant diseases, particularly in the case of multi-host pathogens (Gladieux et al., 2018).

Taken together, the population genomics confirmed what previous microsatellite studies have suggested. These studies have suggested the *C. beticola* exists as a global population with little differentiation, and high levels of gene flow (Groenewald et al., 2008; Knight et al., 2019b; Vaghefi et al., 2017a). These studies did not consider isolates collected from sea beet. Here, it was shown that the same assertions hold true for these accessions, as well.

What Differentiates the Isolates Found on Different Hosts and at Different Locations?

While strong population structure does not exist within the *C. beticola* global population, there are some differences between isolates collected from the different sites. For this, additional population genomic statistics were used to locate regions that could be of biological relevance.

The statistics used in this study can illustrate several biological scenarios. Tajima's D is a statistic that measures the mean number of pairwise differences between sequences and the number of segregating sites (Tajima, 1989). This can be used to evaluate whether a sequence is evolving neutrally or whether it is undergoing non-random evolution. If Tajima's D is close to 0, usually including values from -1 to 1, the sequence is considered to evolve neutrally. If Tajima's D is positive, usually greater than 1, it is due to the lack of rare alleles. This may be caused by balancing selection, or a sudden population contraction. If Tajima's D is negative, usually smaller than -1, there is an excess of rare alleles in the population. This can be caused by due to a recent selective sweep, or recent population expansion. F_{ST} is a measure of relative divergence between individuals developed from Wright's F-statistic (Wright, 1950). An $F_{ST} = 0$ indicates no divergence between individuals while $F_{ST} = 1$ indicates no similarity between individuals. F_{ST} can be used to measure demographic history of selected regions within the genome (Holsinger and Weir, 2009).

As more whole genome data is becoming available for population genomics studies, the methods employed to detect signatures of selection and differentiation change. Until recently, demographic analyses were an important aspect of determining the presence of genes under selection. However, in the past two years, several studies have shown that using a combination of Tajima's D distribution and pairwise F_{ST}, regions that are undergoing selection can be identified accurately (Mercier et al., 2020). This was particularly advantageous for this study as most selection detection approaches rely on very clear population structure. As explained above, *C. beticola* isolates included in this study did not show clear population structure, and as such, most methods cannot be used without overextending the assumptions these models incorporate. However, F_{ST} is a measure of relative divergence, and can also detect signals that are not related to external selection pressures (Schirrmann et al., 2018).

By extracting regions were the F_{ST} was higher than the rest of the genome, genes surrounding the area most impacted by differentiation could be assessed. In other studies that refined the data by using both relative and absolute divergence outliers, the regions surrounding the genes with the highest divergence are often not considered, only the genes that are outliers (e.g. Mercier et al.,

2020). Some studies do, in fact, consider the genomic landscape surrounding the location of a selective sweep (e.g. Derbyshire et al., 2019). Due to time constraints, this study solely made use of the latter approach. Future studies including these isolates should extend to the entire genome, considering regions that are less pronounced than the two regions considered here.

Unsuprisingly, some of the genes that show differences in their F_{ST} when compared to the UK population. An Alpha-L-fucosidase has been shown to be over-expressed when the Dutch elm pathogen, *Ophiostoma novo-ulmi* changes its morphology between a yeast and mycelium stage (Nigg et al., 2015). The putative oxidoreductase encoded for by XM_023597613.1 is an interesting target for future studies. In some populations there was a high rate of synonymous mutations, and no non-synonymous mutations in any individual. While the function of this enzyme has not been characterised in any related species, this may be indicative that there is a strong pressure on this enzyme to maintain its integrity. The O-methyl strigmatocystin oxidoreductase encoded by XM_023597610.1 is homologous to oxidoreductases in *Aspergillus* species that function in the final step in the production of the aflatoxin B_1 (Gengan et al., 2006). The premature stop codon occurred fairly early in the mRNA, and indicated that this enzyme is likely inactive in the accessions that contained it. Further, there was a difference in the distribution of individuals that contained this stop codon. Within the North American individuals, this stop codon was present in a random manner. However, in the isolates from Croatia, the individuals that contained the stop codon clustered separately in the reticulation network while those that did not contain it were interspersed among the North American isolates (Supplementary Figure 7). This may be indicative that this pathway has a functional role within *C. beticola*, as well. It may also be indicative that other genes within this region are under selection, and that this stop codon is merely hitch-hiking and has no functional effect. Further studies would be important to confirm this. These studies include *in planta* characterisation of expression of this gene and regions surrounding it during cross-infection studies on wild and domesticated beet that consider specifically isolates that are closely related, but represent different XM_023597610.1 genotypes.

Conclusions and Perspectives

This study has shown that, as of yet, *C. beticola* does not show strong signs of distinct evolution driven by host domestication. This study confirmed that *C. beticola* exists as a global population rather than distinct populations based on host or location. However, there were some differences between isolates collected from the various locations. The *C. beticola* population sampled in the UK showed higher levels of diversity than other previously sampled populations. This may be due to the genetic diversity of hosts driving higher diversity in the fungal populations. The geographic isolation of these isolates minimises the opportunity for them to undergo admixture with other isolates from mainland Europe. This may allow for the maintenance of standing genetic diversity. While the Croatian isolates were also isolated from sea beet, these isolates were not as diverse as those from the UK, and are admixed with isolates from sugar beet. Whole genome studies on unique isolates showed that there was no strong population structure based on host, but that the clusters from different locations showed that there were different signatures of admixture within the locations. Using population genomics approaches, two target regions that are likely under selection were identified. Genes contained within these regions were often stress-related genes. There was a single oxidoreductase within the highly divergent region that contained a premature stop codon in some isolates.

Following this study, several open questions pertaining to *C. beticola* remain. Firstly, *in planta* experiments should be performed to validate the ability of *C. beticola* isolates from sea beet and sugar beet to infect the other respective host. This would also test whether the divergent region on chromosome CM008502.1 is an important target for emerging host specialisation. Secondly, a survey of *C. beticola* on sugar beet in the UK should be performed to determine whether the *C. beticola* isolates infecting domesticated beet there are significantly different from the isolates infecting sea beet, as well as those infecting sugar beet in Europe and the USA. Thirdly, the direction of gene flow must be examined to establish whether the isolates from sea beet donate genetic material to the isolates in the domesticated fields through sexual reproduction. Finally, *in*

vitro expression analyses of the regions identified by this study should be performed to qualify whether they are functionally important in the lifecycle of *C. beticola*.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B., and Singh, K.B. (2010). Plants versus pathogens: an evolutionary arms race. Funct. Plant Biol. FPB *37*, 499–512.

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

Arnaud, J.-F., Fénart, S., Cordellier, M., and Cuguen, J. (2010). Populations of weedy crop–wild hybrid beets show contrasting variation in mating system and population genetic structure. Evol. Appl. *3*, 305–318.

Auwera, G.A.V. der, Carneiro, M.O., Hartl, C., Poplin, R., Angel, G. del, Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr. Protoc. Bioinforma. *43*, 11.10.1-11.10.33.

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., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. *19*, 455–477.

Baranova, M.A., Logacheva, M.D., Penin, A.A., Seplyarskiy, V.B., Safonova, Y.Y., Naumenko, S.A., Klepikova, A.V., Gerasimov, E.S., Bazykin, G.A., James, T.Y., et al. (2015). Extraordinary Genetic Diversity in a Wood Decay Mushroom. Mol. Biol. Evol. *32*, 2775–2783.

Barrett, R.D.H., and Schluter, D. (2008). Adaptation from standing genetic variation. Trends Ecol. Evol. *23*, 38–44.

Bartsch, D. (2010). Gene flow in sugar beet. Sugar Tech *12*, 201–206.

Bergström, A., Simpson, J.T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A.N., Moses, A.M., Louis, E.J., Mustonen, V., et al. (2014). A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. Mol. Biol. Evol. *31*, 872–888.

Biancardi, E., Panella, L.W., and Lewellen, R.T. (2011). *Beta maritima*: The Origin of Beets (Springer Science & Business Media).

Brown, J.K.M. (2015). Durable Resistance of Crops to Disease: A Darwinian Perspective. Annu. Rev. Phytopathol. *53*, 513–539.

Burdon, J.J., and Thrall, P.H. (2008). Pathogen evolution across the agro-ecological interface: implications for disease management. Evol. Appl. *1*, 57–65.

Burdon, J.J., and Thrall, P.H. (2009). Coevolution of plants and their pathogens in natural habitats. Science *324*, 755–756.

Burdon, J.J., Marshall, D.R., and Oates, J.D. (1992). Interactions between wild and cultivated oats in Australia.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) *6*, 80–92.

Clarke, J.D. (2009). Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation. Cold Spring Harb. Protoc. *2009*, pdb.prot5177.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

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., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

De Gracia, M., Cascales, M., Expert, P., Bellanger, M.-N., Le Cam, B., and Lemaire, C. (2015). How Did Host Domestication Modify Life History Traits of Its Pathogens? PLoS ONE *10*.

Delcher, A.L., Salzberg, S.L., and Phillippy, A.M. (2003). Using MUMmer to Identify Similar Regions in Large Sequence Sets. Curr. Protoc. Bioinforma. *00*, 10.3.1-10.3.18.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Derbyshire, M.C., Denton-Giles, M., Hane, J.K., Chang, S., Mousavi-Derazmahalleh, M., Raffaele, S., Buchwaldt, L., and Kamphuis, L.G. (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*.

Dohm, J.C., Lange, C., Holtgräwe, D., Sörensen, T.R., Borchardt, D., Schulz, B., Lehrach, H., Weisshaar, B., and Himmelbauer, H. (2012). Palaeohexaploid ancestry for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome (Beta vulgaris). Plant J. *70*, 528–540.

Dohm, J.C., Minoche, A.E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sörensen, T.R., Stracke, R., Reinhardt, R., et al. (2014). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature *505*, 546–549.

Eliades, N.G., and Eliades, D.G. (2009). Haplotype Analysis: software for analysis of haplotype data. For. Goettingen Ger. Genet. For. Tree Breed.

Ellstrand, N.C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T.J., Haccou, P., Lu, B.-R., Snow, A.A., Neal Stewart, C., et al. (2013). Introgression of Crop Alleles into Wild or Weedy Populations. Annu. Rev. Ecol. Evol. Syst. *44*, 325–345.

Feurtey, A., Guitton, E., Coquerel, M.D.G., Duvaux, L., Shiller, J., Bellanger, M.-N., Expert, P., Sannier, M., Caffier, V., Giraud, T., et al. (2020). Threat to Asian wild apple trees posed by gene flow from domesticated apple trees and their “pestified” pathogens. Mol. Ecol. *29*, 4925–4941.

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

Frenkel, O., Peever, T.L., Chilvers, M.I., Özkilinc, H., Can, C., Abbo, S., Shtienberg, D., and Sherman, A. (2010). Ecological Genetic Divergence of the Fungal Pathogen *Didymella rabiei* on Sympatric Wild and Domesticated *Cicer* spp. (Chickpea). Appl. Environ. Microbiol. *76*, 30–39.

Frichot, E., and François, O. (2015). LEA: An R package for landscape and ecological association studies. Methods Ecol. Evol. *6*, 925–929.

Gengan, R.M., Chuturgoon, A.A., and Dutton, M.F. (2006). Kinetics of the Oxidoreductase Involved in the Conversion of O-methylsterigmatocystin to Aflatoxin B1. Prep. Biochem. Biotechnol. *36*, 297–306.

Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D.M., and Hood, M.E. (2008). Speciation in fungi. Fungal Genet. Biol. *45*, 791–802.

Giraud, T., Gladieux, P., and Gavrilets, S. (2010). Linking emergence of fungal plant diseases and ecological speciation. Trends Ecol. Evol. *25*, 387–395.

Gladieux, P., Zhang, X.-G., Róldan-Ruiz, I., Caffier, V., Leroy, T., Devaux, M., Glabeke, S.V., Coart, E., and Cam, B.L. (2010). Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. Mol. Ecol. *19*, 658–674.

Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J.L.N., Nhani, A., Chen, L., Terauchi, R., Lebrun, M.-H., Tharreau, D., et al. (2018). Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus *Magnaporthe oryzae*. MBio *9*.

Grandaubert, J., Dutheil, J.Y., and Stukenbrock, E.H. (2019). The genomic determinants of adaptive evolution in a fungal pathogen. Evol. Lett. *3*, 299–312.

Groenewald, M., Groenewald, J.Z., Linde, C.C., and Crous, P.W. (2007). Development of polymorphic microsatellite and single nucleotide polymorphism markers for *Cercospora beticola* (Mycosphaerellaceae). Mol. Ecol. Notes *7*, 890–892.

Groenewald, M., Linde, C.C., Groenewald, J.Z., and Crous, P.W. (2008). Indirect evidence for sexual reproduction in *Cercospora beticola* populations from sugar beet. Plant Pathol. *57*, 25–32.

Hanelt, P., Büttner, R., and Mansfeld, R. (2001). Mansfeld's encyclopedia of agricultural and horticultural crops (except ornamentals) (Springer).

Hirakawa, M.P., Martinez, D.A., Sakthikumar, S., Anderson, M.Z., Berlin, A., Gujja, S., Zeng, Q., Zisson, E., Wang, J.M., Greenberg, J.M., et al. (2015). Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res. *25*, 413–425.

Holsinger, K.E., and Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat. Rev. Genet. *10*, 639–650.

Horner, D.S., Pavesi, G., Castrignanò, T., De Meo, P.D., Liuni, S., Sammeth, M., Picardi, E., and Pesole, G. (2010). Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. Brief. Bioinform. *11*, 181–197.

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

Huson, D.H., and Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. Mol. Biol. Evol. *23*, 254–267.

Jones, P.D., Lister, D.H., Jaggard, K.W., and Pidgeon, J.D. (2003). Future Climate Impact on the Productivity of Sugar Beet (*Beta vulgaris L.*) in Europe. Clim. Change *58*, 93–108.

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., et al. (2018). Gene cluster conservation provides insight into

cercosporin biosynthesis and extends production to the genus *Colletotrichum*. Proc. Natl. Acad. Sci. 201712798.

Kelly, A.C., and Ward, T.J. (2018). Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. PloS One *13*, e0194616.

Knight, N.L., and Pethybridge, S.J. (2020). An improved PCR assay for species-specific detection and quantification of *Cercospora beticola*. Can. J. Plant Pathol. *42*, 72–83.

Knight, N.L., Vaghefi, N., Hansen, Z.R., Kikkert, J.R., and Pethybridge, S.J. (2018). Temporal Genetic Differentiation of *Cercospora beticola* Populations in New York Table Beet Fields. Plant Dis. *102*, 2074–2082.

Knight, N.L., Vaghefi, N., Kikkert, J.R., and Pethybridge, S.J. (2019a). Alternative Hosts of *Cercospora beticola* in Field Surveys and Inoculation Trials. Plant Dis. *103*, 1983–1990.

Knight, N.L., Vaghefi, N., Kikkert, J.R., Bolton, M.D., Secor, G.A., Rivera, V.V., Hanson, L.E., Nelson, S.C., and Pethybridge, S.J. (2019b). Genetic Diversity and Structure in Regional *Cercospora beticola* Populations from *Beta vulgaris* subsp. *vulgaris* Suggest Two Clusters of Separate Origin. Phytopathology® *109*, 1280–1292.

Kohn, L.M. (1995). The clonal dynamic in wild and agricultural plant–pathogen populations. Can. J. Bot. *73*, 1231–1240.

Lawrence, J.S., and Meredith, D.S. (1970). Wind dispersal of conidia of *Cercospora beticola.* Phytopathology *60*, 1076–1078.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760.

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

Linde, C.C., Zhan, J., and McDonald, B.A. (2002). Population Structure of *Mycosphaerella graminicola*: From Lesions to Continents. Phytopathology® *92*, 946–955.

Lischer, H.E.L., and Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics *28*, 298–299.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. Annu. Rev. Plant Biol. *66*, 513–545.

Lofgren, L.A., LeBlanc, N.R., Certano, A.K., Nachtigall, J., LaBine, K.M., Riddle, J., Broz, K., Dong, Y., Bethan, B., Kafer, C.W., et al. (2018). *Fusarium graminearum*: pathogen or endophyte of North American grasses? New Phytol. *217*, 1203–1212.

McDonald, M.C., McDonald, B.A., and Solomon, P.S. (2015). Recent advances in the *Zymoseptoria tritici*–wheat interaction: insights from pathogenomics. Front. Plant Sci. *6*.

Mekonnen, T., Haileselassie, T., Goodwin, S.B., and Tesfayea, K. (2020). Genetic diversity and population structure of *Zymoseptoria tritici* in Ethiopia as revealed by microsatellite markers. Fungal Genet. Biol. *141*, 103413.

Mercier, A., Carpentier, F., Duplaix, C., Auger, A., Pradier, J.-M., Viaud, M., Gladieux, P., and Walker, A.-S. (2019). The polyphagous plant pathogenic fungus *Botrytis cinerea* encompasses host-specialized and generalist populations. Environ. Microbiol. *21*, 4808–4821.

Mercier, A., Simon, A., Lapalu, N., Giraud, T., Bardin, M., Walker, A.-S., Viaud, M., and Gladieux, P. (2020). Population genomics reveals molecular determinants of specialization to tomato in the polyphagous fungal pathogen *Botrytis cinerea.* BioRxiv 2020.07.24.219691.

Michèle, P. (2018). OECD-FAO AGRICULTURAL OUTLOOK 2018-2027. 10.

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

Munkacsi, A.B., Stoxen, S., and May, G. (2007). Domestication of Maize, Sorghum, and Sugarcane Did Not Drive the Divergence of Their Smut Pathogens. Evolution *61*, 388–403.

Nieuwenhuis, B.P.S., and James, T.Y. (2016). The frequency of sex in fungi. Philos. Trans. R. Soc. B Biol. Sci. *371*, 20150540.

Nigg, M., Laroche, J., Landry, C.R., and Bernier, L. (2015). RNAseq Analysis Highlights Specific Transcriptome Signatures of Yeast and Mycelial Growth Phases in the Dutch Elm Disease Fungus *Ophiostoma novo-ulmi*. G3 GenesGenomesGenetics *5*, 2487–2495.

Oates, J.D., Burdon, J.J., and Brouwer, J.B. (1983). Interactions Between Avena and Puccinia Species. II. The Pathogens: *Puccinia coronata* CDA and *P. graminis* Pers. F. Sp. avenae Eriks. & Henn. J. Appl. Ecol. *20*, 585–596.

Petersen, T.N., Brunak, S., Heijne, G. von, and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Plissonneau, C., Hartmann, F.E., and Croll, D. (2018). Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. *16*.

Pool, V.W., and McKay, M.B. (1916). Relation of stomatal movement to infection by *Cercospora beticola*. J. Agric. Res. *5*, 1011–1038.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Auwera, G.A.V. der, Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv 201178.

Power, A.G., and Mitchell, C.E. (2004). Pathogen Spillover in Disease Epidemics. Am. Nat. *164*, S79–S89.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575.

Rodríguez Del Río, Á., Minoche, A.E., Zwickl, N.F., Friedrich, A., Liedtke, S., Schmidt, T., Himmelbauer, H., and Dohm, J.C. (2019). Genomes of the wild beets *Beta patula* and *Beta vulgaris* ssp. *maritima*. Plant J. Cell Mol. Biol. *99*, 1242–1253.

Rouxel, T., Penaud, A., Pinochet, X., Brun, H., Gout, L., Delourme, R., Schmit, J., and Balesdent, M.-H. (2003). A 10-year Survey of Populations of *Leptosphaeria maculans* in France Indicates a Rapid Adaptation Towards the Rlm1 Resistance Gene of Oilseed Rape. Eur. J. Plant Pathol. *109*, 871–881.

Santoni, S., and Bervillé, A. (1992). Characterization of the nuclear ribosomal DNA units and phylogeny of Beta L. wild forms and cultivated beets. Theor. Appl. Genet. *83*, 533–542.

Schirrmann, M.K., Zoller, S., Croll, D., Stukenbrock, E.H., Leuchtmann, A., and Fior, S. (2018). Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. Mol. Ecol. *27*, 3070–3086.

Schneider, K., Kulosa, D., Soerensen, T.R., Möhring, S., Heine, M., Durstewitz, G., Polley, A., Weber, E., Lein, J., and Hohmann, U. (2007). Analysis of DNA polymorphisms in sugar beet (*Beta vulgaris L*.) and development of an SNP-based map of expressed genes. Theor. Appl. Genet. *115*, 601.

Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., Polz, M.F., and Alm, E.J. (2012). Population Genomics of Early Events in the Ecological Differentiation of Bacteria. Science *336*, 48–51.

Shapiro, L.R., Paulson, J.N., Arnold, B.J., Scully, E.D., Zhaxybayeva, O., Pierce, N.E., Rocha, J., Klepac-Ceraj, V., Holton, K., and Kolter, R. (2018). An Introduced Crop Plant Is Driving Diversification of the Virulent Bacterial Pathogen *Erwinia tracheiphila*. MBio *9*.

Skaracis, G.N., Pavli, O.I., and Biancardi, E. (2010). Cercospora leaf spot disease of sugar beet. Sugar Tech *12*, 220–228.

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

Stukenbrock, E.H., and McDonald, B.A. (2008). The Origins of Plant Pathogens in Agro-Ecosystems. Annu. Rev. Phytopathol. *46*, 75–100.

Stukenbrock, E.H., Banke, S., Javan-Nikkhah, M., and McDonald, B.A. (2007). Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. Mol. Biol. Evol. *24*, 398–411.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

Talas, F., and McDonald, B.A. (2015). Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. BMC Genomics *16*, 996.

Tange, O. (2011). Gnu parallel-the command-line power tool.; login: The USENIX Magazine, 36 (1): 42–47 (Feb).

Thrall, P.H., and Burdon, J.J. (1999). The spatial scale of pathogen dispersal: Consequences for disease dynamics and persistence. Evol. Ecol. Res. *1*, 681–701.

Vaghefi, N., Kikkert, J.R., Bolton, M.D., Hanson, L.E., Secor, G.A., and Pethybridge, S.J. (2017a). *De novo* genome assembly of *Cercospora beticola* for microsatellite marker development and validation. Fungal Ecol. *26*, 125–134.

Vaghefi, N., Nelson, S.C., Kikkert, J.R., and Pethybridge, S.J. (2017b). Genetic structure of *Cercospora beticola* populations on Beta vulgaris in New York and Hawaii. Sci. Rep. *7*, 1726.

Vagndorf, N., Heick, T.M., Justesen, A.F., Andersen, J.R., Jahoor, A., Jørgensen, L.N., and Orabi, J. (2018). Population structure and frequency differences of CYP51 mutations in *Zymoseptoria tritici* populations in the Nordic and Baltic regions. Eur. J. Plant Pathol. *152*, 327–341.

Van Geyt, J.P.C., Lange, W., Oleo, M., and De Bock, T.S. (1990). Natural variation within the genus *Beta* and its possible use for breeding sugar beet: a review. Euphytica *49*, 57–76.

Vestal, E.F. (1933). Pathogenicity, host response and control of Cercospora leaf spot of sugar beet. Agr. Exptl Sta. Iowa State College Of Agri. Mechanics Arts. Res. Bull *168*.

Walker, A.-S., Gladieux, P., Decognet, V., Fermaud, M., Confais, J., Roudet, J., Bardin, M., Bout, A., Nicot, P.C., Poncet, C., et al. (2015). Population structure and temporal maintenance of the multihost fungal pathogen *Botrytis cinerea*: causes and implications for disease management. Environ. Microbiol. *17*, 1261–1274.

Weiland, J., and Koch, G. (2004). Sugarbeet leaf spot disease (*Cercospora beticola Sacc*.)†. Mol. Plant Pathol. *5*, 157–166.

Wright, S. (1950). Genetical structure of populations. Nature *166*, 247–249.

Xhaard, C., Fabre, B., Andrieux, A., Gladieux, P., Barrès, B., Frey, P., and Halkett, F. (2011). The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. Mol. Ecol. *20*, 2739–2755.

About Sugar | International Sugar Organization. https://www.isosugar.org/sugarsector/sugar

Supplementary Materials

Supplementary Tables

Supplementary Table 1: Microsatellites and multiplexes from Vaghefi et al. 2017, and comments regarding performance in *C. beticola* isolates from the UK

Primer	Multiplex	Allele Sizes (bp)	Comments
SSRCb22	1	185,188, 191, 199, 203	Use
SSRCb24	1	312,313, 316, 323, 326, 332	1 bp difference in allele sizes between samples makes use difficult
SSRCb25	1	231, 235, 241, 255, 258, 264	Use
SSRCb1	2	208, 210, 221, 223, 225	Use
SSRCb3	2	262, 264, 271, 273, 275, 279, 284, 291, 345	Use
SSRCb21	2	165, 167, 173	Use
SSRCb2	3	195, 197	Good amplification, but does not fit in to existing multiplexes
SSRCb23	3	276, 287, 293, 298, 304	Did not work on many isolates
SSRCb27	3	376, 379, 382, 398, 404, 406, 410, 418, 425, 427, 433, 436, 439, 442	Use
SSRCb4	4	157, 190, 193	Good amplification, but does not fit in to existing multiplexes
SSRCb6	4	228, 232, 234	Did not work on some isolates
SSRCb20	5	157, 165, 168	Did not work on many isolates
SSRCb26	5	354, 355, 359, 368	1 bp difference in allele sizes between samples makes use difficult

Supplementary Table 2: Haplotype information for *C. beticola* isolates from sea beet in the UK established using 6 microsatellites.

Sample	Dataset	Haplotype Code	Haplotype
2A10.1A	Southwold	1	10 245 178 188 14 15
55A3A	Bawdsey Quay	2	10 271 167 185 255 442
3.5.1	Southwold	3	10 271 180 188 255 373
2B10.1A	Southwold	4	10 271 180 199 253 368
4A9.1A	Southwold	5	10 291 173 188 255 410
21.5.1B	Orford	6	221 11 12 199 258 398
3.9.1A	Southwold	7	221 11 167 188 258 376
8B3.2A	Southwold	7	221 11 167 188 258 376
8B3.2B	Southwold	7	221 11 167 188 258 376
4C6.2A	Southwold	8	221 262 167 185 14 418
4C6.2B	Southwold	9	221 262 167 185 235 418
55A11.1	Bawdsey Quay	10	221 264 173 188 231 382
56.12.1B	Bawdsey Quay	10	221 264 173 188 231 382
57.1.2A	Bawdsey Quay	10	221 264 173 188 231 382
57.1.2B	Bawdsey Quay	10	221 264 173 188 231 382
23.2.1B	Orford	11	221 264 173 199 231 382
23.6.2B	Orford	11	221 264 173 199 231 382
26.8.2	Orford	12	221 271 167 185 235 436
24.1.2	Orford	13	221 271 167 199 235 439
24.2.1A	Orford	13	221 271 167 199 235 439
24.2.1B	Orford	13	221 271 167 199 235 439
21.5.1A	Orford	14	221 271 167 199 255 439
24.4.1A	Orford	14	221 271 167 199 255 439
24.4.1B	Orford	14	221 271 167 199 255 439
27.5.1B	Orford	14	221 271 167 199 255 439
27.7.1A	Orford	14	221 271 167 199 255 439
27.7.1B	Orford	14	221 271 167 199 255 439
27.7.2A	Orford	14	221 271 167 199 255 439
27.7.2B	Orford	14	221 271 167 199 255 439
29.5.1B	Orford	14	221 271 167 199 255 439
4A3.2A	Southwold	14	221 271 167 199 255 439
4A8.1A	Southwold	14	221 271 167 199 255 439
1A5.2B	Southwold	15	221 271 167 199 255 442
21.8.1A	Orford	16	221 271 173 185 235 439
21.8.1B	Orford	16	221 271 173 185 235 439
25.1.1A	Orford	17	221 271 180 185 255 373
4A2.1B	Southwold	18	221 271 180 188 261 368
30.5.2B	Orford	19	221 271 180 199 253 368
23.6.2A	Orford	20	221 271 180 199 255 373
25.1.2A	Orford	21	221 273 167 185 14 436

Supplementary Table 2 (continued): Haplotype information for *C. beticola* isolates from sea beet in the UK established using 6 microsatellites.

Sample	Location	Haplotype Code	Haplotype
25.13.1A	Orford	21	221 273 167 185 14 436
54.6.1B	Bawdsey Quay	22	221 273 167 199 255 379
8A11.1	Southwold	22	221 273 167 199 255 379
26.8.1	Orford	23	221 273 167 199 255 442
30.7.2A	Orford	24	221 273 167 199 264 379
30.7.2B	Orford	24	221 273 167 199 264 379
56.2.2	Bawdsey Quay	25	221 275 167 199 255 379
51.7.1	Bawdsey Quay	26	221 284 167 188 255 398
51.7.2B	Bawdsey Quay	26	221 284 167 188 255 398
51.9.2B	Bawdsey Quay	26	221 284 167 188 255 398
2B10.1B	Southwold	27	221 284 167 199 235 442
4A3.1B	Southwold	28	221 291 173 188 255 404
4B9.2A	Southwold	29	221 291 173 188 255 406
4A8.2B	Southwold	30	221 291 173 188 255 410
5.7.2A	Southwold	30	221 291 173 188 255 410
5.7.2B	Southwold	30	221 291 173 188 255 410
4B5.1B	Southwold	31	221 291 173 191 255 410
24.3.1A	Orford	32	221 345 167 185 264 379
53A7.2	Bawdsey Quay	33	223 11 165 188 258 379
8A11.2B	Southwold	34	223 279 165 188 258 433
8A2.2A	Southwold	34	223 279 165 188 258 433
8A2.2B	Southwold	34	223 279 165 188 258 433
10.2.1B	Southwold	35	223 279 173 199 231 382
10.2.2B	Southwold	35	223 279 173 199 231 382
55A9A	Bawdsey Quay	36	225 271 165 185 255 425
9.10.1A	Southwold	37	225 271 165 188 255 427
9.10.1B	Southwold	37	225 271 165 188 255 427
4A8.1B	Southwold	38	237 11 167 188 258 376

Supplementary Table 3: Sequencing information for *C. beticola*. Isolate name, location, host, number of reads per forward and reverse read file for each sequencing run, as well as number of reads mapped to the reference genome after filtering, mapping and merging of reads

Isolate	Country	Location	Host	Number of reads per forward and reverse sequencing file	Number of reads per forward and reverse re-sequencing file	Total of reads mapped to reference genome
1A5.2B	UK	Southwold	Sea beet	4014608	4277957	12878654
2B10.1A	UK	Southwold	Sea beet	4005394	2883070	10121566
2B10.1B	UK	Southwold	Sea beet	4194642	2072976	9794933
3.5.1	UK	Southwold	Sea beet	5538310	1745420	11023518
4A2.1B	UK	Southwold	Sea beet	1048925	7608192	12267388
4A3.1B	UK	Southwold	Sea beet	1048925		12145034
4A3.2A	UK	Southwold	Sea beet	3294464	5157388	14349197
4A8.1B	UK	Southwold	Sea beet	9090071		12761815
4A8.2B	UK	Southwold	Sea beet	7285356		10695207
4A9.1A	UK	Southwold	Sea beet	6972409		9500414
4B5.1B	UK	Southwold	Sea beet	4964279	2007704	10975391
4B9.2A	UK	Southwold	Sea beet	3920290	2188929	9604326
5.7.2A	UK	Southwold	Sea beet	8840454		12561573
8A11.1	UK	Southwold	Sea beet	6862518		9980729
8A11.2B	UK	Southwold	Sea beet	6779565		10512469
8B.3.2B	UK	Southwold	Sea beet	4672286	4222834	14319469
9.10.1A	UK	Southwold	Sea beet	6027799	1164390	8953385
10.2.2B	UK	Southwold	Sea beet	5435092	2797075	13307111
21.5.1A	UK	Orford	Sea beet	4320997	3563404	11712504
21.5.1B	UK	Orford	Sea beet	5915621	1228669	10444621
21.8.1A	UK	Orford	Sea beet	4175894	3980993	12884536
23.2.1B	UK	Orford	Sea beet	7106172		9528468
23.6.2A	UK	Orford	Sea beet	2734860	4371879	10462082
24.2.1A	UK	Orford	Sea beet	5918371	2060355	12250855
24.4.1B	UK	Orford	Sea beet	6326170		9735819
25.1.1A	UK	Orford	Sea beet	4809961	3583927	14491747
25.1.2A	UK	Orford	Sea beet	4248345	4235352	6760468
26.8.1	UK	Orford	Sea beet	5040701	2059658	10519965
27.7.2B	UK	Orford	Sea beet	6645185		10123208
29.5.1B	UK	Orford	Sea beet	4831103	3848328	13558404
30.5.2B	UK	Orford	Sea beet	8816292		12164184
30.7.2B	UK	Orford	Sea beet	908355	7192564	11632926
51.7.1	UK	Bawdsey Quay	Sea beet	6076995		9156094
53A7.2	UK	Bawdsey Quay	Sea beet	4272510	3247948	11891672

Supplementary Table 3 (continued): Sequencing information for *C. beticola*. Isolate name, location, host, number of reads per forward and reverse read file for each sequencing run, as well as number of reads mapped to the reference genome after filtering, mapping and merging of reads

Isolate	Country	Location	Host	Number of reads per forward and reverse sequencing file	Number of reads per forward and reverse re-sequencing file	Total of reads mapped to reference genome
54.6.1B	UK	Bawdsey Quay	Sea beet	6644294		9723179
55A11.1	UK	Bawdsey Quay	Sea beet	5040469	2164109	12388047
55A3A	UK	Bawdsey Quay	Sea beet	5293188	3029257	12185421
55A9A	UK	Bawdsey Quay	Sea beet	4255416	2730803	10964613
56.12.1B	UK	Bawdsey Quay	Sea beet	2913000	2540203	8949071
56.2.2	UK	Bawdsey Quay	Sea beet	2134662	6756842	13464223
57.1.2A	UK	Bawdsey Quay	Sea beet	8735115		11715712
57.1.2B	UK	Bawdsey Quay	Sea beet	6546248		9961563
MB1.10	Croatia	Antenal	Sea beet	8466069		11755318
MB1.11	Croatia	Antenal	Sea beet	7157596		10307355
MB1.14	Croatia	Antenal	Sea beet	8719346		11917096
MB1.18	Croatia	Antenal	Sea beet	9092338		12474262
MB1.23	Croatia	Antenal	Sea beet	8521666		11617333
MB1.4	Croatia	Antenal	Sea beet	8719287		12493618
MB1.6	Croatia	Antenal	Sea beet	8756277		12431381
MB1.8	Croatia	Antenal	Sea beet	7112249		10023792
MB1.9	Croatia	Antenal	Sea beet	7098880		9738204
MB2.25	Croatia	Jadransko	Sea beet	8767278		12185544
MB2.27	Croatia	Jadransko	Sea beet	8682433		12371147
MB2.28	Croatia	Jadransko	Sea beet	6724870		9730075
MB2.30	Croatia	Jadransko	Sea beet	6288138	1674708	11954677
MB2.34	Croatia	Jadransko	Sea beet	8768204		12730012
MB2.36	Croatia	Jadransko	Sea beet	7182364		10542347
MB2.37	Croatia	Jadransko	Sea beet	4701739	3047241	12269765
MB2.39	Croatia	Jadransko	Sea beet	6015158	1606023	11568217
MB2.40	Croatia	Jadransko	Sea beet	6939809		9992150
MB2.42	Croatia	Jadransko	Sea beet	7223839	2344796	14579464
MB2.44	Croatia	Jadransko	Sea beet	5099261		7733998
MB2.47	Croatia	Jadransko	Sea beet	4822687	2722110	11558002
MB2.55	Croatia	Jadransko	Sea beet	6301393	875609	10581206
MB2.57	Croatia	Jadransko	Sea beet	5604576	2148123	11667028
MB2.62	Croatia	Jadransko	Sea beet	5967526	2005039	13003119
22E	Germany	Germany	Sugar beet	9446916		12290189
71E	Germany	Germany	Sugar beet	6515317		8974748
138C	Germany	Germany	Sugar beet	7381168		9871961

Supplementary Table 3 (continued): Sequencing information for *C. beticola*. Isolate name, location, host, number of reads per forward and reverse read file for each sequencing run, as well as number of reads mapped to the reference genome after filtering, mapping and merging of reads

Isolate	Country	Location	Host	Number of reads per forward and reverse sequencing file	Number of reads per forward and reverse re-sequencing file	Total of reads mapped to reference genome
71D	Germany	Germany	Sugar beet	8097640	4126640	11072196
135E	Germany	Germany	Sugar beet	4126640	2272982	6355050
5.8	Italy	Venezia	Sugar beet	4530725	2986716	10992865
5.36	Italy	Venezia	Sugar beet	4408171	2821026	10516845
5.44	Italy	Rovigo	Sugar beet	3625002	4455295	11539328
5.46	Italy	Rovigo	Sugar beet	3865896	4487761	11902725
5.52	Italy	Rovigo	Sugar beet	4641173	3277607	11494393
5.60	Italy	Rovigo	Sugar beet	4979613	2887423	11539100
5.69	Italy	Rovigo	Sugar beet	4723570	2867161	11224094
5.75	Italy	Bologna	Sugar beet	4031636	3656645	11339277
5.81	Italy	Bologna	Sugar beet	4170547	3804380	11483576
5.89	Italy	Bologna	Sugar beet	3801205	4225454	11427745
5.94	Italy	Ferrara	Sugar beet	4158605	3835141	11586957
6.4	Italy	Venezia	Sugar beet	5037477	2643143	11282397
6.14	Italy	Venezia	Sugar beet	4226818	3626363	11432664
6.28	Italy	Venezia	Sugar beet	4353402	2929162	10564701
6.51	Italy	Ferrara	Sugar beet	5003050	2476620	10703934
6.84	Italy	Bologna	Sugar beet	4815756	3125972	11356955
6.94	Italy	Bologna	Sugar beet	8563007		12554776
1	USA	Fargo	Sugar beet	7077611		11184177
2	USA	Fargo	Sugar beet	7405318		11656073
3	USA	Fargo	Sugar beet	5885382		9763662
4	USA	Fargo	Sugar beet	7027467		11346789
5	USA	Fargo	Sugar beet	5216614		8610802
6	USA	Fargo	Sugar beet	6503251		10594080
7	USA	Fargo	Sugar beet	6607268		10722562
9	USA	Fargo	Sugar beet	7136108		11334213
10	USA	Fargo	Sugar beet	6190651		10021786
11	USA	Fargo	Sugar beet	6157368		9827287
12	USA	Fargo	Sugar beet	6255043		10236851
13	USA	Fargo	Sugar beet	4841106		7999765
14	USA	Fargo	Sugar beet	6175383		10087169
15	USA	Fargo	Sugar beet	5433209		8997590
16	USA	Fargo	Sugar beet	7911416		12660879
17	USA	Fargo	Sugar beet	5962635		9867351
18	USA	Fargo	Sugar beet	5450102		9242201
19	USA	Fargo	Sugar beet	6534601		10933059
20	USA	Fargo	Sugar beet	6100778		10111838

Supplementary Table 3 (continued): Sequencing information for *C. beticola*. Isolate name, location, host, number of reads per forward and reverse read file for each sequencing run, as well as number of reads mapped to the reference genome after filtering, mapping and merging of reads

Isolate	Country	Location	Host	Number of reads per forward and reverse sequencing file	Number of reads per forward and reverse re-sequencing file	Total of reads mapped to reference genome
21	USA	Fargo	Sugar beet	6337046		10383771
22	USA	Fargo	Sugar beet	6916031		11117831
23	USA	Fargo	Sugar beet	5683694		9519564
24	USA	Fargo	Sugar beet	4891076		8062310
25	USA	Fargo	Sugar beet	6481401		10809308
31	USA	Fargo	Sugar beet	6815836		11246961
32	USA	Fargo	Sugar beet	7319235		12968319
33	USA	Fargo	Sugar beet	7319235		11754546
34	USA	Fargo	Sugar beet	7693876		12496961
35	USA	Fargo	Sugar beet	7488383		12138580
36	USA	Fargo	Sugar beet	5999096		10059650
37	USA	Fargo	Sugar beet	6715024		11021060
38	USA	Fargo	Sugar beet	6275402		10354920
39	USA	Fargo	Sugar beet	6975840		11470635
40	USA	Fargo	Sugar beet	5905926		9557104
41	USA	Fargo	Sugar beet	5938836		9982141
42	USA	Fargo	Sugar beet	6636661		11025802
43	USA	Fargo	Sugar beet	7397503		11966667
44	USA	Fargo	Sugar beet	6919306		11453157
45	USA	Fargo	Sugar beet	5988285		9819585
46	USA	Fargo	Sugar beet	6912010		11223286
47	USA	Fargo	Sugar beet	6126557		10122940
48	USA	Fargo	Sugar beet	5462873		9101933
49	USA	Fargo	Sugar beet	6346673		10458330
50	USA	Fargo	Sugar beet	6713400		11035426
51	USA	Fargo	Sugar beet	6613118		10629453
52	USA	Fargo	Sugar beet	6041495		10063265
53	USA	Fargo	Sugar beet	6877194		11212381
54	USA	Fargo	Sugar beet	5588335		9362577
55	USA	Fargo	Sugar beet	5304349		8929493
56	USA	Fargo	Sugar beet	6999839		11518847
57	USA	Fargo	Sugar beet	7278717		11827091
58	USA	Fargo	Sugar beet	6126410		9643925
59	USA	Fargo	Sugar beet	6232712		10252265
60	USA	Fargo	Sugar beet	6234287		13586939
61	USA	Fargo	Sugar beet	8481186		13334858
62	USA	Fargo	Sugar beet	8492353		13577410
63	USA	Fargo	Sugar beet	8475883		13510462

Supplementary Table 3 (continued): Sequencing information for *C. beticola*. Isolate name, location, host, number of reads per forward and reverse read file for each sequencing run, as well as number of reads mapped to the reference genome after filtering, mapping and merging of reads

Isolate	Country	Location	Host	Number of reads per forward and reverse sequencing file	Number of reads per forward and reverse re-sequencing file	Total of reads mapped to reference genome
64	USA	Fargo	Sugar beet	8472671		13547048
65	USA	Fargo	Sugar beet	8345351		13404534
66	USA	Fargo	Sugar beet	8460341		12268454
67	USA	Fargo	Sugar beet	7275146		11463379
68	USA	Fargo	Sugar beet	6718936		9886557
69	USA	Fargo	Sugar beet	5697310		11292102
71	USA	Fargo	Sugar beet	6679554		11140309
72	USA	Fargo	Sugar beet	6524046		9794254
73	USA	Fargo	Sugar beet	5704809		9474866
74	USA	Fargo	Sugar beet	5510695		11243195
75	USA	Fargo	Sugar beet	6611631		8630990
76	USA	Fargo	Sugar beet	4981586		10171249
77	USA	Fargo	Sugar beet	6008471		9310905
78	USA	Fargo	Sugar beet	5388088		12216508
79	USA	Fargo	Sugar beet	8538094		12135603
80	USA	Fargo	Sugar beet	8531567		12513638
81	USA	Fargo	Sugar beet	8537917		12275674
82	USA	Fargo	Sugar beet	8535697		12249630
83	USA	Fargo	Sugar beet	8541520		12122710
84	USA	Fargo	Sugar beet	8537971		12305828
85	USA	Fargo	Sugar beet	8537538		12205604
86	USA	Fargo	Sugar beet	8529988		12182138
87	USA	Fargo	Sugar beet	8521064		12254511
88	USA	Fargo	Sugar beet	8527623		13734776
89	USA	Fargo	Sugar beet	8530046		13440317
90	USA	Fargo	Sugar beet	8524568		13552890
91	USA	Fargo	Sugar beet	8537708		13588897
92	USA	Fargo	Sugar beet	8537876		13552545
93	USA	Fargo	Sugar beet	8540768		13494762
94	USA	Fargo	Sugar beet	8538937		13503145
95	USA	Fargo	Sugar beet	8538270		13568916
96	USA	Foxhome	Sugar beet	8540232		7209183
502	USA	Foxhome	Sugar beet	3991089		7215076
503	USA	Foxhome	Sugar beet	4002496		8388291
504	USA	Foxhome	Sugar beet	4713648		7186371
510	USA	Foxhome	Sugar beet	4001223		7189454
511	USA	Foxhome	Sugar beet	4001852		7221083
512	USA	Foxhome	Sugar beet	4006327		8408918

Supplementary Table 3 (continued): Sequencing information for *C. beticola*. Isolate name, location, host, number of reads per forward and reverse read file for each sequencing run, as well as number of reads mapped to the reference genome after filtering, mapping and merging of reads

Isolate	Country	Location	Host	Number of reads per forward and reverse sequencing file	Number of reads per forward and reverse re-sequencing file	Total of reads mapped to reference genome
513	USA	Foxhome	Sugar beet	4728954		8559124
523	USA	Foxhome	Sugar beet	4795755		7289071
524	USA	Foxhome	Sugar beet	4033241		7245515
525	USA	Foxhome	Sugar beet	4021195		7209400
526	USA	Foxhome	Sugar beet	4005147		7230913
527	USA	Foxhome	Sugar beet	3997325		8394806
528	USA	Foxhome	Sugar beet	4762503		8418389
537	USA	Foxhome	Sugar beet	4762536		8505502
538	USA	Foxhome	Sugar beet	4780326		7172045
539	USA	Foxhome	Sugar beet	3995857		7219820
540	USA	New.York	Table beet	4014987		8331799
545	USA	New.York	Table beet	4688259		8258893
546	USA	New.York	Table beet	4647481		8423808
547	USA	New.York	Table beet	4717119		8278308
557	USA	New.York	Table beet	4632251		8200951
558	USA	New.York	Table beet	4626762		8262612
559	USA	New.York	Table beet	4661893		8258239
560	USA	New.York	Table beet	4630565		8269435
561	USA	New.York	Table beet	4627205		8260854
562	USA	New.York	Table beet	4657623		8091231
563	USA	New.York	Table beet	4539038		8143703
564	USA	New.York	Table beet	4587926		8218599
574	USA	New.York	Table beet	4627676		8159283
575	USA	New.York	Table beet	4630593		8390744
576	USA	New.York	Table beet	4690727		8280288
581	USA	New.York	Table beet	4650454		8188128
584	USA	New.York	Table beet	4639312		8351039
585	USA	New.York	Table beet	4687807		8161312
587	USA	New.York	Table beet	4634159		8216585
588	USA	New.York	Table beet	4647238		8290226
589	USA	New.York	Table beet	4690955		7983816
590	USA	New.York	Table beet	4532942		8335352
591	USA	New.York	Table beet	4684869		8223941
592	USA	New.York	Table beet	4646925		8357345
593	USA	New.York	Table beet	4670723		7576533

Supplementary Table 4: Isolate names, country and region of origin, and host, along with percent of genome mapped, mean depth per site, number of variant positions, and number of missing sites

Isolate Name	Country	Location	Host	Percent of Genome Mapped	Mean Depth per Site	Variant Positions	Missing Sites	Percent of Sites Missing
1A5.2B	UK	Southwold	Sea beet	72.81	30.1	190429	36545	4.82
2B10.1A	UK	Southwold	Sea beet	77.35	17.3	179928	46215	6.18
2B10.1B	UK	Southwold	Sea beet	83.13	24.7	181185	45736	6.11
3.5.1	UK	Southwold	Sea beet	77.97	25.5	170840	43263	5.76
4A2.1B	UK	Southwold	Sea beet	76.71	30	183791	40131	5.32
4A3.1B	UK	Southwold	Sea beet	54.05	20.3	183728	41522	5.52
4A3.2A	UK	Southwold	Sea beet	45.3	18.1	191398	35318	4.66
4A8.1B	UK	Southwold	Sea beet	80.77	32.7	183508	39312	5.21
4A8.2B	UK	Southwold	Sea beet	75.02	25.5	182937	41752	5.55
4A9.1A	UK	Southwold	Sea beet	71.69	20.7	175994	46699	6.25
4B5.1B	UK	Southwold	Sea beet	70.36	24	183588	41027	5.45
4B9.2A	UK	Southwold	Sea beet	74.45	22.2	183271	41559	5.52
5.7.2A	UK	Southwold	Sea beet	76.73	31.2	183017	41785	5.55
8A11.1	UK	Southwold	Sea beet	82.13	25.3	185425	35395	4.67
8A11.2B	UK	Southwold	Sea beet	77.92	25.4	166020	44586	5.95
8B.3.2B	UK	Southwold	Sea beet	84.66	40.5	184509	38197	5.05
9.10.1A	UK	Southwold	Sea beet	80.64	26.7	166035	45025	6.01
10.2.2B	UK	Southwold	Sea beet	86.44	36.2	169542	36300	4.79
21.5.1A	UK	Orford	Sea beet	75.21	28.1	190424	36305	4.79
21.5.1B	UK	Orford	Sea beet	69.79	22.9	191963	35237	4.64
21.8.1A	UK	Orford	Sea beet	74.52	30.3	194704	44983	6
23.2.1B	UK	Orford	Sea beet	79.03	23	172336	38764	5.13
23.6.2A	UK	Orford	Sea beet	78.58	24.1	172573	38248	5.06
24.2.1A	UK	Orford	Sea beet	86.19	32.5	189213	45433	6.07
24.4.1B	UK	Orford	Sea beet	84.9	26.4	191184	36667	4.84
25.1.1A	UK	Orford	Sea beet	37.99	16.5	192554	37908	5.01
25.1.2A	UK	Orford	Sea beet	84.42	16.9	191965	39264	5.2
26.8.1	UK	Orford	Sea beet	81.01	27.5	191316	35263	4.65
27.7.2B	UK	Orford	Sea beet	84.48	27.3	189636	44127	5.88
29.5.1B	UK	Orford	Sea beet	71.11	28.2	189998	42723	5.69
30.5.2B	UK	Orford	Sea beet	71.58	26	188901	39386	5.22
30.7.2B	UK	Orford	Sea beet	72.84	26.1	189965	37899	5.01
51.7.1	UK	Bawdsey Quay	Sea beet	84.84	24.2	185610	48676	6.53
53A7.2	UK	Bawdsey Quay	Sea beet	69.42	25.4	178614	38911	5.15
54.6.1B	UK	Bawdsey Quay	Sea beet	75.69	22.4	184133	42474	5.65

Supplementary Table 4 (continued): Isolate names, country and region of origin, and host, along with percent of genome mapped, mean depth per site, number of variant positions, and number of missing sites

Isolate Name	Country	Location	Host	Percent of Genome Mapped	Mean Depth per Site	Variant Positions	Missing Sites	Percent of Sites Missing
55A11.1	UK	Bawdsey Quay	Sea beet	34.71	11.7	168664	39090	5.18
55A3A	UK	Bawdsey Quay	Sea beet	77.93	28.6	167433	42812	5.7
55A9A	UK	Bawdsey Quay	Sea beet	79.38	25	166651	43005	5.73
56.12.1 B	UK	Bawdsey Quay	Sea beet	48.51	11.8	168477	38893	5.15
56.2.2	UK	Bawdsey Quay	Sea beet	73.48	29.4	184288	41114	5.46
57.1.2A	UK	Bawdsey Quay	Sea beet	77.7	28.1	168956	35320	4.66
57.1.2B	UK	Bawdsey Quay	Sea beet	85.55	26.1	168468	37868	5.01
MB1.10	Croatia	Antenal	Sea beet	84.17	30.6	167094	46741	6.25
MB1.11	Croatia	Antenal	Sea beet	87.64	27.7	167110	46032	6.15
MB1.14	Croatia	Antenal	Sea beet	87.16	32.7	167574	46440	6.21
MB1.18	Croatia	Antenal	Sea beet	82.16	31.5	170341	39882	5.29
MB1.23	Croatia	Antenal	Sea beet	82.47	29.3	170621	34661	4.57
MB1.4	Croatia	Antenal	Sea beet	87.95	34	170082	40314	5.35
MB1.6	Croatia	Antenal	Sea beet	84.92	33	169176	41556	5.52
MB1.8	Croatia	Antenal	Sea beet	83.15	25.2	169244	41557	5.52
MB1.9	Croatia	Antenal	Sea beet	84.64	25.2	167055	47304	6.33
MB2.25	Croatia	Jadransko	Sea beet	88.12	29.5	108051	26160	3.41
MB2.27	Croatia	Jadransko	Sea beet	87.65	26.8	110059	35549	4.69
MB2.28	Croatia	Jadransko	Sea beet	90.73	24.6	132638	31295	4.1
MB2.30	Croatia	Jadransko	Sea beet	85.81	28.5	130875	37672	4.98
MB2.34	Croatia	Jadransko	Sea beet	83.88	33	188508	49844	6.7
MB2.36	Croatia	Jadransko	Sea beet	87.29	27.4	170902	43385	5.78
MB2.37	Croatia	Jadransko	Sea beet	85.75	32.4	170189	43319	5.77
MB2.39	Croatia	Jadransko	Sea beet	86.48	29.1	138641	41364	5.5
MB2.40	Croatia	Jadransko	Sea beet	87.68	24	128606	39581	5.25
MB2.42	Croatia	Jadransko	Sea beet	87.44	36	118488	25193	3.28
MB2.44	Croatia	Jadransko	Sea beet	85.97	19.9	192668	38890	5.15
MB2.47	Croatia	Jadransko	Sea beet	87.19	29.8	131799	31102	4.08
MB2.55	Croatia	Jadransko	Sea beet	81.86	23.4	122301	33741	4.44
MB2.57	Croatia	Jadransko	Sea beet	84.89	27.5	122408	33551	4.41
MB2.62	Croatia	Jadransko	Sea beet	68.38	25.5	141682	27501	3.59
22E	Germany	Germany	Sugar beet	74.19	26.7	132596	39145	5.19
71E	Germany	Germany	Sugar beet	71.73	16	125718	34211	4.5
138C	Germany	Germany	Sugar beet	74.5	19.9	140107	36636	4.84

Supplementary Table 4 (continued): Isolate names, country and region of origin, and host, along with percent of genome mapped, mean depth per site, number of variant positions, and number of missing sites

Isolate Name	Country	Location	Host	Percent of Genome Mapped	Mean Depth per Site	Variant Positions	Missing Sites	Percent of Sites Missing
71D	Germany	Germany	Sugar beet	81.01	23.4	112305	33215	4.37
135E	Germany	Germany	Sugar beet	84.49	12.3	94193	37195	4.91
5.8	Italy	Venezia	Sugar beet	82.39	27.1	142975	35979	4.75
5.36	Italy	Venezia	Sugar beet	83.22	26.1	138222	36612	4.83
5.44	Italy	Rovigo	Sugar beet	81.2	27.1	137684	37141	4.91
5.46	Italy	Rovigo	Sugar beet	80.56	28.4	149827	40310	5.35
5.52	Italy	Rovigo	Sugar beet	80.37	27	135565	40117	5.32
5.60	Italy	Rovigo	Sugar beet	79.64	27.4	149461	39778	5.27
5.69	Italy	Rovigo	Sugar beet	77.74	25.7	142735	41636	5.53
5.75	Italy	Bologna	Sugar beet	83.64	27.9	133224	40494	5.37
5.81	Italy	Bologna	Sugar beet	82.84	27	127151	36514	4.82
5.89	Italy	Bologna	Sugar beet	80.37	28.2	148579	39407	5.22
5.94	Italy	Ferrara	Sugar beet	80.08	27.4	133574	31842	4.18
6.4	Italy	Venezia	Sugar beet	79.03	26.9	146526	34925	4.6
6.14	Italy	Venezia	Sugar beet	81.82	27.9	141596	46745	6.25
6.28	Italy	Venezia	Sugar beet	82.92	23.5	106941	26928	3.51
6.51	Italy	Ferrara	Sugar beet	78.06	25	139206	32090	4.21
6.84	Italy	Bologna	Sugar beet	79.85	26.6	139115	34245	4.51
6.94	Italy	Bologna	Sugar beet	81.83	31.9	145004	38932	5.16
1	USA	Fargo	Sugar beet	87.02	18	133545	50791	6.83
2	USA	Fargo	Sugar beet	83.31	18.1	133890	50269	6.76
3	USA	Fargo	Sugar beet	87.48	15.9	140060	40486	5.37
4	USA	Fargo	Sugar beet	90.69	18.3	133233	51555	6.94
5	USA	Fargo	Sugar beet	90.31	13.3	142704	52214	7.04
6	USA	Fargo	Sugar beet	91.37	15.2	98814	32979	4.33
7	USA	Fargo	Sugar beet	89.29	17.4	133718	51454	6.93
9	USA	Fargo	Sugar beet	86.95	18.2	133463	50602	6.81
10	USA	Fargo	Sugar beet	86.47	16.3	146246	43942	5.86
11	USA	Fargo	Sugar beet	85.91	14.9	123037	48303	6.48
12	USA	Fargo	Sugar beet	88.82	16.2	133550	51830	6.98
13	USA	Fargo	Sugar beet	91.8	11.4	132006	56174	7.61
14	USA	Fargo	Sugar beet	90.86	16	132497	53256	7.19
15	USA	Fargo	Sugar beet	91.2	14	133301	54248	7.33
16	USA	Fargo	Sugar beet	91.69	20.3	133712	49748	6.68
17	USA	Fargo	Sugar beet	89.06	15.5	125610	44245	5.9
18	USA	Fargo	Sugar beet	90.51	13.4	118381	49896	6.7

Supplementary Table 4 (continued): Isolate names, country and region of origin, and host, along with percent of genome mapped, mean depth per site, number of variant positions, and number of missing sites

Isolate Name	Country	Location	Host	Percent of Genome Mapped	Mean Depth per Site	Variant Positions	Missing Sites	Percent of Sites Missing
19	USA	Fargo	Sugar beet	84.42	16.2	133195	52287	7.05
20	USA	Fargo	Sugar beet	90.33	16.2	133138	52653	7.1
21	USA	Fargo	Sugar beet	88.63	16.9	133717	48766	6.54
22	USA	Fargo	Sugar beet	87.84	17.3	119112	41359	5.49
23	USA	Fargo	Sugar beet	92.83	11.9	70321	17374	2.24
24	USA	Fargo	Sugar beet	91.93	11.6	120712	52882	7.13
25	USA	Fargo	Sugar beet	91.01	15	109802	30300	3.97
31	USA	Fargo	Sugar beet	88.11	18.7	141959	48309	6.48
32	USA	Fargo	Sugar beet	88.02	20.7	118979	41612	5.53
33	USA	Fargo	Sugar beet	84.71	19.1	142419	46450	6.21
34	USA	Fargo	Sugar beet	90.97	20.8	133552	50743	6.83
35	USA	Fargo	Sugar beet	89.51	18.2	117996	42236	5.62
36	USA	Fargo	Sugar beet	90.96	14.4	111306	40611	5.39
37	USA	Fargo	Sugar beet	91.17	15.9	111403	40730	5.41
38	USA	Fargo	Sugar beet	90.11	14.7	97913	34531	4.55
39	USA	Fargo	Sugar beet	89.97	15.6	95924	31781	4.17
40	USA	Fargo	Sugar beet	91.8	9	41155	17922	2.31
41	USA	Fargo	Sugar beet	92.58	9.4	40996	17486	2.25
42	USA	Fargo	Sugar beet	91.16	17.5	132615	44854	5.99
43	USA	Fargo	Sugar beet	87.59	14.4	66786	28785	3.76
44	USA	Fargo	Sugar beet	91.7	14.4	87768	35275	4.65
45	USA	Fargo	Sugar beet	89.4	12.3	82510	35548	4.69
46	USA	Fargo	Sugar beet	91.04	15.5	101057	42510	5.66
47	USA	Fargo	Sugar beet	88.88	14.1	112079	41755	5.55
48	USA	Fargo	Sugar beet	87.48	12.6	108942	39228	5.2
49	USA	Fargo	Sugar beet	91.26	14.1	97923	35507	4.68
50	USA	Fargo	Sugar beet	92.1	13.3	73358	25534	3.32
51	USA	Fargo	Sugar beet	89.49	16.9	132613	45429	6.07
52	USA	Fargo	Sugar beet	92.11	9.6	40990	17216	2.22
53	USA	Fargo	Sugar beet	88.3	16.6	107047	34310	4.52
54	USA	Fargo	Sugar beet	90.57	12.4	102731	39312	5.21
55	USA	Fargo	Sugar beet	90.81	9.9	63606	15827	2.04
56	USA	Fargo	Sugar beet	91.13	13.1	64373	17435	2.25
57	USA	Fargo	Sugar beet	91.81	17	51099	13997	1.8
58	USA	Fargo	Sugar beet	84.92	13.2	116903	43728	5.83
59	USA	Fargo	Sugar beet	88.81	15.9	122681	49370	6.63
61	USA	Fargo	Sugar beet	86.99	18.2	94778	40230	5.34
62	USA	Fargo	Sugar beet	83.88	19.2	121665	48742	6.54
63	USA	Fargo	Sugar beet	85.39	20.6	128746	48328	6.48

Supplementary Table 4 (continued): Isolate names, country and region of origin, and host, along with percent of genome mapped, mean depth per site, number of variant positions, and number of missing sites

Isolate Name	Country	Location	Host	Percent of Genome Mapped	Mean Depth per Site	Variant Positions	Missing Sites	Percent of Sites Missing
64	USA	Fargo	Sugar beet	85.63	21	134997	54761	7.41
65	USA	Fargo	Sugar beet	80.73	16.5	93228	27584	3.6
66	USA	Fargo	Sugar beet	84.24	18.9	115274	33407	4.39
67	USA	Fargo	Sugar beet	87.43	19.7	131947	40826	5.42
68	USA	Fargo	Sugar beet	87.02	18.2	139416	51963	7
69	USA	Fargo	Sugar beet	87.67	14.8	133974	57049	7.74
71	USA	Fargo	Sugar beet	87.5	17.4	134397	56141	7.61
72	USA	Fargo	Sugar beet	87.07	17.1	134186	55556	7.52
73	USA	Fargo	Sugar beet	85.66	14.4	133855	56212	7.62
74	USA	Fargo	Sugar beet	87.7	14.2	134078	56819	7.7
75	USA	Fargo	Sugar beet	87.86	17.5	134543	55106	7.46
76	USA	Fargo	Sugar beet	85.36	12.3	133584	57427	7.79
77	USA	Fargo	Sugar beet	88.2	15.3	131961	50152	6.74
78	USA	Fargo	Sugar beet	88.84	14.1	132982	45926	6.14
79	USA	Fargo	Sugar beet	84.01	14.4	78579	34893	4.6
80	USA	Fargo	Sugar beet	86.34	14.9	82620	29021	3.8
81	USA	Fargo	Sugar beet	84.99	19.2	141162	45550	6.08
82	USA	Fargo	Sugar beet	82.22	17.9	128479	47125	6.31
83	USA	Fargo	Sugar beet	86.05	15.9	94048	41803	5.56
84	USA	Fargo	Sugar beet	84.69	17	117594	40437	5.37
85	USA	Fargo	Sugar beet	85.59	17.2	113271	40267	5.34
86	USA	Fargo	Sugar beet	85.61	18.5	128500	48616	6.52
87	USA	Fargo	Sugar beet	85.35	18.2	129969	52170	7.03
88	USA	Fargo	Sugar beet	86.06	18.7	141167	51589	6.95
89	USA	Fargo	Sugar beet	82.86	20.8	135082	54132	7.31
90	USA	Fargo	Sugar beet	84.81	20.7	135224	53778	7.26
91	USA	Fargo	Sugar beet	86.64	21.4	134955	54194	7.32
92	USA	Fargo	Sugar beet	88.42	21.4	125680	47616	6.38
93	USA	Fargo	Sugar beet	87.59	18	90004	38456	5.09
94	USA	Fargo	Sugar beet	87.48	21.4	134388	48463	6.5
95	USA	Fargo	Sugar beet	87.42	20.5	114097	29771	3.9
96	USA	Fargo	Sugar beet	86.42	19.1	115818	44994	6.01
502	USA	Foxhome	Sugar beet	90.37	16.3	132250	43357	5.78
503	USA	Foxhome	Sugar beet	90.83	16.6	139490	38959	5.16
504	USA	Foxhome	Sugar beet	90.5	17.4	103960	23132	3
510	USA	Foxhome	Sugar beet	89.85	16.9	145947	35508	4.68
511	USA	Foxhome	Sugar beet	90.37	16.8	141538	42033	5.59
512	USA	Foxhome	Sugar beet	90.48	16.2	132598	38559	5.1
513	USA	Foxhome	Sugar beet	90.77	19.2	133598	41215	5.47

Supplementary Table 4 (continued): Isolate names, country and region of origin, and host, along with percent of genome mapped, mean depth per site, number of variant positions, and number of missing sites

Isolate Name	Country	Location	Host	Percent of Genome Mapped	Mean Depth per Site	Variant Positions	Missing Sites	Percent of Sites Missing
523	USA	Foxhome	Sugar beet	90.1	19.7	132567	42793	5.7
525	USA	Foxhome	Sugar beet	91.14	16.7	139196	40559	5.38
526	USA	Foxhome	Sugar beet	90.58	13.8	99052	27250	3.56
527	USA	Foxhome	Sugar beet	91.02	16.5	136944	42795	5.7
528	USA	Foxhome	Sugar beet	88.63	18.7	132945	42832	5.7
537	USA	Foxhome	Sugar beet	89.54	20.4	146492	34814	4.59
538	USA	Foxhome	Sugar beet	89.51	20.1	141720	42446	5.65
539	USA	Foxhome	Sugar beet	90.14	16.7	144769	45184	6.03
540	USA	Foxhome	Sugar beet	90.03	16.8	144947	45038	6.01
545	USA	New.York	Table beet	91.19	20.1	135584	44800	5.98
546	USA	New.York	Table beet	92.05	19.9	135088	44032	5.87
547	USA	New.York	Table beet	89.4	21.4	175091	43618	5.81
557	USA	New.York	Table beet	89.56	21	174423	36709	4.85
558	USA	New.York	Table beet	89.47	20.2	165627	43598	5.81
559	USA	New.York	Table beet	90.29	21.6	172590	38148	5.05
560	USA	New.York	Table beet	90.22	19.6	134894	36531	4.82
561	USA	New.York	Table beet	90.68	21	166557	39995	5.3
562	USA	New.York	Table beet	88.77	20.5	153175	34873	4.59
563	USA	New.York	Table beet	90.81	20.3	164301	38234	5.06
564	USA	New.York	Table beet	90.11	20.9	177170	43052	5.73
574	USA	New.York	Table beet	89.58	20.2	153452	42610	5.67
575	USA	New.York	Table beet	89.55	20.3	155888	34132	4.49
576	USA	New.York	Table beet	89.39	21.4	167350	39994	5.3
581	USA	New.York	Table beet	90.47	20.8	168412	44560	5.94
584	USA	New.York	Table beet	88.97	19.6	153874	41848	5.56
585	USA	New.York	Table beet	88.77	19.2	133132	42282	5.62
587	USA	New.York	Table beet	88.9	19.7	151954	34776	4.58
588	USA	New.York	Table beet	89.38	20.7	174658	45885	6.13
589	USA	New.York	Table beet	88.57	20.3	165371	46066	6.16
590	USA	New.York	Table beet	89.86	19.6	169111	45605	6.09
591	USA	New.York	Table beet	90.97	21.8	172487	38570	5.11
592	USA	New.York	Table beet	89.74	21	171760	45496	6.08
593	USA	New.York	Table beet	90.5	21.6	167805	39125	5.18

Supplementary Table 5: Chromosome names, size, and number of SNPs per chromosome for each sampling location of *C. beticola*

Chromosome	1	2	3	4	5	6	7	8	9	10
Annotated name	CM008499.1	CM008500.1	CM008501.1	CM008502.1	CM008503.1	CM008504.1	CM008505.1	CM008506.1	CM008507.1	CM008508.1
Size (Mb)	6.19	4.22	4.2	4.18	4.17	3	3	2.63	2.17	1.56
Bawdsey Quay	52266	40630	40507	29894	30586	34253	35041	24981	20441	13543
Orford	41751	36470	29851	28706	23321	27724	26920	22691	20246	10657
Southwold	56684	41829	45635	35711	36452	40290	40601	28967	25354	15328
Antenal	36311	29406	27478	30161	25206	21720	23699	17794	12232	10382
Jadransko	61321	49552	47255	47201	37462	39062	43037	27622	24976	18214
Italy	46226	42335	34742	36835	27643	30778	38309	25256	17392	15282
Fargo	54946	43650	45774	39861	36757	37612	43088	26449	22537	15477
Foxhome	36943	29869	23130	30256	24875	31502	33040	21930	15881	12495
New York	59659	46412	37095	38336	37212	35621	42161	26946	22436	14484

Supplementary Table 6: Minimum and maximum R^2 values for *C. beticola* each population for each chromosome

	Southwold	Orford	Bawdsey Quay	Antenal	Jadransko	Italy	Fargo	Foxhome	New York
CM00499.1									
Maximum R^2	0.43	0.8	0.50	0.88	0.34	0.44	0.35	0.6	0.37
Minimum R^2	0.1	0.11	0.13	0.28	0.03	0.04	0.002	0.03	0.04
CM00500.1									
Maximum R^2	0.47	0.67	0.49	0.8	0.29	0.38	0.31	0.55	0.36
Minimum R^2	0.003	0.1	0.14	0.16	0.05	0.05	0.02	0.03	0.03
CM00501.1									
Maximum R^2	0.4	0.99	0.51	0.7	0.36	0.44	0.27	0.66	0.44
Minimum R^2	0.07	0.25	0.12	0.21	0.03	0.03	0.001	0.02	0.03
CM00502.1									
Maximum R^2	0.45	1.00	0.54	0.76	0.35	0.46	0.36	0.63	0.47
Minimum R^2	0.06	0.09	0.18	0.14	0.04	0.03	0.003	0.1	0.06
CM00503.1									
Maximum R^2	0.46	0.99	0.61	0.77	0.39	0.73	0.35	0.74	0.42
Minimum R^2	0.02	0.02	0.11	0.13	0.02	0.01	0.003	0.04	0.005
CM00504.1									
Maximum R^2	0.41	0.78	0.52	0.82	0.34	0.46	0.28	0.53	0.42
Minimum R^2	0.09	0.06	0.18	0.29	0.07	0.02	0.02	0.02	0.02
CM00505.1									
Maximum R^2	0.39	1.00	0.7	0.92	0.33	0.42	0.27	0.56	0.38
Minimum R^2	0.07	0.22	0.24	0.16	0.04	0.02	0.001	0.08	0.05
CM00506.1									
Maximum R^2	0.41	0.67	0.59	0.73	0.32	0.41	0.29	0.46	0.41
Minimum R^2	0.1	0.19	0.15	0.32	0.05	0.03	0.007	0.04	0.02
CM00507.1									
Maximum R^2	0.46	1.00	0.59	0.8	0.36	0.53	0.36	0.74	0.49
Minimum R^2	0.02	0.52	0.16	0.33	0.07	0.04	0.03	0.04	0.05
CM00508.1									
Maximum R^2	0.44	0.66	0.65	0.87	0.29	0.34	0.31	0.52	0.41
Minimum R^2	0.06	0.05	0.11	0.08	0.04	0.02	0.02	0.04	0.02

Supplementary Table 7: ID of genes 50 kb up- and down-stream of the 650 kb F_{ST} outlier region on CM008499.1, as well length and GO terms

ID	Best Match	Length	E-value	GO term
XM_023592888.1	Amycolatopsis mediterranei S699, complete genome	975	0	F:hydrolase activity
XM_023592889.1	Cercospora beticola Beta-1,2-xylosyltransferase 1 (CB0940_00257), mRNA	1833	0	no GO terms
XM_023592890.1	Cercospora beticola MAP kinase kinase MKK1/SSP32 (CB0940_00258), mRNA	1512	0	P:protein phosphorylation; F:protein kinase activity; F:ATP binding
XM_023592891.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1539	0	no GO terms
XM_023592892.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1104	0	no GO terms
XM_023592893.1	Cercospora beticola DNA polymerase lambda (CB0940_00261) mRNA	2214	0	P:DNA repair; F:DNA binding; F:DNA-directed DNA polymerase activity; F:DNA polymerase activity
XM_023592894.1	Cercospora beticola Pumilio domain-containing protein (CB0940_00262), mRNA	2064	0	F:RNA binding
XM_023592895.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	597	0	no GO terms
XM_023592896.1	Cercospora beticola putative HC-toxin efflux carrier TOXA (CB0940_00264) mRNA	1884	0	P:transmembrane transport; F:transmembrane transporter activity
XM_023592897.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1665	0	no GO terms
XM_023592898.1	---NA---	1794		no GO terms
XM_023592899.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	969	0	no GO terms
XM_023592900.1	---NA---	2706		P:carbohydrate metabolic process; F:hydrolase activity, hydrolyzing O-glycosyl compounds
XM_023592901.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1338	0	F:oxidoreductase activity

Supplementary Table 7 (continued): ID of genes 50 kb up- and down-stream of the 650 kb F_{ST} outlier region on CM008499.1, as well length and GO terms

ID	Best Match	Length	E-value	GO term
XM_023592902.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1158	0	no GO terms
XM_023592903.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1626	0	no GO terms
XM_023592904.1	Cercospora beticola Aflatoxin B1 aldehyde reductase member 2 (CB0940_00271), mRNA	1137	0	P:obsolete oxidation-reduction process; F:oxidoreductase activity; F:D-threo-aldose 1-dehydrogenase activity
XM_023592905.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	441	1.29E-115	no GO terms
XM_023592906.1	Cercospora beticola Tubulin gamma chain (CB0940_00273), mRNA	1422	0	P:microtubule-based process; P:microtubule nucleation; P:cytoplasmic microtubule organization; F:GTPase activity; C:gamma-tubulin complex; C:microtubule
XM_023592907.1	---NA---	447		no GO terms
XM_023592908.1	Cercospora beticola 60S ribosomal protein L3 (CB0940_00275), mRNA	1179	0	P:translation; F:structural constituent of ribosome; C:ribosome
XM_023592909.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	585	0	no GO terms
XM_023592910.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	576	0	P:regulation of translational fidelity
XM_023592911.1	Cercospora beticola ATP synthase subunit d, mitochondrial (CB0940_00278), mRNA	525	0	P:ATP synthesis coupled proton transport; F:proton transmembrane transporter activity; C:mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)
XM_023592912.1	Cercospora beticola Zinc finger protein (CB0940_00279), mRNA	1182	0	no GO terms
XM_023592913.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	540	0	no GO terms

Supplementary Table 7 (continued): ID of genes 50 kb up- and down-stream of the 650 kb F_{ST} outlier region on CM008499.1, as well length and GO terms

ID	Best Match	Length	E-value	GO term
XM_023592914.1	Cercospora beticola MAP kinase kinase kinase wis4 (CB0940_00281), mRNA	4251	0	P:protein phosphorylation; P:stress-activated MAPK cascade; F:protein kinase activity; F:MAP kinase kinase kinase activity; F:ATP binding; C:cytoplasm
XM_023592915.1	Cercospora beticola putative WD repeat-containing protein (CB0940_00282), mRNA	1155	0	P:mRNA export from nucleus; F:protein binding
XM_023592916.1	Cercospora beticola Vegetative incompatibility protein HET-E-1 (CB0940_00283) mRNA	2427	0	F:protein binding
XM_023592917.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1623	0	no GO terms
XM_023592918.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	552	0	no GO terms
XM_023592919.1	Cercospora beticola alcohol dehydrogenase (CB0940_00286), mRNA	1083	0	P:obsolete oxidation-reduction process
XM_023592920.1	Cercospora beticola WHI2-like protein (CB0940_00287), mRNA	909	0	no GO terms
XM_023592921.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	534	0	no GO terms
XM_023592922.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	1170	0	no GO terms
XM_023592923.1	Cercospora sojina strain RACE15 chromosome III, complete sequence	987	0	no GO terms
XM_023592924.1	Cercospora beticola 60S ribosomal protein L14-B (CB0940_00291), mRNA	438	0	P:translation; F:RNA binding; F:structural constituent of ribosome; C:ribosome
XM_023592925.1	Cercospora beticola Nascent polypeptide-associated complex subunit alpha (CB0940_00292), mRNA	606	0	C:nascent polypeptide-associated complex
XM_023592926.1	Cercospora beticola Phosphatidylserine decarboxylase proenzyme 3 (CB0940_00293), mRNA	3450	0	P:phospholipid biosynthetic process; F:phosphatidylserine decarboxylase activity; F:calcium ion binding

Supplementary Table 8: ID of genes 50 kb up- and down-stream of the 2.95 Mb F_{ST} outlier region on CM008502.1, as well length and GO terms

ID	Best Match	Length	E-value	GO term
XM_023597582.1	Cercospora beticola putative glycosidase (CB0940_04868), mRNA	2424	0	P:carbohydrate metabolic process; F:carbohydrate binding
XM_023597583.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	387	0	no GO terms
XM_023597584.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1695	0	no IPS match
XM_023597585.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	318	9.57E-160	no GO terms
XM_023597586.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	2928	0	no GO terms
XM_023597587.1	Cercospora beticola Protein get1 (CB0940_04873), mRNA	633	0	P:tail-anchored membrane protein insertion into ER membrane
XM_023597588.1	Cercospora beticola Ethanolamine-phosphate cytidylyltransferase (CB0940_04874), mRNA	1362	0	P:biosynthetic process; F:catalytic activity
XM_023597589.1	Cercospora beticola Delta-aminolevulinic acid dehydratase (CB0940_04875), mRNA	1146	0	P:tetrapyrrole biosynthetic process; F:catalytic activity; F:porphobilinogen synthase activity; F:metal ion binding
XM_023597590.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	879	0	F:zinc ion binding; F:ubiquitin protein ligase activity
XM_023597591.1	Cercospora beticola Acyl-coenzyme A oxidase (CB0940_04877), mRNA	2133	0	P:fatty acid metabolic process; P:fatty acid beta-oxidation; P:obsolete oxidation-reduction process; F:acyl-CoA oxidase activity; F:oxidoreductase activity, acting on the CH-CH group of donors; F:flavin adenine dinucleotide binding; F:FAD binding; C:peroxisome
XM_023597592.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1380	0	no GO terms
XM_023597593.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	678	0	P:glutathione metabolic process; F:protein binding

Supplementary Table 8 (continued): ID of genes 50 kb up- and down-stream of the 2.95 Mb F_{ST} outlier region on CM008502.1, as well length and GO terms

ID	Best Match	Length	E-value	GO term
XM_023597594.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	915	0	no IPS match
XM_023597595.1	Cercospora beticola putative transporter (CB0940_04881) mRNA	1584	0	P:transmembrane transport; F:transmembrane transporter activity
XM_023597596.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1101	0	no GO terms
XM_023597597.1	Cercospora beticola putative SWI/SNF-related regulator of chromatin (CB0940_04883) mRNA	1326	0	F:ATP binding; F:nucleosome-dependent ATPase activity
XM_023597598.1	---NA---	558		F:ATP binding; F:nucleosome-dependent ATPase activity
XM_023597599.1	Cercospora beticola hypothetical protein (CB0940_04885), mRNA	1443	0	no GO terms
XM_023597600.1	Cercospora beticola hypothetical protein (CB0940_04886), mRNA	2544	0	no GO terms
XM_023597601.1	Cercospora beticola ATP-dependent RNA helicase mtr4 (CB0940_04887), mRNA	3246	0	P:RNA catabolic process; F:nucleic acid binding; F:RNA binding; F:RNA helicase activity; F:ATP binding
XM_023597602.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	939	0	no GO terms
XM_023597603.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1071	0	no GO terms
XM_023597604.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	2031	0	C:Ada2/Gcn5/Ada3 transcription activator complex
XM_023597605.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	531	0	no GO terms
XM_023597606.1	Cercospora beticola Branched-chain-amino-acid aminotransferase, mitochondrial (CB0940_04892), mRNA	1314	0	P:branched-chain amino acid metabolic process; F:catalytic activity; F:branched-chain-amino-acid transaminase activity

Supplementary Table 8 (continued): ID of genes 50 kb up- and down-stream of the 2.95 Mb F_{ST} outlier region on CM008502.1, as well length and GO terms

ID	Best Match	Length	E-value	GO term
XM_023597607.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1617	0	no GO terms
XM_023597608.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	3021	0	no GO terms
XM_023597609.1	Cercospora beticola Alpha-L-fucosidase (CB0940_04895) mRNA	1839	0	P:carbohydrate metabolic process; P:fucose metabolic process; F:alpha-L-fucosidase activity
XM_023597610.1	Cercospora beticola O-methylsterigmatocystin oxidoreductase (CB0940_04896) mRNA	1575	0	P:obsolete oxidation-reduction process; F:iron ion binding; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding
XM_023597611.1	Cercospora beticola Mediator of RNA polymerase II transcription subunit 5 (CB0940_04897), mRNA	3111	0	P:regulation of transcription by RNA polymerase II; F:transcription coregulator activity; C:mediator complex
XM_023597612.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	558	0	no GO terms
XM_023597613.1	Cercospora beticola putative oxidoreductase (CB0940_04899) mRNA	777	0	P:obsolete oxidation-reduction process
XM_023597614.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1056	0	no GO terms
XM_023597615.1	Cercospora beticola Ribonucleoside-diphosphate reductase large chain (CB0940_04901), mRNA	2742	0	P:DNA replication; P:obsolete oxidation-reduction process; F:ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor; F:ATP binding
XM_023597616.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	2730	0	no GO terms

Supplementary Table 8 (continued): ID of genes 50 kb up- and down-stream of the 2.95 Mb F_{ST} outlier region on CM008502.1, as well length and GO terms

ID	Best Match	Length	E-value	GO term
XM_023597617.1	Cercospora beticola Ornithine carbamoyltransferase, mitochondrial (CB0940_04903), mRNA	1092	0	P:cellular amino acid metabolic process; F:amino acid binding; F:carboxyl- or carbamoyltransferase activity
XM_023597618.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1041	0	F:nucleic acid binding
XM_023597619.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	1056	0	P:lipid biosynthetic process; P:obsolete oxidation-reduction process; F:iron ion binding; F:oxidoreductase activity
XM_023597620.1	Cercospora beticola Acetolactate synthase small subunit, mitochondrial (CB0940_04906), mRNA	1020	0	P:branched-chain amino acid biosynthetic process; F:acetolactate synthase regulator activity
XM_023597621.1	Cercospora beticola hypotheticalsprotein (CB0940_04907), mRNA	1587	0	F:RNA binding; C:nucleus
XM_023597622.1	Cercospora beticola hypotheticalsprotein (CB0940_04907), mRNA	1074	0	F:RNA binding; C:nucleus
XM_023597623.1	Cercospora sojina strain RACE15 chromosome XII, complete sequence	960	0	no GO terms

Supplementary Table 9: Further characterisation of the outlier regions on CM008499.1, including whether proteins are effectors, whether these contain signal peptide regions, Tajima's D, and pairwise F_{ST} between the UK and the other four sampling sites

					Mean pairwise F_{ST} of each gene			
Gene ID	**Effector Status**	**Signal Peptide**	**Number of SNPS per gene**	**Tajima's D**	**UK vs Croatia**	**UK vs Italy**	**UK vs ND**	**UK vs NY**
XM_023592888.1	Non-Effector	No	7	-0.61	0.15	0.133	0.21	0.19
XM_023592889.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592890.1	Non-Effector	No	5	3.25	0.1	0.65	0.11	0.53
XM_023592891.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592892.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592893.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592894.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592895.1	Non-Effector	Yes	0	NA	NA	NA	NA	NA
XM_023592896.1	Non-Effector	No	18	1.83	0.24	0.47	0.29	0.48
XM_023592897.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592898.1	Non-Effector	No	2	0.55	0.1	0.65	0.13	0.5
XM_023592899.1	Non-Effector	Yes	22	0.25	0.2	0.01	0.31	0.16
XM_023592900.1	Non-Effector	Yes	48	2.42	0.04	0.008	0.28	0.04
XM_023592901.1	Non-Effector	No	9	0.74	0.09	0.37	0.24	0.39
XM_023592902.1	Non-Effector	No	5	0.94	0.19	0.15	0.32	0.17
XM_023592903.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592904.1	Non-Effector	No	19	3.23	0.08	0.07	0.44	0.07
XM_023592905.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592906.1	Non-Effector	No	73	-1.05	0.39	0.43	0.66	0.46
XM_023592907.1	Effector	No	6	-1.9	0.02	NA	NA	0.08
XM_023592908.1	Non-Effector	No	9	-2.1	0.02	NA	NA	NA
XM_023592909.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592910.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592911.1	Non-Effector	No	1	-0.95	0.02	NA	NA	NA
XM_023592912.1	Non-Effector	No	4	0.03	0.57	0.15	0.35	0.57
XM_023592913.1	Non-Effector	No	11	2.8	0.38	0.5	0.51	0.21
XM_023592914.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592915.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592916.1	Non-Effector	No	23	0.2	0.56	0.53	0.76	0.56
XM_023592917.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023592918.1	Non-Effector	No	3	-0.13	0	0.4	0.07	0.14
XM_023592919.1	Non-Effector	No	3	-0.01	0.11	0.13	0.28	0.13
XM_023592920.1	Non-Effector	No	2	0.25	0	0.02	0	0
XM_023592921.1	Non-Effector	No	16	-0.82	0.24	0.13	0.33	0.22
XM_023592922.1	Non-Effector	No	40	3.24	0.2	0.06	0.19	0.14
XM_023592923.1	Non-Effector	No	0	NA	NA	NA	NA	NA

Supplementary Table 9 (continued): Further characterisation of the outlier regions on CM008499.1, including whether proteins are effectors, whether these contain signal peptide regions, Tajima's D, and pairwise F_{ST} between the UK and the other four sampling sites

					Mean pairwise F_{ST} of each gene			
Gene ID	**Effector Status**	**Signal Peptide**	**Number of SNPS per gene**	**Tajima's D**	**UK vs Croatia**	**UK vs Italy**	**UK vs ND**	**UK vs NY**
XM_023592924.1	Effector	No	3	0.16	0.206	0.45	0.26	0.28
XM_023592925.1	Non-Effector	No	5	-1.09	0.19	0.17	0.38	0.19
XM_023592926.1	Non-Effector	No	18	4.52	0.11	0.61	0.13	0.2

Supplementary Table 10: Further characterisation of the outlier regions on CM008502.1, including whether proteins are effectors, whether these contain signal peptide regions, Tajima's D, and pairwise F_{ST} between the UK and the other four sampling sites

					Mean pairwise F_{ST} of each gene			
Gene ID	**Effector Status**	**Signal Peptide**	**Number of SNPS per gene**	**Tajima's D**	**UK vs Croatia**	**UK vs Italy**	**UK vs ND**	**UK vs NY**
XM_023597582.1	Non-Effector	Yes	9	-0.56	0.29	0.31	0	0.02
XM_023597583.1	Non-Effector	No	1	-0.94	NA	NA	0	NA
XM_023597584.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597585.1	Non-Effector	Yes	8	-1.73	0.14	0.06	0.01	0.02
XM_023597586.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597587.1	Effector	Yes	2	-1.28	0.005	0.03	0.02	0
XM_023597588.1	Non-Effector	No	3	0.22	0.29	0.37	0.02	0.41
XM_023597589.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597590.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597591.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597592.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597593.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597594.1	Non-Effector	Yes	29	-2.17	0.02	0.23	0.15	0.06
XM_023597595.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597596.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597597.1	Non-Effector	No	23	-0.48	0.22	0.27	0.16	0.12
XM_023597598.1	Effector	No	3	-1.47	0.07	0.03	0	0
XM_023597599.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597600.1	Non-Effector	No	20	-1.38	0.06	0.36	0.15	NA
XM_023597601.1	Non-Effector	No	12	-1.5	0.14	0.36	0	0
XM_023597602.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597603.1	Non-Effector	No	66	2.15	0.29	0.21	0.25	0.07
XM_023597604.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597605.1	Effector	No	0	NA	NA	NA	NA	NA
XM_023597606.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597607.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597608.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597609.1	Non-Effector	Yes	70	4.36	0.69	0.003	0.4	0.17
XM_023597610.1	Non-Effector	No	3	2.9	0.75	0.04	0.35	0.14
XM_023597611.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597612.1	Non-Effector	Yes	4	0.19	0.25	0.28	0.1	0
XM_023597613.1	Effector	No	48	2.52	0.75	0.14	0.17	0.14
XM_023597614.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597615.1	Non-Effector	No	6	-1.75	0.01	0.01	0.05	0.15
XM_023597616.1	Non-Effector	No	9	-1.21	0.27	0.04	0.37	0.14
XM_023597617.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597618.1	Non-Effector	No	0	NA	NA	NA	NA	NA

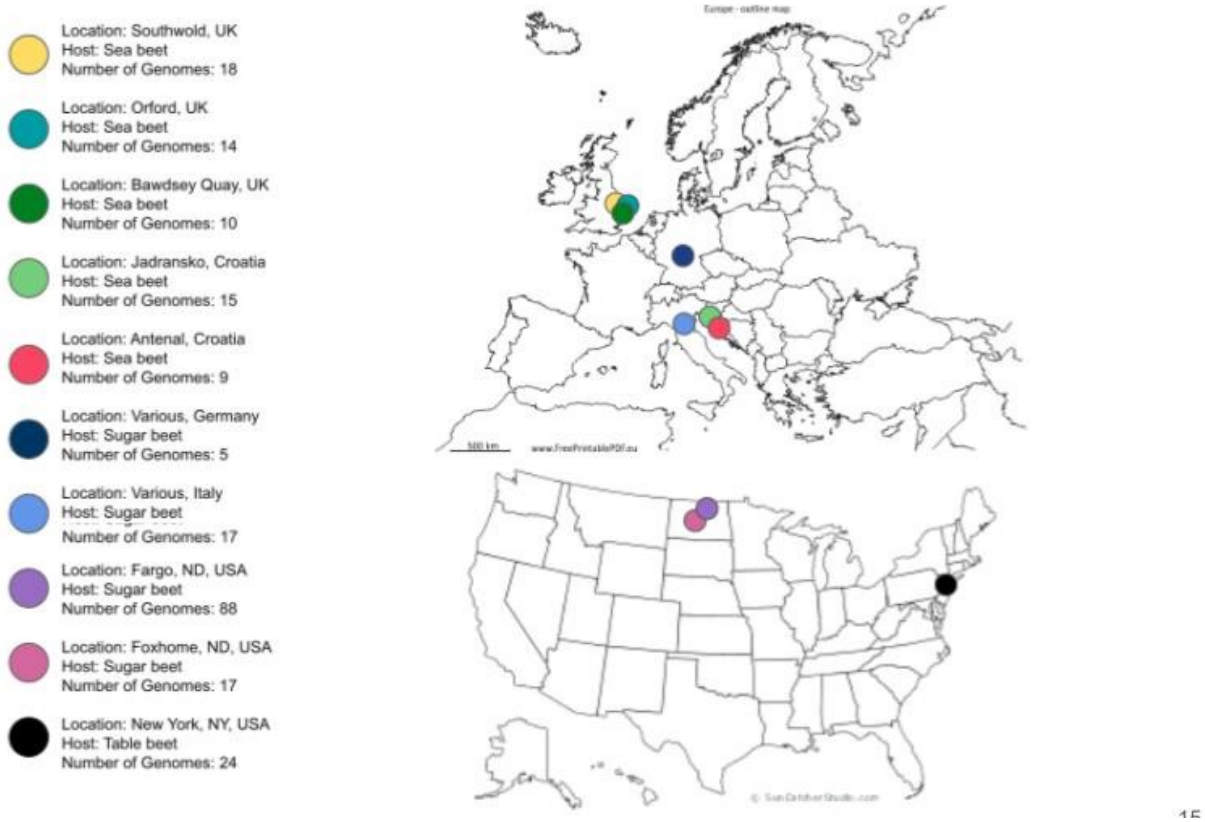
Supplementary Table 10 (continued): Further characterisation of the outlier regions on CM008502.1, including whether proteins are effectors, whether these contain signal peptide regions, Tajima's D, and pairwise F_{ST} between the UK and the other four sampling sites

					Mean pairwise F_{ST} of each gene			
Gene ID	**Effector Status**	**Signal Peptide**	**Number of SNPS per gene**	**Tajima's D**	**UK vs Croatia**	**UK vs Italy**	**UK vs ND**	**UK vs NY**
XM_023597619.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597620.1	Non-Effector	No	0	NA	NA	NA	NA	NA
XM_023597622.1	Non-Effector	No	4	-0.414	0.04	0.02	0.41	0.21
XM_023597621.1	Non-Effector	No	4	-0.414	0.04	0.02	0.41	0.21
XM_023597623.1	Non-Effector	No	3	-0.06	0.04	0.006	0.42	0.14

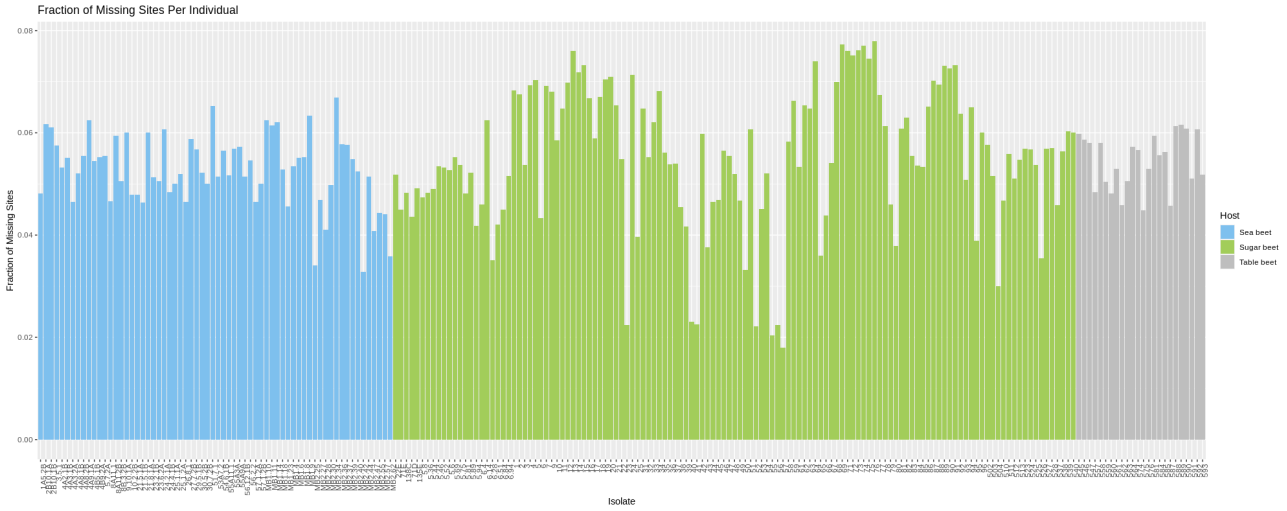
Supplementary Table 11: Functional Role of Mutations in the Differentiated Region on CM008502.1

ID	Type of Mutation	Croatia	Italy	North Dakota	New York	UK
XM_023597609.1	Missense	13	7	16	10	9
	Synonymous	51	27	51	51	51
XM_023597610.1	Missense	2	2	2	2	2
	Synonymous	0	0	0	0	0
	Stop codon gained	1	1	1	1	1
	Frequency of stop codon	0.79	0.06	0.51	0.25	0.05
XM_023597612.1	Missense	2	2	2	1	1
	Synonymous	1	1	1	0	1
XM_023597613.1	Synonymous	48	46	47	2	2

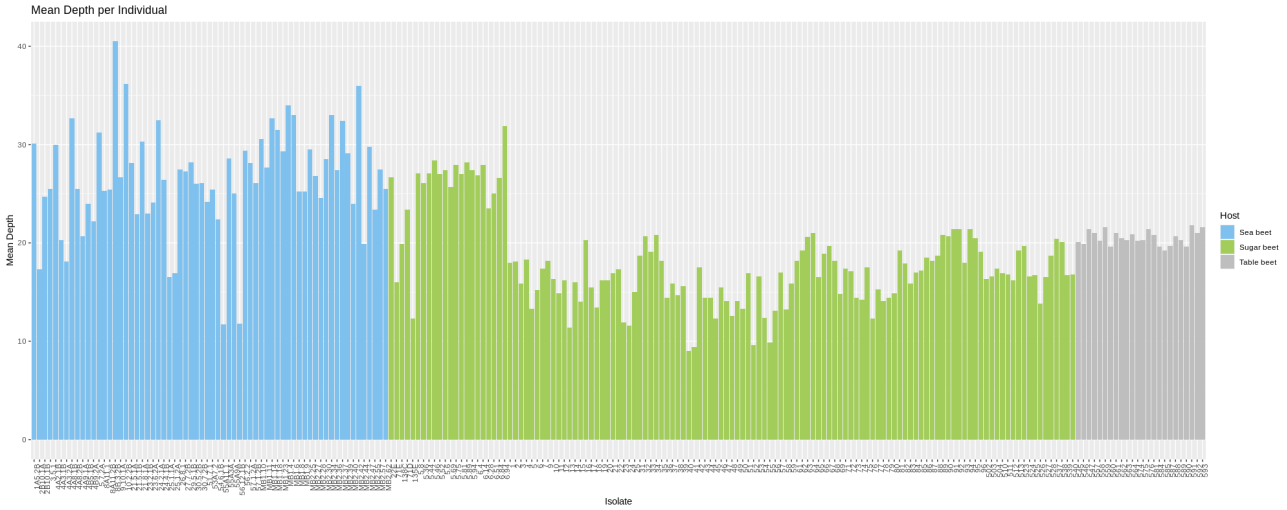
Supplementary Figures



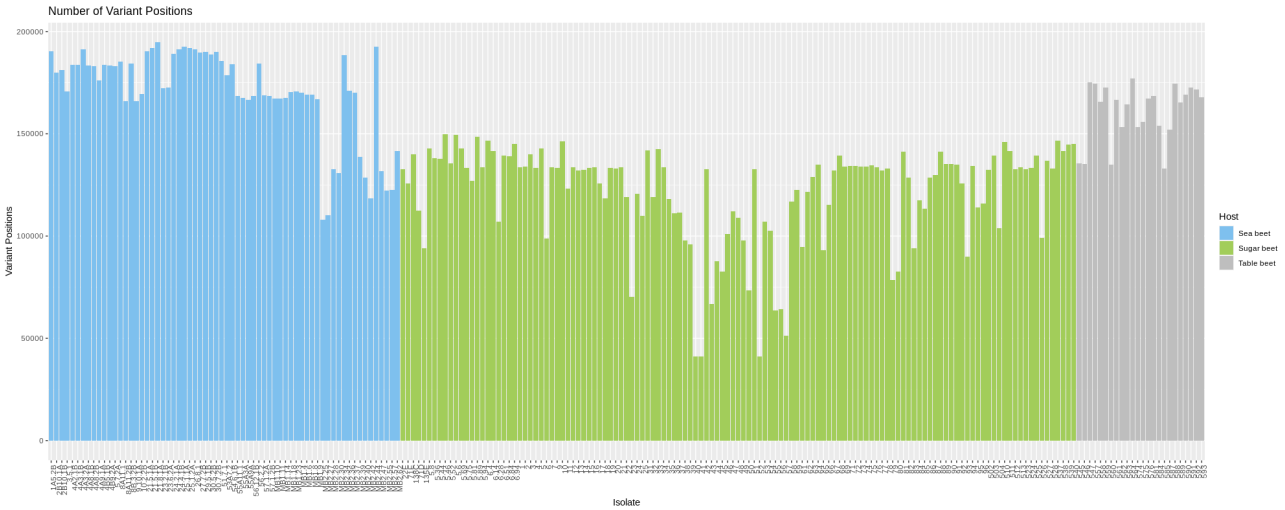
Supplementary Figure 1: Map of where *C. betcola* isolates were collected, as well as the host, and the number of isolates from each sampling location.



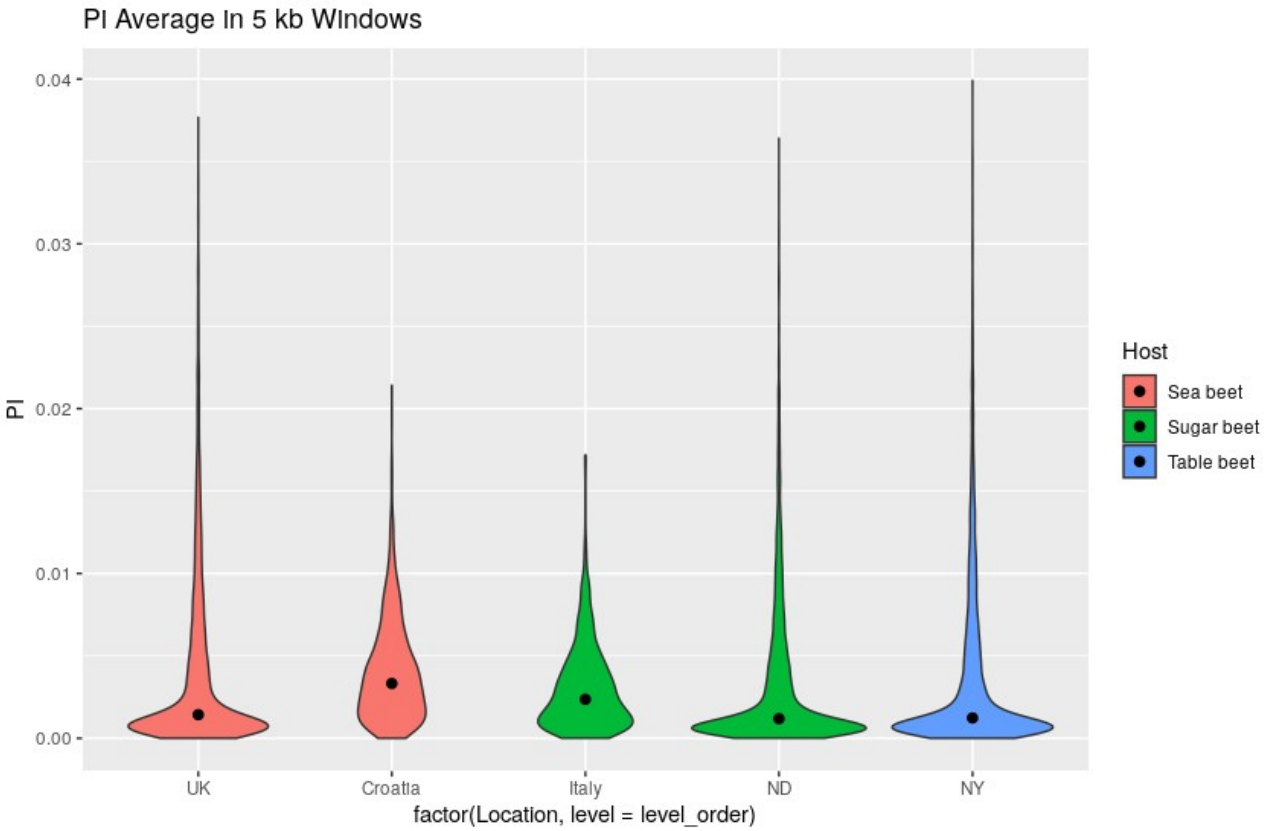
Supplementary Figure 2: The proportion of missing sites per individual of the *C. beticola* isolates. Isolates had different proportions of missingness, and no individual contained more than 8% missing sites. *C. beticola* isolates from sea beet were represented by blue, sugar beet by green, and from table beet by grey bars.



Supplementary Figure 3: The average sequencing depth per individual *C. beticola* isolate. There is some variation between individuals, and all individuals were sequenced to a depth of at least 10x. *C. beticola* isolates from sea beet were represented by blue, sugar beet by green, and from table beet by grey bars.

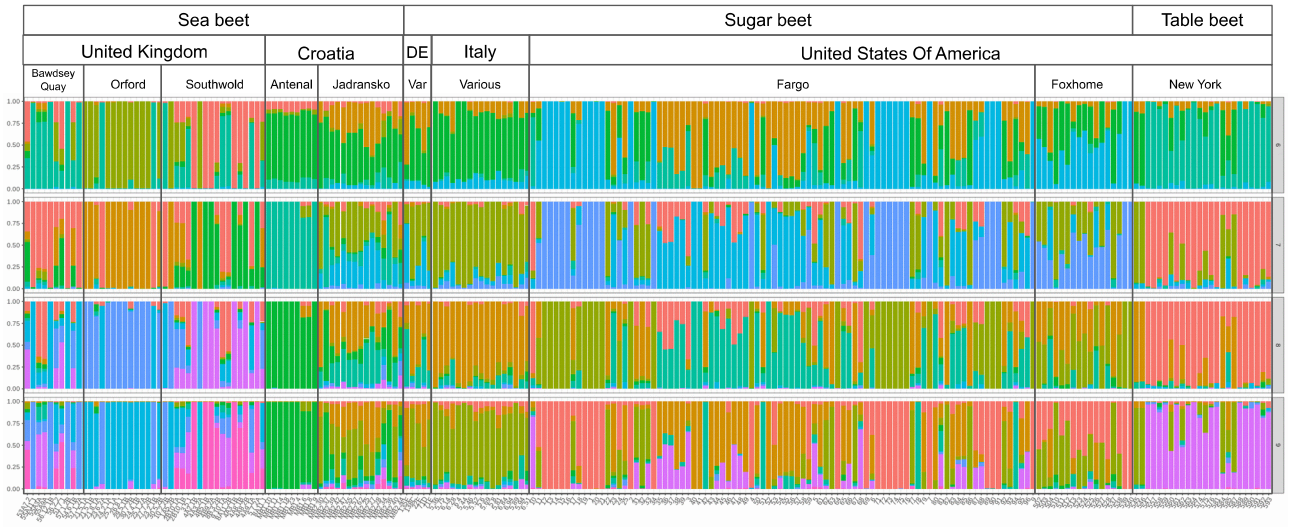


Supplementary Figure 4: Number of SNPs per individual in each *C. beticola* individual. Individuals from different sampling locations showed different numbers of variant positions. Isolates from sea beet were more diverse than those from domesticated beets. *C. beticola* isolates from sea beet were represented by blue, sugar beet by green, and from table beet by grey bars.

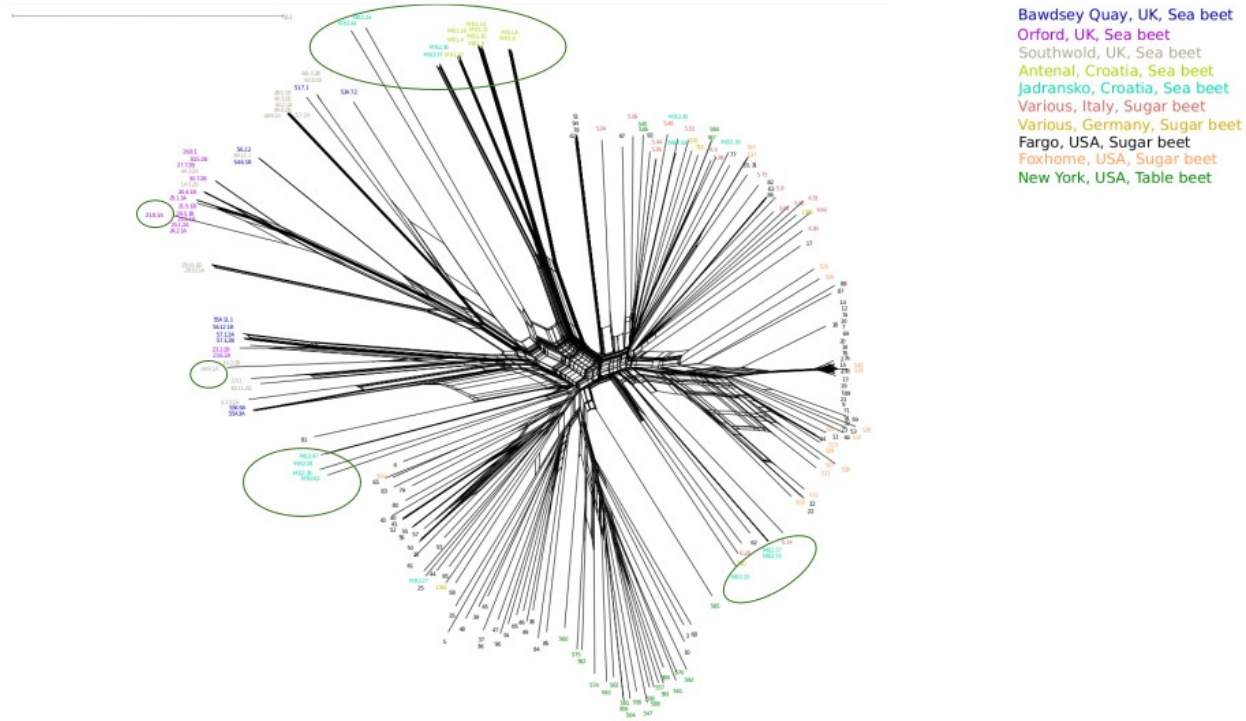


Supplementary Figure 5: Violin plots of the Tajima's D distribution across the genome for *C. beticola* from the various sampling sites in 5 kb windows. Hosts are indicated by colours.

Supplementary Figure 6: R^2 decay over the entire chromosome for individuals from each sampling site. The different colours indicate different sampling sites. There was variation that R^2 decay occurred along each chromosome, with some chromosomes from the Orford individuals showing virtually no R^2 decay. This indicated that there may be differences in recombination rates between chromosomes in *C. beticola.*



Supplementary Figure 6: Population structure of the *C. beticola* isolates. SNPs were thinned to 5 kb distance of one another for K = 6-9. The isolates from the UK showed higher levels of admixture than isolates from mainland Europe and the USA.



Supplementary Figure 7: Network showing which individuals contain the premature stop codon in the XM_023597610.1 gene. Only individuals in the UK and Croatian accessions that contained the stop codon are indicated on this network by circles. The isolates within the North American accessions were interspersed randomly among isolates that did not contain the stop codon.

Supplementary Materials for Protocols in the Laboratory

Media used to culture *C. beticola*

Water agar plates • 7.5 g Agar • 500 ml H_2O
Tomato agar plates • 250 ml Tomato juice • 250 ml H2O • 7.5 g Agar • 0.1 g Streptomycin-sulfate in 500 µl sterile H2O added to media after sterilisation
Tomato liquid media • 250 ml Tomato juice • 250 ml H_2O

Supplementary Figure 8: Tomato juice from the local supermarket used to produce the media to culture *C.beticola*

Chemicals for CTAB Extraction

Extraction Buffer for 100ml

- 10 ml 1M Tris-Hcl pH8
- 28 ml 5M NaCl
- 4 ml 0.5M EDTA
- 2 g CTAB
- 58 ml H_2O
- 500 µl B-mercaptoethanol

Phenol/chloroform 1:1

Isopropanol

TE buffer, pH8

Rnase (10mg/ml)

3 M Sodium acetate (NaOAc)

100% Ethanol

75% Ethanol

Scripts

count.reads.sh

```
parallel “echo {} && gunzip -c {} | wc -l | awk ‘{d=\$1; print d/4;}” ::: *.gz
```

Reference based mapping pipeline in various bash scripts

```
#read trimming
for f1 in your_directory/*fastq
do
	fastq_quality_trimmer -t 20 -l 30 -i ${f1} -o ${f1}.trimmed
done

#repairing trimmed reads
for f1 in your_directory/*_R1_001.fastq.trimmed.gz
do
	f2=${f1%%_R1_001.fastq.trimmed.gz}"_R2_001.fastq.trimmed.gz"
	cd your_programs/bbmap/
	./repair.sh in1=${f1} in2=${f2} out1=${f1}fixed.fq out2=${f2}fixed.fq outsingle=${f1}.single.fq
done

#mapping with bwa-mem
for f1 in your_directory/*_1.fq.gz
do
	f2=${f1%%_1.fq.gz}"_2.fq.gz"
	bwa mem -t 4 -I -P /media/lizel/82AEF34EAEF3396D/chapter2/ref/core.fna ${f1} ${f2} > $
	{f1}.sam
	samtools view -S -b ${f1}.sam > ${f1}.bam
	rm *sam
done

#sort and index bam files
for f1 in your_directory/*bam
do
	samtools sort -@ 20 ${f1} > ${f1}.sorted
done

#if there were resequencing runs, merge the two (or more) bam files
	samtools merge ${run1} ${run2} > ${run1}.merged

#remove duplicates and index bam files
for f1 in your_directory/*bam.sorted.bam
do
	samtools rmdup -S ${f1} ${f1}_rmdup.bam
	samtools index ${f1}_rmdup.bam
done

#Reference genome must be formatted for GATK
	samtools faidx ref.fasta
```

```
#GATK HaplotypeCaller to produce GVCFs for each accession
for f1 in /media/lizel/82AEF34EAEF3396D/chapter2/italy.sugar/*rg
do
      gatk HaplotypeCaller -R ref.fasta -I ${f1} -O ${f1}.g.vcf.gz -ERC GVCF -ploidy 1
done

#Combine all of the GVCF files
      gatk CombineGVCFs –variant 1 –variant 2 -O combined.g.vcf.gz

#Genotype GVCFs
      gatk GenotypeGVCFs -R ref.fasta -V combined.g.vcf.gz -O raw.vcf

#raw.vcf can be further processed with VCFtools, and any downstream analyses!
```

cercospora.plots.R

```
library(tidyverse)
library(ggplot2)
library(dplyr)
library(ggpubr)
library(rstatix)
library(nortest)
library(Hmisc)
library(readODS)

setwd("your_directory")
master = read_ods("you master file with all info in different columns")

#Depth per site
ggplot(depth, aes(x=POS, y= MEAN_DEPTH)) +
  geom_line() +
  facet_wrap(~CHROM)

#Depth per individual
ggplot(master, aes(x=reorder(Isolate.Name, Row.Number), y =Mean.Depth, fill=Host)) +
  geom_bar(stat = "identity") +
  facet_wrap(~Country, scales = "free_x") +
  theme(legend.position = "right",axis.text.x=element_text(angle=90,hjust=1)) +
  scale_color_manual(labels = c("Sea beet", "Sugar beet", "Table beet"),
                     values = c("skyblue2", "darkolivegreen3", "grey"), aesthetics = "fill") +
  labs(title="Mean Depth per Individual", x = "Isolate", y = "Mean Depth", color = "Host")

#plot number of snps per individual
ggplot(master, aes(x=reorder(Isolate.Name, Row.Number), y = Variant, fill = Host)) +
  geom_bar(stat = "identity") +
  facet_wrap(~Country, scales = "free_x") +
  theme(legend.position = "right",axis.text.x=element_text(angle=90,hjust=1)) +
  scale_color_manual(labels = c("Sea beet", "Sugar beet", "Table beet"),
                     values = c("skyblue2", "darkolivegreen3", "grey"), aesthetics = "fill") +
  labs(title="Number of Variant Positions Per Individual", x = "Isolate", y = "Variant Positions", color
= "Host")
```

```
#plot number of missingness per individual
ggplot(master, aes(x=reorder(Isolate.Name, Row.Number), y = Percent.Missing, fill = Host)) +
  geom_bar(stat = "identity") +
  facet_wrap(~Country, scales = "free_x") +
  theme(legend.position = "right",axis.text.x=element_text(angle=90,hjust=1)) +
  scale_color_manual(labels = c("Sea beet", "Sugar beet", "Table beet"),
                     values = c("skyblue2", "darkolivegreen3", "grey"), aesthetics = "fill") +
  labs(title="Fraction of Missing Sites Per Individual", x = "Isolate", y = "Fraction of Missing Sites",
color = "Host")

#Prepare environment for pi plot
uk = read.table("uk.recode.vcf.windowed.pi", header = T)
nd = read.table("nd.recode.vcf.windowed.pi", header = T)
ny = read.table("ny.recode.vcf.windowed.pi", header = T)
croatia = read.table("croatia.recode.vcf.windowed.pi", header = T)
italy = read.table("italy.recode.vcf.windowed.pi", header = T)

uk = mutate(uk, Location = "UK")
uk = mutate(uk, Host = "Sea beet")
nd = mutate(nd, Location = "ND")
nd = mutate(nd, Host = "Sugar beet")
ny = mutate(ny, Location = "NY")
ny = mutate(ny, Host = "Table beet")
croatia = mutate(croatia, Location = "Croatia")
croatia = mutate(croatia, Host = "Sea beet")
italy = mutate(italy, Location = "Italy")
italy = mutate(italy, Host = "Sugar beet")

combined = c('uk', 'nd', 'ny', 'croatia', "italy")
x.list <- lapply(combined, get)


combined.pi = do.call(rbind, x.list)

#Prepare environment for Tajima's D plot
uk = read.table("uk.recode.vcf.Tajima.D", header = T)
nd = read.table("nd.recode.vcf.Tajima.D", header = T)
ny = read.table("ny.recode.vcf.Tajima.D", header = T)
croatia = read.table("croatia.recode.vcf.Tajima.D", header = T)
italy = read.table("italy.recode.vcf.Tajima.D", header = T)

uk = mutate(uk, Location = "UK")
uk = mutate(uk, Host = "Sea beet")
nd = mutate(nd, Location = "ND")
nd = mutate(nd, Host = "Sugar beet")
ny = mutate(ny, Location = "NY")
ny = mutate(ny, Host = "Table beet")
croatia = mutate(croatia, Location = "Croatia")
croatia = mutate(croatia, Host = "Sea beet")
italy = mutate(italy, Location = "Italy")
italy = mutate(italy, Host = "Sugar beet")

combined = c('uk', 'nd', 'ny', 'croatia', "italy")
```

```
x.list <- lapply(combined, get)


combined.tajd = do.call(rbind, x.list)

level_order <- c('UK', 'Croatia', 'Italy', "ND", "NY")

#Plot Tajima's D and pi
ggplot(combined.pi, aes(x= factor(Location, level = level_order), y = PI, fill = Host)) +
  geom_violin() +
  ggtitle("Pi Average in 5 kb Windows") +
  stat_summary(fun=median, geom="point", size=2, color="black")
  stat_summary(fun.data="mean_sdl", mult=1,
               geom="pointrange", width=0.2 )


ggplot(combined.tajd, aes(x= factor(Location, level = level_order), y = TajimaD, fill = Host)) +
    geom_violin() +
    ggtitle("Tajima's D Average in 5 kb Windows") +
  xlab("Location") + ylab("Tajima's D") +
    stat_summary(fun=median, geom="point", size=2, color="black")
  stat_summary(fun.data="mean_sdl", mult=1,
               geom="pointrange", width=0.2 )
```

anova.of.tajd.R

```
library(tidyverse)
library(ggplot2)
library(dplyr)
library(ggpubr)
library(rstatix)
library(nortest)
library(Hmisc)

setwd("your_directory")

#Prepare environment for pi
uk = read.table("uk.recode.vcf.windowed.pi", header = T)
nd = read.table("nd.recode.vcf.windowed.pi", header = T)
ny = read.table("ny.recode.vcf.windowed.pi", header = T)
croatia = read.table("croatia.recode.vcf.windowed.pi", header = T)
italy = read.table("italy.recode.vcf.windowed.pi", header = T)

uk = mutate(uk, Location = "UK")
uk = mutate(uk, Host = "Sea beet")
nd = mutate(nd, Location = "ND")
nd = mutate(nd, Host = "Sugar beet")
ny = mutate(ny, Location = "NY")
ny = mutate(ny, Host = "Table beet")
croatia = mutate(croatia, Location = "Croatia")
croatia = mutate(croatia, Host = "Sea beet")
```

```
italy = mutate(italy, Location = "Italy")
italy = mutate(italy, Host = "Sugar beet")

combined = c('uk', 'nd', 'ny', 'croatia', "italy")
x.list <- lapply(combined, get)


combined.pi = do.call(rbind, x.list)

#Prepare environment for Tajima's D
uk = read.table("uk.recode.vcf.Tajima.D", header = T)
nd = read.table("nd.recode.vcf.Tajima.D", header = T)
ny = read.table("ny.recode.vcf.Tajima.D", header = T)
croatia = read.table("croatia.recode.vcf.Tajima.D", header = T)
italy = read.table("italy.recode.vcf.Tajima.D", header = T)

uk = mutate(uk, Location = "UK")
uk = mutate(uk, Host = "Sea beet")
nd = mutate(nd, Location = "ND")
nd = mutate(nd, Host = "Sugar beet")
ny = mutate(ny, Location = "NY")
ny = mutate(ny, Host = "Table beet")
croatia = mutate(croatia, Location = "Croatia")
croatia = mutate(croatia, Host = "Sea beet")
italy = mutate(italy, Location = "Italy")
italy = mutate(italy, Host = "Sugar beet")

combined = c('uk', 'nd', 'ny', 'croatia', "italy")
x.list <- lapply(combined, get)

combined.tajd = do.call(rbind, x.list)

level_order <- c('UK', 'Croatia', 'Italy', "ND", "NY")

group_by(combined.tajd, Location) %>%
  summarise(
    count = n(),
    mean = mean(TajimaD, na.rm = TRUE),
    sd = sd(TajimaD, na.rm = TRUE)
  )

#HOST
#check outliers
temp = combined.tajd %>%
  group_by(Host) %>%
  identify_outliers(TajimaD)

#build linear model
model  <- lm(TajimaD ~ Host, data = combined.tajd)
ggqqplot(residuals(model))

# Compute Shapiro-Wilk test of normality
shapiro_test(residuals(model))
ad.test(combined.tajd$TajimaD)
```

```
ggqqplot(combined.tajd, "TajimaD", facet.by = "Host")

plot(model, 1)

res.aov <- combined.tajd %>% anova_test(TajimaD ~ Host)
res.aov

pwc <- combined.tajd %>% tukey_hsd(TajimaD ~ Host)
pwc

pwc <- pwc %>% add_xy_position(x = "Host")
ggboxplot(combined.tajd, x = "Host", y = "TajimaD") +
  stat_pvalue_manual(pwc, hide.ns = TRUE) +
  labs(
    subtitle = get_test_label(res.aov, detailed = TRUE),
    caption = get_pwc_label(pwc)
  )


#LOCATION

#check outliers
temp = combined.tajd %>%
  group_by(Location) %>%
  identify_outliers(TajimaD)

#build linear model
model  <- lm(TajimaD ~ Location, data = combined.tajd)
ggqqplot(residuals(model))

# Compute Shapiro-Wilk test of normality
shapiro_test(residuals(model))
ad.test(combined.tajd$TajimaD)


ggqqplot(combined.tajd, "TajimaD", facet.by = "Location")

plot(model, 1)

res.aov <- combined.tajd %>% anova_test(TajimaD ~ Location)
res.aov

pwc <- combined.tajd %>% tukey_hsd(TajimaD ~ Location)
pwc

pwc <- pwc %>% add_xy_position(x = "Location")
ggboxplot(combined.tajd, x = "Location", y = "TajimaD") +
  stat_pvalue_manual(pwc, hide.ns = TRUE) +
  labs(
    subtitle = get_test_label(res.aov, detailed = TRUE),
    caption = get_pwc_label(pwc)
  )
```

```
res.aov2 <- aov(TajimaD ~ Location + Host, data = combined.tajd)
summary(res.aov2)
```

ld.R

```
#from https://www.biostars.org/p/300381/

library(dplyr)
library(stringr)
library(ggplot2)
dfr <-read.delim("/media/lizel/82AEF34EAEF3396D/chapter2/2_vcf/34_rehh/chr/
uk.recode.vcf.chr2.recode.vcf.biallelic.nomissing.recode.vcf.75k.recode.vcf.summary",sep="",hea
der=F,check.names=F,stringsAsFactors=F)
colnames(dfr) <- c("dist","rsq")
dfr$distc <- cut(dfr$dist,breaks=seq(from=min(dfr$dist)-1,to=max(dfr$dist)+1,by=10000))
dfr1 <- dfr %>% group_by(distc) %>% summarise(mean=mean(rsq),median=median(rsq))
dfr1 <- dfr1 %>% mutate(start=as.integer(str_extract(str_replace_all(distc,"[\\(\\)\\[\\]]",""),"^[0-9-
e+.]+")),
                        end=as.integer(str_extract(str_replace_all(distc,"[\\(\\)\\[\\]]",""),"[0-9-e+.]+$")),
                        mid=start+((end-start)/2))

ggplot(dfr1, aes(x=start,y=mean))+
  geom_line()
```

vcf2geno.sh

```
for f1 in /media/lizel/82AEF34EAEF3396D/chapter2/2_vcf/45_pop/*GT.FORMAT
do
#vcftools --vcf $VCFNAME.vcf --extract-FORMAT-info GT
cat ${f1} | cut -f 3- > ${f1}.2
head -n 1 ${f1}.2 | sed "s/\t/\n/g" > ${f1}.ind
sed "s/\t//g" ${f1}.2 | tail -n +2 > ${f1}.geno
done
```

snmf.R

```
library(LEA)
library(tidyverse)

### Example of analysis using snmf ###
# Creation of the genotype file: genotypes.geno.
# The data contain 400 SNPs for 50 individuals.
#data("tutorial")
#write.geno(tutorial.R, "genotypes.geno")
```

```
#################
# running snmf #
#################
project.snmf = snmf("yourfiles.GT.FORMAT.geno",
                    K = 1:20,
                    entropy = TRUE,
                    ploidy = 1,
                    repetitions = 10,
                    project = "new")
# plot cross-entropy criterion of all runs of the project
plot(project.snmf, cex = 1.2, col = "lightblue", pch = 19)
# get the cross-entropy of the 10 runs for K = 4
ce = cross.entropy(project.snmf, K = 4)
# select the run with the lowest cross-entropy for K = 4
best = which.min(ce)

#add names
metadata = read.table("your .d file", header =T)
indv_snmf = read_tsv("your isolate name filet", col_names = F)
names(indv_snmf) = "Sample"
datalist = list()
for (i in c(6, 7, 8, 9)){
  best = which.min(cross.entropy(project.snmf, K = i))
  temp = as.data.frame(Q(project.snmf, i, best))
  temp= cbind(indv_snmf, temp)
  temp = temp %>%
    gather("Cluster", "Admix_coef", -"Sample") %>%
    mutate(K=i)
  datalist[[i]] = as.tibble(temp)
}
snmf_results_per_K = bind_rows(datalist)%>%
  inner_join(., metadata, by = c("Sample" = "Isolate")) %>%
  unite(ID, Country, Location, Host, col = "for_display", remove = F)

ggplot(snmf_results_per_K, aes(x = reorder(Sample, ID), y = Admix_coef, fill = Cluster,
                               text = for_display)) +
  geom_bar(position = "stack", stat = "identity", show.legend = F) +
  facet_grid(K~.) +
  theme_bw() +
  theme(axis.title = element_blank(),
                    axis.text.x = element_text(angle = 60, hjust = 1),
                    legend.title = element_blank())

# display the Q-matrix
#my.colors <- c("tomato", "lightblue",
              "olivedrab", "gold")
#barchart(project.snmf, K = 4, run = best,
         border = NA, space = 0, col = my.colors,
         xlab = "Individuals", ylab = "Ancestry proportions",
         main = "Ancestry matrix") -> bp
#axis(1, at = 1:length(bp$order),
     labels = bp$order, las = 3, cex.axis = .4)
```


Chapter 3:

Genome content of the *Cercospora* genus predicts host range

Chapter 3: Genome content of the *Cercospora* genus predicts host range

Lizel Potgieter[1,2], Laura C. Ruppert[2], Alice Feurtey[1] , Eva H. Stukenbrock[1,2]

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

[2]Environmental Genomics, Christian-Albrechts University of Kiel, Kiel, Germany

Abstract

Members of the genus *Cercospora* are ascomycete fungi that cause various leaf spot diseases on plants. The number of fully sequenced and assembled genomes of these species has increased over the past years. The availability of these genome resources allows for the application of comparative analyses to investigate the underlying basis of virulence, for example, by the identification of gene gains and losses between species. In other fungal species, as well as among species of *Cercospora*, gene gains and losses have been correlated with host range and virulence. Many ascomycetes have been shown to have plastic genomes with regions undergoing structural mutations whereby, for example, regions of chromosomes can translocate to other chromosomes. We included five *Cercospora* species for an in-depth comparison of genome composition and structure. We show that there are substantial gene gain and loss events in the genomes of the *Cercospora* species, and propose that these may reflect distinct host ranges. Finally, we show that the genomes of two closely related *Cercospora* species, *C. beticola* and *C. cf. flagellaris*, are highly syntenic while the genome of their most recent common ancestor may have undergone large translocations. We conclude that *Cercospora* species follow evolutionary trajectories of other ascomycete plant pathogenic fungi with a high extent of genome plasticity.

Introduction

Fungal plant pathogens are threats to global food safety and security (Strange and Scott, 2005). Some of the most destructive pathogens threatening global food production are fungi, including species belonging to the *Cercospora* genus (Fones et al., 2020; Groenewald et al., 2013). Plant pathogenic *Cercospora* species as these fungi infect a wide variety of crops including sugar beet, soybean, maize, and wheat (Holtschulte, 2000; Soares et al., 2015). *Cercospora* species are often associated with leaf spot diseases, lesions in flowers, and post-harvest fruit rot (Silva and Pereira, 2008; To-Anun et al., 2011). Some species have a very narrow host range while others have a very broad host range (Groenewald et al., 2013).

Phytopathogenic fungi make use of a variety of carbohydrate active enzymes (CAZymes) and effectors during their infection of plants (e.g. Barrett et al., 2020; Kameshwar and Qin, 2018; Lo Presti et al., 2015). Effectors are small secreted proteins that interact with plants during infection while CAZymes aid in the degradation of the cell walls (Cantarel et al., 2009; Rep, 2005). CAZymes fulfill various biochemical functions, and are grouped according to their predicted function (Garron and Henrissat, 2019). The functions of these enzymes can be classified broadly as glycosyl transferases, glycoside hydrolases, polysaccharide lyases, carbohydrate esterases, and carbohydrate binding molecules (Davies and Henrissat, 1995; The CAZypedia Consortium, 2018). The enzyme families included in the CAZyme classification continues to grow as more genomes become available and annotation tools are improved.

For many fungal pathogen species, the diversity in the CAZyme and effector repertoire has been correlated with their host range and lifestyle (Barrett et al., 2020; Kameshwar and Qin, 2018). In the genus *Colletotrichum*, a decrease in CAZyme content was strongly associated with a decreased host range (Baroncelli et al., 2016). *Rhizoctonia solani* isolates with different host ranges also showed that differences in effector repertoire was strongly linked to host range of isolates, with isolates lacking a particular effector gene having a significantly reduced host range (Anderson et al., 2017). Studies like these have demonstrated the power of *in silico* prediction tools to associate CAZyme and effector content with host range.

Cercospora species infect a variety of plants with some species being generalists while other species are more host-specific (Groenewald et al., 2013). Morphologically, species with wide host ranges have been distinguished from species that have narrow host ranges by differences in conidiophores (Groenewald et al., 2013). Species with wide host ranges have circumspersed loci on thin-walled conidiophores, while the species with narrow host ranges had fewer apical or lateral loci on moderately to thick walled, to thick-walled condiophores. We suggest that species with narrow and broad host ranges can be more accurately distinguished using comparative genomics approaches. In this study we included *C. beticola*, *C. cf. flagellaris*, *C. cf. sigesbeckiae*, *C. sojina,* and *C. zeina* for whole genome comparisons. *C. beticola, C. cf. flagellaris* and *C. cf. sigesbeckiae* are generalists, while *C. sojina* and *C. zeina* are more specific to their agricultural hosts. *C. beticola* is the causal agent of Cercospora Leaf Spot (CLS) on many wild and domesticated plants including *Beta* species, *Acanthus, Amaranthus, Apium, Atriplex, Chenopodium, Chrysanthemum, Cycloloma, Goniolimon, Limonium, Malva, and Plantago* (Bobev et al., 2009; Chupp, 1953; Groenewald et al., 2006; Jacobsen and Franc, 2009; Pool and McKay, 1916; Rooney-Latham et al., 2010; Vestal, 1933). It has been proposed that *C. beticola* colonises many of these plants as a saprobe or a secondary invader (Crous and Groenewald, 2005) . These wild plants may serve as a reservoir of inoculum for infections of sugar beet. The ability of *C. beticola* to colonise weeds is likely indicative of the genomic variation of the fungus. *C. cf. flagellaris* causes CLS on melons, soybean, *Amaranthus, Cosmos sulphureus, Hydrangea, Phytolacca americana, Citrus* sp., *Populus deltoides, Bromus* sp., *Eichhornia crassipes*, and many others (Albu et al., 2016; Groenewald et al., 2013; Park et al., 2020). *C. cf. sigesbeckiae* causes disease on *Begonia* sp., *Sigesbeckia glabescens, Dioscorea tokoro, Persicaria orientalis, Pilea pumila, Paulownia coreana, Malva verticillata,* soybean as well as several other plants (Albu et al., 2016; Groenewald et al., 2013). *C. sojina* is more of a specialist, and primarily grows on soybean and causes Frogeye Leaf Spot (FLS) (Soares et al., 2015; Zeng et al., 2017a). *C. zeina* primarily grows on maize where it causes grey leaf spot (GLS) (Tehon and Daniels, 1925).

Phylogenetic studies have provided insights into the diversity and phylogenetic relationships of *Cercospora* species. The monophyletic *Cercospora* genus contains 659 different species (Crous and Braun, 2003). However, the phylogenetic relationships of *Cercospora* species is difficult to establish as housekeeping “marker” genes contain very few informative sites (Groenewald et al., 2013). The use of whole genome data provides an opportunity to construct more accurate phylogenetic relationships (Yu and Reva, 2018). There are two primary methodologies used to construct phylogenies using whole genome data, namely alignment-based and alignment-free methods (reviewed by Zielezinski et al., 2017). In this study, we made use of alignment-based methods, although we found alignment-free methods to be as accurate (unpublished data). Alignment-based methods can either be performed on whole genome data, or by computing single trees for all orthologous genes, and combining these into a single species tree (de Queiroz and Gatesy, 2007). A sound phylogeny is crucial for downstream comparative genomics approaches.

To date, there has not been a comparative genomics study on *Cercospora* species to establish the diversity in effector and CAZyme content. It would be informative to compare these repertoires between species to draw an association with host range. A species tree computed from whole genome data has also not been compiled for the *Cercospora* genus, and the possibility of strengthening existing phylogenies is important. Additionally, synteny between closely related *Cercospora* species has not been investigated. Therefore, this study aims to establish a species tree using whole genome data, and to track CAZyme and effector gains and losses in the *Cercospora* genus. Genome synteny of two closely related *Cercospora* species will also be compared to investigate genome plasticity. This will allow us to correlate genome content with host range of various *Cercospora* species. It will highlight the necessity of well-assembled genome for comparative genomics studies.

Methods and Materials

Description of *Cercospora* Genomes

To reconstruct the phylogenetic relationship of species in the *Cercospora* genus, we established a genome data set based on available genome data. We included already published genomes and genomes available through collaborations (Table 1, Supplementary Table 1). We used a more complete genome assembly of *C. zeina* based on PacBio sequencing, collaborating with D. Berger, University of Pretoria, South Africa. To this end we first generated high quality DNA for sequencing. The *C. zeina* isolate was grown on tomato juice agar plates, and DNA was extracted following the CTAB protocol (Clarke, 2009). Sporulating conidia were washed off the plates, and ground with a mortar and pestle with liquid nitrogen. The resulting powder was transferred to 2 ml Eppendorf tubes and 1.2 ml of warm CTAB extraction buffer was added. Samples were incubated in a 65°C water bath for an hour, and shaken gently every 15 minutes. Samples were spun in a centrifuge for 10 minutes at 13500 g, and the supernatant was transferred. The supernatant was washed twice with a 1:1 phenol/chloroform mixture, and spun at 13500 g for 15 minutes between washes. The aqueous phase was transferred to 800 µl ice cold isopropanol after the second wash and incubated at -20°C for 20 minutes. Samples were spun at 13500 g for 10 minutes, and the pellet was suspended in 250 µl TE buffer and 25 µl RNase and incubated at 37°C for 30 minutes. After spinning samples in a centrifuge for 10 minutes, the pellet was suspended in 25 µl NaOAc and 600 µl ice cold 100% ethanol. Following an overnight incubation step, the pellet was washed with 70% ethanol twice, and suspended in 100 µl TE buffer. DNA concentration was determined by fluorometry, and quality was confirmed by gel electrophoresis. This was repeated until 7 ng of *C. zeina* DNA was extracted. The DNA was sequenced at the Max Planck Genome Center in Cologne using the Pacific Biosciences platform.

Raw reads were assembled with HGAP 4, a part of SMRTLink version 6, with default parameters and an estimated genome size of 40 Mb (Chin et al., 2013). The draft assembly was polished using arrow, another function of SMRTLink. Contigs that had sequencing coverage that deviated more

than 1.5 X from the average across all contigs weighted by contig length were discarded (Filter_cover_PacBio_assemblies_SMRTLink6.R). To identify the CCCTAA telomeric repeat, bowtie was used (Langmead et al., 2009) (Identify_telomeres.sh). All scripts are available in the supplementary materials online (https://github.com/lpotgieter/phd.scripts).

Table 1: Genomes of *Cercospora* species available on public databases as of September 2019

Species	**Source and Accession**	**Reference**
C. beticola	NCBI (LKMD00000000)	(de Jonge et al., 2018)
C. cf. flagellaris	NCBI (RJLU00000000)	Unpublished
C. cf. sigesbeckiae	NCBI (RQIF00000000)	(Albu et al., 2017)
C. sojina	NCBI (GCA_004299825)	(Luo et al., 2018)
C. zeina	Unpublished	This study

Analysis of CAZymes and Effectors

For a detailed comparison of gene content, we focused our analyses on the five genomes with high quality assemblies: *C. beticola*, *C. cf. flagellaris, C. sojina*, *C. cf. sigesbeckiae*, and *C. zeina* (Albu et al., 2017; de Jonge et al., 2018; Luo et al., 2018). Contigs smaller than 1 Mb were filtered out using a custom script (removesmalls.pl) .

To standardise the annotation of the genomes, an *ab initio* gene prediction approach was applied. All assemblies were annotated with WebAUGUSTUS v3.3.3 with default paramters as above (Stanke and Morgenstern, 2005). The predicted proteome of *Cercospora* species was used to predict the secretome, effectors, and CAZymes of the isolates. Proteins containing signal peptides were determined with SignalP 5.0 (Almagro Armenteros et al., 2019). Proteins containing a signal peptide region were further established to be effectors with EffectorP 2.0 (Sperschneider et al., 2018). CAZymes were identified by the metaserver dbCAN v2.0.0 (Zhang et al., 2018). CAZymes identified by HMMER v3.3, DIAMOND, and Hotpep (Buchfink et al., 2015; Busk, 2020; Finn et al., 2011). CAZymes identified by at least two of the search tools were used for further analyses. The heatmap of CAZyme family frequency was generated with a custom script (Heatmap.R).

Gene Family Expansions and Contractions in *Cercospora*

To reconstruct the history of gene gains and losses, we constructed a phylogenetic tree. To this end we first identified all single copy orthologous genes were determined using OrthoMCL v2.0.9-4 with default parameters (Li et al., 2003). The corresponding protein sequences were aligned with MUSCLE v3.8.1551 (Edgar, 2004). A maximum likelihood phylogenetic tree was constructed for each alignment with IQ-Tree with 1000 bootstrap replicates and automatic selection of the best fit model (Nguyen et al., 2015). A species tree of all single trees that corresponded with the largest number of quartet trees was constructed with ASTRAL v5.7.3 (Mirarab et al., 2014). The species tree was rooted by mid-point rooting in the programme FigTree v1.4.4 (Rambaut, 2012). The topology was converted to an ultrametric tree with a custom script (Ultrametric_tree.R).

To determine whether any CAZyme or effector families had undergone expansions or contractions within the genus, their history was plotted against the species tree. The predicted effectors and CAZymes were extracted and compared to the species tree with CAFE v4.2.1 (De Bie et al., 2006). Gene family contractions and expansions are tracked along the species tree using a stochastic birth and death process.

Whole Genome Synteny Between *C. beticola* and *C. cf. flagellaris*

Based on the species tree, it was shown that *C. beticola* and *C. flagellaris* were most closely related, and were used to investigate differences in overall genome organisation. The assemblies were aligned with PROmer, a part MUMmer v4.0.0 (Delcher et al., 2002). The alignment was visualised with Circos v0.69-9 (Krzywinski et al., 2009).

Results

PacBio Sequencing of *C. zeina*

The long read assembly of *C. zeina* was more complete than the *C. zeina* assembly produced by short read sequencing. The PacBio filtered assembly contained 22 contigs, 21 of which were larger than 10 kb (Table 2). The assembly was 41.7 Mbp long, and had an N50 of 3.97 Mbp. Three contigs were assembled telomere-to-telomere, five contained telomeres at the start of the contig, and seven contained telomeres only at the end of the contig. The remaining seven contigs did not contain any telomeric regions.

Table 2: Assembly statistics of *C. zeina* after filtering the long read data compared to the previous short read assembly

	***C. zeina* (PacBio) Polished Assembly**	***C. zeina* (PacBio) Filtered Assembly**	***C. zeina* (Illumina HiSeq)**
Total sequence length	41.8 Mbp	41.7 Mbp	40.8 Mbp
Number of scaffolds	22	17	10,027
N50	3.98 Mbp	3.98 Mbp	161 Kbp
L50	5	5	61

Analysis of CAZymes and Effectors

The five *Cercospora* species showed substantial variation in the number of protein-coding genes as well as genome size (Table 3). Genome lengths ranged from 33 to 41 Mb. The *C. beticola* contained over 11000 protein-coding genes, while the genome of *C. cf. sigesbeckiae* contained less than 6500. This may be due to assembly quality, or the species merely contains fewer genes.

Table 3: Metrics of included *Cercospora* genome assemblies and annotations

Species	***C. beticola***	***C. cf. flagellaris***	***C. cf. sigesbeckiae***	***C. sojina***	***C. zeina***
Assembly	GCA_002742065.1	GCA_005356885.1	GCA_005356805.1	GCA_002534735.1	Czeina_filtered
Accession number	PRJNA270309	PRJNA503907	PRJNA503907	PRJNA371568	
Origin	USA: North Dakota	USA: Arkansas	USA: Louisiana	China: Heilongjiang	South Africa
Host isolated from	*Beta vulgaris*	*Glycine max*	*Glycine max*	*Glycine max*	*Zea maydis*
Year (of isolation)	2009	2011	2012	2010	
Contig number	252	57	335	62	17
Total length (bp)	37057033	33240740	33720866	40836407	41717156
Largest contig (bp)	6188355	4579182	2125636	6706376	5193669
N50	4173231	2848694	1308006	1594415	3977222
L50	4	5	10	6	5
Number of genes	11339	9983	6419	9967	9985

CAZyme families varied in their frequency among the different *Cercospora* species (Figure 1). Overall, *C. cf. sigesbeckiae* contained the lowest number of CAZymes, and *C. beticola* the highest numbers within each family. CAZymes with auxillary activities were present with the highest frequency relative to other CAZyme families within the genome.

To reconstruct the gain and loss events related to genes encoding CAZymes and effectors in *Cercospora*, a phylogenetic tree was constructed (Figure 2A). The tree was constructed with 2781 single-copy orthologous genes from the *Cercospora* genus. There was a variable number of CAZymes represented by each species (Figure 2B, Supplementary Table 2). The *C. cf. sigesbeckiae* genome was the smallest, contained the lowest number of predicted genes, as well

as CAZymes and effectors. The *C. beticola* genome contained the most genes, and genes of each category. The total number of genes predicted for each of the remaining three species was similar, but *C. sojina* and *C. zeina* had comparable secretome sizes, as well as numbers of effectors and CAZymes. These two species also form their own clade. *C. cf. flagellaris* had a few number of predicted genes than *C. beticola*, as well as smaller secretome. However, these two species contained similar numbers of effectors and CAZymes.

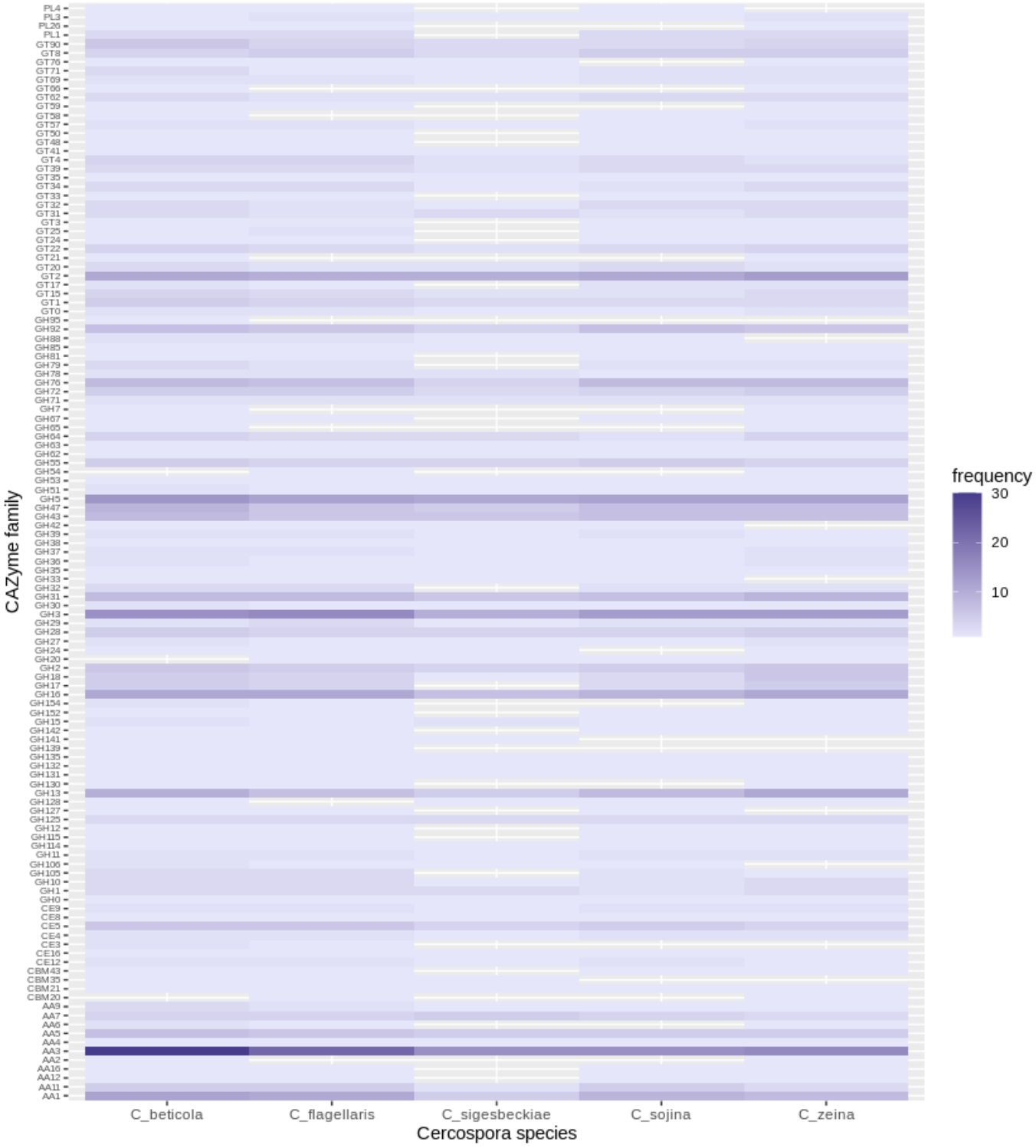


Figure 1: Heatmap of CAZyme family frequencies in the five *Cercospora* species. The intensity of the colour corresponds to the frequency of the family with lighter families having a lower frequency than those indicated by darker colours. CAZyme functions are annotated as polysaccharide lyases (PL), glycosyl transferases (GL), glycoside hydrolases (GL), carbohydrate esterases (CE), and auxiliary activities (AA).

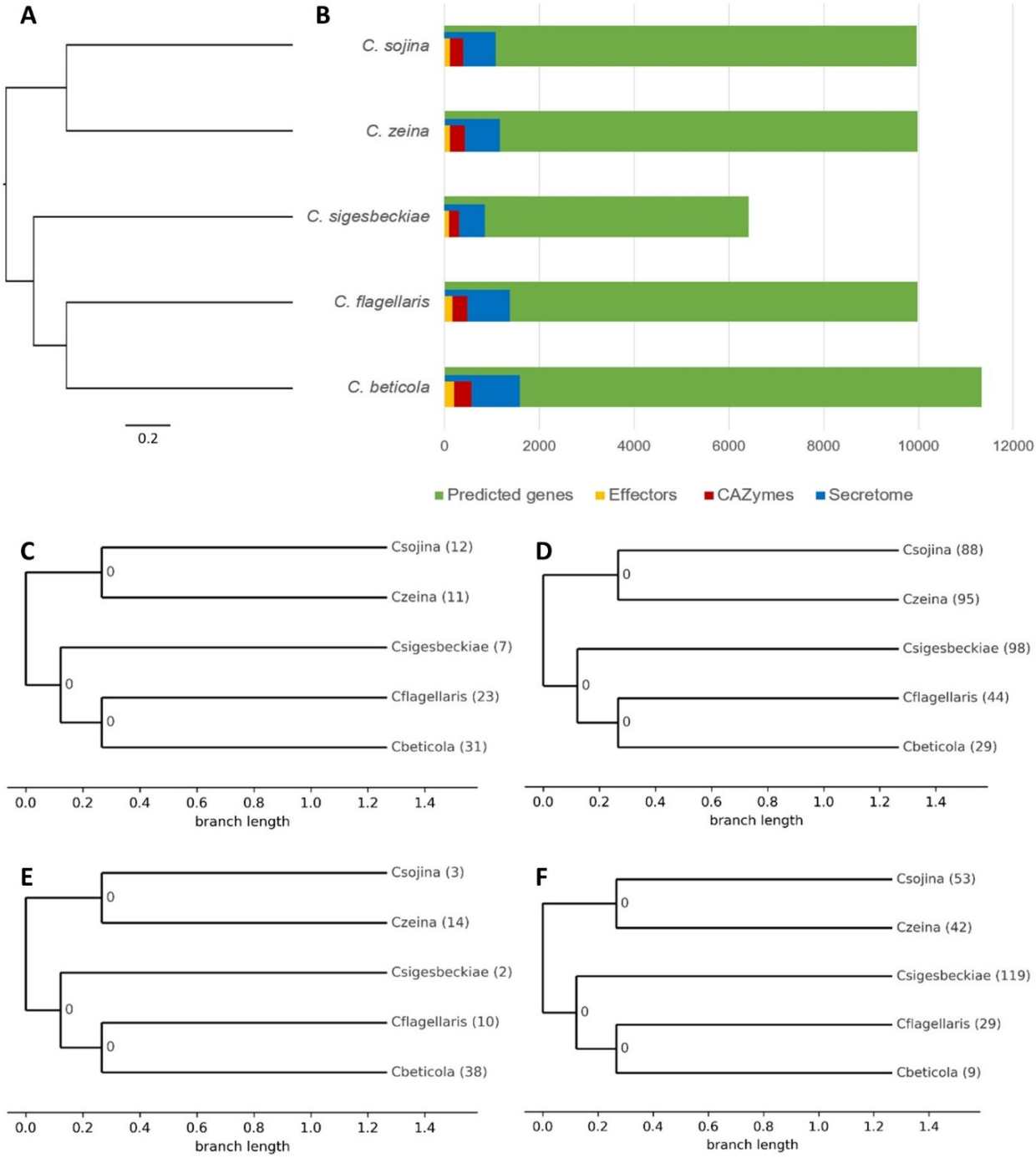


Figure 2: Genome characteristics the *Cercospora* genus. **A**: The ML phylogeny based on single-copy orthologous genes, **B**: The number of predicted protein-coding genes (green), secretome (blue), CAZymes (red), and effectors (yellow) in each *Cercospora* species, **C**: Effector expansions in each *Cercospora* species, **D:** Effector contractions in each *Cercospora*

species, **E**: CAZyme expansions in each *Cercospora* species, **F**: CAZyme contractions in each *Cercospora* species. Figures in brackets in **C-F** indicate gains or losses, respectively.

Within the *Cercospora* genus, several genes encoding effectors and CAZymes have undergone expansions and contractions (Figure 2C-F). The gene content in *C. beticola* has undergone the most expansions and the fewest contractions compared to the other four species. Although *C. beticola* and *C. cf. flagellaris* contained similar numbers of CAZymes and effectors, *C. cf. flagellaris* has undergone a net total loss of effectors and CAZymes whereas *C. beticola* has a net gain in each classification. The genome of *C. beticola* gained 31 effectors and 38 CAZymes, while losing 29 effectors and only 9 CAZymes. While *C. cf. flagellaris* is also a generalist *Cercospora* species, the gain of CAZymes and effectors occurred at lower frequencies, and losses of these genes occurred more often. *C. cf. flagellaria* gained 23 effectors and 10 CAZymes, and lost 44 effectors and 29 CAZymes. *C. cf. sigesbeckiae* showed large contractions in both effectors and CAZymes. *C. sojina* and *C. zeina* also showed nett contractions in both CAZyme and effector repertoire. Interestingly, our analysis showed all gains and losses occurred after speciation.

Whole Genome Synteny Between *C. beticola* and *C. cf. flagellaris*

Genome synteny analyses can detect translocations that are common in ascomyctes. Substantial synteny was observed between the *C. beticola* and *C. cf. flagellaris* genomes (Figure 3). For the majority of the chromosomes, the whole chromosome was present in both species without large translocations. Effectors and CAZymes were found evenly spaced across the genomes. Unique effectors weren't strongly associated with the ends of chromosomes, but the unique CAZymes were. The two largest chromosomes in each species showed different patterns of synteny compared to the chromosomes of smaller size. Chromosome 1 in *C. beticola* was larger than the corresponding chromosome in *C. cf. flagellaris*, with parts of the chromosome showing synteny to *C. cf. flagellaris* chromosome 7, and a region unique to *C. beticola*. Within this unique region, there were also CAZymes unique to *C. beticola* as well as a unique effector. Chromosome 2 of *C. beticola* was syntenic with a large region of *C. cf. flagellaris* chromosome 7, and the entirety of *C. cf. flagellaris* chromosome 9. *C. cf. flagellaris* chromosome 2 was syntenic to *C. beticola*

chromosomes 2 and 10, and in both cases, was the only major syntenic region. This is may show that during the speciation of *C. beticola* and *C. cf. flagellaris*, the ancestral chromosome 2 could have undergone chromosome fission, forming the two differentiated chromosomes in *C. beticola*. The same may have occurred in the ancestral chromosomes 1 and 2 that are present in *C. beticola* to produce in the separate chromosomes 7 and 9 in *C. cf. flagellaris.*

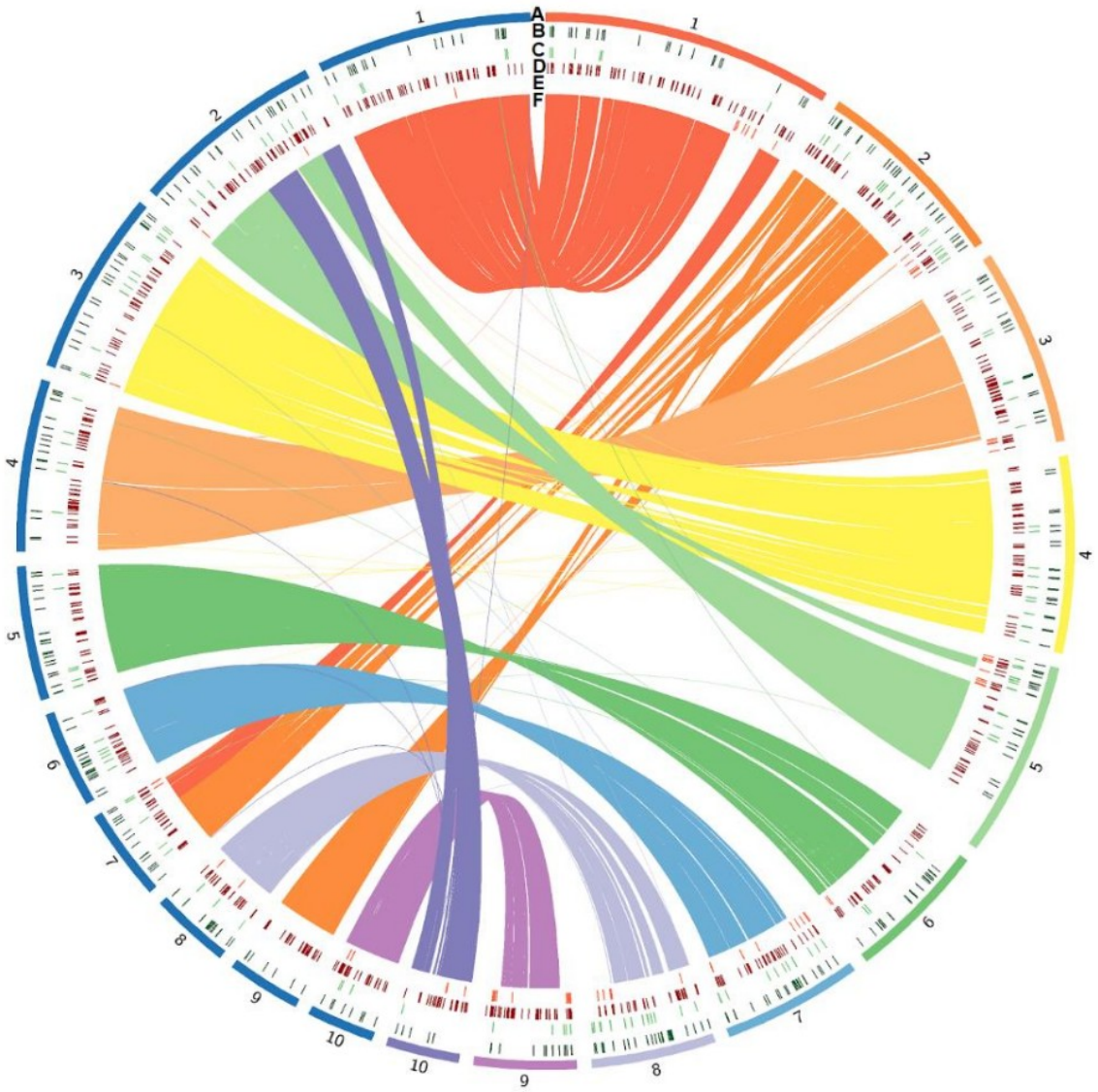


Figure 3: Whole genome comparison of *C. beticola* and *C. cf. flagellaris*. **A**: the 10 *C. cf. flagellaris* chromosomes in dark blue on the left, and the 10 *C. beticola* chromosomes in rainbow colours on the right, **B**: Location of effectors, **C**: Effector singletons **D**: Location of CAZymes, **E**: CAZyme singletons, **F**: Regions blocks of synteny between *C. beticola* and *C. cf. flagellaris*

Discussion

Fungal genomes contain a lot of variation that results in phenotypic changes such as host range shifts or expansions. This phenotypic diversity is mediated by genomic changes. Here we considered structural variation in the form of gene gains and losses, and how these may affect host range in *Cercospora*.

In this study we analysed the genomes of five *Cercospora* species to compare and quantify the extent of structural variants. We document that there have been substantial gains and losses of effectors and CAZymes when comparing species the two species that are specialists to the two species that are generalists. We have shown contractions of genes encoding CAZymes and effectors in specialist species. We have also shown that synteny has been maintained between two of the generalist species we considered.

Comparative genomics relies on well assembled genomes. We included *C. cf sigesbeckiae* for the species tree as the assembly statistics suggested that the genome was assembled to a satisfactory threshold. The reduced number of predicted genes in *C. cf. sigesbeckiae* may be due to the assembly indeed not being as for good as the other four species. Currently, the smallest fungal genome is *Peltaster fructicola* (Wang et al., 2020). The genome of *P. fructicola* is only 18.99 Mb long, and is predicted to contain around 8000 genes. As such, the remaining discussion will consider only the two generalists, *C. beticola* and *C. cf. flagellaris,* and the two more specialised species, *C. sojina* and *C. zeina* due to the lack of confidence in the *C. cf sigesbeckiae* predictions. Comparative genomics studies between *C. sojina* isolates, and other Ascomycetes have been performed and our results will be incorporated with these studies (Gu et al., 2020; Luo et al., 2018).

Gene gains and losses are often attributed to the lifestyle of the fungus. *Blumeria graminis* is an obligate biotrophic plant pathogen that shows much retrotransposon accumulation, genome expansion, and gene loss (Spanu et al., 2010). Specific gene losses are strongly correlated with a biotrophic lifestyle. This is likely due to the redundancy of metabolic pathways, transporters, and CAZymes as the growth on plants is not always required, with the exception of obligate biotrophs

(Spanu, 2012). Additionally, co-evolution with the host species and host jumps further drive genome evolution (Menardo et al., 2017). In *Colletotrichum* spp. a reduction in CAZyme and protease repertoire was linked to a reduced host range (Baroncelli et al., 2016).

In studies where different *Cercospora* species with differences in virulence on soybean, comparative genomics showed that gene gains and losses likely influenced the phenotype (Gu et al., 2020). The comparison of the *C. sojina* genomes of two strains with different virulence phenotypes showed a difference of 245 specific genes, with five candidates related to the regulation of host resistance and self-toxicity. Additionally, different *C. sojina* races with differences in virulence also have different CAZyme profiles (Luo et al., 2018). Less virulent isolates had fewer carbohydrate binding molecule (CBM) genes, a class of CAZymes. CBM enzymes anchor the enzymes to cellulose of the plant cell wall, allowing the fungus to potentially improve the efficacy degradation of the plant cell wall (BORASTON et al., 2004). It has been proposed that a reduced set of CBMs may slow the infection process, resulting in a less virulent pathogen (Wang et al., 2011). In this study we also showed that there is variation in the CBM profiles of *Cercospora* species, with differences in abundance in the generalist and the specialist species. The specialist species have a reduced number of CBMs compared to the generalist species, which may indicate that a slower infection strategy may not be detrimental to the infection and reproduction of the fungus.

There was substantial variation in the frequency of CAZyme families between the generalist and the specialist species. Plant cell walls of monocotyledonous and dicotyledonous plants have different components, and differ in their carbohydrate composition (King et al., 2011; Lagaert et al., 2009; Vogel, 2008). Most importantly, monocotyledonous plant cell walls contain less pectin than the cell walls of dicotyledonous plants (Vogel, 2008). Therefore, pathogens of strictly dicotyledonous plants, often contain, among others, more pectinases (Zhao et al., 2013). Within the dicotyledonous and monocotyledonous plants, there is also substantial variation between the composition of carbohydrates that make up plant cell walls, and these different components may drive the diversification of CAZymes (Blanco-Ulate et al., 2014).

The comparison between the generalist and specialist species showed differences in CAZyme family composition. We found that an “auxiliary action” CAZyme family, AA3, was the most common in the two generalist species, and occurred at a lower frequency in the two specialist species. The AA3 family catalyses the oxidation of alcohols or carbohydrates (Levasseur et al., 2013). These enzymes fulfill a variety of functions. For instance, in insects, AA3 enzymes are associated with immunity and development, while in yeast they are responsible for alcohol oxidation that releases hydrogen peroxide (Goswami et al., 2013; Iida et al., 2007; Sun et al., 2012). AA3 and AA9, a lytic monooxygenase, are co-regulated and co-secreted (Miyauchi et al., 2017). Interestingly, the AA9 family did not show the same copy number variation as the AA3 family. The AA3 family can be further subdivided into four separate groups, however, we did not investigate which family contributed to the copy number variation (The CAZypedia Consortium, 2018).

In addition to CAZymes, we also considered the diversity of effectors among the *Cercospora* species. Effectors are small secreted proteins, with a signal peptide at the N-terminal of the enzyme that is cleaved upon secretion (Lo Presti et al., 2015). Highly dynamic repertoires of effectors has been proposed to be strongly associated with dynamic host ranges of phytopathogens (e.g. Guyon et al., 2014; Schulze-Lefert and Panstruga, 2011). Durable plant resistance to phytopathogens can be mediated through effector recognition, indicating the necessity of these proteins during infection (Depotter and Doehlemann, 2020). We have shown that in the specialist species, there have been effectors lost than in the more generalist *Cercospora* species. The total number of effectors in the generalist species is also higher compared to the the specialist species of *Cercospora*. A dynamic content of genes encoding effectors and CAZymes may drive the differences in host range. Future studies should investigate whether the effectors that have been lost are orthologous, or whether the different species lost unique effectors. The pangenome of each species can be considered to this end. Pangenomic studies have shown incredible variation of genes represented by a species, with some pangenome

of some species containing up to 6600 genes that were specific to a subset of isolates (Plissonneau et al., 2018).

Transciption assays of *C. sojina* showed an enrichment of GH109 in comparison to other fungi (Luo et al., 2018). While not the cause of the difference in virulence between *C. sojina* isolates, these enzymes are important in lectin-mediated resistance in soybean (Gu et al., 2020). In our strict filtering approach, we have not detected a representative of the GH109 family in any of the *Cercospora* species. While *in silico* studies have great potential to explore the genomic potential of a species or a genus, caution should be exercised. We have applied very strict filters to ensure that predictions made are reliable, and this may have resulted in some genes being filtered out of the analyses. Additionally, it is also possible that the single individuals considered in this comparative study have lost these genes during random processes as *Cercospora* species readily undergo gene gains and losses. Genome analyses should be accompanied with experimental studies to validate the predictions of functional relevance.

Within the Ascomycota, mesosynteny is observed between distantly related as there are many inversions within the genomes, while translocations are more rare (Hane et al., 2011). The *Verticillium* genus has signatures many of genomic rearrangements between various *V. dahliae* strains (Faino et al., 2015, 2016; Jonge et al., 2012). Further, large-scale genomic rearrangements are common among different *Verticillium* species based on ancestral genomic reconstruction with varying frequencies in pathogenic and non-pathogenic species (Shi-Kunne et al., 2018). Synteny within the *Mycosphaerella* genus showed a high level of synteny between *C. sojina* and *Zymoseptoria tritici*, but less between *C. sojina* and the remaining *Mycosphaella* species considered (Zeng et al., 2017b). The comparison of the *C. beticola* and *C. cf. flagerllaris* showed that within the *Cercospora* genus these events may have occurred as well. The synteny between the smaller chromosomes of one species and the larger chromosomes of the others may be indicative of this. It is known that the *C. beticola* genome is plastic and that there are chromosomal rearrangements between isolates (Weiland and Koch, 2004). Ancestral reconstructions of these genomes like those done in *Verticillium* would be necessary to confirm the chromosomal

composition of the most recent common ancestor of *C. beticola* and *C. cf. flagellaris* and may also give more documentation to gene gain and loss processes.

Conclusion

This comparative genomics study has shown that the genomes of the genus hold the potential to answer many different questions relating to host range and genome evolution. Here we showed that species of *Cercospora* have a conserved *CTB* cluster. We showed the function of using whole genome data to construct species trees. We also showed that there was a difference in the effector and CAZyme repertoire of different *Cercospora* species. We showed that the generalist *Cercospora* species lost fewer CAZymes and effectors than the specialist species. We showed that several CAZyme families that may be important during infection are more frequent in the generalist *Cercospora* species. Finally, we showed that two very closely related *Cercospora* species, *C. beticola* and *C. cf. flagellaris* shared substantial synteny, and that the genome of their most recent common ancestor likely had a plastic genome where two chromosomes split to form the 10 chromosomes seen in the species at present.

References

Albu, S., Schneider, R.W., Price, P.P., and Doyle, V.P. (2016). *Cercospora cf. flagellaris* and *Cercospora cf. sigesbeckiae* Are Associated with Cercospora Leaf Blight and Purple Seed Stain on Soybean in North America. Phytopathology® *106*, 1376–1385.

Albu, S., Sharma, S., Bluhm, B.H., Price, P.P., Schneider, R.W., and Doyle, V.P. (2017). Draft Genome Sequence of *Cercospora cf. sigesbeckiae*, a Causal Agent of Cercospora Leaf Blight on Soybean. Genome Announc. *5*.

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

Anderson, J.P., Sperschneider, J., Win, J., Kidd, B., Yoshida, K., Hane, J., Saunders, D.G.O., and Singh, K.B. (2017). Comparative secretome analysis of *Rhizoctonia solani* isolates with different host ranges reveals unique secretomes and cell death inducing effectors. Sci. Rep. *7*, 10410.

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

Barrett, K., Jensen, K., Meyer, A.S., Frisvad, J.C., and Lange, L. (2020). Fungal secretome profile categorization of CAZymes by function and family corresponds to fungal phylogeny and taxonomy: Example *Aspergillus* and *Penicillium*. Sci. Rep. *10*, 5158.

Blanco-Ulate, B., Morales-Cruz, A., Amrine, K.C.H., Labavitch, J.M., Powell, A.L.T., and Cantu, D. (2014). Genome-wide transcriptional profiling of Botrytis cinerea genes targeting plant cell walls during infections of different hosts. Front. Plant Sci. *5*.

Bobev, S.G., Groenewald, J.Z., and Crous, P.W. (2009). First Report of *Cercospora beticola* as a Pathogen of German Statice (*Goniolimon tataricum*) in Bulgaria. Plant Dis. *93*, 553–553.

BORASTON, A.B., BOLAM, D.N., GILBERT, H.J., and DAVIES, G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. *382*, 769–781.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

Busk, P.K. (2020). Accurate, automatic annotation of peptidases with hotpep-protease. Green Chem. Eng.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. *37*, D233–D238.

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

Chupp, C.C. (1953). A monograph of the fungus genus *Cercospora*. Monogr. Fungus Genus Cercospora.

Clarke, J.D. (2009). Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation. Cold Spring Harb. Protoc. *2009*, pdb.prot5177.

Crous, P.W., and Braun, U. (2003). Mycosphaerella and its anamorphs: Names published in *Cercospora* and *Passalora. Mycosphaerella* Its Anamorphs Names Publ. *Cercospora Passalora*.

Crous, P.W., and Groenewald, J.Z. (2005). Hosts, species and genotypes: opinions versus data. Australas. Plant Pathol. *34*, 463–470.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853–859.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271.

Delcher, A.L., Phillippy, A., Carlton, J., and Salzberg, S.L. (2002). Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. *30*, 2478–2483.

Depotter, J.R.L., and Doehlemann, G. (2020). Target the core: durable plant resistance against filamentous plant pathogens through effector recognition. Pest Manag. Sci. *76*, 426–431.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Faino, L., Seidl, M.F., Datema, E., Berg, G.C.M. van den, Janssen, A., Wittenberg, A.H.J., and Thomma, B.P.H.J. (2015). Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. MBio *6*.

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

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.

Fones, H.N., Bebber, D.P., Chaloner, T.M., Kay, W.T., Steinberg, G., and Gurr, S.J. (2020). Threats to global food security from emerging fungal and oomycete crop pathogens. Nat. Food *1*, 332–342.

Garron, M.-L., and Henrissat, B. (2019). The continuing expansion of CAZymes and their families. Curr. Opin. Chem. Biol. *53*, 82–87.

Goswami, P., Chinnadayyala, S.S.R., Chakraborty, M., Kumar, A.K., and Kakoti, A. (2013). An overview on alcohol oxidases and their potential applications. Appl. Microbiol. Biotechnol. *97*, 4259–4275.

Groenewald, J.Z., Nakashima, C., Nishikawa, J., Shin, H.-D., Park, J.-H., Jama, A.N., Groenewald, M., Braun, U., and Crous, P.W. (2013). Species concepts in *Cercospora*: spotting the weeds among the roses. Stud. Mycol. *75*, 115–170.

Groenewald, M., Groenewald, J.Z., Braun, U., and Crous, P.W. (2006). Host range of *Cercospora apii* and *C. beticola* and description of *C. apiicola*, a novel species from celery. Mycologia *98*, 275–285.

Gu, X., Ding, J., Liu, W., Yang, X., Yao, L., Gao, X., Zhang, M., Yang, S., and Wen, J. (2020). Comparative genomics and association analysis identifies virulence genes of *Cercospora sojina* in soybean. BMC Genomics *21*, 172.

Guyon, K., Balagué, C., Roby, D., and Raffaele, S. (2014). Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. BMC Genomics *15*, 336.

Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H., Goodwin, S.B., and Oliver, R.P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biol. *12*, R45.

Holtschulte, B. (2000). Cercospora beticola - worldwide distribution and incidence. *Cercospora Beticola Sacc* Biol. Agron. Influ. Control Meas. Sugar Beet 5–16.

Iida, K., Cox-Foster, D.L., Yang, X., Ko, W.-Y., and Cavener, D.R. (2007). Expansion and evolution of insect GMC oxidoreductases. BMC Evol. Biol. *7*, 75.

Jacobsen, B., and Franc, G.D. (2009). Compendium of Beet Diseases and Pests (APS Press, St. Paul, MN).

de Jonge, R. de, Esse, H.P. van, Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K.V., et al. (2012). Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc. Natl. Acad. Sci. *109*, 5110–5115.

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., et al. (2018). Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus *Colletotrichum*. Proc. Natl. Acad. Sci. 201712798.

Kameshwar, A.K.S., and Qin, W. (2018). Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. Mycology *9*, 93–105.

King, B.C., Waxman, K.D., Nenni, N.V., Walker, L.P., Bergstrom, G.C., and Gibson, D.M. (2011). Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. Biotechnol. Biofuels *4*, 4.

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

Lagaert, S., Beliën, T., and Volckaert, G. (2009). Plant cell walls: Protecting the barrier from degradation by microbial enzymes. Semin. Cell Dev. Biol. *20*, 1064–1073.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol. Biofuels *6*, 41.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. *13*, 2178–2189.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. Annu. Rev. Plant Biol. *66*, 513–545.

Luo, X., Cao, J., Huang, J., Wang, Z., Guo, Z., Chen, Y., Ma, S., and Liu, J. (2018). Genome sequencing and comparative genomics reveal the potential pathogenic mechanism of *Cercospora sojina* Hara on soybean. DNA Res. *25*, 25–37.

Menardo, F., Wicker, T., and Keller, B. (2017). Reconstructing the Evolutionary History of Powdery Mildew Lineages (*Blumeria graminis*) at Different Evolutionary Time Scales with NGS Data. Genome Biol. Evol. *9*, 446–456.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics *30*, i541–i548.

Miyauchi, S., Navarro, D., Grisel, S., Chevret, D., Berrin, J.-G., and Rosso, M.-N. (2017). The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. PLOS ONE *12*, e0175528.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. *32*, 268–274.

Park, M.-J., Back, C.-G., and Park, J.-H. (2020). Occurrence of Cercospora Leaf Spot Caused by *Cercospora cf. flagellaris* on Melon in Korea. Mycobiology *48*, 418–422.

Plissonneau, C., Hartmann, F.E., and Croll, D. (2018). Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. *16*.

Pool, V.W., and McKay, M.B. (1916). Relation of stomatal movement to infection by *Cercospora beticola*. J. Agric. Res. *5*, 1011–1038.

de Queiroz, A., and Gatesy, J. (2007). The supermatrix approach to systematics. Trends Ecol. Evol. *22*, 34–41.

Rambaut, A. (2012). FigTree v1. 4.

Rep, M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. FEMS Microbiol. Lett. *253*, 19–27.

Rooney-Latham, S., Scheck, H.J., and Walber, T.M. (2010). First Report of *Cercospora beticola* Causing a Leaf Spot Disease on *Acanthus mollis* in California. Plant Dis. *95*, 224–224.

Schulze-Lefert, P., and Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends Plant Sci. *16*, 117–125.

Shi-Kunne, X., Faino, L., Berg, G.C.M. van den, Thomma, B.P.H.J., and Seidl, M.F. (2018). Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene loss. Environ. Microbiol. *20*, 1362–1373.

Silva, M., and Pereira, O.L. (2008). Postharvest *Cercospora apii* fruit rot disease on *Cucurbita maxima* (Cucurbitaceae). Australas. Plant Dis. Notes *3*, 21.

Soares, A.P.G., Guillin, E.A., Borges, L.L., da Silva, A.C.T., de Almeida, Á.M.R., Grijalba, P.E., Gottlieb, A.M., Bluhm, B.H., and de Oliveira, L.O. (2015). More *Cercospora* Species Infect Soybeans across the Americas than Meets the Eye. PLoS ONE *10*.

Spanu, P.D. (2012). The Genomics of Obligate (and Nonobligate) Biotrophs. Annu. Rev. Phytopathol. *50*, 91–109.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., Themaat, E.V.L. van, Brown, J.K.M., Butcher, S.A., Gurr, S.J., et al. (2010). Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. Science *330*, 1543–1546.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B., and Taylor, J.M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. *19*, 2094–2110.

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. *33*, W465–W467.

Strange, R.N., and Scott, P.R. (2005). Plant disease: a threat to global food security. Annu. Rev. Phytopathol. *43*, 83–116.

Sun, W., Shen, Y.-H., Yang, W.-J., Cao, Y.-F., Xiang, Z.-H., and Zhang, Z. (2012). Expansion of the silkworm GMC oxidoreductase genes is associated with immunity. Insect Biochem. Mol. Biol. *42*, 935–945.

Tehon, L.R., and Daniels, E. (1925). Notes on the Parasitic Fungi of Illinois—II. Mycologia *17*, 240–249.

The CAZypedia Consortium (2018). Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. Glycobiology *28*, 3–8.

To-Anun, C., Hidayat, I., and Meeboon, J. (2011). Genus Cercospora in Thailand: taxonomy and phylogeny (with a dichotomous key to species). Plant Pathol. Quar. *1*, 11–87.

Vestal, E.F. (1933). Pathogenicity, host response and control of Cercospora leaf spot of sugar beet. Agr. Exptl Sta. Iowa State College Of Agri. Mechanics Arts. Res. Bull *168*.

Vogel, J. (2008). Unique aspects of the grass cell wall. Curr. Opin. Plant Biol. *11*, 301–307.

Wang, B., Liang, X., Gleason, M.L., Hsiang, T., Zhang, R., and Sun, G. (2020). A chromosome-scale assembly of the smallest Dothideomycete genome reveals a unique genome compaction mechanism in filamentous fungi. BMC Genomics *21*, 321.

Wang, Y., Wu, J., Park, Z.Y., Kim, S.G., Rakwal, R., Agrawal, G.K., Kim, S.T., and Kang, K.Y. (2011). Comparative Secretome Investigation of *Magnaporthe oryzae* Proteins Responsive to Nitrogen Starvation. J. Proteome Res. *10*, 3136–3148.

Weiland, J., and Koch, G. (2004). Sugarbeet leaf spot disease (*Cercospora beticola Sacc.*)†. Mol. Plant Pathol. *5*, 157–166.

Yu, X., and Reva, O.N. (2018). SWPhylo – A Novel Tool for Phylogenomic Inferences by Comparison of Oligonucleotide Patterns and Integration of Genome-Based and Gene-Based Phylogenetic Trees. Evol. Bioinforma. Online *14*.

Zeng, F., Wang, C., Zhang, G., Wei, J., Bradley, C.A., and Ming, R. (2017a). Draft genome sequence of *Cercospora sojina* isolate S9, a fungus causing frogeye leaf spot (FLS) disease of soybean. Genomics Data *12*, 79–80.

Zeng, F., Lian, X., Zhang, G., Yu, X., Bradley, C.A., and Ming, R. (2017b). A comparative genome analysis of *Cercospora sojina* with other members of the pathogen genus *Mycosphaerella* on different plant hosts. Genomics Data *13*, 54–63.

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

Zhao, Z., Liu, H., Wang, C., and Xu, J.-R. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics *14*, 274.

Zielezinski, A., Vinga, S., Almeida, J., and Karlowski, W.M. (2017). Alignment-free sequence comparison: benefits, applications, and tools. Genome Biol. *18*, 186.

Supplementary Materials

Supplementary Tables

Supplementary Table 1: Assembly statistics for the *Cercospora* species included from NCBI

	C. beticola	***C. cf. flagellaris***	***C. cf. sigesbeckiae***	***C. sojina***
Total sequence length (Mb)	36.55	33.24	34.33	31.11
Scaffolds	252	57	335	12
N50	173 kb	2.9 Mb	1.3 Mb	Full assembly
L50	62	5	10	Full assembly

Supplementary Table 2: Number of predicted genes, secretome effectors, CAZymes, and CAZyme families of *Cercospora* species

	C. beticola	***C. cf. flagellaris***	***C. cf. sigesbeckiae***	***C. sojina***	***C. zeina***
Predicted genes	11339	9983	6419	9967	9985
Secretome	1016	891	550	680	735
Effectors	204	169	99	117	118
CAZymes	371	319	206	281	310
Number of CAZyme families	120	115	84	104	104

Scripts

Filter_cover_PacBio_assemblies_SMRTLink6.R

```
library(ggplot2)
library(tidyverse)
library(seqinr)


dir_name = "your_directory"
file_names=c("your_polished_assembly")
polishing_algo="arrow"

suffix_cov="-alignment_summary.gff"
#Note: in the new format, the gff gives several information about the coverage (based on the
python code found at https://github.com/PacificBiosciences/pbreports/blob/master/pbreports/
report/summarize_coverage/summarize_coverage.py):
#cov=(min_cov, median_cov, max_cov)))
#cov2=(mean_cov, sd_cov)))
#Because for the old version of SMRT assembly, we had only the average, I will use the average
with the new version too.
suffix_fasta=".fasta"


for ( i in file_names) {
  #File names
  fasta_input_name = paste(dir_name, i, suffix_fasta, sep = "")
  fasta_output_name = paste(dir_name, i, ".filtered_on_cov", suffix_fasta, sep = "")
  cov_file_name = paste(dir_name, i, suffix_cov, sep = "")
  print(fasta_input_name)

  #Get the average coverage data from the gff file
  names_T = c("Unitig", "Source", "Feature", "Start",
              "End", "Score", "Strand", "Phase", "Attributes")
  T = read_tsv(cov_file_name, col_names = names_T, comment = "#")
  T_sep = separate(data = T, col = Attributes, into = c("Cov", "Cov2", "Gaps"), sep=";")
  T_sep$Cov2 = str_remove(T_sep$Cov2, "cov2=")
  T = separate(data = T_sep, col = Cov2, into = c("Mean_coverage", "Sd_cov"), sep=",", convert =
TRUE)

  #Extracting the length of the unitig and the average coverage
  long = aggregate(T$End, by = list(T$Unitig), FUN = max)
  mean = aggregate(T$Mean_coverage, by = list(T$Unitig), FUN = mean)
  E = merge(long, mean, by= "Group.1")
  names(E) = c("Unitig", "Length", "Mean")
  med_value = mean(rep(E$Mean, times=E$Length))
  med_max = med_value * 1.5
  med_min = med_value / 1.5
  kept = E[E$Mean > med_min & E$Mean < med_max,]
  E$Pass = ifelse(E$Unitig %in% kept$Unitig, "pass", "fail")
  T$Pass = ifelse(T$Unitig %in% kept$Unitig, "pass", "fail")
  print(dim(kept))
  print(dim(E))
```

```
 len_kept=sum(kept$Length)
 len_all = sum(E$Length)
 prop_kept = round(len_kept*100/len_all, 2)

 #Now, that we have the thresholds and results, we will print the results in graphs and tables
 update_geom_defaults("point", list(colour = NULL))
 value_plot = ggplot(data = T, aes(x=Unitig, y = Mean_coverage, col = Pass, fill = Pass))
 value_plot + geom_boxplot(outlier.alpha = 0.5, outlier.size = 0.6) +
   geom_hline(aes(yintercept=med_min)) +
   geom_hline(aes(yintercept=med_max)) +
   theme_bw() +
   theme(axis.text.x = element_text(angle = 90, vjust = 0, hjust= 1))  +
   labs(title = paste("Coverage filter on PacBio assembly for", i),
        subtitle = paste("The total assembly goes from", len_all,
                         "to", len_kept, "bp (", prop_kept," kept) and from",
                         nrow(E), "contigs to", nrow(kept),"."))
 ggsave(paste(dir_name, i, "_filter.png", sep=""), width = 11, height = 8)
 write.table(E, file = paste(dir_name, i, "_filter.tab", sep=""),
             quote = FALSE, sep = "\t", row.names = FALSE)
 write.table(kept, file = paste(dir_name, i, "_pass_filter.tab", sep=""),
             quote = FALSE, sep = "\t", row.names = FALSE)

 #Let's filter the data, if the fasta files are there
 if (file.exists(fasta_input_name)) {
 fastafile = read.fasta(file = fasta_input_name, seqtype = "DNA",
                        forceDNAtolower = FALSE, as.string = TRUE, set.attributes = FALSE)

 f<-fastafile[names(fastafile) %in% paste(kept$Unitig, "|", polishing_algo, sep = "")]

 write.fasta(f, names(f), file.out=fasta_output_name)
 }else{
   print(paste("WARNING. I can't find the fasta file ", fasta_input_name,
               " so I will ignore the filtereing part", sep =""))
 }


}
```

Identify_telomeres.sh

```
#   <<>><<>><<>><<>>
# |  Needed inputs  |
#   <<>><<>><<>><<>>

bowtie_dir=”bowtie_directory”
work_dir=”your_directory”
polished_fasta="assembly_Czeina_default.fasta"
sample_name="Czeina"
```

```
telom_fn=${polished_fasta%.fasta}_telomeric_repeats
filtered_fn=${polished_fasta%.fasta}.filtered_on_cov.fasta
mt_ref_seq="Pseudocercospora_mori_mitochondrion.fasta" #From NCBI accession MG543071.1


#  -----------------------------
#   Detecting telomeric repeats
#  -----------------------------

${bowtie_dir}bowtie-build ${work_dir}${polished_fasta} ${work_dir}${sample_name}

${bowtie_dir}bowtie \
  ${work_dir}${sample_name} \
  -c CCCTAA \
  --all -v 0  \
  --threads 4 \
  | sort  -k 3 -nk 4  \
 > ${work_dir}${telom_fn}.txt

#Conversion from bowtie output format to bed format
awk 'BEGIN {FS= "\t"; OFS="\t"} {print $3, $4, $4+length($5), $1, 111, $2}' ${work_dir}$
{telom_fn}.txt \
 > ${work_dir}${telom_fn}.bed

#Merge repeats closer than the length of 1 (in case one repeat is mutated or has a sequencing
error, likely at the contig end since the coverage depth drops)
bedtools merge -i ${work_dir}${telom_fn}.bed -d 7 \
 > ${work_dir}${telom_fn}_merged.bed

#Keep only blocks of more than 10 repeats (so longer than 60)
awk 'BEGIN {FS= "\t"; OFS="\t"} {diff=($3 - $2); if (diff > 60) print $1,$2,$3,diff} ' \
   ${work_dir}${telom_fn}_merged.bed \
   > ${work_dir}${telom_fn}_merged_long.bed
```

removesmalls.pl

```
##removesmalls.pl
##from https://www.biostars.org/p/79202/
#!/usr/bin/perl
use strict;
use warnings;

my $minlen = shift or die "Error: `minlen` parameter not provided\n";
{
    local $/=">";
    while(<>) {
        chomp;
        next unless /\w/;
        s/>$//gs;
        my @chunk = split /\n/;
        my $header = shift @chunk;
        my $seqlen = length join "", @chunk;
```

```
        print ">$_" if($seqlen >= $minlen);
    }
    local $/="\n";
}
```

Heatmap.R

```
#heatmap in R:
#plot_cazymes.R
library(ggplot2)
mine.heatmap <- ggplot(data_h, mapping = aes(x = species, y = cazyme_family, fill = frequency))
+
   geom_tile() +
   theme(axis.text.y = element_text(size=6)) +
   xlab(label = "Cercospora species") +
   ylab(label = "CAZyme family") +
   scale_fill_gradient(low = "grey87",
                       high = "grey20")
mine.heatmap
data_h <- read.table("cazy_all_species_heatmap.txt", header = T)
```

Ultrametric_tree.R

```
library(phytools)
library(phangorn)
setwd('/home/laura/Desktop/Laura_Ruppert/1_data_Laura/Cercospora/11_CAFE/
Cercospora_effectors/')
species_tree = read.newick(file = "Cercospora_maxlike_tree_root_midpoint.nwk",text)
force.ultrametric<-function(tree,method=c("nnls","extend")){
  method<-method[1]
  if(method=="nnls") tree<-nnls.tree(cophenetic(tree),tree,
                                     rooted=TRUE,trace=0)
  else if(method=="extend"){
    h<-diag(vcv(tree))
    d<-max(h)-h
    ii<-sapply(1:Ntip(tree),function(x,y) which(y==x),
               y=tree$edge[,2])
    tree$edge.length[ii]<-tree$edge.length[ii]+d
  } else
    cat("method not recognized: returning input tree\n\n")
  tree
}
species_tree_ultra = force.ultrametric(species_tree)
write.tree(species_tree_ultra, file = "tree_ultrametric.txt", append = FALSE, digits = 10,
tree.names = FALSE)
is.ultrametric(species_tree_ultra)
```


Conclusions

and

Perspectives

Conclusions and Perspectives

The aim of this thesis was to investigate the population genomic variation in the fungal pathogen *Cercospora beticola* on wild and domesticated beet. We aimed to determine whether recent host domestication has led to different evolutionary trajectories between populations isolated from wild and domesticated beet.

We show that the commonly used reference-based mapping approach has high accuracy and recall at the read depth available for *C. beticola* isolates. Moreover, we show that genome composition has an effect on the reliability of variant calling using multiple genome alignment-based methods at 25X, currently an average depth for whole genome sequencing of fungi. We also show that repeat content has a significantly smaller effect on the reliability of multiple genome alignment-based methods when read coverage is 100X. We also showed that the use of multiple genome alignment-based methodologies can recover large structural variants with accuracy, and at the correct location without the use of additional software. Based on our study, we conclude that the multiple genome alignment-based approach can be applied in research op population data as sequencing costs decrease and higher depths can be achieved, and as third generation sequencing techniques develop.

Host domestication often results in the pathogens of these plants following different evolutionary trajectories than those on the wild relatives of domesticated crops (Stukenbrock, 2013; Stukenbrock and McDonald, 2008). Many agricultural crops have been domesticated thousands of years ago, making the 300 year domestication history of sugar beet a peculiar case. Previous studies indicated that global genotype flow in *C. beticola* is common, and that non-random mating occurs (Vaghefi et al., 2017). Our results concur with the high level of gene flow with isolates from North America and mainland Europe. Gene flow between the isolates from sea beet in the UK occurs to a lesser extent. As such, the isolates from the UK cluster by separately from the remaining isolates. Surprisingly, the isolates from the UK show substantial genetic variation among themselves.

Genetic differentiation between isolates from sea beet in the UK and Croatia showed a region that was highly differentiated between the two sampling locations. Within this region there was a gene that encoded an enzyme that plays a role in the production of aflatoxin B_1 (Gengan et al., 2006). This gene had a high frequency of premature stop codons in the isolates from Croatian sea beet, but very low in those from UK sea beet. The premature stop codon was present at an equally low frequency in isolates from Italian sugar beet, and intermediate frequencies in those isolates from North American sugar and table beet. The Croatian isolates that harboured the premature stop codon also clustered away from the remaining isolates in a reticulation network. Therefore, this region may be indicative of other regions that undergo similar evolution as the region that we have identified. There may be a biochemical interplay between sea beet and *C. beticola* in Croatia that is not seen elsewhere.

Our survey of the *Cercospora* genus showed that host range is correlated with effector and carbohydrate active enzyme (CAZyme) repertoires. In many fungal plant pathogens, the number of effectors and CAZymes a species harbours is strongly correlated with host range (Baroncelli et al., 2016; Guyon et al., 2014; Schmidt and Panstruga, 2011). We showed that all *Cercospora* species we considered contained the biosynthetic gene cluster for a universal toxin, cercosporin, produced by *Cerospora* species. It has been debated whether cercosporin is indeed produced by all *Cercospora* species due to the slow colonisation rate of some species (Assante et al., 1977; Groenewald et al., 2013; Luo et al., 2018). This gene cluster has also been duplicated and experienced multiple horizontal transfers among other plant pathogenic fungi (de Jonge et al., 2018). While cercosporin is a universal toxin, it is not required for the successful infection and colonisation of plants by all *Cercospora* species (Choquer et al., 2005). Therefore, *Cercospora* species must also rely on other mechanisms to infect the host. To illustrate the diversity in CAZymes and effector repertoire, we considered two specialist *Cercospora* species, *C. zeina* and *C. sojina*, as well as two generalist species, *C. beticola* and *C. cf. flagellaris*. We showed that two more specialised *Cercospora* species contained fewer CAZymes and effectors than the two generalist species. Furthermore, we investigated the genomic synteny between the two generalist

species. Large translocation events within the genomes of Ascomycete fungi are not uncommon (Hane et al., 2011). Therefore, we propose that the most recent common ancestor of these two species underwent chromosomal translocation events in its larger chromosomes.

This thesis provides a starting point for various experimental and further population genomics based questions. Experimental questions could aim to answer whether the Croatian isolates that possess the premature stop codon show host preference when infecting sea or sugar beet, or whether they are specific to sea beet. Further sampling of sugar beet in the UK should be considered to determine whether the substantial genetic diversity of *C. beticola* on sea beet is matched by these isolates. Furthermore, the genetic diversity of sea beet on the East Coast of the UK should be quantified in order to correlate fungal diversity with host diversity. Additionally, the direction of gene flow among *C. beticola* isolates should be identify to confirm whether there is gene flow from isolates found on wild beet to isolates from sugar beet, or vice versa. The identification of this is important to determine whether *C. beticola* on sea beet is reservoir for genetic diversity that can be introduced into isolates on sugar beet. This is an important aspect of anticipating emerging pandemics of sugar beet, and potentially breeding for resistance prior to an outbreak. Quantifying the level of gene flow between isolates from mainland Europe and North America may be insightful with regards to more careful control of international sugar beet seed imports and exports, as this may be a vector by which *C. beticola* is spread across the globe (Knight and Pethybridge, 2020).

In summary, with work presented in this thesis, we have shown that the evolution of *C. beticola* has not yet been strongly influenced by the domestication of its host. We found that *C. beticola* from both wild and domesticated hosts on mainland Europe and North America undergo admixture with one another. Additionally, we show that the *C. beticola* isolates from sea beet in the UK cluster separately from the remaining isolates. This may be due to geographic isolation of these isolates. Therefore, we postulate that geography is currently the primary contributing factor to the different evolutionary trajectories of *C. beticola*.

References for General Introduction and Conclusion

Abe, J., and Shimamoto, Y. (1989). Evolutionary aspects and species relationships. Int. Crop Netw. Ser. IBPGR.

Abe, J., Guan, G.-P., and Shimamoto, Y. (1997). A gene complex for annual habit in sugar beet (*Beta vulgaris L.*). Euphytica *94*, 129–135.

Abegg, F.A. (1936). A genetic factor for the annual habit in beets and linkage relationship (US Government Printing Office).

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

Alfenito, M.R., and Birchler, J.A. (1993). Molecular characterization of a maize B chromosome centric sequence. Genetics *135*, 589–597.

Allaby, R.G., Fuller, D.Q., and Brown, T.A. (2008). The genetic expectations of a protracted model for the origins of domesticated crops. Proc. Natl. Acad. Sci. U. S. A. *105*, 13982–13986.

Almeida, F., Rodrigues, M.L., and Coelho, C. (2019). The Still Underestimated Problem of Fungal Diseases Worldwide. Front. Microbiol. *10*.

Andersen, N.S., Siegismund, H.R., Meyer, V., and Jørgensen, R.B. (2005). Low level of gene flow from cultivated beets (*Beta vulgaris L.* ssp. *vulgaris*) into Danish populations of sea beet (Beta vulgaris L. ssp. maritima (L.) Arcangeli). Mol. Ecol. *14*, 1391–1405.

Anderson, R.M., and May, R.M. (1982). Coevolution of hosts and parasites. Parasitology *85*, 411–426.

Arkhipova, I.R. (2018). Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. Mol. Biol. Evol. *35*, 1332–1337.

Arnaud, J.-F., Fénart, S., Godé, C., Deledicque, S., Touzet, P., and Cuguen, J. (2009). Fine-scale geographical structure of genetic diversity in inland wild beet populations. Mol. Ecol. *18*, 3201–3215.

Arnaud, J.-F., Fénart, S., Cordellier, M., and Cuguen, J. (2010). Populations of weedy crop–wild hybrid beets show contrasting variation in mating system and population genetic structure. Evol. Appl. *3*, 305–318.

Assante, G., Locci, R., Camarda, L., Merlini, L., and Nasini, G. (1977). Screening of the genus *Cercospora* for secondary metabolites. Phytochemistry *16*, 243–247.

Aylward, J., Steenkamp, E.T., Dreyer, L.L., Roets, F., Wingfield, B.D., and Wingfield, M.J. (2017). A plant pathology perspective of fungal genome sequencing. IMA Fungus *8*, 1–15.

Bakhshi, M., Arzanlou, M., and Babai-Ahari, A. (2011). Uneven distribution of mating type alleles in Iranian populations of *Cercospora beticola*, the causal agent of Cercospora leaf spot disease of sugar beet. Phytopathol. Mediterr. *50*, 101–109.

Balter, M. (2007). Seeking Agriculture's Ancient Roots. Science *316*, 1830–1835.

Barnosky, A.D. (2001). Distinguishing the Effects of the Red Queen Court Jester on Miocene Mammal Evolution in the Northern Rocky Mountains. J. Vertebr. Paleontol. *21*, 172–185.

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

Barrett, R.D.H., and Schluter, D. (2008). Adaptation from standing genetic variation. Trends Ecol. Evol. *23*, 38–44.

Bartsch, D. (2010). Gene flow in sugar beet. Sugar Tech *12*, 201–206.

Bartsch, D., and Ellstrand, N.C. (1999). Genetic evidence for the origin of Californian wild beets (genus *Beta*). Theor. Appl. Genet. *99*, 1120–1130.

Bartsch, D., Lehnen, M., Clegg, J., Pohl-Orf, M., Schuphan, I., and Ellstrand, N.C. (1999). Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. Mol. Ecol. *8*, 1733–1741.

Bartsch, D., Cuguen, J., Biancardi, E., and Sweet, J. (2003). Environmental implications of gene flow from sugar beet to wild beet - current status and future research needs. Environ. Biosafety Res. *2*, 105–115.

Beauvais, A., Bozza, S., Kniemeyer, O., Formosa, C., Balloy, V., Henry, C., Roberson, R.W., Dague, E., Chignard, M., and Brakhage, A.A. (2013). Deletion of the α-(1, 3)-glucan synthase genes induces a restructuring of the conidial cell wall responsible for the avirulence of *Aspergillus fumigatus*. PLoS Pathog *9*, e1003716.

Bell, G.D.H., and Bauer, A.B. (1942). Experiments on growing sugar beet under continuous illumination. J. Agric. Sci. *32*, 112–141.

Bergelson, J., Kreitman, M., Stahl, E.A., and Tian, D. (2001). Evolutionary Dynamics of Plant R-Genes. Science *292*, 2281–2285.

Biancardi, E., and De Biaggi, M. (1979). *Beta maritima L.* in the Po Delta. Atti Convegno Tec. Internazionale Sulla Bieticoltura Commem. Ottavio Munerati ISCI Rovigo Italy 183–185.

Biancardi, E., Panella, L.W., and Lewellen, R.T. (2011). *Beta maritima*: The Origin of Beets (Springer Science & Business Media).

Bobev, S.G., Groenewald, J.Z., and Crous, P.W. (2009). First Report of *Cercospora beticola* as a Pathogen of German Statice (Goniolimon tataricum) in Bulgaria. Plant Dis. *93*, 553–553.

Bolton, M.D., Thomma, B.P.H.J., and Nelson, B.D. (2006). *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. Mol. Plant Pathol. *7*, 1–16.

Bolton, M.D., Secor, G.A., Rivera, V., Weiland, J.J., Rudolph, K., Birla, K., Rengifo, J., and Campbell, L.G. (2012). Evaluation of the potential for sexual reproduction in field populations of *Cercospora beticola* from USA. Fungal Biol. *116*, 511–521.

Bolton, M.D., de Jonge, R., Inderbitzin, P., Liu, Z., Birla, K., Van de Peer, Y., Subbarao, K.V., Thomma, B.P.H.J., and Secor, G.A. (2014). The heterothallic sugarbeet pathogen *Cercospora beticola* contains exon fragments of both MAT genes that are homogenized by concerted evolution. Fungal Genet. Biol. *62*, 43–54.

Booker, T.R., Jackson, B.C., and Keightley, P.D. (2017). Detecting positive selection in the genome. BMC Biol. *15*, 98.

Bornscheuer, E., Meyerholz, K., and Wunderlich, K.H. (1993). Seed production and quality. In The Sugar Beet Crop, (Springer), pp. 121–155.

Boyd, L.A., Ridout, C., O'Sullivan, D.M., Leach, J.E., and Leung, H. (2013). Plant-pathogen interactions: disease resistance in modern agriculture. Trends Genet. TIG *29*, 233–240.

Callahan, T.M., Rose, M.S., Meade, M.J., Ehrenshaft, M., and Upchurch, R.G. (1999). CFP, the Putative Cercosporin Transporter of *Cercospora kikuchii*, Is Required for Wild Type Cercosporin Production, Resistance, and Virulence on Soybean. Mol. Plant-Microbe Interactions® *12*, 901–910.

Camacho, J.P.M., Shaw, M.W., López-León, M.D., Pardo, M.C., and Cabrero, J. (1997). Population Dynamics of a Selfish B Chromosome Neutralized by the Standard Genome in the Grasshopper *Eyprepocnemis plorans*. Am. Nat. *149*, 1030–1050.

Cambareri, E.B., Singer, M.J., and Selker, E.U. (1991). Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. Genetics *127*, 699–710.

Capuano, F., Mülleder, M., Kok, R., Blom, H.J., and Ralser, M. (2014). Cytosine DNA Methylation Is Found in *Drosophila melanogaster* but Absent in *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, and Other Yeast Species. Anal. Chem. *86*, 3697–3702.

Carvunis, A.-R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., et al. (2012). Proto-genes and *de novo* gene birth. Nature *487*, 370–374.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A.G., Grigoriev, I.V., et al. (2016). Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. PLoS Genet. *12*.

Chen, H., Lee, M.-H., Daub, M.E., and Chung, K.-R. (2007). Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. Mol. Microbiol. *64*, 755–770.

Choquer, M., Dekkers, K.L., Chen, H.-Q., Cao, L., Ueng, P.P., Daub, M.E., and Chung, K.-R. (2005). The CTB1 Gene Encoding a Fungal Polyketide Synthase Is Required for Cercosporin Biosynthesis and Fungal Virulence of *Cercospora nicotianae*. Mol. Plant-Microbe Interactions® *18*, 468–476.

Chupp, C.C. (1953). A monograph of the fungus genus *Cercospora*. Monogr. Fungus Genus *Cercospora*.

Clay, K., and Kover, P.X. (1996). The Red Queen Hypothesis and Plant/Pathogen Interactions. Annu. Rev. Phytopathol. *34*, 29–50.

Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., Schmutz, J., Taga, M., White, G.J., Zhou, S., et al. (2009). The Genome of *Nectria haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion. PLOS Genet. *5*, e1000618.

Cook, D.E., Mesarich, C.H., and Thomma, B.P.H.J. (2015). Understanding Plant Immunity as a Surveillance System to Detect Invasion. Annu. Rev. Phytopathol. *53*, 541–563.

Croll, D., and McDonald, B.A. (2012). The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. PLOS Pathog. *8*, e1002608.

van Dam, P., Fokkens, L., Ayukawa, Y., van der Gragt, M., ter Horst, A., Brankovics, B., Houterman, P.M., Arie, T., and Rep, M. (2017). A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. Sci. Rep. *7*, 9042.

Dang, Y., Yang, Q., Xue, Z., and Liu, Y. (2011). RNA Interference in Fungi: Pathways, Functions, and Applications. Eukaryot. Cell *10*, 1148–1155.

Daub, M.E. (1982). Cercosporin, a photosensitizing toxin from *Cercospora* species. Phytopathology *72*, 370–374.

Daub, M.E., and Briggs, S.P. (1983). Changes in Tobacco Cell Membrane Composition and Structure Caused by Cercosporin. Plant Physiol. *71*, 763–766.

Daub, M.E., and Chung, K.-R. (2007). Cercosporin: a photoactivated toxin in plant disease. Online APSnet Featur. Doi *10*, 2007–0207.

Daub, M.E., and Ehrenshaft, M. (2000). The Photoactivated *Cercospora* Toxin Cercosporin: Contributions to Plant Disease and Fundamental Biology. Annu. Rev. Phytopathol. *38*, 461–490.

Daub, M.E., Herrero, S., and Chung, K.-R. (2005). Photoactivated perylenequinone toxins in fungal pathogenesis of plants. FEMS Microbiol. Lett. *252*, 197–206.

Denslow, S.A., Walls, A.A., and Daub, M.E. (2005). Regulation of biosynthetic genes and antioxidant properties of vitamin B6 vitamers during plant defense responses. Physiol. Mol. Plant Pathol. *66*, 244–255.

Dent, K.C., Stephen, J.R., and Finch-Savage, W.E. (2004). Molecular profiling of microbial communities associated with seeds of *Beta vulgaris* subsp. *vulgaris* (sugar beet). J. Microbiol. Methods *56*, 17–26.

Diamond, J., and Guns, G. (1997). Steel: The fates of human societies (New York: WW Norton).

Dohm, J.C., Minoche, A.E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sörensen, T.R., Stracke, R., Reinhardt, R., et al. (2014). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature *505*, 546–549.

Dolatabadian, A., Patel, D.A., Edwards, D., and Batley, J. (2017). Copy number variation and disease resistance in plants. Theor. Appl. Genet. *130*, 2479–2490.

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

Dutheil, J.Y., Mannhaupt, G., Schweizer, G., M.K. Sieber, C., Münsterkötter, M., Güldener, U., Schirawski, J., and Kahmann, R. (2016). A Tale of Genome Compartmentalization: The Evolution of Virulence Clusters in Smut Fungi. Genome Biol. Evol. *8*, 681–704.

Ehrenshaft, M., Bilski, P., Li, M.Y., Chignell, C.F., and Daub, M.E. (1999). A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis. Proc. Natl. Acad. Sci. *96*, 9374–9378.

Ellstrand, N.C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T.J., Haccou, P., Lu, B.-R., Snow, A.A., Neal Stewart, C., et al. (2013). Introgression of Crop Alleles into Wild or Weedy Populations. Annu. Rev. Ecol. Evol. Syst. *44*, 325–345.

Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. Trends Ecol. Evol. *21*, 569–575.

Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat. Rev. Genet. *8*, 610–618.

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

Felix, G., Regenass, M., and Boller, T. (1993). Specific perception of subnanomolar concentrations of chitin fragments by tomato cells: induction of extracellular alkalinization, changes in protein phosphorylation, and establishment of a refractory state. Plant J. *4*, 307–316.

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

Fluit, A.C., Visser, M.R., and Schmitz, F.-J. (2001). Molecular Detection of Antimicrobial Resistance. Clin. Microbiol. Rev. *14*, 836–871.

Fones, H.N., Bebber, D.P., Chaloner, T.M., Kay, W.T., Steinberg, G., and Gurr, S.J. (2020). Threats to global food security from emerging fungal and oomycete crop pathogens. Nat. Food *1*, 332–342.

Fulci, V., and Macino, G. (2007). Quelling: post-transcriptional gene silencing guided by small RNAs in *Neurospora crassa*. Curr. Opin. Microbiol. *10*, 199–203.

Fuller, D.Q. (2007). Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World. Ann. Bot. *100*, 903–924.

Gavrilets, S. (2004). Fitness landscapes and the origin of species (MPB-41) (Princeton University Press).

Gengan, R.M., Chuturgoon, A.A., and Dutton, M.F. (2006). Kinetics of the Oxidoreductase Involved in the Conversion of O-methylsterigmatocystin to Aflatoxin B1. Prep. Biochem. Biotechnol. *36*, 297–306.

Giraud, T., Gladieux, P., and Gavrilets, S. (2010). Linking emergence of fungal plant diseases and ecological speciation. Trends Ecol. Evol. *25*, 387–395.

Gisi, U., Chin, K.M., Knapova, G., Küng Färber, R., Mohr, U., Parisi, S., Sierotzki, H., and Steinfeld, U. (2000). Recent developments in elucidating modes of resistance to phenylamide, DMI and strobilurin fungicides. Crop Prot. *19*, 863–872.

Glémin, S. (2003). How Are Deleterious Mutations Purged? Drift Versus Nonrandom Mating. Evolution *57*, 2678–2687.

Godfray, H.C.J., Mason-D'Croz, D., and Robinson, S. (2016). Food system consequences of a fungal disease epidemic in a major crop. Philos. Trans. R. Soc. B Biol. Sci. *371*.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Goodwin, S.B., Dunkle, L.D., and Zismann, V.L. (2001). Phylogenetic Analysis of *Cercospora* and *Mycosphaerella* Based on the Internal Transcribed Spacer Region of Ribosomal DNA. Phytopathology™ *91*, 648–658.

Graham, R.D., and Webb, M.J. (1991). Micronutrients and Disease Resistance and Tolerance in Plants. In Micronutrients in Agriculture, (John Wiley & Sons, Ltd), pp. 329–370.

Groenewald, J.Z., Nakashima, C., Nishikawa, J., Shin, H.-D., Park, J.-H., Jama, A.N., Groenewald, M., Braun, U., and Crous, P.W. (2013). Species concepts in *Cercospora*: spotting the weeds among the roses. Stud. Mycol. *75*, 115–170.

Groenewald, M., Groenewald, J.Z., Braun, U., and Crous, P.W. (2006a). Host range of *Cercospora apii* and *C. beticola* and description of *C. apiicola*, a novel species from celery. Mycologia *98*, 275–285.

Groenewald, M., Groenewald, J.Z., Harrington, T.C., Abeln, E.C.A., and Crous, P.W. (2006b). Mating type gene analysis in apparently asexual *Cercospora* species is suggestive of cryptic sex. Fungal Genet. Biol. *43*, 813–825.

Groenewald, M., Groenewald, J.Z., Linde, C.C., and Crous, P.W. (2007). Development of polymorphic microsatellite and single nucleotide polymorphism markers for *Cercospora beticola* (*Mycosphaerellaceae*). Mol. Ecol. Notes *7*, 890–892.

Groenewald, M., Linde, C.C., Groenewald, J.Z., and Crous, P.W. (2008). Indirect evidence for sexual reproduction in *Cercospora beticola* populations from sugar beet. Plant Pathol. *57*, 25–32.

Gullino, M.L., Leroux, P., and Smith, C.M. (2000). Uses and challenges of novel compounds for plant disease control. Crop Prot. *19*, 1–11.

Guyon, K., Balagué, C., Roby, D., and Raffaele, S. (2014). Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. BMC Genomics *15*, 336.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H.Y., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T., et al. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature *461*, 393–398.

Habig, M., and Stukenbrock, E.H. (2020). 2 Origin, Function, and Transmission of Accessory Chromosomes. In Genetics and Biotechnology, J.P. Benz, and K. Schipper, eds. (Cham: Springer International Publishing), pp. 25–47.

Hahn, M. (2014). The rising threat of fungicide resistance in plant pathogenic fungi: *Botrytis* as a case study. J. Chem. Biol. *7*, 133–141.

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica *156*, 1–13.

Hammer, K. (1984). Das Domestikationssyndrom. Kult. *32*, 11–34.

Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H., Goodwin, S.B., and Oliver, R.P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biol. *12*, R45.

Harlan, J.R. (1976). Genetic Resources in Wild Relatives of Crops1. Crop Sci. *16*, cropsci1976.0011183X001600030004x.

Haueisen, J., Möller, M., Eschenbrenner, C.J., Grandaubert, J., Seybold, H., Adamiak, H., and Stukenbrock, E.H. (2019). Highly flexible infection programs in a specialized wheat pathogen. Ecol. Evol. *9*, 275–294.

Hawkins, N.J., Bass, C., Dixon, A., and Neve, P. (2019). The evolutionary origins of pesticide resistance. Biol. Rev. *94*, 135–155.

Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. Nature *447*, 418–424.

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

Henrissat, B., and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. *293*, 781–788.

Hermisson, J., and Pennings, P.S. (2005). Soft Sweeps: Molecular Population Genetics of Adaptation From Standing Genetic Variation. Genetics *169*, 2335–2352.

Hillman, G., Hedges, R., Moore, A., Colledge, S., and Pettitt, P. (2001). New evidence of Lateglacial cereal cultivation at Abu Hureyra on the Euphrates. The Holocene *11*, 383–393.

Hohmann, S., Kadereit, J.W., and Kadereit, G. (2006). Understanding Mediterranean-Californian disjunctions: molecular evidence from *Chenopodiaceae-Betoideae*. TAXON *55*, 67–78.

Honda, S., Lewis, Z.A., Huarte, M., Cho, L.Y., David, L.L., Shi, Y., and Selker, E.U. (2010). The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. Genes Dev. *24*, 443–454.

Horns, F., Petit, E., Yockteng, R., and Hood, M.E. (2012). Patterns of Repeat-Induced Point Mutation in Transposable Elements of Basidiomycete Fungi. Genome Biol. Evol. *4*, 240–247.

Hotegni Houessou, J.H., Beed, F., Sikirou, R., and Ezin, V. (2011). First report of *Cercospora beticola* on lettuce (*Lactuca sativa*) in Benin. New Dis. Rep. *23*, 16–16.

Houben, A. (2017). B Chromosomes – A Matter of Chromosome Drive. Front. Plant Sci. *8*.

Jacobsen, B., and Franc, G.D. (2009). Compendium of Beet Diseases and Pests (APS Press, St. Paul, MN).

Jeon, J., Choi, J., Lee, G.-W., Park, S.-Y., Huh, A., Dean, R.A., and Lee, Y.-H. (2015). Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. Sci. Rep. *5*, 8567.

John, B., and Lewis, K.R. (1966). Chromosome Variability and Geographic Distribution in Insects. Science *152*, 711–721.

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

Johnston, A.E., Trust, L., and Fellow, S. (2000). Efficient use of nutrients in agricultural production systems. Commun. Soil Sci. Plant Anal. *31*, 1599–1620.

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

Jones, P.D., Lister, D.H., Jaggard, K.W., and Pidgeon, J.D. (2003). Future Climate Impact on the Productivity of Sugar Beet (*Beta vulgaris L.*) in Europe. Clim. Change *58*, 93–108.

de Jonge, R., Esse, H.P. van, Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K.V., et al. (2012). Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc. Natl. Acad. Sci. *109*, 5110–5115.

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., et al. (2018). Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus *Colletotrichum*. Proc. Natl. Acad. Sci. 201712798.

Jullien, P.E., Susaki, D., Yelagandula, R., Higashiyama, T., and Berger, F. (2012). DNA Methylation Dynamics during Sexual Reproduction in *Arabidopsis thaliana*. Curr. Biol. *22*, 1825–1830.

Karr, A.L., and Albersheim, P. (1970). Polysaccharide-degrading Enzymes are Unable to Attack Plant Cell Walls without Prior Action by a “Wall-modifying Enzyme.” Plant Physiol. *46*, 69–80.

Kazazian, H.H. (2004). Mobile Elements: Drivers of Genome Evolution. Science *303*, 1626–1632.

Keightley, P.D., and Eyre-Walker, A. (2010). What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? Philos. Trans. R. Soc. B Biol. Sci. *365*, 1187–1193.

Kent, T.V., Uzunović, J., and Wright, S.I. (2017). Coevolution between transposable elements and recombination. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20160458.

Khan, J., Rio, L.E.D., Nelson, R., Rivera-Varas, V., Secor, G.A., and Khan, M.F.R. (2008). Survival, Dispersal, and Primary Infection Site for *Cercospora beticola* in Sugar Beet. Plant Dis. *92*, 741–745.

Kim, K.-T., Jeon, J., Choi, J., Cheong, K., Song, H., Choi, G., Kang, S., and Lee, Y.-H. (2016). Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. Front. Plant Sci. *7*.

King, R., Urban, M., Lauder, R.P., Hawkins, N., Evans, M., Plummer, A., Halsey, K., Lovegrove, A., Hammond-Kosack, K., and Rudd, J.J. (2017). A conserved fungal glycosyltransferase facilitates pathogenesis of plants by enabling hyphal growth on solid surfaces. PLoS Pathog. *13*, e1006672.

Kleemann, J., Rincon-Rivera, L.J., Takahara, H., Neumann, U., Themaat, E.V.L. van, Does, H.C. van der, Hacquard, S., Stüber, K., Will, I., Schmalenbach, W., et al. (2012). Sequential Delivery of Host-Induced Virulence Effectors by Appressoria and Intracellular Hyphae of the Phytopathogen *Colletotrichum higginsianum*. PLOS Pathog. *8*, e1002643.

Knight, N.L., and Pethybridge, S.J. (2020). An improved PCR assay for species-specific detection and quantification of *Cercospora beticola*. Can. J. Plant Pathol. *42*, 72–83.

Knight, N.L., Vaghefi, N., Hansen, Z.R., Kikkert, J.R., and Pethybridge, S.J. (2018). Temporal Genetic Differentiation of *Cercospora beticola* Populations in New York Table Beet Fields. Plant Dis. *102*, 2074–2082.

Knight, N.L., Vaghefi, N., Kikkert, J.R., and Pethybridge, S.J. (2019a). Alternative Hosts of *Cercospora beticola* in Field Surveys and Inoculation Trials. Plant Dis. *103*, 1983–1990.

Knight, N.L., Vaghefi, N., Kikkert, J.R., Bolton, M.D., Secor, G.A., Rivera, V.V., Hanson, L.E., Nelson, S.C., and Pethybridge, S.J. (2019b). Genetic Diversity and Structure in Regional *Cercospora beticola* Populations from *Beta vulgaris* subsp. *vulgaris* Suggest Two Clusters of Separate Origin. Phytopathology® *109*, 1280–1292.

Knogge, W. (1996). Fungal Infection of Plants. Plant Cell *8*, 1711–1722.

Kondrashov, A.S. (1995). Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? J. Theor. Biol. *175*, 583–594.

Kouzminova, E., and Selker, E.U. (2001). dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in Neurospora. EMBO J. *20*, 4309–4323.

Kubiak-Martens, L. (1999). The plant food component of the diet at the late Mesolithic (Ertebolle) settlement at Tybrind Vig, Denmark. Veg. Hist. Archaeobotany *8*, 117–127.

Kubiak-Martens, L. (2002). New evidence for the use of root foods in pre-agrarian subsistence recovered from the late Mesolithic site at Halsskov, Denmark. Veg. Hist. Archaeobotany *11*, 23–32.

Laricchia, K.M., Zdraljevic, S., Cook, D.E., and Andersen, E.C. (2017). Natural Variation in the Distribution and Abundance of Transposable Elements Across the *Caenorhabditis elegans* Species. Mol. Biol. Evol. *34*, 2187–2202.

Lartey, R.T., Caesar-TonThat, T.C., Caesar, A.J., Shelver, W.L., Sol, N.I., and Bergman, J.W. (2005). Safflower: A New Host of *Cercospora beticola*. Plant Dis. *89*, 797–801.

Latz, E., Eisenhauer, N., Rall, B.C., Allan, E., Roscher, C., Scheu, S., and Jousset, A. (2012). Plant diversity improves protection against soil-borne pathogens by fostering antagonistic bacterial communities. J. Ecol. *100*, 597–604.

Lawrence, J.S., and Meredith, D.S. (1970). Wind dispersal of conidia of *Cercospora beticola*. Phytopathology *60*, 1076–1078.

Levin, D.A., Francisco-Ortega, J., and Jansen, R.K. (1996). Hybridization and the extinction of rare plant species. Conserv. Biol. *10*, 10–16.

Lewis, Z.A., Adhvaryu, K.K., Honda, S., Shiver, A.L., Knip, M., Sack, R., and Selker, E.U. (2010). DNA Methylation and Normal Chromosome Behavior in *Neurospora* Depend on Five Components of a Histone Methyltransferase Complex, DCDC. PLOS Genet. *6*, e1001196.

Li, L., Chang, S., and Liu, Y. (2010). RNA interference pathways in filamentous fungi. Cell. Mol. Life Sci. *67*, 3849–3863.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. Annu. Rev. Plant Biol. *66*, 513–545.

Loewe, L., and Hill, W.G. (2010). The population genetics of mutations: good, bad and indifferent. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1153–1167.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M., and Henrissat, B. (2010). A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem. J. *432*, 437–444.

Londo, J.P., Chiang, Y.-C., Hung, K.-H., Chiang, T.-Y., and Schaal, B.A. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc. Natl. Acad. Sci. U. S. A. *103*, 9578–9583.

Lucas, J.A., Hawkins, N.J., and Fraaije, B.A. (2015). The evolution of fungicide resistance. Adv. Appl. Microbiol. *90*, 29–92.

Luo, X., Cao, J., Huang, J., Wang, Z., Guo, Z., Chen, Y., Ma, S., and Liu, J. (2018). Genome sequencing and comparative genomics reveal the potential pathogenic mechanism of *Cercospora sojina* Hara on soybean. DNA Res. *25*, 25–37.

Ma, Z., and Michailides, T.J. (2005). Advances in understanding molecular mechanisms of fungicide resistance and molecular detection of resistant genotypes in phytopathogenic fungi. Crop Prot. *24*, 853–863.

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

Martínez-Ainsworth, N.E., and Tenaillon, M.I. (2016). Superheroes and masterminds of plant domestication. C. R. Biol. *339*, 268–273.

May, R.M., Anderson, R.M., Bodmer, W.F., and Kingman, J.F.C. (1983). Epidemiology and genetics in the coevolution of parasites and hosts. Proc. R. Soc. Lond. B Biol. Sci. *219*, 281–313.

Mbengue, M., Navaud, O., Peyraud, R., Barascud, M., Badet, T., Vincent, R., Barbacci, A., and Raffaele, S. (2016). Emerging Trends in Molecular Interactions between Plants and the Broad Host Range Fungal Pathogens *Botrytis cinerea* and *Sclerotinia sclerotiorum*. Front. Plant Sci. *7*.

McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G.M., Goff, S., Holub, E.B., and Dangl, J.L. (1998). Intragenic Recombination and Diversifying Selection Contribute to the Evolution of Downy Mildew Resistance at the RPP8 Locus of *Arabidopsis*. Plant Cell *10*, 1861–1874.

McGrath, M.T. (2001). Fungicide Resistance in Cucurbit Powdery Mildew: Experiences and Challenges. Plant Dis. *85*, 236–245.

McLysaght, A., and Guerzoni, D. (2015). New genes from non-coding sequence: the role of *de novo* protein-coding genes in eukaryotic evolutionary innovation. Philos. Trans. R. Soc. B Biol. Sci. *370*, 20140332.

McMahon, T.A., Halstead, N.T., Johnson, S., Raffel, T.R., Romansic, J.M., Crumrine, P.W., and Rohr, J.R. (2012). Fungicide-induced declines of freshwater biodiversity modify ecosystem functions and services. Ecol. Lett. *15*, 714–722.

Melchers, L.S., and Stuiver, M.H. (2000). Novel genes for disease-resistance breeding. Curr. Opin. Plant Biol. *3*, 147–152.

Mercier, A., Carpentier, F., Duplaix, C., Auger, A., Pradier, J.-M., Viaud, M., Gladieux, P., and Walker, A.-S. (2019). The polyphagous plant pathogenic fungus *Botrytis cinerea* encompasses host-specialized and generalist populations. Environ. Microbiol. *21*, 4808–4821.

Mercier, A., Simon, A., Lapalu, N., Giraud, T., Bardin, M., Walker, A.-S., Viaud, M., and Gladieux, P. (2020). Population genomics reveals molecular determinants of specialization to tomato in the polyphagous fungal pathogen *Botrytis cinerea*. BioRxiv 2020.07.24.219691.

Meredith, D.S. (1967). Conidium release and dispersal in *Cercospora beticola*. Phytopathology *57*, 889–893.

Messer, P.W., and Petrov, D.A. (2013). Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol. Evol. *28*, 659–669.

Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S., and Michelmore, R.W. (1998). Receptor-like Genes in the Major Resistance Locus of Lettuce Are Subject to Divergent Selection. Plant Cell *10*, 1833–1846.

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

Migicovsky, Z., and Myles, S. (2017). Exploiting Wild Relatives for Genomics-assisted Breeding of Perennial Crops. Front. Plant Sci. *8*.

Mohanta, T.K., and Bae, H. (2015). The diversity of fungal genome. Biol. Proced. Online *17*, 8.

Molina-Cano, J.-L., Russell, J.R., Moralejo, M.A., Escacena, J.L., Arias, G., and Powell, W. (2005). Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. TAG Theor. Appl. Genet. Theor. Angew. Genet. *110*, 613–619.

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

Möller, M., Habig, M., Freitag, M., and Stukenbrock, E.H. (2018). Extraordinary Genome Instability and Widespread Chromosome Rearrangements During Vegetative Growth. Genetics *210*, 517–529.

Moretti, M., Saracchi, M., and Farina, G. (2004). Morphological, physiological and genetic diversity within a small population of *Cercospora beticola Sacc*. Ann. Microbiol. *54*, 129–150.

Morris, J.J., Lenski, R.E., and Zinser, E.R. (2012). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. MBio *3*.

Moyers, B.T., Morrell, P.L., and McKay, J.K. (2018). Genetic Costs of Domestication and Improvement. J. Hered. *109*, 103–116.

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

Nakayashiki, H., Kadotani, N., and Mayama, S. (2006). Evolution and Diversification of RNA Silencing Proteins in Fungi. J. Mol. Evol. *63*, 127–135.

Ohta, T. (1973). Slightly Deleterious Mutant Substitutions in Evolution. Nature *246*, 96–98.

Olmo, H.P., and Simmonds, N.W. (1976). Evolution of crop plants (Longman: London, UK).

Olsson, O., and Hibbs, D.A. (2005). Biogeography and long-run economic development. Eur. Econ. Rev. *49*, 909–938.

Omrane, S., Audéon, C., Ignace, A., Duplaix, C., Aouini, L., Kema, G., Walker, A.-S., and Fillinger, S. (2018). Correction for Omrane et al., “Plasticity of the MFS1 Promoter Leads to Multidrug Resistance in the Wheat Pathogen *Zymoseptoria tritici.*” MSphere *3*.

Owen, F.V. (1954). The significance of single gene reactions in sugar beets. Proc Am Soc Sugar Beet Technol. *8*, 392–398.

Owen, F.V., Carsner, E., and Stout, M. (1940). Photothermal induction of flowering in sugar beets. J Agric Res *61*, 101–124.

Palma-Guerrero, J., Torriani, S.F.F., Zala, M., Carter, D., Courbot, M., Rudd, J.J., McDonald, B.A., and Croll, D. (2016). Comparative transcriptomic analyses of *Zymoseptoria tritici* strains show complex lifestyle transitions and intraspecific variability in transcription profiles. Mol. Plant Pathol. *17*, 845–859.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B.H., and Jones, J.D.G. (1997). Novel Disease Resistance Specificities Result from

Sequence Exchange between Tandemly Repeated Genes at the Cf-4/9 Locus of Tomato. Cell *91*, 821–832.

Pennings, P.S., and Hermisson, J. (2006). Soft Sweeps II—Molecular Population Genetics of Adaptation from Recurrent Mutation or Migration. Mol. Biol. Evol. *23*, 1076–1084.

Pethybridge, S.J., Kikkert, J.R., Hanson, L.E., and Nelson, S.C. (2018). Challenges and Prospects for Building Resilient Disease Management Strategies and Tactics for the New York Table Beet Industry. Agronomy *8*, 112.

Pimentel, D., Wilson, C., McCullum, C., Huang, R., Dwen, P., Flack, J., Tran, Q., Saltman, T., and Cliff, B. (1997). Economic and Environmental Benefits of Biodiversity. BioScience *47*, 747–757.

Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F.E., and Croll, D. (2017). Using Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal Pathogen Evolution. Front. Plant Sci. *8*.

Pool, V.W., and McKay, M.B. (1916). Relation of stomatal movement to infection by *Cercospora beticola*. J. Agric. Res. *5*, 1011–1038.

Poole, N.F., and Arnaudin, M.E. (2014). The role of fungicides for effective disease management in cereal crops. Can. J. Plant Pathol. *36*, 1–11.

Priest, S.J., Yadav, V., and Heitman, J. (2020). Advances in understanding the evolution of fungal genome architecture. F1000Research *9*.

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

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H.Y., Zody, M.C., Kunjeti, S.G., Donofrio, N.M., et al. (2010). Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage. Science *330*, 1540–1543.

Rangel, L.I., Spanner, R.E., Ebert, M.K., Pethybridge, S.J., Stukenbrock, E.H., de Jonge, R., Secor, G.A., and Bolton, M.D. (2020). *Cercospora beticola*: The intoxicating lifestyle of the leaf spot pathogen of sugar beet. Mol. Plant Pathol. *21*, 1020–1041.

Rep, M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. FEMS Microbiol. Lett. *253*, 19–27.

Robinson, D.E., and Harild, J.A. (2002). Archaeobotany of an early Ertebölle (late Mesolithic) site at Halsskov, Zealand, Denmark. Hunt.-Gatherer Archaeobotany Perspect. North. Temp. Zone Univ. Coll. Lond. Lond. 84–95.

Rolph, G.M. (1917). Something about sugar: Its history, growth, manufacture and distribution (JJ Newbegin).

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. Mol. Microbiol. *6*, 3343–3353.

Rooney-Latham, S., Scheck, H.J., and Walber, T.M. (2010). First Report of *Cercospora beticola* Causing a Leaf Spot Disease on Acanthus mollis in California. Plant Dis. *95*, 224–224.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., et al. (2011). Effector diversification within compartments of

the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. *2*, 202.

Rovira, A.D. (1965). Interactions between plant roots and soil microorganisms. Annu. Rev. Microbiol. *19*, 241–266.

Ruppel, E.G. (1972). Variation among isolates of *Cercospora beticola* from sugar beet. Phytopathology *62*, 134–136.

Russell, P.E. (2005). A century of fungicide evolution. J. Agric. Sci. *143*, 11–25.

Sánchez-Vallet, A., Mesters, J.R., and Thomma, B.P.H.J. (2015). The battle for chitin recognition in plant-microbe interactions. FEMS Microbiol. Rev. *39*, 171–183.

Santoni, S., and Bervillé, A. (1992). Characterization of the nuclear ribosomal DNA units and phylogeny of *Beta L.* wild forms and cultivated beets. Theor. Appl. Genet. *83*, 533–542.

Schmidt, S.M., and Panstruga, R. (2011). Pathogenomics of fungal plant parasites: what have we learnt about pathogenesis? Curr. Opin. Plant Biol. *14*, 392–399.

Schouten, A., Baarlen, P.V., and Kan, J.A.L.V. (2008). Phytotoxic Nep1-like proteins from the necrotrophic fungus *Botrytis cinerea* associate with membranes and the nucleus of plant cells. New Phytol. *177*, 493–505.

Secor, G.A., and Gudmestad, N.C. (1999). Managing fungal diseases of potato. Can. J. Plant Pathol. *21*, 213–221.

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

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

Selker, E.U., and Stevens, J.N. (1985). DNA methylation at asymmetric sites is associated with numerous transition mutations. Proc. Natl. Acad. Sci. *82*, 8114–8118.

Selker, E.U., Cambareri, E.B., Jensen, B.C., and Haack, K.R. (1987). Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell *51*, 741–752.

Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., and Freitag, M. (2003). The methylated component of the *Neurospora crassa* genome. Nature *422*, 893–897.

Shaw, B., Thomas, T.H., and Cooke, D.T. (2002). Responses of sugar beet (*Beta vulgaris L*.) to drought and nutrient deficiency stress. Plant Growth Regul. *37*, 77–83.

Shiu, P.K.T., Raju, N.B., Zickler, D., and Metzenberg, R.L. (2001). Meiotic Silencing by Unpaired DNA. Cell *107*, 905–916.

Silva, M., and Pereira, O.L. (2008). Postharvest *Cercospora apii* fruit rot disease on *Cucurbita maxima* (Cucurbitaceae). Australas. Plant Dis. Notes *3*, 21.

Singh, K.B., and Reddy, M.V. (1991). Advances In Disease-Resistance Breeding In Chickpea11Journal Article No. 1124; joint contribution from the International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria, and the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh 502 324, India. In Advances in Agronomy, N.C. Brady, ed. (Academic Press), pp. 191–222.

Skaracis, G.N., and Biancardi, E. (2000). Breeding for *Cercospora* resistance in sugarbeet. *Cercospora Beticola Sacc* Biol. Agron. Influ. Control Meas. Sugar Beet 177–195.

Skaracis, G.N., Pavli, O.I., and Biancardi, E. (2010). Cercospora leaf spot disease of sugar beet. Sugar Tech *12*, 220–228.

Smith, G.A., and Ruppel, E.G. (1974). Herability of Resistance to Cercospora Leaf Spot in Sugarbeet1. Crop Sci. *14*, cropsci1974.0011183X001400010034x.

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. *23*, 23–35.

Soyer, J.L., Ghalid, M.E., Glaser, N., Ollivier, B., Linglin, J., Grandaubert, J., Balesdent, M.-H., Connolly, L.R., Freitag, M., Rouxel, T., et al. (2014). Epigenetic Control of Effector Gene Expression in the Plant Pathogenic Fungus *Leptosphaeria maculans*. PLOS Genet. *10*, e1004227.

Spann, T.M., and Schumann, A.W. (2010). Mineral Nutrition Contributes to Plant Disease and Pest Resistance. EDIS *2010*.

Stephan, W. (2019). Selective Sweeps. Genetics *211*, 5–13.

Stergiopoulos, I., and de Wit, P.J.G.M. (2009). Fungal Effector Proteins. Annu. Rev. Phytopathol. *47*, 233–263.

Stern, D.L. (2013). The genetic causes of convergent evolution. Nat. Rev. Genet. *14*, 751–764.

Stevanato, P., Biaggi, M.D., Skaracis, G.N., Colombo, M., Mandolino, G., and Biancardi, E. (2001). The sea beet (*Beta vulgaris L.* ssp. *maritima*) of the adriatic coast as source of resistance for sugar beet. Sugar Tech *3*, 77–82.

Stukenbrock, E.H. (2013). Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. New Phytol. *199*, 895–897.

Stukenbrock, E.H., and McDonald, B.A. (2008). The Origins of Plant Pathogens in Agro-Ecosystems. Annu. Rev. Phytopathol. *46*, 75–100.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature *414*, 277–283.

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

Taylor, T.V., Mitchell, T.K., and Daub, M.E. (2006). An Oxidoreductase Is Involved in Cercosporin Degradation by the Bacterium *Xanthomonas campestris* pv. *zinniae*. Appl. Environ. Microbiol. *72*, 6070–6078.

Tedford, S.L., Burlakoti, R.R., Schaafsma, A.W., and Trueman, C.L. (2018). Relationships among airborne *Cercospora beticola* conidia concentration, weather variables and cercospora leaf spot severity in sugar beet (*Beta vulgaris L.*). Can. J. Plant Pathol. *40*, 1–10.

Tellier, A., de Vienne, D.M., Giraud, T., Hood, M., and Refrégier, G. (2010). Theory and examples of reciprocal influence between hosts and pathogens, from short-term to long term interactions: coevolution, cospeciation and pathogen speciation following host shifts (Nova Science Publishers, Inc.; 1 edition (September 1, 2010)).

Thrall, P.H., and Burdon, J.J. (1999). The spatial scale of pathogen dispersal: Consequences for disease dynamics and persistence. Evol. Ecol. Res. *1*, 681–701.

Thynne, E., Mead, O.L., Chooi, Y.-H., McDonald, M.C., and Solomon, P.S. (2019). Acquisition and Loss of Secondary Metabolites Shaped the Evolutionary Path of Three Emerging Phytopathogens of Wheat. Genome Biol. Evol. *11*, 890–905.

Torres, M.A., Jones, J.D.G., and Dangl, J.L. (2006). Reactive Oxygen Species Signaling in Response to Pathogens. Plant Physiol. *141*, 373–378.

Torres-Martínez, S., and Ruiz-Vázquez, R.M. (2017). The RNAi Universe in Fungi: A Varied Landscape of Small RNAs and Biological Functions. Annu. Rev. Microbiol. *71*, 371–391.

Trindade, S., Perfeito, L., and Gordo, I. (2010). Rate and effects of spontaneous mutations that affect fitness in mutator *Escherichia coli*. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1177–1186.

Turgay, E.B., Bakır, M., Özeren, P., Katırcioğlu, Y.Z., and Maden, S. (2010). Detection of pathotypes and genetic diversity of *Cercospora beticola*. Plant Pathol. J. *26*, 306–312.

Upchurch, R.G., Rose, M.S., Eweida, M., and Callahan, T.M. (2002). Transgenic assessment of CFP-mediated cercosporin export and resistance in a cercosporin-sensitive fungus. Curr. Genet. *41*, 25–30.

Vaghefi, N., Nelson, S.C., Kikkert, J.R., and Pethybridge, S.J. (2017a). Genetic structure of *Cercospora beticola* populations on Beta vulgaris in New York and Hawaii. Sci. Rep. *7*, 1726.

Vaghefi, N., Kikkert, J.R., Bolton, M.D., Hanson, L.E., Secor, G.A., Nelson, S.C., and Pethybridge, S.J. (2017b). Global genotype flow in *Cercospora beticola* populations confirmed through genotyping-by-sequencing. PLOS ONE *12*, e0186488.

Valenzeno, D.P., and Pooler, J.P. (1987). Photodynamic Action. BioScience *37*, 270–276.

Van Geyt, J.P.C., Lange, W., Oleo, M., and De Bock, T.S. (1990). Natural variation within the genus *Beta* and its possible use for breeding sugar beet: a review. Euphytica *49*, 57–76.

Van Valen, L. (1977). The Red Queen. Am. Nat. *111*, 809–810.

Van Wyk, S., Wingfield, B.D., De Vos, L., Van Der Merwe, N.A., and Steenkamp, E. (2020). RIP in the Ascomycota Genome-wide analyses of Repeat-Induced Point (RIP) mutations in the Ascomycota. Front. Microbiol. *11*, 3625.

Vestal, E.F. (1933). Pathogenicity, host response and control of Cercospora leaf spot of sugar beet. Agr. Exptl Sta. Iowa State College Of Agri. Mechanics Arts. Res. Bull *168*.

Vinckenbosch, N., Dupanloup, I., and Kaessmann, H. (2006). Evolutionary fate of retroposed gene copies in the human genome. Proc. Natl. Acad. Sci. *103*, 3220–3225.

Voigt, R., Grüger, E., Baier, J., and Meischner, D. (2008). Seasonal variability of Holocene climate: a palaeolimnological study on varved sediments in Lake Jues (Harz Mountains, Germany). J. Paleolimnol. *40*, 1021–1052.

Walker, A.-S., Gladieux, P., Decognet, V., Fermaud, M., Confais, J., Roudet, J., Bardin, M., Bout, A., Nicot, P.C., Poncet, C., et al. (2015). Population structure and temporal maintenance of the multihost fungal pathogen *Botrytis cinerea*: causes and implications for disease management. Environ. Microbiol. *17*, 1261–1274.

Wang, L., Sun, Y., Sun, X., Yu, L., Xue, L., He, Z., Huang, J., Tian, D., Hurst, L.D., and Yang, S. (2020). Repeat-induced point mutation in *Neurospora crassa* causes the highest known mutation rate and mutational burden of any cellular life. Genome Biol. *21*, 142.

Watters, M.K., Randall, T.A., Margolin, B.S., Selker, E.U., and Stadler, D.R. (1999). Action of Repeat-Induced Point Mutation on Both Strands of a Duplex and on Tandem Duplications of Various Sizes in *Neurospora*. Genetics *153*, 705–714.

Weiland, J., and Koch, G. (2004). Sugarbeet leaf spot disease (*Cercospora beticola Sacc.*)†. Mol. Plant Pathol. *5*, 157–166.

Wendel, J.F., Lisch, D., Hu, G., and Mason, A.S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Curr. Opin. Genet. Dev. *49*, 1–7.

Werren, J.H. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc. Natl. Acad. Sci. *108*, 10863–10870.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. *8*, 973–982.

Williamson, B., Tudzynski, B., Tudzynski, P., and Kan, J. a. L.V. (2007). *Botrytis cinerea*: the cause of grey mould disease. Mol. Plant Pathol. *8*, 561–580.

Wilson, E.B. (1907). The supernumerary chromosomes of *Hemiptera*. Sci. NY *26*, 870–871.

van Wyk, S., Wingfield, B.D., De Vos, L., van der Merwe, N.A., Santana, Q.C., and Steenkamp, E.T. (2019). Repeat-Induced Point Mutations Drive Divergence between *Fusarium circinatum* and Its Close Relatives. Pathogens *8*, 298.

van Wyk, S. van, Wingfield, B.D., Merwe, N.A. van der, Vos, L.D., Maphosa, M., Potgieter, L., and Steenkamp, E.T. (2020). Repeat-Induced Point Mutations (RIP) Drives the Formation of Distinct Sub-Genomic Compartments in *Fusarium Circinatum*.

Yoshida, K., Saunders, D.G.O., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue, Y., Chuma, I., Tosa, Y., Cano, L.M., et al. (2016). Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics *17*, 370.

Yu, M.H., Heijbroek, W., and Pakish, L.M. (1999). The sea beet source of resistance to multiple species of root-knot nematode. Euphytica *108*, 151–155.

Zachow, C., Müller, H., Tilcher, R., and Berg, G. (2014). Differences between the rhizosphere microbiome of *Beta vulgaris* ssp. *maritima*—ancestor of all beet crops—and modern sugar beets. Front. Microbiol. *5*.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. Science *328*, 916–919.

Zhao, Z., Liu, H., Wang, C., and Xu, J.-R. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics *14*, 274.

Zilberman, D. (2008). The evolving functions of DNA methylation. Curr. Opin. Plant Biol. *11*, 554–559.

Zipfel, C. (2014). Plant pattern-recognition receptors. Trends Immunol. *35*, 345–351.

Zipfel, C., and Robatzek, S. (2010). Pathogen-Associated Molecular Pattern-Triggered Immunity: Veni, Vidi…? Plant Physiol. *154*, 551–554.

Zohary, D. (1999). Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. Genet. Resour. Crop Evol. *46*, 133–142.

Zohary, D., Hopf, M., and Weiss, E. (2012). Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin (Oxford University Press on Demand).

Zuccaro, A., and Langen, G. (2020). Breeding for resistance: can we increase crop resistance to pathogens without compromising the ability to accommodate beneficial microbes? New Phytol. *227*, 279–282.

Acknowledgements

This thesis is a product of an exceptional three years, and a lifetime of work. It has taken a lot to go from the barefooted child catching bugs and climbing trees in my backyard in South Africa to completing my Ph.D. at the Max Planck Institute. I could not have managed this by myself, and would like to express my heartfelt gratitude to several people.

Firstly, to my supervisor, Eva Stukenbrock, thank you for the opportunity to conduct this research in your lab. I value the life lessons I have learned from you. I appreciate the freedom to explore my ideas, and to manage the responsibilities you have entrusted to me. I also wish to thank the International Max Planck Research School for funding me and my work, and for providing excellent opportunities to grow as a scientist.

Alice Feurtey, thank you for your unwavering support from the day I started the IMPRS rotation with you to the end. Having you for emotional and intellectual support has been my saving grace, and I would have really struggled without your help!

I would like to thank my thesis advisory committee members, Tal Dagan and Tobias Kaiser, for their guidance and continued interest in my project. I also wish to thank my examination committee, Tal, Irene Barnes, and Matthias Leippe for their involvement in the examination of my thesis.

All current and past members of the Environmental Genomics research group, and my friends at the MPI for Evolutionary Biology. You have all made this experience one I will treasure. It feels strange singling specific people out as everyone within the group has played a role in my life during my time here. As such, I will narrow it down to the people that have shared the most beer and conversations with me: Wagner, Eli, Christoph, Michael, and Demetris, thank you for that! My gratitude also goes to Andrea for making everything so easy to deal with, coffees, snacks, and all the penguins! A special thanks also goes to the students I have been fortunate to supervise, and that have contributed to the project, Laura, Johannes, and Ronja. Thank you for allowing me to be a part of your education. Wiola, thank you for always being up for an adventure, regardless of how inane you think it might be!

My family. My mom, Andrea, and my bonus mom, Aunty Ann. Thank you for always believing in me and always encouraging my ambitious dreams. This would most definitely not have been possible without such a strong base to grow from. Both of you have inspired me and helped me become the person I am. Without you two strong, independent women in my life I may not have had the faith in myself to embrace the adventure of moving across the world to pursue this degree. To my maternal grandparents, for instilling a sense of curiosity in the world, and fostering this throughout my childhood.

My closest friend, Stephanie, for all the phone calls, emails, proof-reading, jokes, and the most wonderful friendship. Thank you for never being far away, for talking courage into me when the world was grey, and for celebrating with me when everything was colourful again. We have come a long way from the honours student in questionable shorts that eyed the grouchy lab manager in 2014 to where we are today. You inspire me more than you know.

I have made many memories in Kiel that I hold very dear to me. Many of these occurred within the walls of Unrat, so a great deal of gratitude goes out to everyone there. They have become my home away from home. Of course, there are a few friends that I hold very dear here, too. Thank you for all of the memories in front of stages, in camp sites, in trains, and in the harbour. Finally, I extend gratitude to Nyx, for being a good cat.

Declaration of Author's Contributions

The Ph.D. thesis of **Lizel Potgieter** consisted of three chapters in the form of published and unpublished manuscripts. Specific contributions for the chapters are detailed here. Data for chapters 2 and 3 are available upon request.

Chapter 1: On variant discovery in genomes of fungal plant pathogens

Chapter 1 was published in Frontier in Microbiology in 2020

Lizel Potgieter, Alice Feurtey, Julien Y. Dutheil, and Eva H. Stukenbrock. "On variant discovery in genomes of fungal plant pathogens." Frontiers in microbiology 11 (2020): 626.

LP and AF carried out the implementation of the framework with input from JYD. LP analyzed the data and wrote the manuscript with input from all authors. JYD and EHS conceived the study and were in charge of overall direction and planning.

Chapter 2: Population genomics of *Cercospora beticola* suggests that recent host domestication has not influenced genome evolution

Lizel Potgieter, Alice Feurtey, Kim Hufnagel, Doreen Landermann, Mark McMullan, Melvin D. Bolton, Eva H. Stukenbrock

Conceptualisation: EHS, LP. Collection of fungal isolates: LP, MM, MDB. DNA extraction and clone correction: LP, KH, DL. *In silico* analyses: LP, AF. Preparation and writing of manuscript: LP and EHS.

Chapter 3: Genome content of the *Cercospora* genus predicts host range

Lizel Potgieter, Laura C. Ruppert, Alice Feurtey, Eva H. Stukenbrock

Conceptualisation: LP. Genome sequencing and assembly: LP, AF. Comparative genomics of *Cercospora* genus: LCR, LP. Preparation and writing of manuscript: LP and EHS.

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

Lizel Potgieter

Kiel, 01.03.2021