https://doi.org/10.35662/unine-thesis-2868

Emergence and population genomics of the highly invasive chestnut blight pathogen

A thesis submitted to the
University of Neuchâtel
Faculty of Science - Institute of Biology

for the degree of
DOCTOR OF SCIENCE

presented by

Lea Stauber

PhD jury members
Prof. Dr. Daniel Croll (thesis director)
Dr. Simone Prospero (supervisor)
Dr. Pierre Gladieux (external expert)
Prof. Dr. Niklaus Grünwald (external expert)
Prof. Dr. Jacob Koella (internal expert)

2020

UNIVERSITÉ DE NEUCHÂTEL

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Madame Lea STAUBER

Titre:

"Emergence and population genomics of the highly invasive chestnut blight pathogen"

sur le rapport des membres du jury composé comme suit:

- Prof. ass. Daniel Croll, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Jacob Koella, Université de Neuchâtel, Suisse
- Dr Simone Propero, WSL. Birmensdorf, Suisse
- Dr Pierre Gladieux, INRAE Montpellier, France
- Prof. Nik Grundwald, Oregon State University, USA

Neuchâtel, le 19 janvier 2021

Le Doyen, Prof. A. Bangerter

Table of contents

Summary

Invasive pathogens are a threat to forest and agroecosystems, as well as animal and human health. Identifying genomic determinants of pathogen evolution, as well as investigations into the genetic structure of invasive pathogen populations provide fundamental insights to why species can emerge as invasive pathogens. In this PhD project I investigated the emergence and population genomics of the invasive chestnut blight fungus *Cryphonectria parasitica*, using comparative and population genomic approaches. *C. parasitica* recently emerged as an invasive bark pathogen on non-Asian *Castanea* species in North America and Europe. In the first chapter, I investigated genomic determinants of lifestyle transitions in the genus *Cryphonectria*, by genome comparisons of *C. parasitica* and its sister species. The study uncovered a striking loss of genes associated with carbohydrate metabolism in the invasive pathogen *C. parasitica*, which may have promoted its pathogenicity on *Castanea* species. The second chapter explores the emergence and diversification of a highly invasive chestnut blight lineage across south-eastern Europe. By analyzing the genome-wide diversity of a large set of *C. parasitica* isolates of predominantly European origin, the study showed that a highly successful clonal pathogen lineage can emerge from a recombinant bridgehead population within Europe. Interestingly, the emergence of this clonal lineage was accompanied by an evolutionary transition from mixed mating type populations to single mating type outbreak populations. Lastly, in the third chapter I investigated temporal changes in genetic diversity of established *C. parasitica* populations in southern Switzerland, as well as potential links between the presence of the deleterious hyperparasitic mycovirus Cryphonectria hypovirus 1 (CHV1) and fungal genome-wide diversity. The results indicate increased mating among related fungal individuals, resulting in high genetic similarity of genotypes and facilitated CHV1 transmission. There were no substantial changes in fungal population structure and after ~30 years and no detectable impact of CHV1 presence on fungal genome-wide diversity. Although our results show stable CHV1 incidence in fungal populations over three decades, the short-term interaction dynamics are likely highly volatile. The overall findings of this PhD thesis highlight the relevance of genomic determinants facilitating pathogen emergence and invasions. *C. parasitica* is a useful model to study fundamental questions of pathogen evolution and invasive processes, as well as antagonistic pathogen-hyperparasite interactions.

Keywords: Invasive tree pathogen, *Cryphonectria*, lifestyle evolution, population genomics, pathogen-hyperparasite interactions

Résumé

Les agents pathogènes envahissants constituent une menace pour les forêts et les agroécosystèmes, ainsi que pour la santé animale et humaine. L'identification des déterminants génomiques de l'évolution des agents pathogènes, ainsi que l'analyse de la structure génétique des populations d'agents pathogènes envahissants fournissent des informations fondamentales sur les raisons pour lesquelles des espèces peuvent devenir des agents pathogènes envahissants. Dans ce projet de doctorat, j'ai étudié l'émergence et la génomique des populations de la maladie du chancre de l'écorce du châtaignier qui est provoqué par le champignon envahissant *Cryphonectria parasitica*, en utilisant des approches de génomique comparative et de génomique des populations. *C. parasitica* est récemment apparu comme un pathogène envahissant de l'écorce sur des espèces de *Castanea* non asiatiques en Amérique du Nord et en Europe. Dans le premier chapitre, j'ai étudié les déterminants génomiques des transitions du mode de vie dans le genre *Cryphonectria*, en comparant le génome de *C. parasitica* et de ses espèces sœurs. L'étude a révélé une diminution frappante du nombre de gènes associés au métabolisme des glucides chez le pathogène envahissant *C. parasitica*, qui pourrait avoir favorisé sa pathogénicité sur les espèces de *Castanea*. Le deuxième chapitre explore l'émergence et la diversification d'une lignée de chancre du châtaignier très envahissante dans le sud-est de l'Europe. En analysant la diversité du génome d'un large ensemble d'isolats de *C. parasitica* d'origine principalement européenne, l'étude a montré qu'une lignée clonale envahissante du pathogène peut émerger à partir d'une population recombinante en Europe dite de "tête de pont". Il est intéressant de noter que l'émergence de cette lignée clonale s'est accompagnée d'une transition évolutive d'une reproduction de type sexuée vers une reproduction asexuée au sein de ces populations. Enfin, dans le troisième chapitre, j'ai étudié les changements temporels de la diversité génétique de populations de *C. parasitica* établies dans le sud de la Suisse, ainsi que les liens potentiels entre la présence du mycovirus hyperparasitaire délétère Cryphonectria hypovirus 1 (CHV1) et la diversité du génome fongique. Les résultats indiquent une augmentation du niveau de recombinaison entre individus fongiques apparentés, entraînant une grande similitude génétique des génotypes et facilitant la transmission du mycovirus CHV1. Il n'y a pas eu de changements substantiels dans la structure de cette population fongique du sud de la Suisse, et aucun impact lié à la présence du CHV1 n'a été détecté sur la diversité du génome fongique durant une période d'environ 30 ans. Bien que nos résultats montrent une incidence stable du mycovirus CHV1 au sein des populations fongiques sur trois décennies, les dynamiques d'interaction à court terme sont probablement très volatiles. Les conclusions générales de cette thèse de doctorat soulignent la pertinence des déterminants génomiques qui facilitent l'émergence et les invasions d'agents pathogènes. *C. parasitica* est un modèle utile pour étudier les questions fondamentales de l'évolution des pathogènes et des processus invasifs, ainsi que les interactions antagonistes pathogènes-hyperparasites.

Mots-clés : Pathogène arboricole envahissant, *Cryphonectria*, évolution du mode de vie, génomique des populations, interactions pathogène-hyperparasite

Scientific output

Publications and manuscripts

Stauber L, Prospero S, Croll D. 2020. Comparative genomics analyses of lifestyle transitions at the origin of an invasive fungal pathogen in the genus *Cryphonectria*. *mSphere* 5:e00737-20. https://doi.org/10.1128/mSphere.00737-20.

Stauber L, Badet T, Fleurtey A, Prospero S, Croll D. 2020. Emergence and diversification of a highly invasive chestnut pathogen lineage across south-eastern Europe. *bioRxiv* https://doi.org/10.1101/2020.02.15.950170 (Revisions submitted to eLife)

Stauber L, Croll D, Prospero S. 2020. Analyses of the temporal and spatial genetic structure of the chestnut blight pathosystem. (unpublished manuscript)

Conferences and presentations

Host-microbe genomics meeting Neuchâtel (Switzerland) 28.9.2017
Type of contribution: Flash talk
Title: Studying the invasion history of a widespread clonal chestnut blight (*Cryphonectria parasitica*) strain in south-eastern Europe

IUFRO Phylosphere Diseases Conference Florence (Italy) 6.5.–10.5.2019
Type of contribution: Oral presentation
Title: Emergence and diversification of a highly invasive tree pathogen lineage

Host-microbe genomics meeting Zurich (Switzerland) 24.9.2019
Type of contribution: Oral presentation
Title: Emergence and diversification of a highly invasive tree pathogen lineage

Zurich Mycology Symposium (Switzerland) 24.1.2020
Type of contribution: Oral presentation
Title: Emergence and diversification of a highly invasive tree pathogen lineage

Biology20, Fribourg (Switzerland) 6.2.–7.2.2020

Type of contribution: Poster presentation

Title: Emergence and diversification of a highly invasive tree pathogen lineage

European Conference of Fungal Genetics (ECFG15) Rome (Italy) 17.2. –20.2.2020

Type of contribution: Oral presentation

Title: Emergence and diversification of a highly invasive tree pathogen lineage

Host-microbe genomics meeting Zurich (Switzerland) 11.9.2020

Type of contribution: Oral presentation

Title: Lifestyle transitions in the genus *Cryphonectria*

General introduction

The vast increase in global trade has given rise to biological invasions by non-native species, posing a serious threat to ecosystems, biodiversity, as well as animal and human health (Hobbs, 2000). Prominent examples of invasive species include the Asian Harlequin ladybird beetle (*Harmonia axyridis*) predating the European seven-spot ladybird (*Coccinella septempunctata*) and causing the vast decline of European ladybirds (Ware and Majerus, 2007). Moreover, invasive plants like the common ragweed (*Ambrosia artemisiifolia*) are the cause of severe allergic reactions in humans (Caillaud et al., 2014). Aside from ecological or health related aspects, invasive species my cause immense economic damage, with estimates ranging up to 336 billion US dollars (Marbuah et al., 2014). Thus, for effective management and prevention of new invasions it is essential to investigate factors contributing to the emergence of invasive pathogens, as well as gaining a fundamental understanding of genetic and ecological factors facilitating invasive success.

Traits in pathogen invasions

A key group of invasive species threatening forest ecosystems are fungal pathogens (Desprez-Loustau et al., 2007). The number of introduced invasive forest pathogens has strikingly increased since the second world war (Santini et al., 2013). Reduced transportation time increased the survival of pathogen propagules and their chances for establishment (Hulme, 2009). For successful establishment and spread in new geographic areas, invasive species have to rapidly adapt to novel biotic and abiotic conditions. Some invasive species such as *Seiridium cardinale*, the causal agent of cypress canker, have shown high phenotypic plasticity and rapid genetic adaptation, facilitating its invasive success (Garbelotto et al., 2015). However, invasions are often characterized by severe genetic bottlenecks, caused by low numbers of initial founders, thus restraining genetic adaptation *in situ* (Bock et al., 2015). In the invasive pathogen *Ophiostoma novo-ulmi* causing Dutch elm disease, interspecific hybridization with *O. ulmi* facilitated genetic adaptation and acquisition of fitness associated traits (Hessenauer et al., 2020). However, invasive species might also be preadapted and environmental compatibility could increase establishment success in novel areas (Cadotte et al., 2018; Schlaepfer et al., 2010). Moreover, adapted genotypes might be created through the so-called bridgehead effect, where recombination of genotypes in the area of initial introduction creates adapted lineages, which cause a secondary and more expansive invasion (Lombaert et al., 2010).

Reproductive modes and dispersal abilities have been identified as important factors determining invasion success in invasive forest pathogens (Philibert et al., 2011). Mixed reproductive systems with both sexual and asexual reproduction likely enhance invasive success. Many fungal pathogens can undergo both sexual and asexual reproduction, which not only enhances their adaptive potential, but also allows for establishment of a

few founding individuals in absence of mating partners (Gladieux et al., 2015). The successful invasion *O. novo-ulmi* into Europe was initiated by clones with a single mating type, but introgression of the opposite mating type from *O. ulmi*, subsequently facilitated rapid adaptation and diversification (Paoletti et al., 2006). Reproductive modes can also vary on micro- and macrospatial scales. For instance, on microspatial scale (i.e. on leaf lesions) the invasive *Eucalyptus* spp. leaf pathogen *Teratosphaeria nubilosa* is propagating clonally, but on a larger population-wide scale (i.e. across geographic regions) the species is outcrossing, facilitating population diversification and enhancing its adaptive potential (Perez et al., 2010). Additionally, dispersal abilities are highly linked to reproductive modes. Sexual spores are often highly durable and can be wind dispersed over long distances (Philibert et al., 2011). Contrary, asexual spores are often produced in large quantities but dispersed over short distances (Drenth et al., 2019; Prospero and Cleary, 2017). However, insects or birds might additionally contribute to long distance dispersal of asexual spores and contribute to colonization of new areas (Rigling and Prospero, 2018; Webber, 2000).

Chestnut blight – a prime example for an invasive pathogen

The ascomycete *Cryphonectria parasitica* (Murr.) Barr. is a major invasive tree pathogen on non-Asian *Castanea* (Fagaceae) species. Native to East Asia, the pathogen vastly decimated the American chestnut (*Ca. dentata*) causing the decline of a billion-dollar timber industry (Holmes et al., 2009) and the disease continues to threaten European chestnut (*Ca. sativa*) stands. On its native hosts the Chinese (*Ca. mollissima*) and Japanese chestnut (*Ca. crenata*), *C. parasitica* only appears to be mildly pathogenic which has been attributed to host-pathogen co-evolution (Rigling and Prospero, 2018). Additionally, *C. parasitica* has also shown incidental, largely non-pathogenic growth on other hosts in the native and invasive range, including oaks (*Quercus* spp.), maples (*Acer* spp.) or the European hornbeam (*Carpinus betulus*) (Jiang et al., 2019; Rigling and Prospero, 2018). *Castanea* spp., *Quercus* app., *Acer* spp., as well as *Car. betulus* are also considered as hosts for other *Cryphonectria* species (Bragança et al., 2011; Myburg et al., 2004, C. Cornejo pers. comm.). The saprophyte *C. radicalis* has been reported on *Castanea* and *Quercus* species in Europe, North America, but also in Japan (Hoegger et al., 2002; Myburg et al., 2004). However, rare reports of this species suggest that it may have been displaced in North America and Europe after the introduction of the pathogenic sister species *C. parasitica* (Hoegger et al., 2002). *C. japonica* has been found on Asian *Castanea* and *Quercus* species and has been associated with causing bark cankers (Jiang et al., 2019; Liu et al., 2003; Myburg et al., 2004). Likewise, *C. naterciae* has been reported as a weak pathogen on European *Castanea* and *Quercus* species (Bragança et al., 2011). Moreover, the recently described species *C. carpinicola* has been isolated from declining European hornbeams in Austria, Georgia, Italy, and Switzerland (C. Cornejo, pers. comm.). Other severe bark pathogens in the family of Cryphonectriaceae include pathogens in the genus *Chrysoporthe*, affecting *Eucalyptus* spp. species in the southern hemisphere (Gryzenhout et al., 2004). However, within the genus *Cryphonectria,* only

C. parasitica has emerged as a severe pathogen, raising important questions on the evolution of pathogenicity and underlying genomic adaptations.

As a necrotrophic bark pathogen, *C. parasitica* only affects the above-ground parts of trees such as stems, branches and twigs. Although mainly infecting living trees, the fungus also survives and grows on the bark of fresh dead wood and can infect abandoned galls of the chestnut gall wasp (*Dryocosmus kuriphilus*) (Meyer et al., 2015; Prospero et al., 2006). The pathogen enters through bark wounds and starts growing in and under the bark. Perennial necrotic lesions (so-called cankers) form on the bark, which may develop over years before causing mortality. In virulent cankers the cambium is killed and the bark sink inwards (Anagnostakis, 2012). Once the canker has girdled the entire stem or branch leaves on branches distal to the canker start wilting and remain hanging on the tree, while several epicormic shoots (i.e. side shoots) can form below the canker (Rigling and Prospero, 2018). The pathogen produces pale brown mycelial fans that grow between bark cells and cambium. This causes splitting of tree cells and induces the host to produce lignified cell walls and wound periderm (Rigling and Prospero, 2018). The mycelial fans can penetrate these structures by killing the cells with toxins and cell wall degrading enzymes (Rigling and Prospero, 2018). On the bark surface, sexual (perithecia) and asexual (pycnidia) fruiting bodies may be produced, which are embedded in an orange to reddish-brownish stomata (Anagnostakis, 2012; Rigling and Prospero, 2018). Spore release is promoted under moist conditions. Asexual conidiospores (conidia) spread mostly over short distances via splash-dispersal, as well as birds and insects. Sexual ascospores are actively ejected from the sexual fruiting body (perithecia) and tend to be wind-dispersed (Anagnostakis, 2012; Griffin, 1986). Both spore types can cause new infection on susceptible hosts.

Sexual reproduction in *C. parasitica* controlled by a mating type locus with two alleles (MAT-1 and MAT-2), as found in many other ascomycetes (Peters et al., 2012). Mating can only occur if individuals have opposite alleles (idiomorphs) at the mating type locus (Marra and Milgroom, 2001). In absence of the opposite mating type the fungus can propagate asexually (*i.e.* clonally), facilitating spread into new areas without available mating partners (Milgroom et al., 2008; Rigling and Prospero, 2018). However, some individuals are self-fertile by harboring both mating types, facilitating sexual reproduction and ascospore formation in populations with only one dominating mating type (Bragança et al., 2007; Hoegger et al., 2000).

The importance of vegetative incompatibility in *Cryphonectria parasitica*

Vegetative incompatibility is a self/non-self recognition system (allorecognition) of many filamentous fungi (Glass et al., 2000). It prevents hyphal fusion of two individuals through programmed cell death, thereby limiting the exchange of deleterious cytoplasmic material, including parasitic mycoviruses. In *C. parasitica* vegetative incompatibility is genetically controlled by at least six unlinked, biallelic vegetative incompatibility (*vic*) loci, with either of two alleles (Cortesi and Milgroom, 1998). Only when two individuals have the same

alleles at all *vic* loci, they are vegetatively compatible and individuals can fuse through hyphal anastomosis (Cortesi and Milgroom, 1998; Peters et al., 2012). A total of 64 vegetative compatibility (vc) types (2^6 = 64 *vic* genotypes) have been identified and a subsequent collection of European vc type tester strains (EU-01 to EU-64) has enabled research on global vc type diversity (Cortesi and Milgroom, 1998; Rigling and Prospero, 2018). However, vc types that are incompatible with all 64 EU testers have been found in Asia, the USA, as well as Europe, suggesting that vegetative incompatibility might be controlled by more than six *vic* loci (Bragança et al., 2007; Cortesi and Milgroom, 1998; Rigling and Prospero, 2018). Alternatively, additional undiscovered alleles at known *vic* loci might present (Rigling and Prospero, 2018). Identification of vc types can easily be assayed, by pairing an unknown strain with a known vc type tester strain on agar plates (Anagnostakis, 1988). When strains are compatible they merge into a single culture, whilst incompatible strains form a visible barrage. However, this method is time consuming and the merging/barrage response is not always clearly identifiable. Therefore, Cornejo et al. (2019) recently developed a molecular approach for identifying vc types in a multiplex PCR. This new method allows for rapid identification of compatibility types of a large number of isolates.

The vegetative compatibility system in *C. parasitica* has a crucial influence on the success of biocontrol of chestnut blight. In the 1950s, approximately 15 years after the first official report of chestnut blight in Europe, healing chestnut blight cankers were first observed in Italy and France (Heiniger and Rigling, 1994). This specific canker phenotype was later discovered to be caused by a hypovirulence inducing mycovirus of the family Hypoviridae (Choi and Nuss, 1992). The virus consists of a protein coat-free double stranded RNA molecule, which is associated with fungal membrane vesicles (Fahima et al., 1993). Currently, there are four known *Cryphonectria* hypovirus species, namely CHV-1, CHV-2, CHV-3 CHV-4, of which CHV-1 is efficiently used as a biocontrol agent against chestnut blight in Europe (Milgroom and Cortesi, 2004; Rigling and Prospero, 2018). CHV-1 infected fungal strains only cause superficial cankers growth which generally do not kill the tree and mycelial fan formation is significantly reduced (Heiniger and Rigling, 1994). Furthermore, hypovirulent fungal strains show reduced sporulation and white pigmentation (instead of orange pigmentation in virus-free strains) when grown on PDA medium (Heiniger and Rigling, 1994). The hypovirus can be transmitted horizontally via hyphal anastomosis and transmission is consequently inhibited by vegetative incompatibility barriers, *i.e.* virus spread is drastically reduced between individuals of different vc types (Liu and Milgroom, 1996; Papazova-Anakieva et al., 2008). Virus transmission can also occur vertically through asexual conidiospores, while sexual ascospore are virus-free (Prospero et al., 2006). In Europe, where vc type diversity is comparatively low, natural and artificially introduced hypovirulence has mitigated the detrimental effects of chestnut blight (Anagnostakis, 1982). Contrarily, in North America restauration efforts of the American chestnut using hypovirulence have largely failed due to high vc type diversity limiting virus dissemination, as well as higher susceptibility of the American chestnut towards the pathogen (Anagnostakis, 2012; Rigling and Prospero, 2018). To efficiently control chestnut blight, scientists in the USA have started

resistance breeding programs, through backcrosses of blight resistance from the Chinese chestnut into the genome of the American chestnut (Diskin et al., 2006).

Vegetative compatibility markers have in the past been widely used to study the diversity of *C. parasitica* in its native and introduced range (Bragança et al., 2007; Cortesi and Milgroom, 1998; Milgroom et al., 2008). Native Asian *C. parasitica* populations are highly diverse and consist of vc types that are largely incompatible with all EU tester strains (Liu and Milgroom, 2007). Furthermore, vc type diversity in the invasive US populations is much higher than in Europe, though European vc type diversity is highly variable between countries. Countries with high diversity include Spain, France, Italy and southern Switzerland (Robin and Heiniger, 2001), i.e. regions of first introduction of the pathogen. Contrary, populations at the disease front (e.g. Macedonia, Greece and Turkey) show rather low diversity (Milgroom et al., 2008; Sotirovski et al., 2004). Vc type diversity can also vary within a country, as exemplified by Italy where higher diversity is found in the North compared to southern Italy (Robin and Heiniger, 2001). Generally, low vc type diversity is found in areas where *C. parasitica* has only recently been introduced or when no perithecia are found, indicating an asexual spread of the pathogen (Robin and Heiniger, 2001). However, when sexual reproduction occurs, polymorphic *vic* loci are recombined and vc type diversity may increase (Cortesi and Milgroom, 1998). Interestingly, vc type diversity of sexually reproducing populations in Europe has consistently remained lower than expected under random mating (Ježić et al., 2018, Ježić et al. 2020, submitted). For thorough population genetic analyses, vc type diversity provides limited information. Some individuals may have the same vc type, but genotypic analyses can display significant differences (Ježić et al., 2012). Therefore, tools such as sequence-characterized amplified regions (SCAR) or microsatellite markers have been used complementarily for characterizing the structure of *C. parasitica* populations (Milgroom et al., 2008; Prospero and Rigling, 2012). However, the markers lack the necessary resolution to investigate the invasion history and pathways of spread of asexually (i.e. clonally) reproducing populations.

Invasion history of *Cryphonectria parasitica*

Cryphonectria parasitica is native to Asia, but was first officially described in the United States where it was discovered in 1904 in the Zoological Park of New York City on the American chestnut (*Ca. dentata*) (Rigling and Prospero, 2018). The pathogen was likely accidentally introduced to the US with imported chestnut plants from Japan (Anagnostakis, 2012; Dutech et al., 2012). In the following years the disease rapidly spread over eastern North America, which led to the fast decimation of *Ca. dentata*, eliminating its dominance as a canopy tree species in local forests (Elliott and Swank, 2008). In 1938, the fungus was first detected in Europe, near Genoa (Italy) on the European chestnut *Ca. sativa*. Successively, the pathogen spread further to France,

Switzerland and Slovenia, as well as other regions in eastern and south-eastern Europe and can now be found in all major European chestnut growing areas (Rigling and Prospero, 2018).

Compared to North America, European chestnut blight populations display relatively low genetic diversity (Milgroom and Cortesi, 2004). According to Dutech et al. (2012), several distinct introductions of *C. parasitica* have taken place in Europe. The invasion of south-eastern France, Switzerland and Italy predominantly occurred through North American genotypes, while western Europe was largely colonized by Asian lineages (Dutech et al., 2012). The invasion of chestnut blight in Europe has particularly been characterized by the expansion of a few highly successful clonal lineages (Demené et al., 2019; Milgroom et al., 2008). Although both mating types are found in south-western Europe and multiple introductions generally imply an increase of genetic diversity, clonal population structure is still a commonly observed (Dutech et al., 2010). Even more strikingly, south-eastern Europe the seems to have predominantly invaded by a single clonal lineage (Dutech et al., 2012; Milgroom et al., 2008). However, the origin of clonal outbreak lineages, as well as potential adaptive genetic processes behind the emergence of successful invasive lineages are largely unknown.

Main objectives of this thesis

This doctoral thesis aims to identify genomic determinants in the evolution of pathogenic lifestyles and resolve population genomic structures of an invasive fungal pathogen.

***Chapter 1. Comparative genomics analyses of lifestyle transitions at the origin of an invasive fungal pathogen in the genus* Cryphonectria**

Pathogen emergence is associated with evolving lifestyle adaptations, which are enabled by specific changes in the gene content of a species. How pathogens evolve from nonpathogenic ancestors is still poorly understood, making the prediction of future outbreaks challenging. Additionally, host shifts or geographic range expansions frequently promote emergence of new pathogens (Giraud et al., 2010). Numerous species have recently have emerged as invasive pathogens, raising the question whether genomic features contribute to pathogenic lifestyle transitions of formerly unnoticed pathogens. In this chapter I investigated whether genomic determinants facilitated the emergence of a highly pathogenic invasive species, by comparing it with genomes of sister species.

Chapter 2: Emergence and diversification of a highly invasive chestnut pathogen lineage across south-eastern Europe

Invasions are often dominated by one or a small number of genotypes, resulting in invasive populations of low genetic diversity, which reduces adaptive genetic variation. However, many species invasions were successful

despite low genetic diversity within founding populations (Fontaine et al., 2013; Raboin et al., 2007). Population genomics can give valuable insight into dynamics shaping diversity of invasive populations, as well as the origin of successful invasive lineages. In this chapter, I investigated the emergence and diversification of a highly invasive clonal pathogen lineage. Using whole-genome sequencing data of a global sample set of the invasive chestnut blight fungus *C. parasitica*, the study explored the most likely origin of a highly successful clonal chestnut blight lineage in south-eastern Europe.

Chapter 3: Analyses of the temporal and spatial genetic structure of the chestnut blight pathosystem

Co-evolution in pathosystems determines disease dynamics through selection in pathogen virulence and host resistance. Yet, co-evolutionary trajectories might additionally be shaped by hyperparasites which infect the pathogens themselves, thereby influencing stability of both host and pathogen populations. In this chapter I investigated whether presence of a deleterious hyperparasitic mycovirus shapes the spatial and temporal population structure of its fungal pathogen host. For this, three local *C. parasitica* populations in southern Switzerland were sampled at two time points approximately 30 years apart. Genomes of all fungal isolates were sequenced and presence of the hyperparasitic mycovirus Cryphonectria hypovirus 1 (CHV1) in fungal cultures was determined with RT-PCR. Additionally, spatial distribution of fungal genotypes and CHV1 in 2019 was analyzed, by recording the GPS location of sampled trees in all three populations.

References

Anagnostakis SL. 2012. Chestnut breeding in the United States for disease and insect resistance. *Plant Dis* **96**:1392–1403.

Anagnostakis SL. 1988. *Cryphonectria parasitica*, Cause of Chestnut Blight, Advances in Plant Pathology. Academic Press Limited. doi:10.1016/B978-0-12-033706-4.50011-6

Anagnostakis SL. 1982. Biological control of chestnut blight. *Science (80-)* **215**:466–471. doi:10.1126/science.215.4532.466

Bock DG, Caseys C, Cousens RD, Hahn MA, Heredia SM, Hübner S, Turner KG, Whitney KD, Rieseberg LH. 2015. What we still don't know about invasion genetics. *Mol Ecol* **24**:2277–2297.

Bragança H, Rigling D, Diogo E, Capelo J, Phillips A, Tenreiro R. 2011. *Cryphonectria naterciae*: A new species in the *Cryphonectria-Endothia* complex and diagnostic molecular markers based on microsatellite-primed PCR. *Fungal Biol* **115**:852–861.

Bragança H, Simões S, Onofre N, Tenreiro R, Rigling D. 2007. *Cryphonectria parasitica* in Portugal: diversity of vegetative compatibility types, mating types, and occurrence of hypovirulence. *For Pathol* **37**:391–402.

Cadotte MW, Campbell SE, Li S, Sodhi DS, Mandrak NE. 2018. Preadaptation and naturalization of nonnative species: Darwin's two fundamental insights into species invasion. *Annu Rev Plant Biol* **69**:661–684.

Caillaud D, Thibaudon M, Martin S, Ségala C, Besancenot JP, Clot B, François H. 2014. Short-term effects of airborne ragweed pollen on clinical symptoms of hay fever in a panel of 30 patients. *J Investig Allergol Clin Immunol* **24**:249–256.

Choi GH, Nuss DL. 1992. Hypovirulence of chestnut blight fungus conferred by an infectious viral cDNA. *Science (80-)* **257**:800–803.

Cornejo C, Šever B, Kupper Q, Prospero S, Rigling D. 2019. A multiplexed genotyping assay to determine vegetative incompatibility and mating type in *Cryphonectria parasitica*. *Eur J Plant Pathol* 1–11.

Cortesi P, Milgroom MG. 1998. Genetics of Vegetative Incompatibility in *Cryphonectria parasitica*. *Appl Environ Microbiol* **64**:2988–2994.

Demené A, Legrand L, Gouzy J, Debuchy R, Saint-Jean G, Fabreguettes O, Dutech C. 2019. Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of *Cryphonectria parasitica* in western Europe. *Fungal Genet Biol*.

Desprez-Loustau ML, Robin C, Buée M, Courtecuisse R, Garbaye J, Suffert F, Sache I, Rizzo DM. 2007. The fungal dimension of biological invasions. *Trends Ecol Evol* **22**:472–480. doi:10.1016/j.tree.2007.04.005

Diskin M, Steiner KC, Hebard F V. 2006. Recovery of American chestnut characteristics following hybridization and backcross breeding to restore blight-ravaged *Castanea dentata*. *For Ecol Manage* **223**:439–447.

Drenth A, McTaggart AR, Wingfield BD. 2019. Fungal clones win the battle, but recombination wins the war. *IMA Fungus* **10**:18.

Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V. 2012. The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Mol Ecol* **21**:3931–3946.

Dutech C, Fabreguettes O, Capdevielle X, Robin C. 2010. Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity (Edinb)* **105**:220–228. doi:10.1038/hdy.2009.164

Elliott KJ, Swank WT. 2008. Long-term changes in forest composition and diversity following early logging (1919-1923) and the decline of American chestnut (*Castanea dentata*). *Plant Ecol* **197**:155–172.

Fahima T, Kazmierczak P, Hansen DR, Pfeiffer P, Van Alfen NK. 1993. Membrane-associated replication of an unencapsidated double-strand RNA of the fungus, *Cryphonectria parasitica*. *Virology* **195**:81–89.

Fontaine MC, Gladieux P, Hood ME, Giraud T. 2013. History of the invasion of the anther smut pathogen on *Silene latifolia* in North America. *New Phytol* **198**:946–956.

Garbelotto M, Rocca G Della, Osmundson T, di Lonardo V, Danti R. 2015. An increase in transmission-related traits and in phenotypic plasticity is documented during a fungal invasion. *Ecosphere* **6**:1–16.

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

Gladieux P, Feurtey A, Hood ME, Snirc A, Clavel J, Dutech C, Roy M, Giraud T. 2015. The population biology of fungal invasions. *Mol Ecol* **24**:1969–1986.

Glass NL, Jacobson DJ, Shiu PKT. 2000. The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi. *Annu Rev Genet* **34**:165–186.

Griffin GJ. 1986. Chestnut blight and its control. *Hortic Rev (Am Soc Hortic Sci)* **8**:291–336.
Gryzenhout M, Myburg H, der Merwe NA, Wingfield BD, Wingfield MJ. 2004. *Chrysoporthe*, a new genus to accommodate *Cryphonectria cubensis*. *Stud Mycol* **50**:119–142.
Heiniger U, Rigling D. 1994. Biological control of chestnut blight in Europe. *Annu Rev Phytopathol* **32**:581–599.
Hessenauer P, Fijarczyk A, Martin H, Prunier J, Charron G, Chapuis J, Bernier L, Tanguay P, Hamelin RC, Landry CR. 2020. Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. *Nat Ecol Evol* **4**:626–638.
Hobbs RJ. 2000. Invasive species in a changing world. Island Press.
Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. 2002. *Cryphonectria radicalis*: rediscovery of a lost fungus. *Mycologia* **94**:105–115.
Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. 2000. Genetic structure of newly established populations of *Cryphonectria parasitica*. *Mycol Res* **104**:1108–1116.
Holmes TP, Aukema JE, Von Holle B, Liebhold A, Sills E. 2009. Economic impacts of invasive species in forest past, present, and future. *Year Ecol Conserv Biol 2009 Ann NY Acad Sci 1162 18-38* **1162**:18–38.
Hulme PE. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J Appl Ecol* **46**:10–18.
Ježić M, Krstin L, Rigling D, Ćurković-Perica M. 2012. High diversity in populations of the introduced plant pathogen, *Cryphonectria parasitica*, due to encounters between genetically divergent genotypes. *Mol Ecol* **21**:87–99.
Ježić M, Mlinarec J, Vuković R, Katanić Z, Krstin L, Nuskern L, Poljak I, Idžojtić M, Tkalec M, Ćurković-Perica M. 2018. Changes in *Cryphonectria parasitica* populations affect natural biological control of chestnut blight. *Phytopathology* **108**:870–877. doi:10.1094/PHYTO-07-17-0252-R
Jiang N, Fan XL, Tian CM. 2019. Identification and pathogenicity of Cryphonectriaceae species associated with chestnut canker in China. *Plant Pathol* **68**:1132–1145.
Liu Y-C, Linder-Basso D, Hillman BI, Kaneko S, Milgroom MG. 2003. Evidence for interspecies transmission of viruses in natural populations of filamentous fungi in the genus *Cryphonectria*. *Mol Ecol* **12**:1619–1628.
Liu Y-C, Milgroom MG. 2007. High diversity of vegetative compatibility types in *Cryphonectria parasitica* in Japan and China. *Mycologia* **99**:279–284.
Liu YC, Milgroom MG. 1996. Correlation Between Hypovirus Transmission and the Number of Vegetative Incompatibility (*vic*) Genes Different Among Isolates from a Natural Population of *Cryphonectria parasitica*. *Phytopathology* **86**:79–86.
Lombaert E, Guillemaud T, Cornuet J-M, Malausa T, Facon B, Estoup A. 2010. Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS One* **5**:1–7.
Marbuah G, Gren I-M, McKie B. 2014. Economics of harmful invasive species: a review. *Diversity* **6**:500–523.
Marra RE, Milgroom MG. 2001. The mating system of the fungus *Cryphonectria parasitica*: selfing and self-incompatibility. *Heredity (Edinb)* **86**:134–143.
Meyer JB, Gallien L, Prospero S. 2015. Interaction between two invasive organisms on the European chestnut: does the chestnut blight fungus benefit from the presence of the gall wasp? *FEMS Microbiol Ecol* **91**:fiv122. doi:/10.1093/femsec/fiv122
Milgroom MG, Cortesi P. 2004. Biological control of chestnut blight with hypovirulence: a critical analysis. *Annu Rev Phytopathol* **42**:311–338.
Milgroom MG, Sotirovski K, Spica D, Davis JE, Brewer MT, Milev M, Cortesi P. 2008. Clonal population structure of the chestnut blight fungus in expanding ranges in southeastern Europe. *Mol Ecol* **17**:4446–4458. doi:10.1111/j.1365-294X.2008.03927.x
Myburg H, Gryzenhout M, Wingfield BD, Milgroom MG, Kaneko S, Wingfield MJ. 2004. DNA sequence data and morphology define *Cryphonectria* species in Europe, China, and Japan. *Can J Bot* **82**:1730–1743.
Paoletti M, Buck KW, Brasier CM. 2006. Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote *Ophiostoma novo-ulmi*. *Mol Ecol* **15**:249–262.
Papazova-Anakieva I, Sotirovski K, Cortesi P, Milgroom MG. 2008. Horizontal transmission of hypoviruses between vegetative compatibility types of *Cryphonectria parasitica* in Macedonia. *Eur J Plant Pathol* **120**:35–42.
Perez G, Slippers B, Wingfield BD, Hunter GC, Wingfield MJ. 2010. Micro-and macrospatial scale analyses illustrates mixed mating strategies and extensive geneflow in populations of an invasive haploid pathogen. *Mol Ecol* **19**:1801–1813.
Peters FS, Holweg CL, Rigling D, Metzler B. 2012. Chestnut blight in south-western Germany: multiple

introductions of *Cryphonectria parasitica* and slow hypovirus spread. *For Pathol* **42**:397–404.
Philibert A, Desprez-Loustau M-L, Fabre B, Frey P, Halkett F, Husson C, Lung-Escarmant B, Marçais B, Robin C, Vacher C, others. 2011. Predicting invasion success of forest pathogenic fungi from species traits. *J Appl Ecol* **48**:1381–1390.
Prospero S, Cleary M. 2017. Effects of host variability on the spread of invasive forest diseases. *Forests* **8**:80.
Prospero S, Conedera M, Heiniger U, Rigling D. 2006. Saprophytic activity and sporulation of *Cryphonectria parasitica* on dead chestnut wood in forests with naturally established hypovirulence. *Phytopathology* **96**:1337–1344.
Prospero S, Rigling D. 2012. Invasion genetics of the chestnut blight fungus *Cryphonectria parasitica* in Switzerland. *Phytopathology* **102**:73–82. doi:10.1094/PHYTO-02-11-0055
Raboin L-M, Selvi A, Oliveira KM, Paulet F, Calatayud C, Zapater M-F, Brottier P, Luzaran R, Garsmeur O, Carlier J, others. 2007. Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*. *Fungal Genet Biol* **44**:64–76.
Rigling D, Prospero S. 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: Invasion history, population biology and disease control. *Mol Plant Pathol* **19**:7–20. doi:10.1111/mpp.12542
Robin C, Heiniger U. 2001. Chestnut blight in Europe: diversity of *Cryphonectria parasitica*, hypovirulence and biocontrol. *For Snow Landsc Res* **76**:361–367.
Santini A, Ghelardini L, De Pace C, Desprez-Loustau M-L, Capretti P, Chandelier A, Cech T, Chira D, Diamandis S, Gaitniekis T, others. 2013. Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytol* **197**:238–250.
Schlaepfer DR, Glättli M, Fischer M, van Kleunen M. 2010. A multi-species experiment in their native range indicates pre-adaptation of invasive alien plant species. *New Phytol* **185**:1087–1099.
Sotirovski K, Papazova-Anakieva I, Grünwald NJ, Milgroom MG. 2004. Low diversity of vegetative compatibility types and mating type of *Cryphonectria parasitica* in the southern Balkans. *Plant Pathol* **53**:325–333.
Ware RL, Majerus MEN. 2007. Intraguild predation of immature stages of British and Japanese coccinellids by the invasive ladybird *Harmonia axyridis*. From Biological Control to Invasion: The Ladybird *Harmonia axyridis* as a Model Species. Springer. pp. 169–188.
Webber JF. 2000. Insect vector behavior and the evolution of Dutch elm diseaseThe Elms. Springer. pp. 47–60.

Chapter 1

Comparative genomics analyses of lifestyle transitions at the origin of an invasive fungal pathogen in the genus *Cryphonectria*

Lea Stauber[1,2], Simone Prospero[1], Daniel Croll[2]

[1] Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

[2] Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel

Published in mSphere 5.5 (2020). doi: https://doi.org/10.1128/mSphere.00737-20

Abstract

Emerging fungal pathogens are a threat to forest and agroecosystems, as well as animal and human health. How pathogens evolve from non-pathogenic ancestors is still poorly understood making the prediction of future outbreaks challenging. Most pathogens have evolved lifestyle adaptations, which were enabled by specific changes in the gene content of the species. Hence, understanding transitions in the functions encoded by genomes gives valuable insight into the evolution of pathogenicity. Here, we studied lifestyle evolution in the genus *Cryphonectria*, including the prominent invasive pathogen *C. parasitica*, the causal agent of chestnut blight on *Castanea* species. We assembled and compared the genomes of pathogenic and putatively non-pathogenic *Cryphonectria* species, as well as sister group pathogens in the family Cryphonectriaceae (Diaporthales, Ascomycetes) to investigate the evolution of genome size and gene content. We found a striking loss of genes associated with carbohydrate metabolism (CAZymes) in *C. parasitica* compared to other Cryphonectriaceae. Despite substantial CAZyme gene loss, experimental data suggests that *C. parasitica* has retained wood colonization abilities shared with other *Cryphonectria* species. Putative effectors substantially varied in number, cysteine content and protein length among species. In contrast, secondary metabolite gene clusters show a high degree of conservation within the genus. Overall, our results underpin the recent lifestyle transition of *C. parasitica* towards a more pathogenic lifestyle. Our findings suggest that a CAZyme loss may have promoted pathogenicity of *C. parasitica* on chestnuts. Analyzing gene complements underlying key nutrition modes can facilitate the detection of species with the potential to emerge as pathogens.

Importance

Forest and agroecosystems, as well as animal and human health are threatened by emerging pathogens. Following decimation of chestnuts in the US, the fungal pathogen *Cryphonectria parasitca* colonized Europe. After establishment, the pathogen population gave rise to a highly successful lineage that spread rapidly across the continent. Core to our understanding what makes a successful pathogen is the genetic repertoire enabling the colonization and exploitation of host species. Here, we have assembled >100 genomes across two related genera to identify key genomic determinants leading to the emergence of chestnut blight. We found subtle yet highly specific changes in the transition from saprotrophy to latent pathogenicity mostly determined by enzymes involved in carbohydrate metabolism. Large-scale genomic analyses of genes underlying key nutrition modes can facilitate the detection of species with the potential to emerge as pathogens.

Introduction

Across the fungal kingdom, species have evolved the ability to persist either as symbionts, commensals or pathogens on a wide range of living insect, animal and plant hosts. This variety of fungal lifestyles requires complex adaptations encoded in the genome. Lifestyle-associated adaptations have been of particular interest as pathogen emergence is frequently associated with a significant gain in virulence of a formerly weak pathogen (Giraud et al., 2010). This has been shown for *Pyrenophora tritici-repentis*, a former saprophyte or weak pathogen on grass species including wheat, which became highly pathogenic on wheat through acquisition of the virulence gene *ToxA* from the wheat pathogen *Stagonospora nodorum* (Friesen et al., 2006). Moreover, pathogen emergence can be promoted through host jumps or geographic range expansions (Byrnes III et al., 2010) or complete host shifts (Giraud et al., 2010). Such host shifts can occur across kingdoms, as shown for insect pathogens from the genus *Metarhizium*, which likely evolved from plant endophytes or pathogens (Moonjely et al., 2016). Interestingly, phylogenomic analyses have shown that pathogens can emerge repeatedly within fungal clades such as *Dothideomycetes* or even at the genus level (*e.g. Aspergillus*) (Haridas et al., 2020; Rokas et al., 2020). Hence, many pathogenic fungi have non-pathogenic ancestors. This suggests that the emergence and evolution of pathogenic lifestyles is coupled with the acquisition of specific traits distinct from non-pathogenic relatives.

To be successful, pathogens must overcome physical and chemical barriers deployed by the host (Hardham, 2001). Plant pathogenic fungi have evolved specific lifestyles (*i.e.* biotrophy, hemi-biotrophy, necrotrophy) to exploit the host and each lifestyle requires distinct sets of genes (De Wit et al., 2012; O'Connell et al., 2012; Spanu et al., 2010; Stergiopoulos and de Wit, 2009). The gene repertoire of pathogens evolved through gene gains or losses, proliferation of transposable elements, as well as expansions or contractions of entire gene families, sometimes resulting in increased genome sizes, compared to related non-pathogenic species (Raffaele and Kamoun, 2012; Sánchez-Vallet et al., 2018). Gene families notably associated with fungal plant pathogenicity include enzymes for cell-wall degradation, small secreted proteins (*i.e.* effectors) and secondary metabolite gene clusters (Chooi and Solomon, 2014; De Jonge et al., 2011; Gaulin et al., 2018; Lelwala et al., 2019; Lyu et al., 2015; Macheleidt et al., 2016). Cell walls are an important physical barrier against pathogens but can be broken down and used as carbon sources by a variety of fungi. Carbohydrate active enzymes (CAZymes) specific for cellulose, hemi-cellulose or pectin degradation, are typically classified into the superfamilies of glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), as well as enzymes with auxiliary activities (AAs) and carbohydrate-binding modules (CBMs) (Kubicek et al., 2014). The types and number of CAZyme encoding genes varies among species and likely reflects adaptation to different nutritional niches (Glass et al., 2013). Most notably, necrotrophic pathogens tend to deploy cell-wall degrading enzymes to promote host damage and colonization (Rodriguez-Moreno et al., 2018). In contrast, biotrophic pathogens tend to have fewer enzymes involved in cell-

wall degradation (Kubicek et al., 2014; Rodriguez-Moreno et al., 2018). Saprotrophic fungi feeding on decaying plant matter often show an overall reduced CAZyme complement compared to necrotrophic fungi (Hane et al., 2020), but specific expansions in CAZymes related to cellulose degradation (Ohm et al., 2012).

The emergence of pathogenic lifestyles has often required the ability to secrete effector proteins and secondary metabolites during contact with the host. Effectors are characterized as quickly evolving small, cysteine-rich secreted proteins, which are produced to manipulate plant host immune responses (Stukenbrock et al., 2011; Wang and Wang, 2017). Biotrophic and hemi-biotrophic pathogens secrete effector proteins to suppress host immunity and manipulate host cell physiology (Lo Presti et al., 2015). Necrotrophs deploy effectors also as host-specific toxins (Lo Presti et al., 2015; Tan et al., 2010). However, small secreted proteins resembling effectors are also expressed by saprophytic fungi and may be involved in degradative processes (Feldman et al., 2017). Virulence factors in pathogenic fungi can also include secondary metabolites, which are often low molecular weight compounds not essential for fungal growth. Polyketides, non-ribosomal peptides, terpenes and indole alkaloids are the main bioactive compounds acting as cytotoxins, antimicrobials or enzyme inhibitors (Keller et al., 2005). Genes underlying secondary metabolite biosynthesis pathways are often clustered in the genome (Nielsen et al., 2017). Secondary metabolites are produced by fungi of various lifestyles but may be more relevant virulence factors for necrotrophs, while biotrophs tend to lose the underlying genes (Spanu et al., 2010). Beyond pathogenicity-related functions, saprophytic or endophytic fungi produce secondary metabolites with important antimicrobial activity (Santos et al., 2017; Schalchli et al., 2011).

The family Cryphonectriaceae (Diaporthales, Ascomycetes) includes mainly bark-inhabiting species ranging from weak to severe pathogens (Gryzenhout et al., 2006; Jiang et al., 2020). The most aggressive pathogens include *Chrysoporthe* species affecting hosts in the order Myrtales (*e.g. Eucalyptus* spp.), as well as *Cryphonectria parasitica* (Murr.) Barr., the causal agent of chestnut blight on *Castanea* (Fagaceae) species (Gryzenhout et al., 2004; Rigling and Prospero, 2018). *C. parasitica* is native to East Asia (*i.e.* China, Korea, Japan) where it occurs as a weak pathogen on Chinese (*Ca. mollissima* Blume) and Japanese (*Ca. crenata* Siebold & Zucc.) chestnuts. However, *C. parasitica* was first described after its discovery in 1904 on American chestnut (*Ca. dentata* (Marsh.) Borkh.) in the United States (Rigling and Prospero, 2018). The rapid spread of the pathogen following its introduction resulted in the ecological extinction of *Ca. dentata* throughout its native distribution range in North America (Elliott and Swank, 2008). In Europe, chestnut blight was first observed in the 1930s and is nowadays present in all major chestnut growing areas (Rigling and Prospero, 2018). Following the colonization of Europe, *C. parasitica* has rapidly spread through most of southeastern Europe driven by the emergence of a highly successful lineage (Stauber et al., 2020). The invasion success likely stems from the establishment of a highly diverse European bridgehead population and a switch to asexual reproduction (Stauber et al., 2020). Besides host species in the genus *Castanea*, *C. parasitica* has been occasionally reported on oaks (*Quercus* spp.), maples (*Acer* spp.) and European hornbeam (*Carpinus betulus* L.) (Rigling and Prospero, 2018).

Both in the native and the invasive range, *C. parasitica* has closely related sister species, which are considered weak pathogens or saprophytes (Dennert et al., 2020). Among these, *C. japonica* Tak. Kobay. & Kaz. Itô (previously named *C. nitschkei*) was isolated from *Ca. crenata* in Japan (Liu et al., 2003; Myburg et al., 2004) and from oaks in China, on which it causes bark cankers (Jiang et al., 2019). The European species *C. naterciae* M.H. Bragança (syn. *C. decipiens*, (Gryzenhout et al., 2009)) was isolated from *Ca. sativa* and *Quercus* spp. in Portugal, Sardinia and Algeria (Bragança et al., 2011; Pinna et al., 2019; Smahi et al., 2018). Inoculation experiments showed that both *C. japonica* and *C. naterciae* are significantly less virulent on *Ca. sativa*, *Quercus robur* L. and *Fagus sylvatica* L. than *C. parasitica* (Dennert et al., 2020; Jiang et al., 2019). Two other *Cryphonectria* species occurring in Europe are *C. radicalis* and *C. carpinicola*. The former is also present in North America and considered to be a saprophyte on dead wood of *Castanea* and *Quercus* species (Hoegger et al., 2002). Interestingly, the low prevalence may be the result of a displacement that occurred when the pathogenic sister species *C. parasitica* was first introduced to both continents (Hoegger et al., 2002). *C. carpinicola* is a recently described species isolated from declining European hornbeams in Austria, Georgia, Italy and Switzerland (C. Cornejo, pers. comm.). The diversity of lifestyles within the Cryphonectriaceae, including the emergence of new pathogens, raises important questions whether genetic factors facilitate pathogenic lifestyles.

In this study, we assembled and analyzed 104 genomes of the Cryphonectriaceae family including the major representatives *C. parasitica*, *C. radicalis*, *C. naterciae*, *C. japonica* and a recently detected European *Cryphonectria* species named *C. carpinicola* (C. Cornejo, pers. comm.). We analyzed orthology among the gene sets of the species and constructed a robust phylogenomic tree. We find that Cryphonectriaceae share similar trophic lifestyle traits. However, the chestnut pathogen *C. parasitica* has a substantially reduced complement in CAZymes. In contrast, the capacity to produce secondary metabolites is reduced among *Cryphonectria* species but is broadly conserved within the genus. Effector candidate proteins show genus and species specificity consistent with faster evolvability of the underlying genes.

Results

Genome assemblies for the *Cryphonectria* genus

We assembled draft genomes of 100 *Cryphonectria* spp. isolates of Asian, European and North American origin, in addition to the previously assembled genome of *C. parasitica* reference genome EP155. As a near outgroup to the genus *Cryphonectria,* we analyzed previously assembled draft genomes of 3 *Chrysoporthe* species from South Africa, Colombia and Indonesia. To assemble *Cryphonectria* genomes *de novo*, we used Illumina sequencing data at 9-53X coverage (Table 1). All *Cryphonectria* and *Chrysoporthe* genome assemblies showed >95% completeness for BUSCO genes (ascomycota_odb9 database) with the *C. parasitica* isolate M7832 having the lowest score at 95.9% (Table 1, Fig. 1A). Based on the assembly size, we estimated that non-

pathogenic species had smaller genomes ranging from 38.6Mb (*C. japonica)* to 41.9 Mb (*C. carpinicola*). Pathogenic species had slightly larger genomes ranging from 43.7 Mb (*C. parasitica* and *Chr. austroafricana*) to 45 Mb (*Chr. cubensis*; Table 1; Fig. 1A). We found no apparent correlation between the estimated genome size and the completeness in BUSCO genes (Fig. 1A-B). Similarly, we detected no correlation between the sequencing depth and the assembled genome size (Fig. 1C). This shows that the short-read based assemblies are expected to reliably represent the gene content across species.

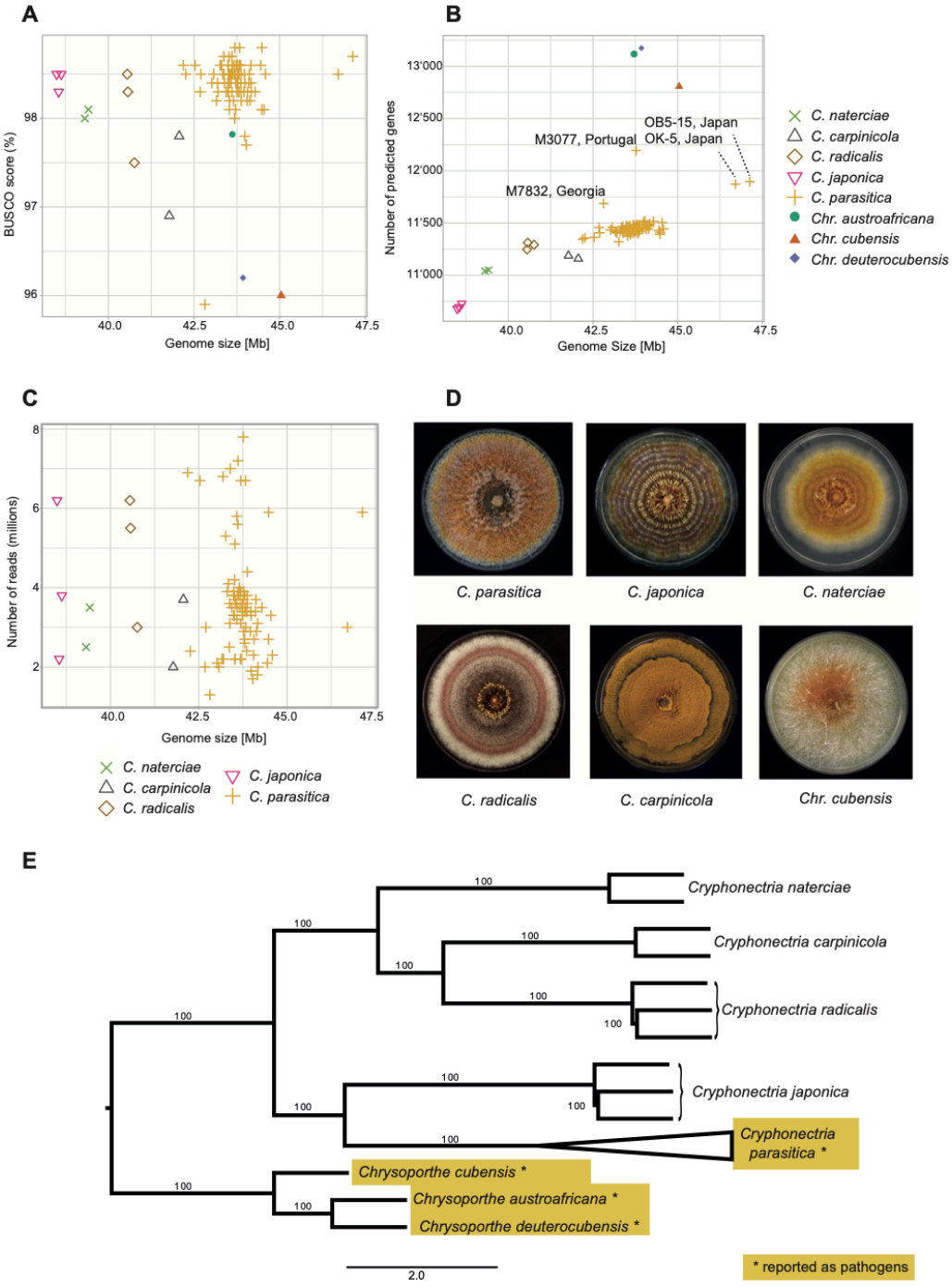


Figure 1: **Assembly statistics and phylogenetic reconstruction.** Estimated genome size in Mb correlated with A) assembly completeness assessed by BUSCO scores, B) number of predicted genes, and C) sequencing depth

of assembled Cryphonectriaceae genomes. D) Culture morphology of the studied Cryphonectriaceae (cultures of *Chr. deuterocubensis* and *Chr. austroafricana* are not shown, as isolates were unavailable for documentation). E) Maximum-likelihood consensus tree based on 6770 single-copy ortholog genes showing the phylogenetic relationship of Cryphonectriaceae species.

Gene annotation and phylogenetic reconstruction

We predicted between ~10'700–12'200 genes in genomes of *Cryphonectria* species compared to ~12'800 – 13'170 genes in *Chrysoporthe* spp. (Table 1, Fig. 1B). Overall, gene content among species was correlated with genome size except for *C. carpinicola* and *Chr. cubensis*, which have fewer predicted genes as expected from their genome size (Fig. 1B). Among *C. parasitica* isolates, M3077 had a higher gene content compared with isolates of similar genome size (Fig. 1B). Moreover, assembled genomes of *C. parasitica* isolates OB5-15 and OK-7 showed increased genome sizes while only having slightly higher gene content compared to other *C. parasitica* isolates (Fig. 1B).

***Table 1*: Genome assembly statistics for *Cryphonectria* spp. and *Chrysoporthe* spp.** The *C. parasitica* reference genome is not included in the summary of *C. parasitica* genomes.* = pathogenic species.

Species	mean size [Mb]	mean N50	mean complete BUSCO (%)	coverage (min-max)	N° predicted genes (min-max)	N° of isolates
*C. parasitica**	43.7	125'044	98.40	9–53X	11'321–12'195	91
C. japonica	38.6	364'390	98.43	17–48X	10'680–10'729	3
C. radicalis	40.6	197'843	98.1	22–46X	11'247–11'312	3
C. naterciae	39.4	120'703	98.05	19–27X	11'041–11'050	2
C. carpinicola	41.9	58'390	97.35	14–26X	11'159–11'187	2
*Chr. austroafricana**	43.7	48'708	97.8	NA	13'125	1
*Chr. cubensis**	45.0	345'702	96	NA	12'807	1
*Chr. deuterocubensis**	43.9	83'661	96.2	NA	13'174	1

The gene ortholog analyses revealed 6770 single-copy orthologs among all species. We found 85 species-specific orthologs, of which 22 were specific for *C. parasitica*. Additionally, we found between 1–10 isolate-specific orthologs among the *C. parasitica* isolates TA51, M7832, DU5, OB5-15, OK-17 and M4030. Moreover, one ortholog was specific for *C. carpinicola*, while no species-specific orthologs were detected in all other *Cryphonectria* species. Within *Chrysoporthe*, we found 19 orthologs specific to *Chr. deuterocubensis*, as well as 12 and 5 orthologs specific to *Chr. cubensis* and *Chr. austroafricana*, respectively. To reconstruct the evolutionary history of *Cryphonectria* and *Chrysoporthe* species, we generated a consensus maximum-likelihood tree based on 6770 single-copy ortholog genes. We found 100% bootstrap branch support between species and a clear divergence at the genus level (Fig. 1E). Furthermore, *Cryphonectria* species were grouping by geographic origin, with *C. naterciae*, *C. radicalis* and *C. carpinicola* being of European origin and *C. japonica* and *C. parasitica* being of Asian descent. Overall, our consensus tree is in accordance with phylogenetic studies on the genera *Cryphonectria* and *Chrysoporthe* (C. Cornejo, pers. comm., van der Merwe et al., 2010).

Lifestyle prediction and capacity for carbohydrate metabolism across species

In order to degrade plant cell walls for nutrition or infection, fungi produce a variety of enzymes involved in carbohydrate metabolism (CAZymes) (Zhao et al., 2014). We analyzed the predicted proteome of Cryphonectriaceae species and other tree-associated fungi to assess trophic lifestyles according to CAZyme content. All *Cryphonectria* species were identified as hemi-biotrophs by CATAStrophy, while *Chrysoporthe* species were classified as necrotrophs. However, the PCA shows close proximity of analyzed *Cryphonectria* and *Chrysoporthe* species, clustering at the verge with other hemi-biotrophic and necrotrophic species (Fig. 2A). Lifestyles of most fungi outside Cryphonectriaceae family matched with predicted lifestyles according to CAZyme content, except for *V. mali* (Fig. 2A).

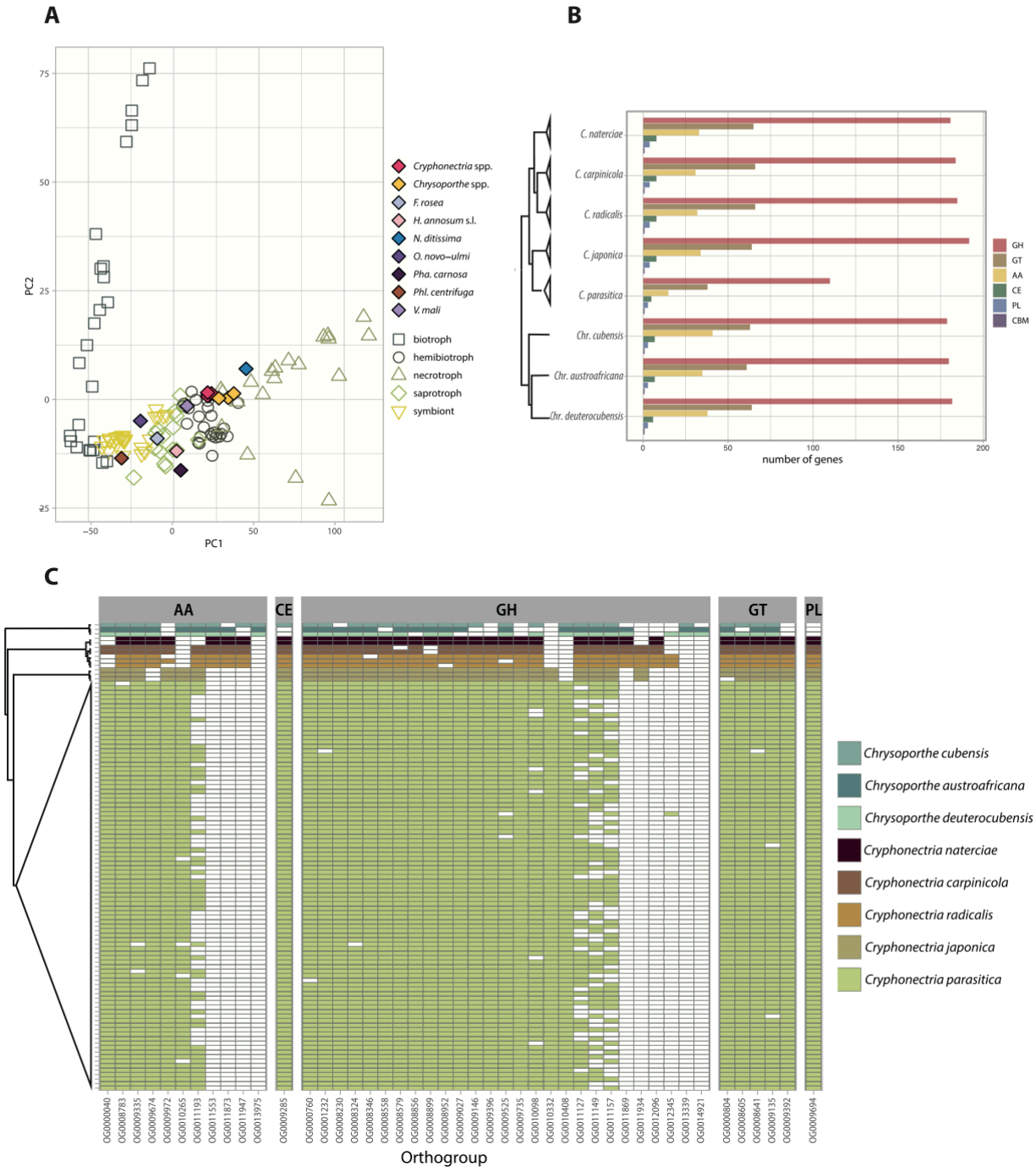


Figure 2: **Carbohydrate Active EnZyme (CAZyme) content among Cryphonectriaceae.** A) Principal component analysis (PCA) of fungal lifestyle predictions, as inferred by CATAStrophy. The plot incorporates 85 reference species of fungi with different lifestyles (*i.e.* biotroph, hemibiotroph, nectrotroph, saprotroph and symbiont) used as a training set by CATAStrophy and shows the CAZyme inferred phenotypic trophy of Cryphonectriaceae and other pathogenic and non-pathogenic tree-associated fungi. B) Number of detected CAZyme genes per species grouped according to CAZyme superfamily: glycoside hydrolase (GH), glycosyl transferase (GT), auxiliary activity (AA), carbohydrate esterase (CE), polysaccharide lyase activity (PL) and carbohydrate-binding modules (CBM). C) Ortholog presence/absence of CAZyme superfamilies for which at least one species is missing an ortholog (the CBM superfamily is not shown, as orthologs were found in all species).

We further assessed CAZyme gene content among Cryphonectriaceae and found a striking gene loss in the chestnut blight pathogen *C. parasitica* (Fig. 2B). The gene loss particularly affected the group of glycoside hydrolases (GH), glycosyl transferases and enzymes with auxiliary activity (AA). Overall, all non-pathogenic *Cryphonectria*, as well as the pathogenic *Chrysoporthe* species encoded between 38.5–42.7% more GH, 37.7–42.4% more GT and 51.6–63.4% more AA than *C. parasitica* (Fig. 2B). We identified gene losses in *C. parasitica* across most CAZyme categories. GH5 associated with hemicellulose degradation showed a particularly remarkable reduction (Supp. Fig. 2). We found between 12–13 GH5 genes in saprophytic *Cryphonectria* and 11 GH5 genes in *Chrysoporthe* spp., while *C. parasitica* had only four GH5 genes. Moreover, slightly fewer GH28 genes involved in pectin degradation, were detected in *C. parasitica* (n = 11), compared to *Chrysoporthe* spp. (n = 12–14) and saprophytic *Cryphonectria* (n = 15–16). Analyzing CAZymes for which at least one species is missing an ortholog, *Cryphonectria* species share a relatively conserved set of ortholog CAZyme genes as expected from their short phylogenetic distance (Fig. 2C). We found one PL orthogroup encoding for pectate lyase (OG0009694), only shared among *Cryphonectria* species. Moreover, we detected one GH orthogroup belonging to the sialidase superfamily (OG0010332), which is only present in Asian *Cryphonectria* species, as well as a single GH orthogroup (OG0012096: GH3) only present in European *Cryphonectria* species (Fig. 2C). *C. parasitica* displayed a particularly high degree of intra-specific presence/absence variation for four auxiliary activity (AA) and GH enzymes, which are otherwise well conserved (OG0011193: GMC oxidoreductase, OG0011127: GH76, OG0011149: GH43, OG0011157: GH76; Fig. 2C). The four orthogroups likely underwent recent gene losses in *C. parasitica.*

To assess the wood colonizing capabilities of different *Cryphonectria* species and a member of the genus *Chrysoporthe* (*Chr. cubensis*), we conducted an inoculation experiment on dormant and healthy chestnut stems. We performed the experiment with and without prior removal of the bark. None of the species were able to colonize dormant chestnut logs without artificial wound induction. After two weeks of incubation, *C. japonica* showed signs of mycelial growth on the bark at max. 1 cm beyond the inoculation point. No bark penetration was detected. For inoculations with bark removal, *C. parasitica* expectedly showed the fastest and most extensive lesion growth. Other *Cryphonectria* species, with the exception of one *C. radicalis* isolate (M4733), developed only minimal lesions (Supp. Fig. 1). We found intraspecific variance in lesion growth, possibly attributed to varying isolate vigor (*e.g. C. radicalis* isolate M283 was isolated in 1953) or varying substrate conditions (*e.g.* state of dormancy, stem thickness) (Supp. Fig. 1). The eucalyptus pathogen *Chr. cubensis* showed growth on non-host chestnut (*Ca. sativa*) logs, however, lesion developed at a comparatively slow pace (Supp. Fig. 1). After four weeks of incubation, mycelial fans were only found in lesions caused by *C. parasitica*.

Variation in secondary metabolite production potential among species

Secondary metabolites (SM) can play important roles in pathogenicity and the interaction with microbes (Brakhage, 2013; Scharf et al., 2014). We investigated variation in biosynthetic core genes as an indicator for metabolite production potential among species. Loss of a biosynthetic core gene from a cluster invariably leads to loss of cluster function. Overall, biosynthetic core gene counts were only variable between genera. *Cryphonectria* species had comparatively fewer biosynthetic core genes than *Chrysoporthe* species (Fig. 3A). Among the detected biosynthetic core genes, both genera shared similar proportions of different gene cluster classes with type 1 PKS (T1PKS) being the most abundant gene cluster class (Fig. 3A). The class of beta-lactone production clusters, which can produce potent anti-bacterial and anti-fungal compounds (Robinson et al., 2019) were exclusively found in *Chrysoporthe* species (Fig. 3A). The presence/absence analyses of biosynthetic core genes per gene clusters (n = 47) revealed 28 clusters conserved among all analyzed Cryphonectriaceae (Fig. 3B). Additionally, core genes in five clusters were conserved in *Cryphonectria*. The same clusters showed partial or complete absence in *Chrysoporthe*. The largest cluster was found on scaffold 4, containing four T1PKS biosynthetic core genes. All four core orthologs of the cluster were retained in *C. parasitica.* Other species lost between one (*C. japonica*) and all four core genes (*Chr. cubensis*) (Fig. 3B). Overall, core genes were highly conserved among *C. parasitica*, except for three T1PKS, NRPS-like and NRPS-T1PKS clusters on scaffolds 4, 6 and 11 (OG0010469, OG0005403 and OG0009181) (Fig. 3B). The clusters showed gene losses in *C. parasitica* isolates from China (TA51), Georgia (M7776, M7832), Japan (WB-3) and USA (MD-1). Generally, we only identified weak homology with secondary metabolite clusters in other species. A notable exception includes two gene clusters potentially underlying emodin production (Supp. Fig. 3).

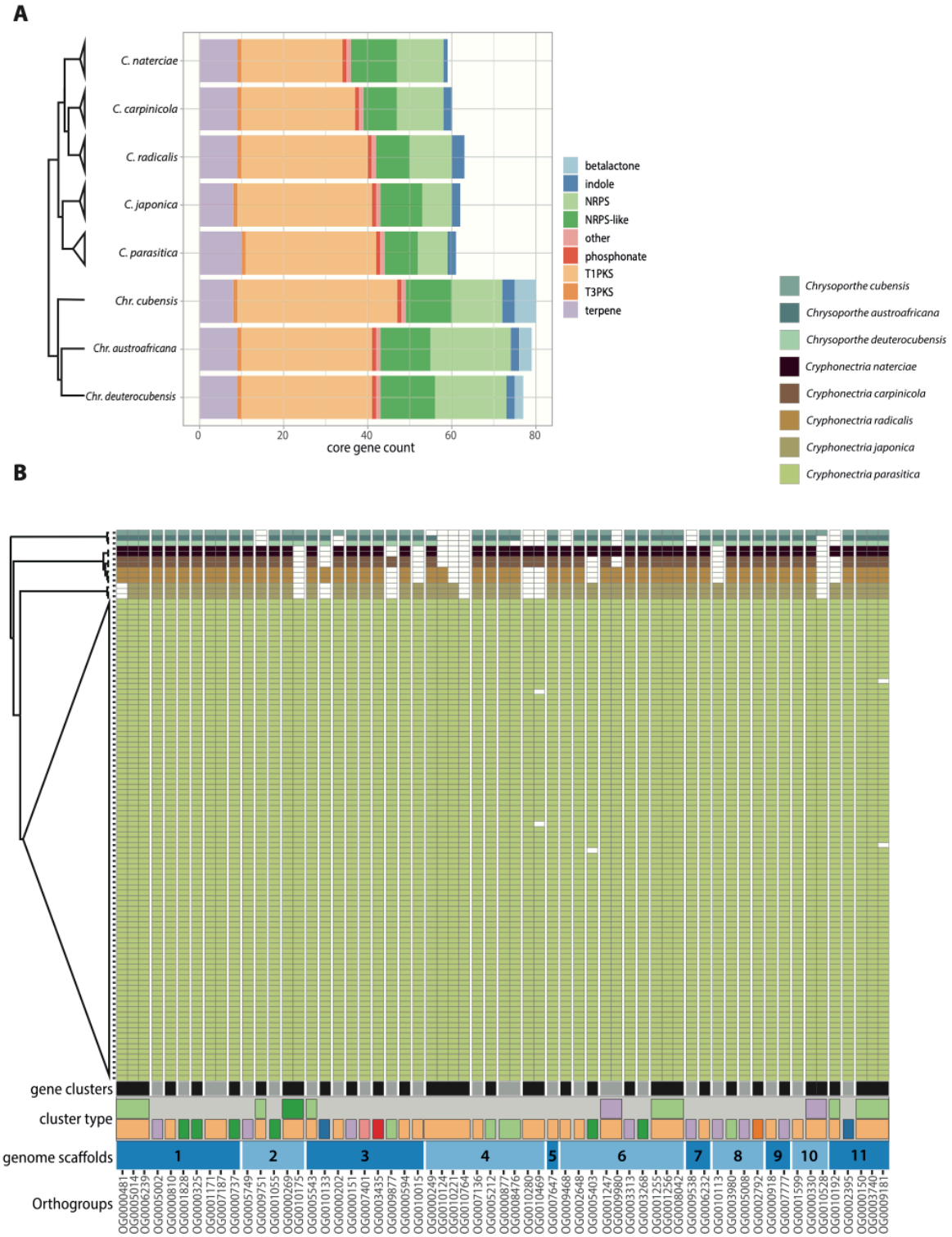


Figure 3: **Secondary metabolite core gene content among Cryphonectriaceae.** A) Count of detected biosynthetic core genes categories across species as identified by antiSMASH. B) Presence/absence of biosynthetic core gene orthologs among species. The plot shows the number of biosynthetic core genes within a gene cluster, the cluster type (color codes are as in A) and the location of clusters according to *C. parasitica* reference genome scaffolds.

Predicted effector genes among Cryphonectriaceae, effector orthologs and cysteine content

Effectors are mostly secreted, cysteine-rich proteins, which play a major role in fungal virulence to overcome host immune defenses (Stergiopoulos and de Wit, 2009). We predicted effector genes with a machine-learning approach and found that neither the number of putative secreted proteins nor the predicted effector content correlated with genome size (Fig. 4A). Saprophytic *Cryphonectria* species encode slightly more putatively secreted proteins (n = 777–796) compared to pathogenic *Chrysoporthe* spp. (n = 751–772). Surprisingly, *C. parasitica* encodes markedly fewer secreted proteins (n = 619) than all other species (Fig. 4A). However, despite the low amount of secreted proteins, *C. parasitica* had the highest ratio of predicted effectors among all species with 7.8% of all secreted proteins predicted to function as effectors (Fig. 4A). Overall, the pathogenic versus saprophytic lifestyle did not correlate with predicted effector content. For example, we found that pathogenic *Chr. deuterocubensis* encoded the smallest number of predicted effectors of all analyzed species (Fig. 4A). The cysteine content of predicted Cryphonectriaceae effectors ranged from 0–12.9% (Fig. 4B). The predicted effectors among *Cryphonectria* contained 53–348 amino acids with one outlier of only 33 amino acids in *C. radicalis.* Predicted *Chrysoporthe* effectors contained 67–436 amino acids (Fig. 4B). The divergence in candidate effector gene content among Cryphonectriaceae matches the divergence in cysteine content and protein length. Analysis of predicted effector ortholog presence/absence among Cryphonectriaceae revealed 41.5% (n = 59) conserved orthologs in all Cryphonectriaceae and 91 orthologs showed presence/absence variation among species (Fig. 4C). We found several orthologs unique to a single species (Fig. 4C). Interestingly, the species-specific *C. parasitica* orthologs OG0010999, OG0010973 and OG0010938, showed presence-absence variation with orthologs missing in isolates from China and South Korea (LB86, M8510, S35; Fig. 4C). For eight candidate effectors, we could not find a corresponding ortholog annotation with OrthoFinder (grey area in Fig. 4C).

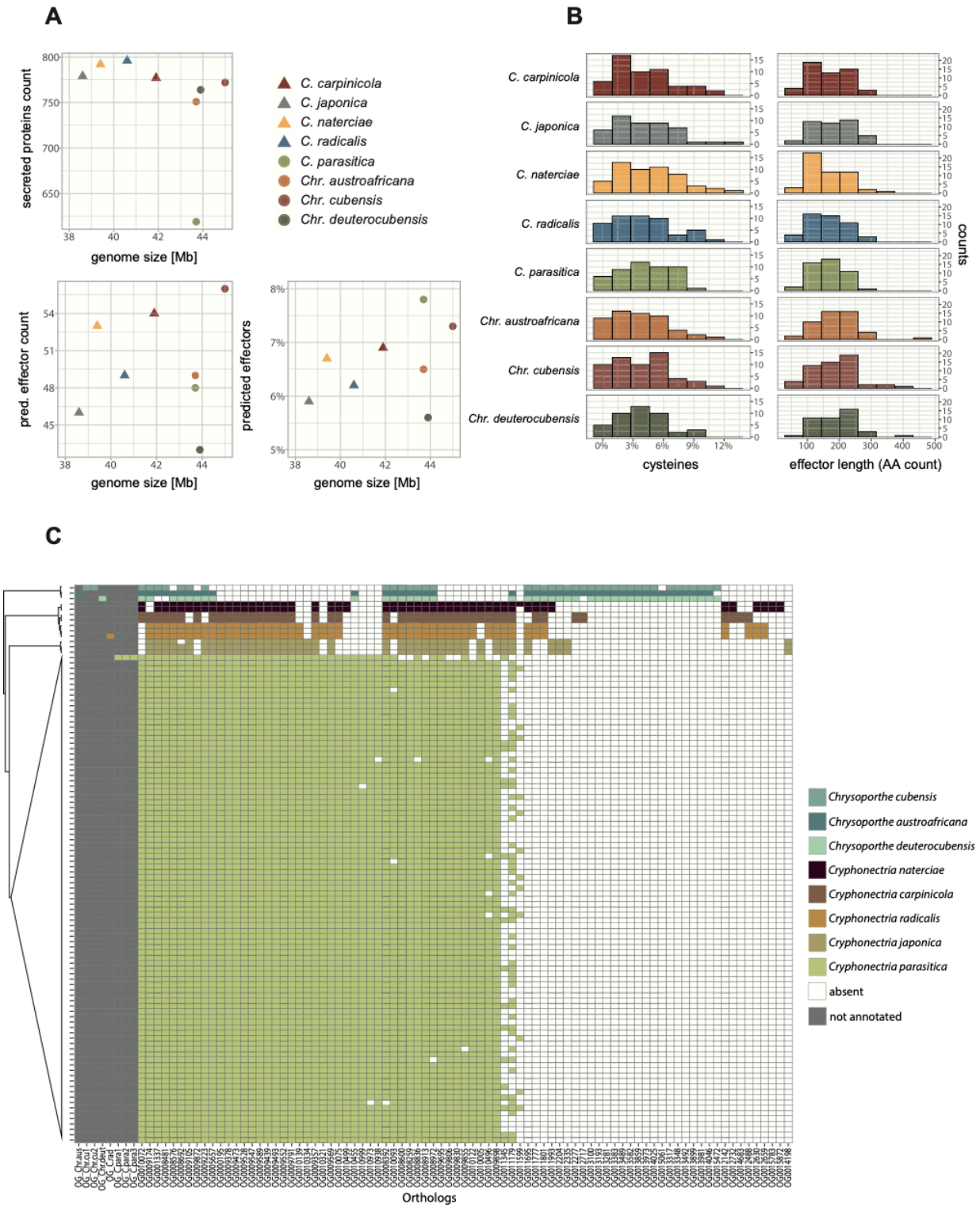


Figure 4: **Predicted secretome and putative effectors among Cryphonectriaceae.** Conserved orthologs (*i.e.* effector genes shared among all species) were omitted. A) Genome size correlations with secreted proteins and predicted effectors (identified by EffectorP). Saprophytic species are shown with triangles, pathogens with circles. B) Histograms showing the cysteine content (%) and the size of predicted effectors per species. C) Presence/absence of predicted effector orthologs among species. Areas in grey show orthologs for which we found no corresponding ortholog.

Discussion

We assembled and analyzed genomes of eight bark-inhabiting Cryphonectriaceae species to retrace the evolution of genome size and gene content. Based on CAZyme content, all analyzed species are predicted to share a similar trophic lifestyle. In the genus *Cryphonectria*, we detected striking CAZyme gene loss in the invasive pathogen *C. parasitica*. In spite of the substantial CAZyme gene loss, *C. parasitica* shares wood colonization strategies with the other *Cryphonectria* species and has retained the ability for early saprotrophic wood decay. In contrast, secondary metabolite gene clusters diverged at the genus level, but were largely conserved among *Cryphonectria* species. Putative effector content varied substantially among species with differences in cysteine content and protein length.

Distinct CAZyme gene loss in a pathogenic species

The CAZyme profiles of the Cryphonectriaceae species analyzed in this study match those of other hemi-biotrophic or necrotrophic fungi. Thus, albeit substantial difference in pathogenicity (Dennert et al., 2020), Cryphonectriaceae species seem to share trophic lifestyle traits, which challenges previous classifications of *C. japonica*, *C. naterciae* and *C. radicalis* as predominantly saprotrophic species (Rigling and Prospero, 2018). Nonetheless, the distinct CAZyme loss in *C. parasitica* coincides with an increased pathogenicity towards non-native (*i.e.* non-Asian) *Castanea* species, which seems to be absent in other *Cryphonectria* species. Many CAZymes play a role in plant cell-wall degradation and can be important virulence factors in necrotrophic fungi. Reductions in CAZymes genes have been observed in biotrophic pathogens and are thought to be an adaptation to reduce the exposure of molecular patterns, which can trigger host defenses (Duplessis et al., 2014; Goulet and Saville, 2017). Moreover, CAZyme loss can occur during host shifts, such as from plant to animal or insect hosts (Zhang et al., 2018). In *C. parasitica,* the CAZyme loss may be an adaptation facilitating an increased pathogenic lifestyle. At the intraspecific level, fewer CAZymes are expressed during pathogenic growth compared to saprotrophic wood decay in the conifer pathogen *H. annosum s.l.,* which has plastic lifestyles (Olson et al., 2012). Similarly, *C. parasitica* may have undergone a transitory phase in the evolution of the predominant pathogenic lifestyle favoring reduced CAZyme expression and ultimately gene losses. Moreover, *H. annosum s.l.* produces more secondary metabolites including phytotoxins during the pathogenic lifestyle (Olson et al., 2012). These findings suggest that necrotrophic pathogens of trees have evolved different wood degradation strategies compared to saprotrophic relatives. The most significant gene loss in *C. parasitica* was found in the CAZyme subfamily GH5, which is underlying hemicellulose degradation. Consistent with this, GH5 expression is lower during pathogenic growth in *H. annosum s.l.* (Olson et al., 2012). In contrast, genomes of saprotrophic wood decayers such as *Pha. carnosa* have expanded GH5 repertoires (Suzuki et al., 2012). Despite extensive CAZyme loss in *C. parasitica*, our experimental data show that all *Cryphonectria* species, including *C. parasitica*, have retained similar wood colonization capabilities through bark wounds. Moreover, *C. parasitica* appears to have retained CAZymes suitable for early wood decay. This confirms field observations

indicating that the fungus is able to survive a few years on the bark of fresh dead chestnut wood (Prospero et al., 2006). In parallel to GH5, pectin degrading enzymes of GH28 are also slightly reduced in *C. parasitica.* However, polygalacturonases belonging to the GH28 family are suggested to contribute to virulence in *C. parasitica* (Lovat and Donnelly, 2019). Similarly, in other necrotrophic pathogens GH28 are also associated with pathogenicity showing expansions in the GH28 family (Sprockett et al., 2011). Similar to GH5, *C. parasitica* may have lost GH28 enzymes triggering host defenses through molecular pattern recognition by the host (Mengiste, 2012).

Potential virulence-associated traits in *C. parasitica*

In contrast to the evolution of CAZymes, secondary metabolite production capabilities are largely conserved within the genus *Cryphonectria*. *C. parasitica* produces virulence-associated compounds including oxalic acid, tannases, laccases and phytotoxins such as cryparin and diaporthin but the genetic basis is only partially resolved (Lovat and Donnelly, 2019). The diaporthin production pathway is encoded by a PKS gene cluster in *Aspergillus oryzae* (Chankhamjon et al., 2016). However, we identified no clearly orthologous cluster in *C. parasitica*. The conservation of gene clusters across *Cryphonectria* species suggests that secondary metabolites played no particular role in the evolution of pathogenicity by *C. parasitica*. However, many fungi can modulate metabolite production depending on environmental conditions (Shwab and Keller, 2008). Hence, even if all *Cryphonectria* species share a core set of gene clusters, lifestyle transitions may induce differential expression depending on biotic or abiotic conditions. In addition to secondary metabolites, small secreted proteins (*i.e.* effectors) can play key roles in the emergence of new pathogens. We identified a broad pool of putative effector orthologs among Cryphonectriaceae. The size of the effector gene pool did not correlate with genome size or lifestyle as seen in other clades of plant pathogens (Lo Presti et al., 2015; Lowe and Howlett, 2012). Effector homologs in the orthogroups OG0010999, OG0010973 and OG0010938 are particularly interesting candidates because the genes are both unique to *C. parasitica* and show presence/absence variation within the species. The recent gene gains in the pathogen lineage and the presence/absence variation within the species could explain variation in pathogenicity between and within the species, respectively. Combining analyses of positive selection, gene expression and targeted gene deletion assays of effector candidates in *C. parasitica* will be needed to elucidate the role of effectors in causing chestnut blight.

Lifestyle and the role of hosts

Cryphonectriaceae species represent a useful model to retrace how lifestyle transitions towards pathogenicity impact the evolution of gene content. On its native Asian hosts (*Ca. crenata* and *Ca. molissima*), the chestnut blight fungus *C. parasitica* causes only mild symptoms, which has been attributed to host-pathogen co-evolution (Rigling and Prospero, 2018). In contrast, on the naïve American and European chestnut species (*Ca. dentata, Ca. sativa*) the pathogen causes lethal bark cankers (Anagnostakis, 1992). In the invasive range, *C. parasitica* might also be a weak pathogen on *Quercus* spp., *Acer* spp. or *Carpinus betulus* (Rigling and Prospero, 2018).

This suggests that *C. parasitica* has the genetic repertoire of a broad host range pathogen and that chestnut species may be the least able to resist pathogen invasion. In diverse forest ecosystems, disease incidence is often negatively correlated with host species richness (the "dilution effect", (Haas et al., 2011; Nguyen et al., 2017; Prospero and Cleary, 2017)). Hence, growth on largely resistant hosts may be a bet hedging strategy of *C. parasitica* to survive and spread in the absence of the primary host (Trapero-Casas and Kaiser, 2009). The weak pathogenicity of other *Cryphonectria* species may be facilitated by environmental conditions, such as abiotic stress on the host or disturbance of the host microbiome (Brader et al., 2017). Subsequently, these *Cryphonectria* species may be considered as latent pathogens similar to some endophytes (Sakalidis et al., 2011; Vaz et al., 2018; Zakaria et al., 2010). Latent pathogenicity has been observed in other Cryphonectriaceae. For example, (Granados et al., 2020) found that the *Eucalyptus* pathogen *Chr. cubensis* is an endophyte on Colombian Melastomataceae trees. Moreover, the pathogen *Chr. austroafricana* occurs as an endophyte in its native range but is pathogenic on non-native *Eucalyptus* trees (Mausse-Sitoe et al., 2016). Hence, host jumps likely facilitated the switch from endophytic to pathogenic lifestyle in both species (Granados et al., 2020; Mausse-Sitoe et al., 2016). *C. parasitica* has recently emerged as a major pathogen on non-Asian chestnut species. To what degree the extensive CAZyme loss increased the pathogenic potential prior to the emergence as an invasive pathogen remains to be investigated. Comparative genomics combined with gene function analyses provide a powerful approach to study lifestyle evolution and changes in the underlying genome architecture.

Materials and Methods

Genome sequencing

We sequenced whole genomes of 90 *C. parasitica*, 3 *C. japonica*, 3 *C. radicalis*, 2 *C. naterciae* and 2 *C. carpinicola*. isolates covering the global distribution range (Supp. Table 1). All isolates were prepared for sequencing as described in (39). Sequencing was conducted using the Illumina HiSeq4000 and Illumina NovaSeq 6000 platforms (Illumina, San Diego, USA) at the Functional Genomics Center Zurich (FGCZ). By choosing the Illumina NovaSeq SP flowcell, NovaSeq reads were compatible with HiSeq4000 reads for downstream analysis.

Genome assembly and gene prediction

All 100 *Cryphonectria* spp. sequences were assembled with SPAdes v3.13.0 (Bankevich et al., 2012), using the --careful option and choosing the k-mers 21, 33, 45, 57 and 69 for the iterative assembly process. Genome sizes and assembly quality of *Cryphonectria* spp. *de novo* assemblies, as well as *Chrysoporthe* spp. draft genomes were assessed with QUAST v5.0.2 and BUSCO v3.0.2 (Mikheenko et al., 2018; Waterhouse et al., 2018). Gene models were predicted using BRAKER2 v2.1.4 (Hoff et al., 2019, 2016; Stanke et al., 2008, 2006). Briefly, we set up gene annotation training using the existing *C. parasitica* v2 reference genome annotation (available from

http://jgi.doe.gov/, (Crouch et al., 2020)) using the BRAKER2 options --alternatives-from-evidence=false, --fungus, --gff3 and --skip_fixing_broken_genes. For splice site hints, intron information was extracted from the reference genome annotation using the construct_introns function from the R package gread v0.99.3 (Srinivasan, 2019). After the training, genes were predicted in all assembled genomes using BRAKER2 adding coding sequence hints of the *C. parasitica* reference genome obtained using gffread v0.11.0 (Trapnell et al., 2012) and EMBOSS v6.6.0 tool transseq (Rice et al., 2011). We set the BRAKER 2 options --alternatives-from-evidence=false, --gff3, --useexisting, --prg=gth, --trainFromGth.

Identification of orthologs and secondary metabolite gene clusters

To identify orthologs among all *Cryphonectria* and *Chrysoporthe* isolates we used OrthoFinder v2.3.7 (Emms and Kelly, 2019). We selected all single-copy ortholog groups and generated sequence alignments using MAFFT v7.429 (Katoh and Standley, 2013). Aligned sequences were used for phylogenetic tree building by generating 100 maximum-likelihood (ML) trees using the GTRCAT model with RAxML v8.2.12 (Stamatakis, 2014). The RAxML generated tree and bootstrap files were subsequently used to build a consensus tree with Astral v5.14.2 (Yin et al., 2019). The obtained consensus tree was visualized with FigTree v1.4.3 (Rambaut and Drummond, 2016). We used the antiSMASH fungal version v5.1.0 (Blin et al., 2019) to identify secondary metabolite gene clusters using one isolate per species; the reference genome EP155 for *C. parasitica*, IF-6 for *C. japonica*, M283 for *C. radicalis*, M3664 for *C. naterciae*, CS3 for *C. carpinicola* and the three NCBI *Chrysoporthe* draft genomes. We used a custom Python script to extract the biosynthetic core genes from the antiSMASH regions.js file. The number of core genes per species was then plotted in R with the packages tidyverse (Wickham, 2017), reshape2 (Wickham, 2007) and ggplot2 (Wickham, 2016). Additionally, we identified biosynthetic core genes per cluster of the EP155 genome and searched for orthologs in all species. Secondary metabolite core gene ortholog presence/absence in each cluster was plotted in R, using the packages reshape2, stringr (Wickham, 2018) and ggplot2.

Classification of fungal lifestyles according to CAZyme content

We inferred trophic lifestyles of Cryphonectriaceae according to carbohydrate-active enzyme (CAZyme) gene content using the CAZyme-Assisted Training And Sorting of -trophy (CATAStrophy) prediction tool (Hane et al., 2020). CATAStrophy annotates CAZymes with HMMER 3.0 (Eddy, 2010) and dbCAN (Yin et al., 2012) and predicts trophic classes based on a multivariate analysis (Hane et al., 2020). To run CATAStrophy, we selected the same Cryphonectriaceae isolates as described above and added additional tree-associated fungi of different lifestyles. For non-pathogenic saprophytes associated with wood degradation, we selected the proteomes of *Fomitopsis rosea* (BioProject PRJNA518053), *Phanerochaete carnosa* (Suzuki et al., 2012) and *Phlebia centrifuga* (Mäkelä et al., 2018). Moreover, we included *Heterobasidion annosum* s.l. (Olson et al., 2012), associated with both saprotrophic and pathogenic lifestyles, and the bark pathogens *Neonectria ditissima* (Nectria canker on apple and pear trees) (Gómez-Cortecero et al., 2015), *Ophiostoma novo-ulmi* (Dutch elm

disease) (Comeau et al., 2015; Forgetta et al., 2013) and *Valsa mali* (Valsa canker on apple trees) (Yin et al., 2015). For trophic lifestyle inferences, we used the catastrophy-pipeline (https://github.com/ccdmb/catastrophy-pipeline), choosing the options -profile conda and --dbcan_version 8. The CATAStrophy literature-derived nomenclature (*i.e.* classification into biotrophs, hemibiotrophs, nectrotrophs, saprotrophs and symbionts) was used for defining trophic lifestyles of species included in the CATAStrophy training set. Moreover, we selected principal components PC1 and PC2, which separate most training set species according to lifestyle for visualization (Hane et al., 2020).

Analysis of carbohydrate-active enzymes genes (CAZymes) and inoculation experiments

For the identification of CAZyme genes we ran dbCAN v2.0.0 (Yin et al., 2012) on the same isolates as in the secondary metabolite analysis (*i.e.* one isolate per species). Only CAZymes which were identified by all three tools (HMMER, diamond, hotpep) were then selected for further analysis. CAZyme orthologs were extracted using Python and plots were generated in R.

To analyze the wood colonization capabilities of the different species, we set up an inoculation experiment. We selected 26 dormant chestnut logs (*Ca. sativa*; length: 50cm, diameter: 3.3-6.7cm), which were cut in a healthy state during winter from chestnut stands in Ticino, Switzerland, a week prior to the experiment. The logs were surface sterilized with 70% ethanol and sealed on both ends with paraffin to prevent desiccation. We selected 3 *C. parasitica* (XA19, CR03, EP155), 2 *C. japonica* (M9249, IF-6), 2 *C. naterciae* (M3664, M3656), 2 *C. radicalis* (M4733, M283), 2 *C. carpinicola* (M9290, CS3) and 1 *Chr. cubensis* (CBS115724) isolate for inoculation. For all isolates except CBS115724, full genome sequences were available for this study. Prior to inoculation, all isolates were freshly inoculated from glycerol stocks onto potato dextrose agar (PDA; 39 ml/L; BD Becton, Dickinson & Company; Franklin Lakes, USA) and incubated at 25°C in complete darkness for five days to induce mycelial growth. For inoculation of the first batch of chestnut logs (n = 13), we removed the bark on 5 equally distanced spots (diameter: 4mm) on each log, placed a mycelial plug (diameter: 4mm) into the wound and sealed it with tape. For the second batch of chestnut logs (n = 13), we directly placed five mycelial plugs onto the bark of each log with equal distance (*i.e.* no wound induction) and sealed the inoculation spots with tape. For each treatment batch with and without wound, we selected 5 replicates per isolate and 5 negative controls (mycelium-free agar plugs), resulting in a total of 130 completely randomized inoculation spots (n = 65 per treatment). All chestnut logs were randomly placed onto racks in plastic containers, separated by treatment, filled with 2L of demineralized water to avoid drying-out and sealed with plastic lids (Dennert et al., 2020). Incubation was at 20°C for both treatments. Logs with wounds were incubated for four weeks in complete darkness and longitudinal lesion size was assessed once a week. Logs without wounds were incubated for 12 weeks and lesion size was assessed at the end of the experiment.

Prediction of effector genes

We performed effector gene prediction with the same isolates as in the secondary metabolite and CAZyme analysis, using a machine-learning approach. First, secreted proteins were predicted with SignalP v5.0b (Armenteros et al., 2019), choosing the options -org euk and -format short. Only proteins with a likelihood probability > 0.5 were selected for further analysis. Next, protein sequences with a predicted secretion signal were extracted with samtools v1.9 (Li et al., 2009) and used as input for effector prediction with EffectorP v2.0 (Sperschneider et al., 2018). Presence/absence variation analyses of predicted effector gene orthologs across species and plotting was performed in Python and R as described above. The cysteine content and protein length of predicted effector genes in all species were determined with EMBOSS pepstats v6.6.0.

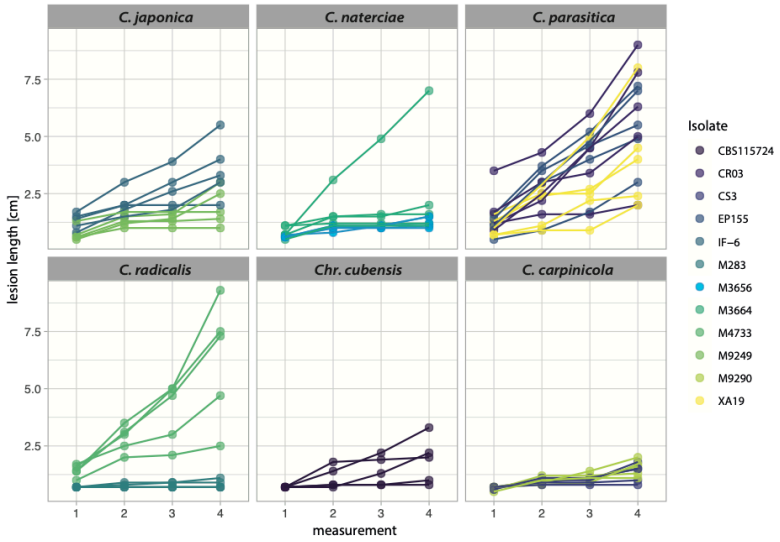
Data availability

All sample accession numbers for the NCBI Short Read Archive, as well as NCBI accessions for all *de novo* assemblies of *Cryphonectria* spp. genomes are available in Supplementary Table S1. Contigs of <400bp were trimmed prior to NCBI genome submission. Outgroup genomes to the genus *Cryphonectria* were obtained for *Chrysoporthe cubensis*, *Chr. deuterocubensis* and *Chr. austroafricana* from NCBI BioProjects PRJNA279968, PRJNA265023 and PRJNA263707, respectively (82, 83).

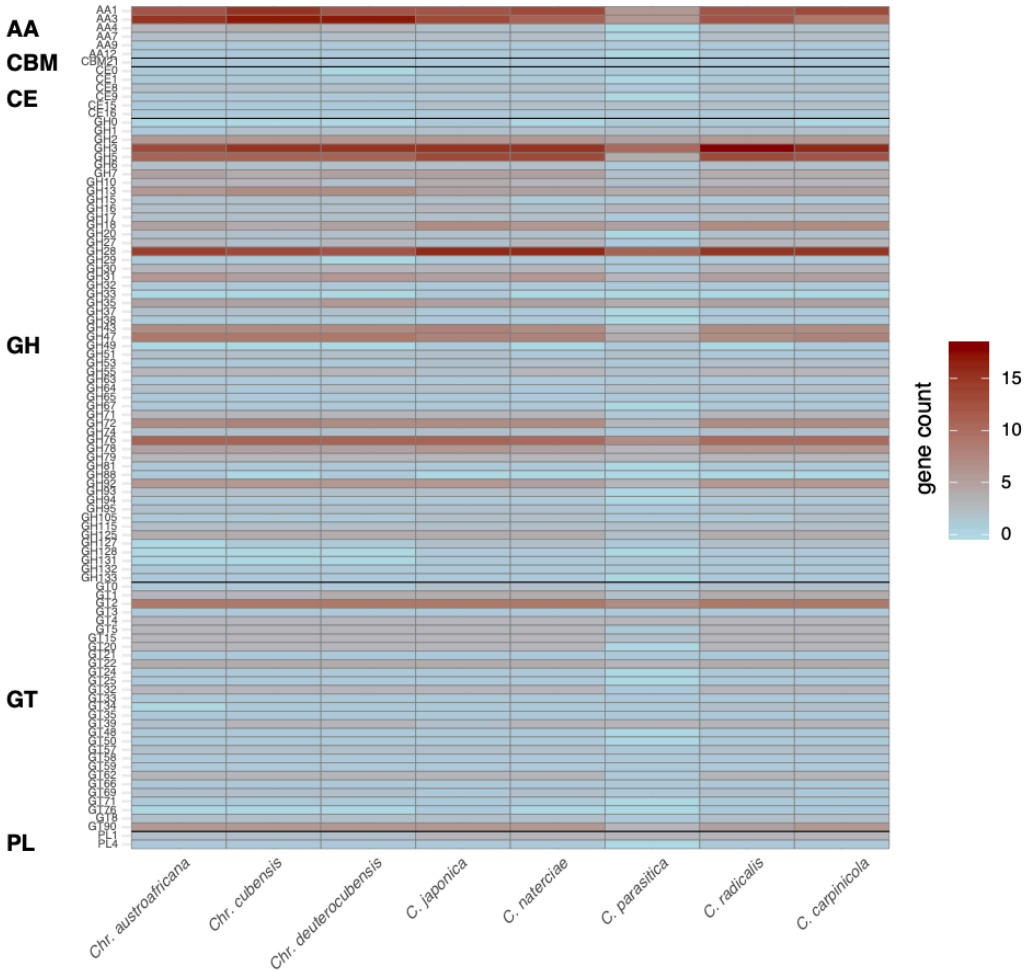
Acknowledgements

We are grateful to Thomas Badet for helpful suggestions on a previous version of this manuscript. We thank Eva Augustiny, Silvia Kobel, Aria Minder, Quirin Kupper and Hélène Blauenstein for laboratory assistance. We acknowledge the Genetic Diversity Centre (GDC), ETH Zurich, and the Functional Genomics Center Zurich (FGCZ) for technical support and facility access. We also thank Martin Wrann for helping with photographic documentation. Paolo Cortesi, Michael Milgroom, Kiril Sotirovski, Mihajlo Risteski, Marin Ježić, Sang Hyun Lee and Seçil Akilli kindly provided samples. Michael and Brenda Wingfield gave valuable insight on phylogenetic relationships among Cryphonectriaceae. We kindly thank Sabina Moser Tralamazza for insightful discussions on secondary metabolites and for sharing scripts. LS was supported by the Swiss National Science Foundation (grant 170188 to SP).

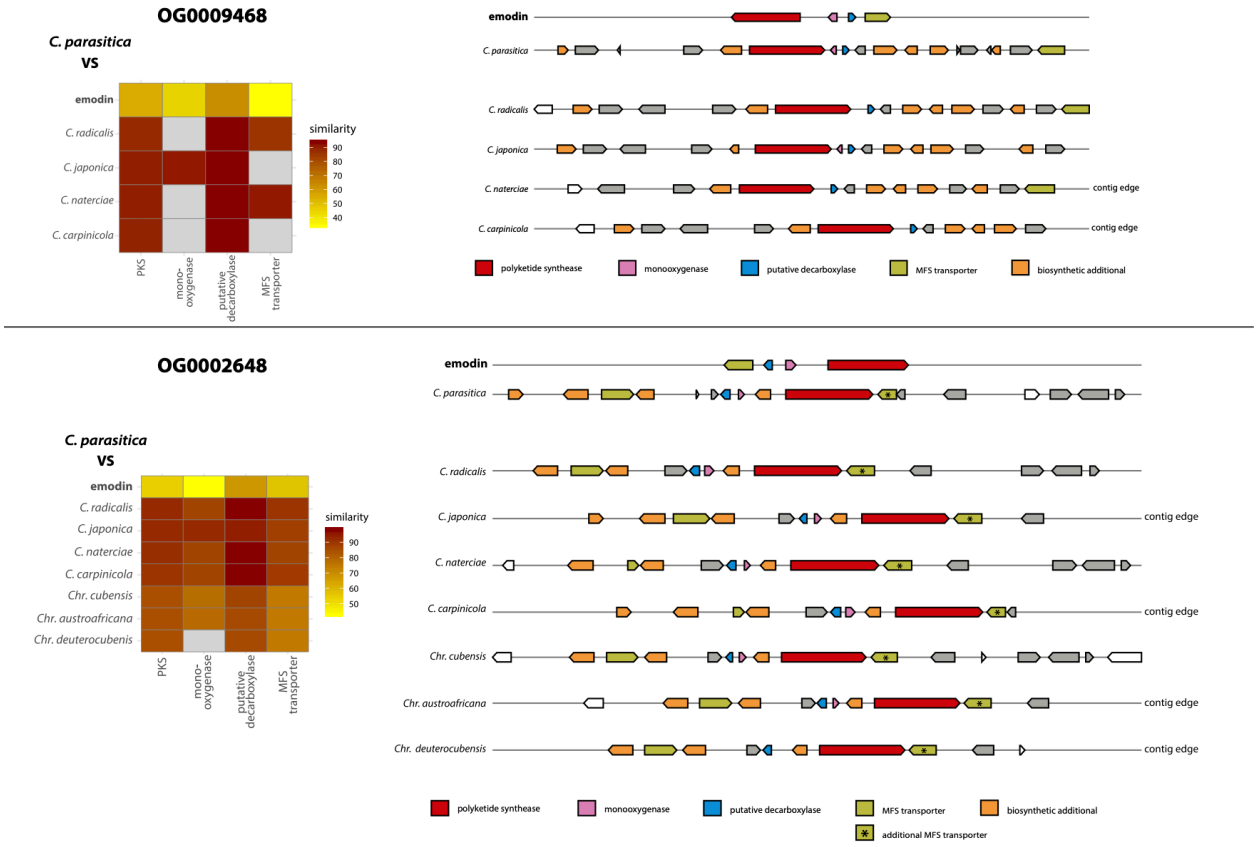
Supplementary Information



Supplementary Figure 1: Lesion length on dormant chestnut stems (*Castanea sativa*) with bark removal. Measurements were taken once per week during four weeks.



Supplementary Figure 2: Gene count of all identified CAZyme families among Cryphonectriaceae.



Supplementary Figure 3: Two PKS gene clusters (biosynthetic core gene orthologs OG0009468 and OG0002648) potentially underlying emodin production. Synteny plots of emodin and identified homologous gene clusters among species. Heatmaps show similarity (%) of genes in the *C. parasitica* reference genome cluster to emodin and orthologs in other Cryphonectriaceae (grey = gene is absent).

Supplementary Table 1: Overview of genomes assembled in the genus *Cryphonectria*. Country, region and host identify the location and host species of collection. See: https://msphere.asm.org/content/5/5/e00737-20/figures-only

References

Anagnostakis SL. 1992. Measuring resistance of chestnut trees to chestnut blight. *Can J For Res* **22**:568–571.

Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**:420–423.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, others. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**:455–477.

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**:W81-W87.

Brader G, Compant S, Vescio K, Mitter B, Trognitz F, Ma L-J, Sessitsch A. 2017. Ecology and genomic insights into plant-pathogenic and plant-nonpathogenic endophytes. *Annu Rev Phytopathol* **55**:61–83.

Bragança H, Rigling D, Diogo E, Capelo J, Phillips A, Tenreiro R. 2011. *Cryphonectria natercia*: A new species in the Cryphonectria-Endothia complex and diagnostic molecular markers based on microsatellite-primed PCR. *Fungal Biol* **115**:852–861.

Brakhage AA. 2013. Regulation of fungal secondary metabolism. *Nat Rev Microbiol* **11**:21–32.

Byrnes III EJ, Li W, Lewit Y, Ma H, Voelz K, Ren P, Carter DA, Chaturvedi V, Bildfell RJ, May RC, others. 2010. Emergence and pathogenicity of highly virulent *Cryptococcus gattii* genotypes in the northwest United States. *PLoS Pathog* **6**.

Chankhamjon P, Tsunematsu Y, Ishida-Ito M, Sasa Y, Meyer F, Boettger-Schmidt D, Urbansky B, Menzel K-D, Scherlach K, Watanabe K, others. 2016. Regioselective Dichlorination of a Non-Activated Aliphatic Carbon Atom and Phenolic Bismethylation by a Multifunctional Fungal Flavoenzyme. *Angew Chemie Int Ed* **55**:11955–11959.

Chooi Y-H, Solomon PS. 2014. A chemical ecogenomics approach to understand the roles of secondary metabolites in fungal cereal pathogens. *Front Microbiol* **5**:640.

Comeau AM, Dufour J, Bouvet GF, Jacobi V, Nigg M, Henrissat B, Laroche J, Levesque RC, Bernier L. 2015. Functional annotation of the *Ophiostoma novo-ulmi* genome: insights into the phytopathogenicity of the fungal agent of Dutch elm disease. *Genome Biol Evol* **7**:410–430.

Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, Hillman B, Milgroom MG, Pangilinan J, Smith M, others. 2020. Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: A fundamental resource for an archetypical invasive plant pathogen. *Phytopathology*.

De Jonge R, Bolton MD, Thomma BPHJ. 2011. How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr Opin Plant Biol* **14**:400–406.

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

Dennert F, Rigling D, Meyer JB, Schefer C, Augustiny E, Prospero S. 2020. Testing the pathogenic potential of *Cryphonectria parasitica* and related species on three common European Fagaceae. *Front For Glob Chang* **3**:52.

Duplessis S, Spanu PD, Schirawski J. 2014. Biotrophic Fungi (Powdery Pildews, Rusts and Smuts). *Ecol Genomics Fungi Plant-Interacting Fungi Sect* 149–168.

Eddy S. 2010. HMMER3: a new generation of sequence homology search software. *URL http//hmmer janelia Org*.

Elliott KJ, Swank WT. 2008. Long-term changes in forest composition and diversity following early logging (1919-1923) and the decline of American chestnut (*Castanea dentata*). *Plant Ecol* **197**:155–172.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**:1–14.

Feldman D, Kowbel DJ, Glass NL, Yarden O, Hadar Y. 2017. A role for small secreted proteins (SSPs) in a saprophytic fungal lifestyle: Ligninolytic enzyme regulation in *Pleurotus ostreatus*. *Sci Rep* **7**:1–13.

Forgetta V, Leveque G, Dias J, Grove D, Lyons Jr R, Genik S, Wright C, Singh S, Peterson N, Zianni M, others. 2013. Sequencing of the Dutch elm disease fungus genome using the Roche/454 GS-FLX Titanium System in a comparison of multiple genomics core facilities. *J Biomol Tech JBT* **24**:39.

Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet* **38**:953–956.

Gaulin E, Pel MJC, Camborde L, San-Clemente H, Courbier S, Dupouy M-A, Lengellé J, Veyssiere M, Le Ru A, Grandjean F, others. 2018. Genomics analysis of Aphanomyces spp. identifies a new class of oomycete effector associated with host adaptation. *BMC Biol* **16**:43.
Giraud T, Gladieux P, Gavrilets S. 2010. Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol Evol* **25**:387–395.
Glass NL, Schmoll M, Cate JHD, Coradetti S. 2013. Plant cell wall deconstruction by ascomycete fungi. *Annu Rev Microbiol* **67**:477–498.
Gómez-Cortecero A, Harrison RJ, Armitage AD. 2015. Draft genome sequence of a European isolate of the apple canker pathogen *Neonectria ditissima*. *Genome Announc* **3**.
Goulet KM, Saville BJ. 2017. Carbon acquisition and metabolism changes during fungal biotrophic plant pathogenesis: insights from *Ustilago maydis*. *Can J Plant Pathol* **39**:247–266.
Granados GM, McTaggart AR, Rodas CA, Roux J, Wingfield MJ. 2020. Species of Cryphonectriaceae occupy an endophytic niche in the Melastomataceae and are putative latent pathogens of Eucalyptus. *Eur J Plant Pathol* **156**:273–283.
Gryzenhout M, Myburg H, der Merwe NA, Wingfield BD, Wingfield MJ. 2004. *Chrysoporthe*, a new genus to accommodate *Cryphonectria cubensis*. *Stud Mycol* **50**:119–142.
Gryzenhout M, Myburg H, Wingfield BD, Wingfield MJ. 2006. Cryphonectriaceae (Diaporthales), a new family including *Cryphonectria, Chrysoporthe, Endothia* and allied genera. *Mycologia* **98**:239–249.
Gryzenhout M, Wingfield BD, Wingfield MJ, others. 2009. Taxonomy, phylogeny, and ecology of bark-inhabiting and tree-pathogenic fungi in the Cryphonectriaceae. American Phytopathological Society (APS Press).
Haas SE, Hooten MB, Rizzo DM, Meentemeyer RK. 2011. Forest species diversity reduces disease risk in a generalist plant pathogen invasion. *Ecol Lett* **14**:1108–1116.
Hane JK, Paxman J, Jones DAB, Oliver RP, de Wit P. 2020. "CATAStrophy," a Genome-Informed Trophic Classification of Filamentous Plant Pathogens-How Many Different Types of Filamentous Plant Pathogens Are There? *Front Microbiol* **10**:3088.
Hardham AR. 2001. Cell biology of fungal infection of plants. Biology of the Fungal Cell. Springer. pp. 91–123.
Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G, others. 2020. 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. *Stud Mycol*.
Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. 2002. *Cryphonectria radicalis*: rediscovery of a lost fungus. *Mycologia* **94**:105–115.
Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**:767–769.
Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER Gene Prediction. Springer. pp. 65–95.
Jiang N, Fan X, Tian C, Crous PW. 2020. Reevaluating Cryphonectriaceae and allied families in Diaporthales. *Mycologia* **112**:267–292.
Jiang N, Fan XL, Tian CM. 2019. Identification and pathogenicity of Cryphonectriaceae species associated with chestnut canker in China. *Plant Pathol* **68**:1132–1145.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.
Keller NP, Turner G, Bennett JW. 2005. Fungal secondary metabolism — from biochemistry to genomics. *Nat Rev Microbiol* **3**:937–947.
Kubicek CP, Starr TL, Glass NL. 2014. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annu Rev Phytopathol* **52**:427–451.
Lelwala R V, Korhonen PK, Young ND, Scott JB, Ades PK, Gasser RB, Taylor PWJ. 2019. Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in *Colletotrichum tanaceti*. *PLoS One* **14**.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079.
Liu Y-C, Linder-Basso D, Hillman BI, Kaneko S, Milgroom MG. 2003. Evidence for interspecies transmission of viruses in natural populations of filamentous fungi in the genus *Cryphonectria*. *Mol Ecol* **12**:1619–1628.
Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol* **66**:513–545.
Lovat C-A, Donnelly DJ. 2019. Mechanisms and metabolomics of the host-pathogen interactions between Chestnut (*Castanea* species) and Chestnut blight (*Cryphonectria parasitica*). *For Pathol* **49**:e12562.

Lowe RGT, Howlett BJ. 2012. Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. *PLoS Pathog* **8**:e1002515.
Lyu X, Shen C, Fu Y, Xie J, Jiang D, Li G, Cheng J. 2015. Comparative genomic and transcriptional analyses of the carbohydrate-active enzymes and secretomes of phytopathogenic fungi reveal their significant roles during infection and development. *Sci Rep* **5**:15565.
Macheleidt J, Mattern DJ, Fischer J, Netzker T, Weber J, Schroeckh V, Valiante V, Brakhage AA. 2016. Regulation and role of fungal secondary metabolites. *Annu Rev Genet* **50**:371–392.
Mäkelä MR, Peng M, Granchi Z, Chin-A-Woeng T, Hegi R, van Pelt SI, Ahrendt S, Riley R, Hainaut M, Henrissat B, others. 2018. Draft genome sequence of the Basidiomycete white-rot fungus *Phlebia centrifuga*. *Genome Announc* **6**.
Mausse-Sitoe SND, Rodas CA, Wingfield MJ, Chen S, Roux J. 2016. Endophytic Cryphonectriaceae on native Myrtales: Possible origin of *Chrysoporthe* canker on plantation-grown Eucalyptus. *Fungal Biol* **120**:827–835.
Mengiste T. 2012. Plant immunity to necrotrophs. *Annu Rev Phytopathol* **50**:267–294.
Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**:i142--i150.
Moonjely S, Barelli L, Bidochka MJ. 2016. Insect pathogenic fungi as endophytes. Advances in Genetics. Elsevier. pp. 107–135.
Myburg H, Gryzenhout M, Wingfield BD, Milgroom MG, Kaneko S, Wingfield MJ. 2004. DNA sequence data and morphology define *Cryphonectria* species in Europe, China, and Japan. *Can J Bot* **82**:1730–1743.
Nguyen D, Boberg J, Cleary M, Bruelheide H, Hönig L, Koricheva J, Stenlid J. 2017. Foliar fungi of Betula pendula: impact of tree species mixtures and assessment methods. *Sci Rep* **7**:1–11.
Nielsen JC, Grijseels S, Prigent S, Ji B, Dainat J, Nielsen KF, Frisvad JC, Workman M, Nielsen J. 2017. Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in Penicillium species. *Nat Microbiol* **2**:1–9.
O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N, others. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat Genet* **44**:1060–1065.
Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, others. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog* **8**:e1003037.
Olson Å, Aerts A, Asiegbu F, Belbahri L, Bouzid O, Broberg A, Canbäck B, Coutinho PM, Cullen D, Dalman K, others. 2012. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytol* **194**:1001–1013.
Pinna C, Linaldeddu BT, Deiana V, Maddau L, Montecchio L, Lentini A. 2019. Plant pathogenic fungi associated with *Coraebus florentinus* (Coleoptera: Buprestidae) attacks in declining oak forests. *Forests* **10**:488.
Prospero S, Cleary M. 2017. Effects of host variability on the spread of invasive forest diseases. *Forests* **8**:80.
Prospero S, Conedera M, Heiniger U, Rigling D. 2006. Saprophytic activity and sporulation of *Cryphonectria parasitica* on dead chestnut wood in forests with naturally established hypovirulence. *Phytopathology* **96**:1337–1344.
Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* **10**:417–430.
Rambaut A, Drummond AJ. 2016. FigTree v1. 4.3. 2016.
Rice P, Bleasby A, Ison J. 2011. EMBOSS user's guide: practical bioinformatics.
Rigling D, Prospero S. 2018.s*Cryphonectria parasitica*, the causal agent of chestnut blight: Invasion history, population biology and disease control. *Mol Plant Pathol* **19**:7–20. doi:10.1111/mpp.12542
Robinson SL, Christenson JK, Wackett LP. 2019. Biosynthesis and chemical diversity of β-lactone natural products. *Nat Prod Rep* **36**:458–475.
Rodriguez-Moreno L, Ebert MK, Bolton MD, Thomma BPHJ. 2018. Tools of the crook-infection strategies of fungal plant pathogens. *Plant J* **93**:664–674.
Rokas A, Mead ME, Steenwyk JL, Oberlies NH, Goldman GH. 2020. Evolving moldy murderers: *Aspergillus* section *Fumigati* as a model for studying the repeated evolution of fungal pathogenicity. *PLoS Pathog* **16**:e1008315.
Sakalidis ML, Hardy GES, Burgess TI. 2011. Endophytes as potential pathogens of the baobab species*Adansonia gregorii*: a focus on the Botryosphaeriaceae. *Fungal Ecol* **4**:1–14.
Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. 2018. The genome biology of effector gene evolution in filamentous plant pathogens. *Annu Rev Phytopathol*.

Santos MS, Orlandelli RC, Polonio JC, dos Santos Ribeiro MA, Sarragiotto MH, Azevedo JL, Pamphile JA. 2017. Endophytes isolated from passion fruit plants: molecular identification, chemical characterization and antibacterial activity of secondary metabolites. *J Appl Pharm Sci* **7**:38–43.
Schalchli H, Hormazabal E, Becerra J, Birkett M, Alvear M, Vidal J, Quiroz A. 2011. Antifungal activity of volatile metabolites emitted by mycelial cultures of saprophytic fungi. *Chem Ecol* **27**:503–513.
Scharf DH, Heinekamp T, Brakhage AA. 2014. Human and plant fungal pathogens: the role of secondary metabolites. *PLoS Pathog* **10**.
Shwab EK, Keller NP. 2008. Regulation of secondary metabolite production in filamentous ascomycetes. *Mycol Res* **112**:225–230.
Smahi H, Belhoucine-Guezouli L, Bouhraoua RT, Franceschini A, Linaldeddu BT. 2018. First report of branch canker and dieback caused by *Cryphonectria naterciae* on *Quercus suber* in Algeria. *Plant Dis* **102**:251.
Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM, Butcher SA, Gurr SJ, others. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science (80-)* **330**:1543–1546.
Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol* **19**:2094–2110.
Sprockett DD, Piontkivska H, Blackwood CB. 2011. Evolutionary analysis of glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions in necrotrophic fungal pathogens. *Gene* **479**:29–36.
Srinivasan A. 2019. gread: Fast Reading and Processing of Common Gene Annotation and Next Generation Sequencing Format Files.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313.
Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**:637–644.
Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**:62.
Stauber L, Badet T, Prospero S, Croll D. 2020. Emergence and diversification of a highly invasive chestnut pathogen lineage across south-eastern Europe. *bioRxiv*.
Stergiopoulos I, de Wit PJGM. 2009. Fungal effector proteins. *Annu Rev Phytopathol* **47**:233–263.
Stukenbrock EH, Bataillon T, Dutheil JY, Hansen TT, Li R, Zala M, McDonald BA, Wang J, Schierup MH. 2011. The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Res* **21**:2157–2166.
Suzuki H, MacDonald J, Syed K, Salamov A, Hori C, Aerts A, Henrissat B, Wiebenga A, Barry K, Lindquist E, others. 2012. Comparative genomics of the white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. *BMC Genomics* **13**:444.
Tan K-C, Oliver RP, Solomon PS, Moffat CS. 2010. Proteinaceous necrotrophic effectors in fungal virulence. *Funct Plant Biol* **37**:907–912.
Trapero-Casas A, Kaiser WJ. 2009. Alternative hosts and plant tissues for the survival, sporulation and spread of the Ascochyta blight pathogen of chickpea. *Eur J Plant Pathol* **125**:573.
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**:562–578.
van der Merwe NA, Gryzenhout M, Steenkamp ET, Wingfield BD, Wingfield MJ. 2010. Multigene phylogenetic and population differentiation data confirm the existence of a cryptic species within *Chrysoporthe cubensis*. *Fungal Biol* **114**:966–979.
Vaz ABM, Fonseca PLC, Badotti F, Skaltsas D, Tomé LMR, Silva AC, Cunha MC, Soares MA, Santos VL, Oliveira G, others. 2018. A multiscale study of fungal endophyte communities of the foliar endosphere of native rubber trees in Eastern Amazon. *Sci Rep* **8**:1–11.
Wang C, Wang S. 2017. Insect pathogenic fungi: genomics, molecular interactions, and genetic improvements. *Annu Rev Entomol* **62**:73–90.
Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva E V, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**:543–548.
Wickham H. 2018. stringr: Simple, Consistent Wrappers for Common String Operations.
Wickham H. 2017. tidyverse: Easily Install and Load the "Tidyverse."
Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham H. 2007. Reshaping Data with the reshape Package. *J Stat Softw* **21**:1–20.
Yin J, Zhang C, Mirarab S. 2019. ASTRAL-MP: Scaling ASTRAL to very large datasets using randomization and parallelization. *Bioinformatics* **35**:3961–3969.
Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**:W445-W451.
Yin Z, Liu H, Li Z, Ke X, Dou D, Gao X, Song N, Dai Q, Wu Y, Xu J-R, others. 2015. Genome sequence of Valsa canker pathogens uncovers a potential adaptation of colonization of woody bark. *New Phytol* **208**:1202–1216.
Zakaria L, Yaakop AS, Salleh B, Zakaria M. 2010. Endophytic fungi from paddy. *Trop life Sci Res* **21**:101.
Zhang W, Zhang X, Li K, Wang C, Cai L, Zhuang W, Xiang M, Liu X. 2018. Introgression and gene family contraction drive the evolution of lifestyle and host shifts of hypocrealean fungi. *Mycology* **9**:176–188.
Zhao Z, Liu H, Wang C, Xu J-R. 2014. Erratum to: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **15**:6.

Chapter 2

Emergence and diversification of a highly invasive chestnut pathogen lineage across south-eastern Europe

Lea Stauber[1,2], Thomas Badet[2], Alice Feurtey[2,3], Simone Prospero[1], Daniel Croll[2]

[1] Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

[2] Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

[3] Plant Pathology, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland

Revisions submitted to eLife

Previous manuscript version on *bioRxiv* doi: https://doi.org/10.1101/2020.02.15.950170

Data availability: All raw sequencing data is available on the NCBI Short Read Archive (BioProject PRJNA604575 and PRJNA644891)

Abstract

Invasive microbial species constitute a major threat to biodiversity, agricultural production and human health. Invasions are often dominated by one or a small number of genotypes, yet the underlying factors driving invasions are poorly understood. The chestnut blight fungus *Cryphonectria parasitica* first decimated the North American chestnut and a more recent outbreak threatens Europe chestnut stands. To unravel the chestnut blight invasion of south-eastern Europe, we sequenced 230 genomes of predominantly European strains. Genotypes outside of the invasion zone showed high levels of diversity with evidence for frequent and ongoing recombination. The invasive lineage emerged from the highly diverse European genotype pool rather than a secondary introduction from Asia. The expansion across south-eastern Europe was mostly clonal and is dominated by a single mating type suggesting a fitness advantage of asexual reproduction. Our findings show how an intermediary, highly diverse bridgehead population gave rise to an invasive, largely clonally expanding pathogen.

Introduction

Over the past century, a multitude of invasive species have emerged as threats to forest and agricultural ecosystems worldwide (Santini et al., 2013; Wingfield et al., 2010). Increased human activities, such as global trade, enabled invasive species to cause economic damage through reduced agricultural production, degradation of ecosystems and negative impacts on human health (Marbuah et al., 2014). A key group of invasive species in forests are fungal pathogens, which are often accidentally introduced via living plants and plant products (Rossman, 2001). To successfully colonize a new environment, fungal pathogens have to overcome several invasion barriers including effective dispersal abilities, changes in available hosts, competition with other fungi, and niche availability (Gladieux et al., 2015; Hayes and Barry, 2008). This may be achieved with plastic phenotypic changes followed by rapid genetic adaptation (Garbelotto et al., 2015). The invasive potential is also influenced by pre-adaptation arising outside of the invasion zone. Pre-adaptation can be a critical factor for environmental filtering by abiotic or host-related factors in the new environment (Krasnov et al., 2015). As the number of initial founders is often low, invasive populations are frequently of low genetic diversity which reduces adaptive genetic variation (Allendorf and Lundquist, 2003; Yang et al., 2012). Yet, many fungal plant pathogen invasions were successful despite low genetic diversity within founding populations (Fontaine et al., 2013; Raboin et al., 2007; Wuest et al., 2017).

A major model explaining the successful expansion of invasive populations despite low initial genetic diversity is the so-called bridgehead effect (Lombaert et al., 2010). In this model, highly adapted lineages emerge through recombination among genotypes established in an area of first introduction. Hence, the primary introduction serves as the bridgehead for a secondary and more expansive invasion. Although the bridgehead effect has been proposed as a scenario for many biological invasions (Gau et al., 2013; van Boheemen et al., 2017), empirical evidence for the creation of highly adaptive genotypes within bridgehead populations is still largely missing (Bertelsmeier and Keller, 2018). An alternative explanation suggests that primary bridgehead populations simply serve as repeated sources of inoculum for secondary invasions without selecting for the invasive (*i.e.* adapted) genotypes *in situ* (Bertelsmeier and Keller, 2018). This alternative scenario implies that initial populations are already composed of genotypes adapted to the new environment or have high phenotypic plasticity (Bock et al., 2015; Gladieux et al., 2015; Vuković et al., 2019). A conceptually similar model to the bridgehead effect was proposed for bacteria, where the success of epidemic clones would be contingent on the generation of recombinants in the original populations (Smith et al., 1993). Dissecting whether invasive species were pre-adapted to the new environment or gained adaptation through a bridgehead effect is crucial for effective containment strategies, but requires a deep sampling of genotypes during the early invasion process.

The pace of adaptive evolution leading to successful invasions is determined by several life-history traits including the mating system (*i.e.* sexual, asexual or mixed). Sexual reproduction can potentially promote the emergence of more virulent strains, enhancing invasive success (Philibert et al., 2011). In contrast, a switch to asexual reproduction can limit genetic diversification and adaptive potential in newly introduced species (Drenth et al., 2019; Taylor et al., 2015). However, the low availability of mating partners at the invasion front can be a major cost for sexual reproduction. To avoid the cost of mate search, invasive fungal pathogens often switch from sexual to predominantly asexual reproduction (Heitman et al., 2013; Suehs et al., 2004). Consequently, asexual reproduction can promote rapid colonization and dissemination of adapted genotypes (Gladieux et al., 2011). Beyond invasive species, many fungal pathogens undergo asexual reproduction during the colonization of habitats, but switch to sexual reproduction for the production of survival propagules prior to unfavorable environmental conditions (Schoustra et al., 2010). Populations of some highly successful invaders, such as the ascomycete *Ophiostoma novo-ulmi* causing Dutch elm disease (Paoletti et al., 2006), or the oomycete *Phytophthora ramorum* (Grünwald et al., 2012), are dominated by a single mating type. Low diversity in invasive lineages may be breached by secondary invasions introducing the opposite mating type as observed in *Phytophthora infestans* in Europe. Prior to the 1980s only the presence of the A1 mating type was clearly established, followed by the secondary introduction of the A2 mating type leading to sexual reproduction (Mariette et al., 2016). Introgression from closely related species can also reintroduce a missing mating type. *O. novo-ulmi* acquired the missing mating type from *O. ulmi* (Brasier and Webber, 2019; Paoletti et al., 2006). Although switches in reproductive modes can be a key factor for invasion success (Philibert et al., 2011), mechanisms underlying such switches remain poorly understood (Billiard et al., 2012). The reproductive mode is the primary factor shaping genetic diversity in invasive pathogen populations. Beyond the impact on genetic diversity, reproductive modes are often tied to specific environmental adaptations including the mode of spore dispersal. Sexual spores are often highly durable and can be wind dispersed over long distances (Philibert et al., 2011; Rigling and Prospero, 2018). Even rare long-distance dispersal events can contribute to the geographic expansion of invasive pathogens (Prospero and Cleary, 2017). In contrast, asexual spores serve mostly short distance dispersal in fungal pathogens with mixed reproduction (Mehrotra, 2013; Prospero and Cleary, 2017).

Some invasive pathogens were discovered to harbor specific characteristics in the architecture of the genome, which may have facilitated the generation of adaptive genetic variation. For instance, genomes of *Ophiostoma* species causing Dutch elm disease show signals of frequent hybridization and introgression events, giving rise to lineages with increased virulence or growth at higher temperatures (Hessenauer et al., 2020). Genomes of a number of crop pathogens including the oomycete *Phytophthora infestans* and the fungal pathogens *Leptosphaeria maculans* and *Zymoseptoria tritici* show compartmentalization into gene-dense and gene-poor regions with virulence factor being preferentially encoded in gene-poor regions (Dong et al., 2015; Frantzeskakis et al., 2019; Torres et al., 2020). The gene-poor regions are enriched in transposable elements (TEs) and the encoded virulence factors (*i.e.* effectors) show epigenetic regulation, rapid sequence turnover and

weak phylogenetic conservation (Fouché et al., 2018). In pathogen systems such as *Cryptococcus gattii* or *P. infestans*, copy number variation has created highly successful outbreak lineages (Knaus et al., 2020; Steenwyk et al., 2016). Beyond potential roles in adaptation, the genome structure of plant pathogens is likely influenced by neutral processes including migratory bottlenecks possibly affecting the activity of TEs (Oggenfuss et al., 2020). Whether virulence factors are associated with genome compartmentalization, and how chromosomal rearrangements could underpin the creation of invasive pathogen genotypes remain largely unknown.

The ascomycete *Cryphonectria parasitica* (Murr.) Barr. is the causal agent of chestnut blight, a lethal bark disease of *Castanea* species (Rigling and Prospero, 2018). The pathogen is native to eastern Asia and was first observed in 1904 in North America on the American chestnut (*Castanea dentata*). In the following years, the disease rapidly spread throughout the distribution range of *C. dentata*, causing the vast decimation of this native tree species (Elliott and Swank, 2008). In 1938, the fungus was first detected on the European chestnut (*C. sativa*) near Genoa (Italy) and is now established in all major chestnut-growing countries in Europe (Rigling and Prospero, 2018). The damage to the European chestnut may have been reduced by the presence of the *Cryphonectria hypovirus 1* (CHV1) which acts as a biological control agent of chestnut blight (Rigling and Prospero, 2018). The virus can be transmitted both vertically to asexual spores (conidia) or horizontally through hyphal anastomoses between virus-infected and virus-free strains (Heiniger and Rigling, 1994). Hyphal anastomoses are controlled by a vegetative compatibility system and the virus spreads most effectively between strains of the same vegetative compatibility type (Cortesi et al., 2001).

Population genetic analyses showed that the invasion of Europe occurred through multiple introductions both from native populations in Asia and from bridgehead populations in North America (Dutech et al., 2012). Invasive European *C. parasitica* populations are characterized by lower vegetative compatibility diversity than North American and Asian populations (Milgroom and Cortesi, 1999). Furthermore, European populations exhibit lower recombination rates (Dutech et al., 2010; González-Varela et al., 2011; Prospero and Rigling, 2012). Mating in *C. parasitica* is controlled by a single mating type (MAT) locus (Marra and Milgroom, 2001) and natural populations can reproduce both sexually and asexually (Marra et al., 2004). Previous analyses of south-eastern European *C. parasitica* populations based on sequence characterized amplified region (SCAR) markers suggested that the region was largely colonized by a single and likely asexual lineage also identified as S12 (Milgroom et al., 2008). The lineage belongs to the vegetative compatibility type EU-12. Within the distribution range of the lineage, sexual structures (*i.e.* perithecia) are rarely found (Milgroom et al., 2008; Sotirovski et al., 2004). Based on SCAR marker and field records, Milgroom et al. (2008) suggested that the invasive S12 lineage originated in northern Italy and a subsequently spread across south-eastern Europe (Avolio, 1978; Biraghi, 1946; Buccianti and Feliciani, 1966; Karadžić et al., 2019; Myteberi et al., 2013; Robin and Heiniger, 2001). However, due to the low molecular marker resolution and challenges in relying on observational data, the origin, invasion route and genetic diversification of *C. parasitica* in south-eastern Europe remain largely unknown.

In this study, we sequenced complete genomes of a comprehensive collection of European *C. parasitica* strains with a fine-scaled sampling throughout the S12 invasion zone. Based on high-confidence genome-wide single nucleotide polymorphisms (SNPs), we identified the most likely origin and recapitulated the invasion process across south-eastern Europe. We show that the invasive *C. parasitica* lineage arose through an intermediary, highly diverse bridgehead population. During the expansion, the lineage became dominated by a single mating type but retained the ability to reproduce sexually.

Results

Genome-wide polymorphism analyses and phylogenomic reconstruction

We analyzed complete genomes of 230 *C. parasitica* isolates covering the global chestnut blight distribution range, as well as the European outbreak region of the invasive S12 lineage (Fig. 1A, Supp. Table 1). Isolates were sequenced at a mean depth of 8–53X to detect high-confidence genome-wide SNPs. A region of 179'501 -2'084'312 bp on scaffold 2 was associated to the mating type locus based on association mapping *p*-values (Supp. Fig. 2). This region encoded known mating type associated genes and was characterized by a high SNP density consistent with observations in other fungi (Idnurm et al., 2015; Taylor et al., 2015). We removed SNPs within the mating type associated region to avoid confounding genetic structure with mating type divergence. We retained 80'700 SNPs and performed a principal component analysis (PCA) and constructed a SplitsTree phylogenetic network (Fig. 1B-C). The PCA revealed four major clusters (CL1-CL4, Fig. 1B), showing weak geographic clustering except for isolates of Chinese origin (cluster CL4 in Fig. 1B). The largest cluster contained the majority of sequenced European isolates, including the S12 lineage, as well as all North American and two Georgian isolates (cluster CL1 in Fig. 1B). The phylogenetic network analysis revealed large phylogenetic distances outside the main European/North American CL1 cluster (Fig. 1B-C).

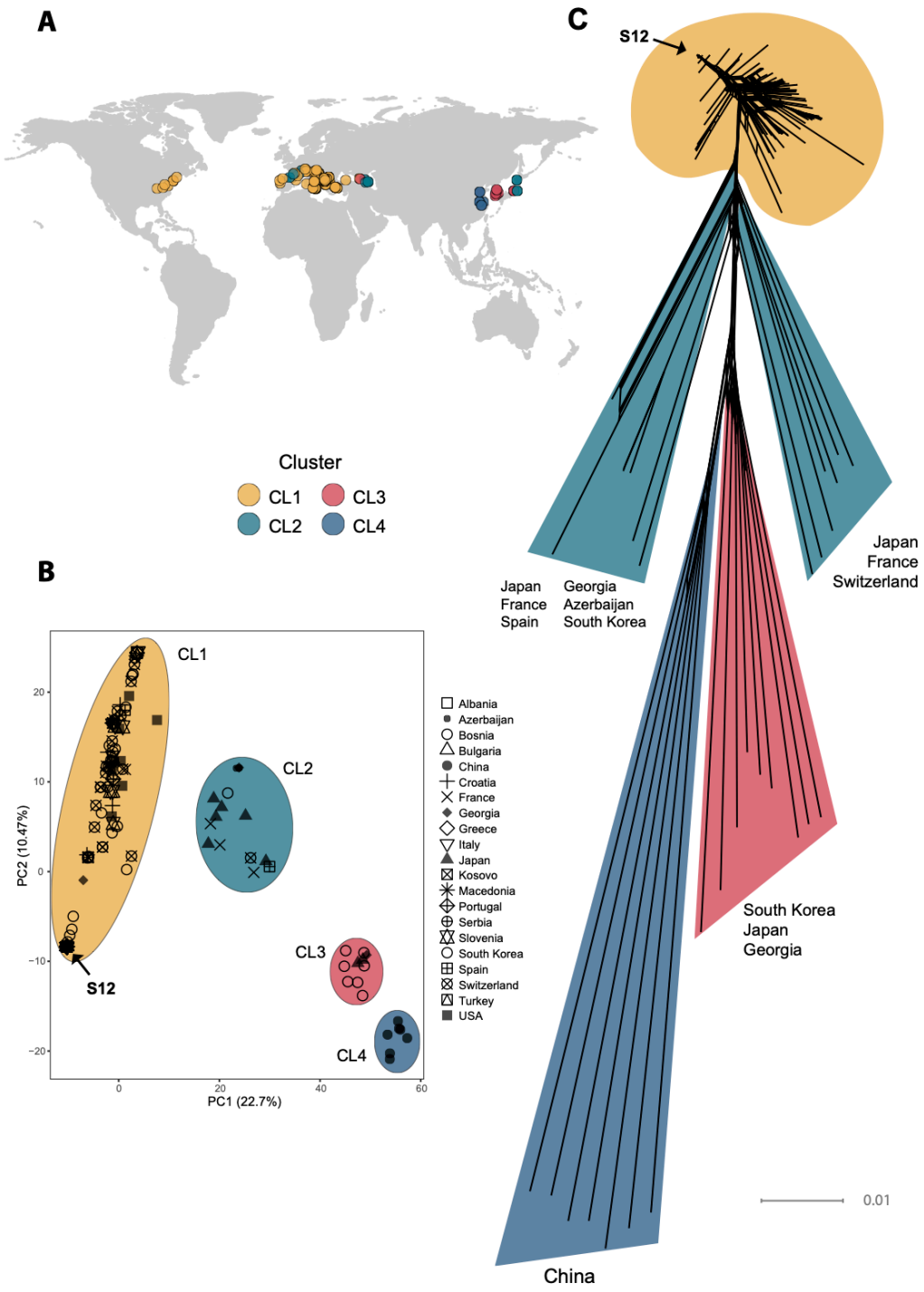


Figure 1: **Genome-wide analyses of global *Cryphonectria parasitica* isolates**. A) Map of global sampling locations and B) Principal component analysis (PCA) of the 230 sequenced *C. parasitica* isolates. Colors indicate PCA clusters (CL1-CL4) and are as in A). C) SplitsTree phylogenetic network of the global *C. parasitica* sample set. Colors are as in A) and B). Isolates belonging to the S12 lineage are marked with an arrow in B) and C).

We implemented a demographic model based on each cluster identified by the PCA (Fig. 1B), merging isolates from the two smaller clusters (*i.e.* CL3 and CL4, Fig. 1B) to obtain a sufficient sample size. The folded site frequency spectrum (SFS) for each of the three groups contained an excess of singletons suggesting a population expansion (Fig. 2). We used the software *dadi* (Gutenkunst et al., 2009) to identify the best fitting model considering a neutral model, as well as population expansions and contractions. For the isolates outside the main European/North American group, we found support for a recent population expansion (CL2-CL4 clusters, Fig. 1B, Fig. 2). However, the large European/North American CL 1 cluster could not be fitted to any particular demographic model most likely due to the unusual shape of the frequency spectra. As shown in Fig. 2, the observed folded SFS for the European/North American cluster (CL1 Fig. 1B, center row in Fig. 2) is W-shaped (red line in Fig. 2), instead of a monotonously declining curve as is the case in the empirical spectra for other populations as well as in the simulations. (W-shaped red line, Fig. 2).

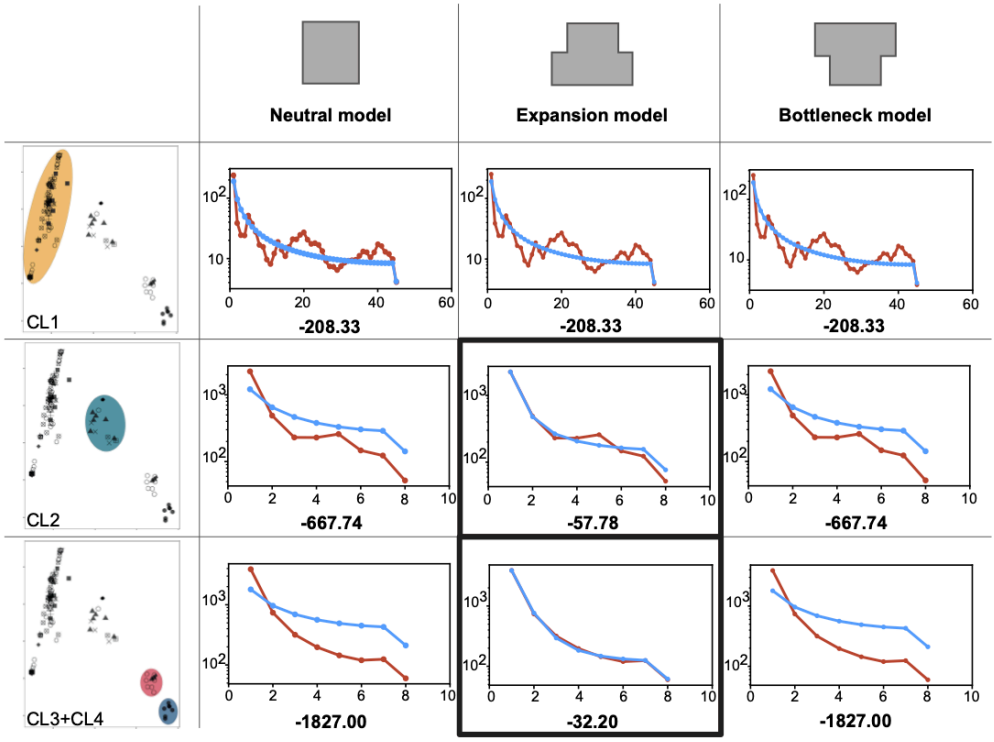


Figure 2: **Demographic models for global *Cryphonectria parasitica* populations.** Neutral, expansion and bottleneck models were tested for three clusters (*i.e.* European/North American CL1, mixed European/Asian CL2 and Asian CL3 and CL4 clusters shown by color highlights). Clusters were inferred by PCA (Fig. 1B). The two smaller Asian clusters (bottom left PCA panel) highlighted in red and blue were combined for demographic analyses to obtain a sufficient sample size. Panels in columns 2-4 show model-data comparison plots for the three tested scenarios (*i.e.* neutral, expansion and bottleneck) as inferred by *dadi*. Values at the bottom of each panel represent likelihood scores. Blue data points and line represent the model, red lines show the empirical data. Best fitting models with highest likelihood values are marked in black boxes.

To further investigate the European/North American CL1 cluster containing the S12 lineage (Fig. 1B), we performed a SplitsTree phylogenetic network analysis to account for reticulation caused by recombination in this cluster specifically. The network showed a substantial diversification, with both long branching and reticulation (Fig. 3A, Supp. Fig. 3). Genotypes within the cluster show significant evidence for recombination (PHI-test; $p < 0.0001$). Despite the high level of genetic diversity, we found no evidence for geographic structure. Moreover, we found no clustering of isolates belonging to the same vegetative compatibility type with the exception of some EU-01, EU-02 and EU-12 (S12) genotypes of predominantly Balkan origin (Fig. 3A, Supp. Fig. 3). Nearly all *C. parasitica* isolates representing the S12 lineage showed almost identical genotypes and tight clustering. The most tightly clustered S12 genotypes were all of mating type MAT-1 ($n = 105$). Consistent with analyses by (Milgroom et al., 2008), this group represents the invasive S12 lineage at the origin of the expansion of *C. parasitica* across south-eastern Europe. Additionally, the phylogenetic network revealed closely related but not identical S12 genotypes of mating type MAT-2 ($n = 7$, Fig. 3A). Hence, S12 outbreak strains of MAT-2 connect the nearly uniform cluster of S12 MAT-1 strains with the remaining genetic diversity of the major European subgroup of *C. parasitica*. The S12 cluster was furthermore connected with the remaining genotypes of the major clade by two EU-12 isolates from Bosnia (M1808 with MAT-1) and Georgia (MAK23 with MAT-2) (Fig. 3A, Supp. Fig. 3). Consistent with previous findings (Milgroom et al., 2008), we also found non-S12 genotypes in the invasion zone of the S12 outbreak including the vegetative compatibility types EU-01 and EU-02 in southern Italy, Greece and Turkey (Fig. 3A-B). Additionally, we found the S12 lineage in otherwise more diverse regions of Croatia and Bosnia (Fig. 3A-B). Surprisingly, we also detected one Georgian (GEZ45) and one Portuguese (M2135) isolate within the S12 cluster suggesting human mediated dissemination of the lineage outside of south-eastern Europe (Fig. 3B).

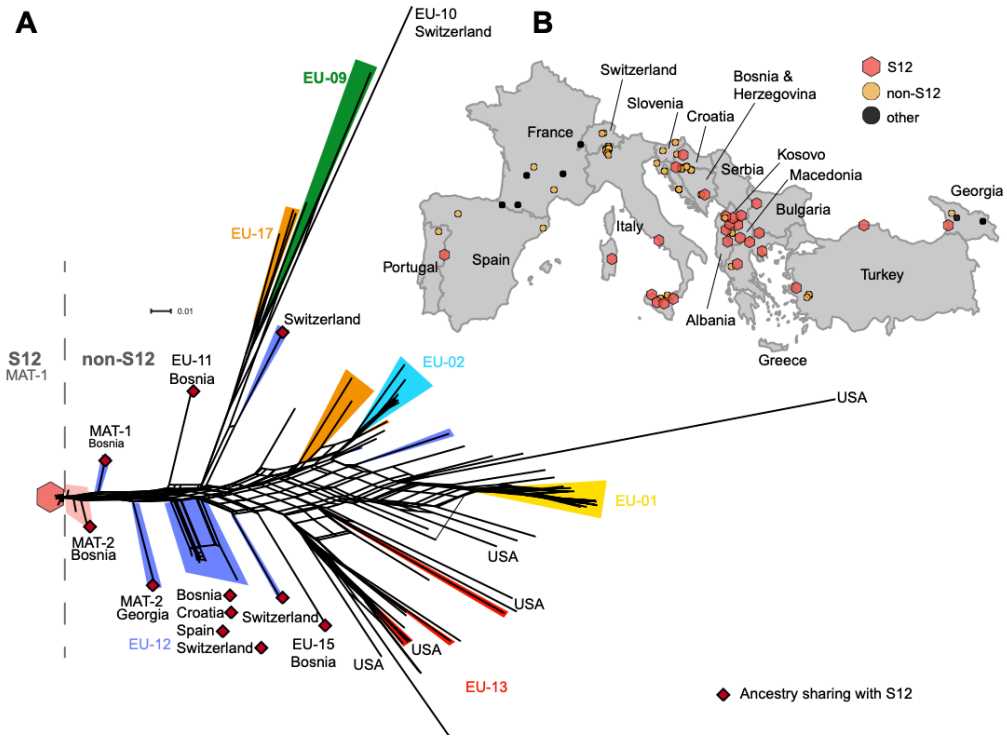


Figure 3: **Phylogenetic network structure of the main European/North American subgroup and spatial distribution of S12 across Europe.** A) The highlighted branches represent the most abundant vegetative compatibility types. Isolates belonging to the S12 outbreak lineage (EU-12; mating type MAT-1, $n = 105$) are marked with a red hexagon and match symbols in B). S12 isolates of mating type MAT-2 are highlighted in light red. Additional EU-12 isolates not belonging to the S12 lineage are highlighted in blue with information on the country of origin. Isolates sharing ancestry with the S12 lineage as inferred by fineSTRUCTURE (Fig. 4, Supp. Fig. 4) are marked with dark red diamonds. B) Geographic map of European sampling locations. Light red hexagons mark locations where S12 was found, yellow circles indicate locations containing other genotypes from the main European/North American CL1 cluster (Fig. 1B). Black circles show the location of highly distinct genotypes outside of the CL1 cluster (Fig. 1A-B).

Analyses of coancestry for the invasive lineage

The SplitsTree network revealed that the invasive S12 lineage has closely related genotypes occurring in Europe. Thus, to dissect the genetic ancestry of S12, we performed a coancestry matrix analysis using fineSTRUCTURE (Lawson et al., 2012) considering all isolates of the major European/North American cluster (CL1, Fig. 1B), including the S12 lineage ($n = 197$). Within this cluster we identified K=43 populations, showing weak geographic structure (Supp. Fig. 4). The highest level of coancestry for any given S12 isolate was found within the S12 population (Fig. 4). Moreover, we identified the highest degree of shared ancestry between S12 and populations from the northern Balkans (Bosnia and Croatia), southern Switzerland, but also Georgia and Spain, with coancestry values between 35.3-49.7 (Fig. 4, Supp. Fig. 4).

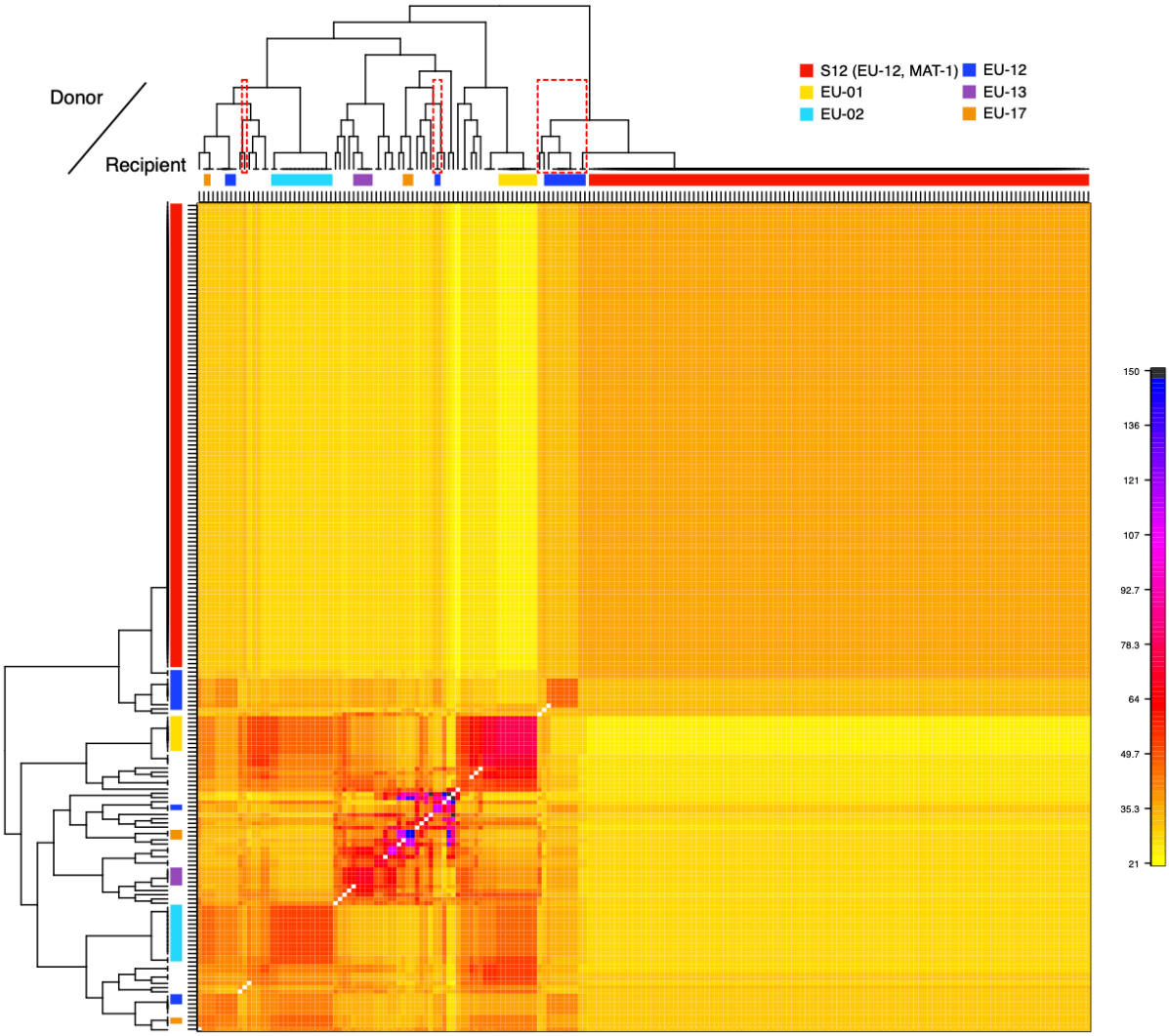


Figure 4: **Analysis of donors to the S12 lineage genotypes**. Averaged co-ancestry matrix and population tree of the North American/European subgroup estimated by fineSTRUCTURE. The heatmap indicates averaged coancestry between populations. Populations sharing ancestry with S12 are marked in red-dashed boxes. S12 ancestry sharing populations are additionally marked in Fig. 3A. Detailed co-ancestry matrix with extensive population information is shown in Supp. Fig. 4.

Transposable element landscape and copy-number variation in the S12 lineage

Invasive pathogen lineages may have undergone crucial genomic rearrangements producing more fit genotypes. Here, we generated a *de novo* identification and annotation of transposable elements (TEs) for the *C. parasitica* genome. We found that 12% of the genome was composed of TEs with striking variation along the assembled scaffolds (*i.e.* quasi-chromosomes; Fig. 5A). In particular, regions on scaffold 2 matching the mating type locus are highly enriched in TEs suggesting that the large non-recombining region has undergone substantial degeneration (Fig. 5A). In contrast, the vegetative incompatibility (*vic*) loci are located in regions devoid of TEs. In fungal pathogens, effector genes and TEs are often co-localized in fast-evolving compartments of the so-called "two-speed genome" (Dong et al., 2015). However, the *C. parasitica* genome shows no apparent compartmentalization into gene-sparse and gene-rich regions. We used machine learning to predict secreted

proteins most likely acting as effectors to manipulate the host. In contrast to other plant pathogens, effector gene candidates showed no tendency to localize in gene-sparse regions of the genome (Fig. 5B). Non-repressed TEs can potentially create additional copies in the genome leading to intra-species variability in TE content. To detect such TE activity, we performed genome-wide scans of *C. parasitica* isolates for presence or absence of TEs based on split read and target site duplication information. At the loci with detectable TE presence/absence polymorphism, we found an over two-fold variation in total TE counts across the genome (Fig. 5C). The total TE count variation among the genetically diverse non-S12 isolates (North America and Europe only, Fig. 3A) was larger than the clonal S12. Nevertheless, the count of TE insertions at polymorphic TE insertion sites ranged between 12-52 among S12 isolates, which was surprisingly high given their recent emergence and extremely high similarity across the genome (Fig. 5C). Across TE insertion loci, we found only a marginally significant difference in the TE frequency among S12 compared to non-S12 isolates (p=0.03, Supp. Fig. 6). This is in strong contrast to segregating SNPs. We found pairwise differences of 0-45 SNPs among S12 isolates and 0-3'770 among non-S12 isolates (test statistic t-test mean difference $p < 0.001$; Supp. Fig. 7). This suggests that TE activity may have continued after the emergence of the S12 lineage and has created *de novo* genetic variation. We also found that the minor allele frequency spectrum of S12 was significantly skewed towards low frequency variants compared to non-S12 (Fisher's exact test, $p < 0.001$; Supp. Fig. 7). The excess of low frequency variants among S12 isolates is consistent with a recent expansion after a bottleneck.

Repetitive sequences such as TEs can trigger non-homologous recombination leading to copy-number variation (CNV) including deletions. We used normalized read coverage to assess copy-number variation across the species including the S12 lineage (Fig. 5A). Compared to the reference genome, we found that most CNVs in the S12 lineage are nearly fixed. Non-S12 isolates segregate most CNVs at intermediate frequencies (comparison between S12 and non-S12; Fisher's exact test $p < 0.001$, Supp. Fig. 6). Genes tend to overlap duplications rather than deletions (odds ratio of 0.37 and 0.13, respectively; two-sided Fisher test, p-value < 0.001). TEs tend to overlap deletions rather than duplications (odds ratio of 5.03 and 1.76, respectively; two-sided Fisher's test, p-value < 0.001) (Fig. 5D). The mating type region on scaffold 2 and the rDNA locus on scaffold 6 show particularly high levels of copy-number variation (Fig. 5A). In a joint analysis of all isolates from the main European/North American cluster (CL1, Fig. 1B), we found that coding sequences overlapping with duplications and deletions are enriched for gene ontology terms associated with protein binding functions and protein phosphorylation activity, respectively (Fig. 5F). Overall, our results show that the S12 lineage underwent specific gene deletion and duplication patterns compared to the broad diversity of non-S12 isolates.

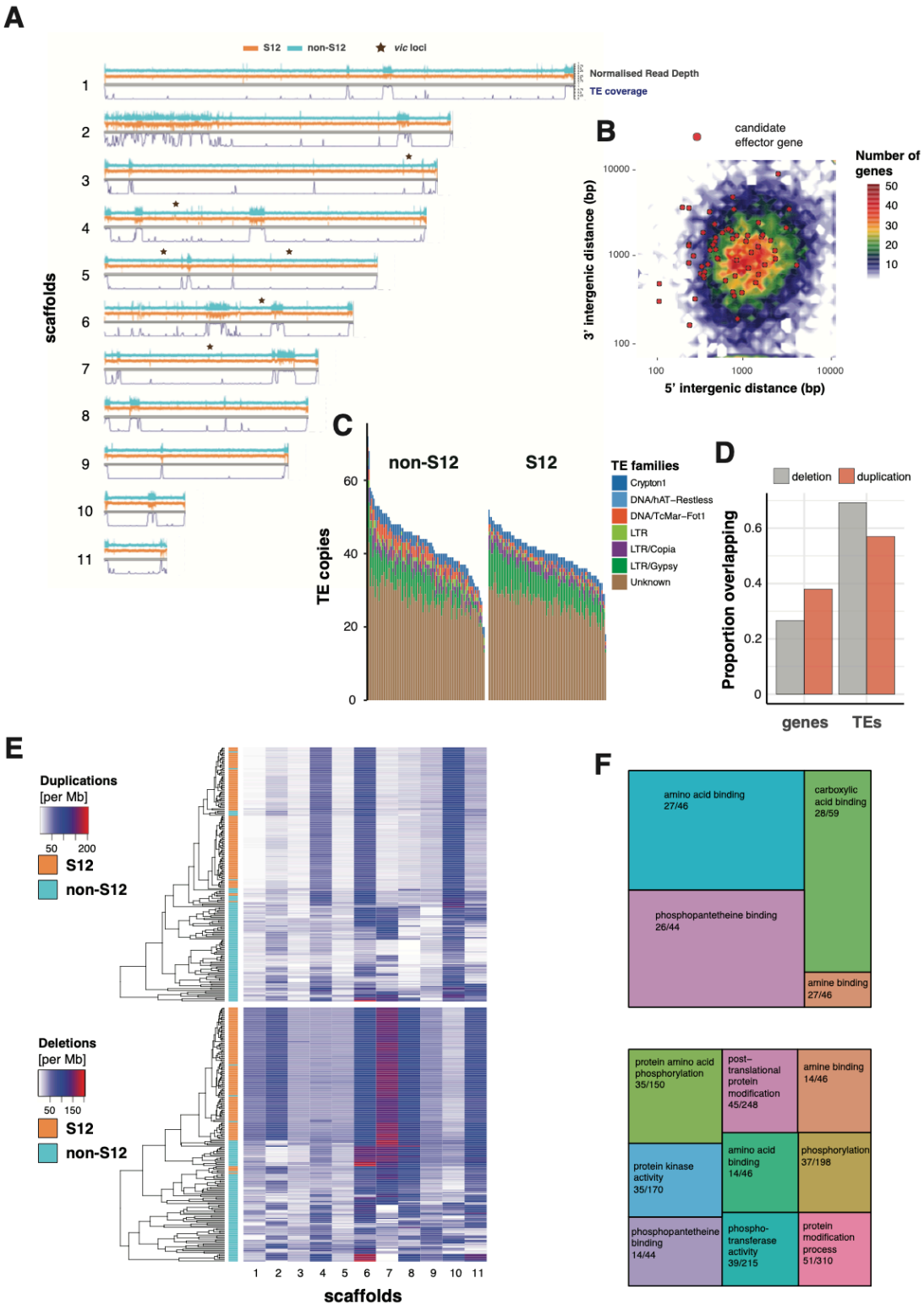


Figure 5: **Transposable element (TE) landscape and copy-number variation.** A) Genome-wide coverage of transposable elements (in 10kb windows) matched by with normalized read depth for the S12 and non-S12 lineages (North America and Europe only). *vic* loci: vegetative incompatibility loci. B) Genome-wide distribution of intergenic distances according to the length of 5' and 3' flanking regions. Red dots represent genes encoding predicted effector proteins. C) Counts of detected TE sequences across S12 and non-S12 isolates

using split reads and target site duplication information. D) Proportion of normalized read depth windows (800 bp) with evidence for duplications (normalized read depth > 1.6) or deletions (< 0.4) overlapping with genes and TEs. E) Heatmap showing the number of windows (800 bp) with duplications and deletions. The dendrogram shows the similarity in duplication or deletion profiles for S12 and non-S12 isolates (North America and Europe only). F) Molecular functions (based on gene ontology) enriched in duplicated and deleted regions (upper and lower panel, respectively). Enrichment was tested by hypergeometric tests and significance was established for a Bonferroni threshold at alpha = 0.05. The numbers represent the number of genes with the matching gene ontology term in a duplicated or deleted region, and across the genome, respectively.

Evolutionary history and evidence for selection in European populations

To gain insights into evolutionary forces shaping polymorphism in the outbreak S12 mating type MAT-1 lineage versus other populations in the European/North American cluster (non-S12, CL1 cluster; Fig. 1B, Fig. 3A), we first analyzed allele frequencies across the genome in both groups (Fig. 6A). The S12 lineage segregated virtually no intermediate allele frequencies in the range of 0.05–0.95. In contrast, the non-S12 genotypes showed overall a wide spectrum of allele frequencies across the genome. Genome-wide nucleotide diversity was extremely reduced in the S12 compared to non-S12 populations (Fig. 6B). We analyzed the predicted impact on protein functions of segregating mutations in the S12 lineage and non-S12 populations (Fig. 6C). We found 4 highly and 104 moderately deleterious mutations segregating within S12 compared to 32 high and 972 moderately deleterious mutations in non-S12 groups (Fig. 6C). Three of the high impact SNPs in the S12 lineage were classified as stop gain mutations, as well as one splice acceptor variant (insertion variant). Two of these high impact mutations affect proteins of the major facilitator superfamily, as well as a protein containing a LCCL domain and an ecdysteroid kinase. Non-S12 populations showed an over-representation of low frequency high-impact mutations (Fisher contingency table test odds ratio of 3.7 and p = 0.002028; Fig. 6C). This is consistent with purifying selection reducing the frequency of these mutations due to fitness costs. Within the S12 lineage nearly all segregating mutations were at very low frequency. We found only modifier (*i.e.* nearly neutral) mutations rising to higher frequency within the lineage. The extremely low level of polymorphism segregating within the S12 lineage prevents strong inferences of selective sweeps since the origin of the lineage. Hence, we focused on potential selective sweeps in the broader European and North American populations. To avoid a bias by the deep sampling of the S12 lineage, we excluded all but one of the S12 MAT-1 isolates (see Methods). We used RAiSD (Alachiotis and Pavlidis, 2018) to produce a composite score of selective sweep signals and identified two strong outlier loci. To control for effects due to the demographic history of these populations, we simulated datasets following the most likely expansion model in addition to the strong bottleneck and neutral models. With 10'000 simulations, the highest values of μ were obtained for the simulations with the expansion scenarios (μ = 37). Consequently, we used this value as a threshold to determine what signatures of selection can be confidently retrieved from the genome scan. Hence, the two outlier loci with μ values above 37 carry most likely true signatures of selection, while loci with lower μ statistic are possibly false positives. The strongest sweep locus is located at the boundary of the mating type locus on scaffold 2 (Fig. 6D). The second sweep locus is on scaffold 1 and encompasses a ~120 kb locus at positions 2'099-3'309 kb. The region contains 335 genes of which 107 encode conserved protein domains (Fig. 6D, Supp. Table 2).

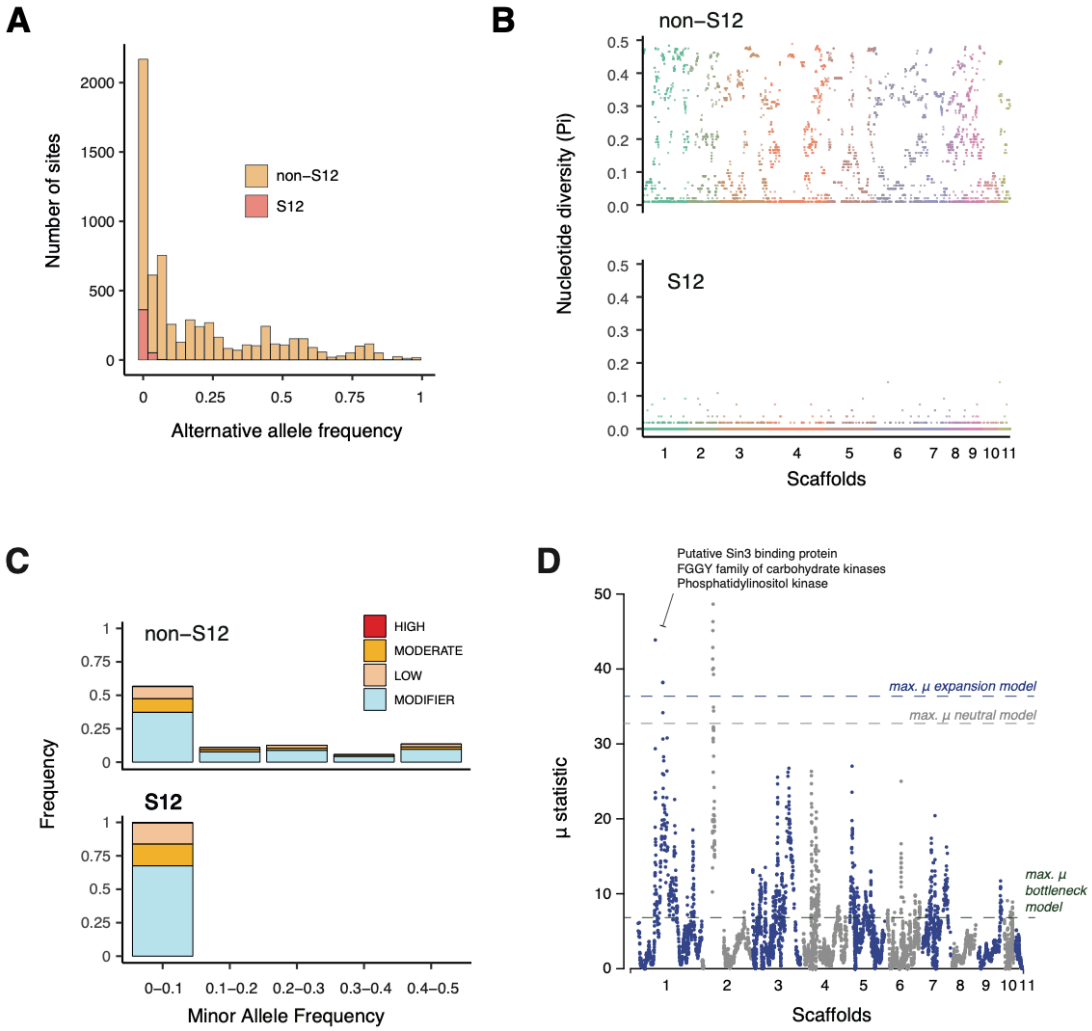


Figure 6: **Polymorphism segregating in the S12 lineage.** A) Alternative allele frequencies spectra across the genome for the S12 lineage (n = 105) compared to all other analyzed European (non-S12; n = 92). B) Genome-wide nucleotide diversity (pi) for the S12 lineage and for the non-S12 lineages in 10 kb windows. C) Minor allele frequency spectra of high, moderate and modifier (*i.e.* near neutral) impact mutations as identified by SnpEff. D) Genome-wide scan for selective sweeps (RAiSD). The three dashed lines show the maximum values observed in the simulated datasets following neutral (grey) expansion (blue) and bottleneck (green) demographic models. Encoded protein functions overlapping with the top regions are shown as summaries.

Retracing S12 invasion routes and retention of mating competence

To infer potential invasion routes of the S12 outbreak lineage, we investigated intra-lineage genetic diversity across south-eastern Europe. We focused only on S12 isolates of mating type MAT-1 to delimit the closest genotypes contributing to the outbreak (n = 105; Fig. 3A). The closely related genotypes segregated 529 high-confidence SNPs across the genome, of which 459 were identified as singletons. The phylogenetic network revealed a star-like structure, consistent with a predominantly clonal population structure (Phi-test p > 0.05; Fig. 7A). The genetic structure assessed by a principal component analysis showed three main groups, of which two groups showed some geographic association (Fig. 7B). The majority of isolates were found in a cluster along the principal component axis 2 (Fig. 7B) without apparent geographic structure within the group. The second group consisted of 8 isolates from

Turkey and one isolate from Georgia consistent with the geographic proximity of the countries (Fig. 7B). The most diverse group consisted exclusively of isolates from Greece showing reticulation in the SplitsTree (Fig. 7A). This suggests that these isolates might have recently undergone sexual reproduction. As we identified non-S12 genotypes in Greece, either within-lineage recombination or exchange with other lineages may underlie the reticulation and increased diversity. We additionally calculated nucleotide diversity for S12 outbreak lineage (n = 105), as well as the two differentiated subgroups identified by PCA, excluding the diverse group of Greek isolates (Fig. 7B). Nucleotide diversity for the S12 outbreak lineage (n = 105) was $\pi = 0.024$, compared to $\pi = 0.022$ for the larger subgroup (n = 89) and $\pi = 0.015$ for the Turkish/Georgian subgroup (n=8).

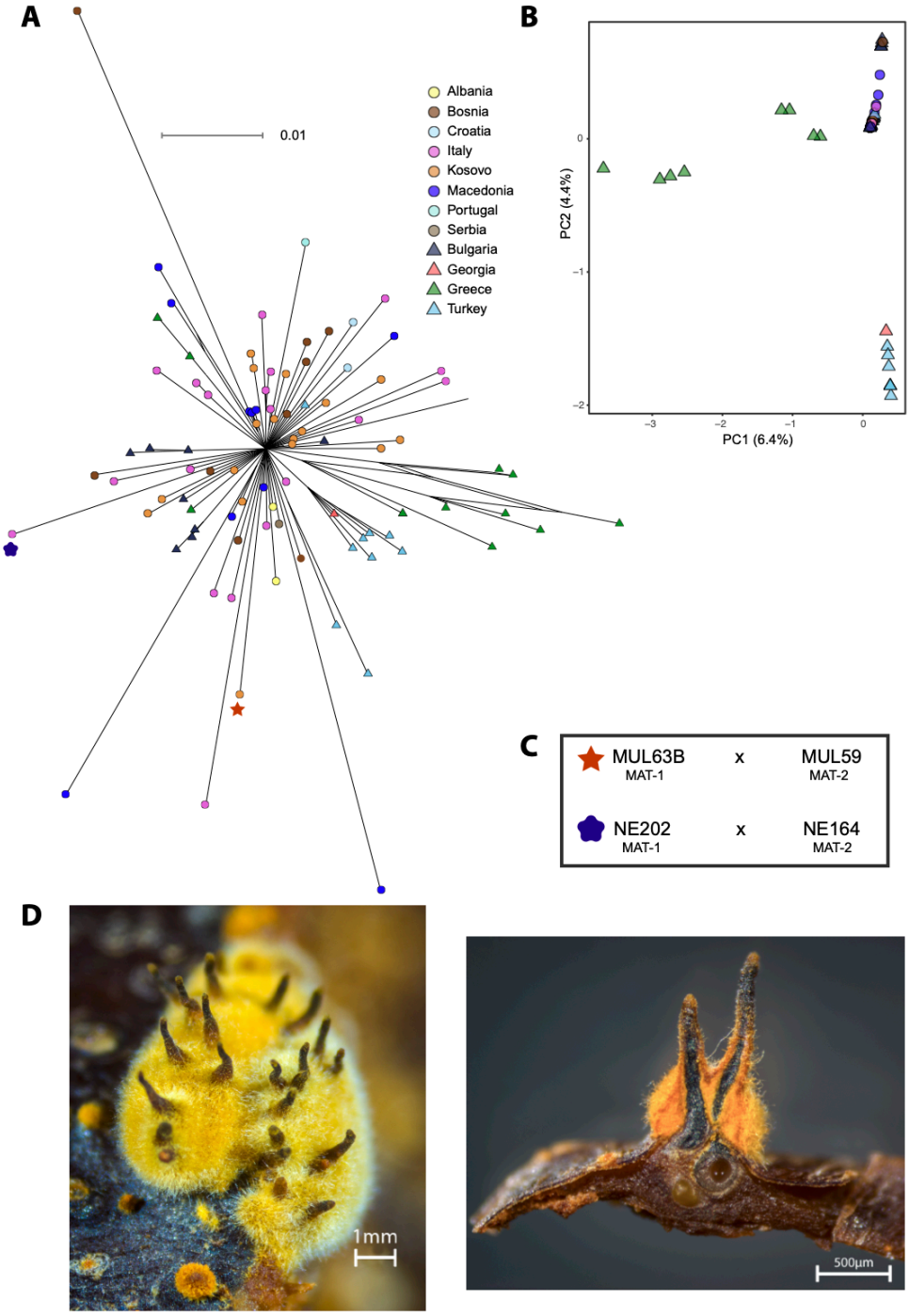


Figure 7: **Fine-scale genetic diversity analyses of the S12 lineage**. A) SplitsTree and B) principal component analysis (PCA) of the S12 mating type MAT-1 outbreak isolates (*n*=105). Symbols and colors are as in A). C) Scheme of successful mating pairs of S12 mating type MAT-1 isolates crossed with isolates from the opposite mating type, of the same geographic origin. Symbols are as in A). D) Photographic images of sexual *C. parasitica* fruiting bodies (*i.e.* perithecia) emerging from crosses of S12 mating type MAT-1 isolates with isolates of the opposite mating type after five months of incubation under controlled conditions. Left: Perithecia embedded in a yellow-orange stromatic tissue. Right: Cross-section of perithecia and chestnut bark. Flask-shaped structures with a long cylindrical neck develop in yellow-orange stromatic tissue and are embedded in

the bark (except for the upper part). The ascospores are formed in sac-like structures (asci) in the basal part of the perithecium. When mature, the ascospores are actively ejected into the air through a small opening (ostiole) at the end of the perithecial neck.

Experimental testing of sexual reproduction in the S12 lineage

We tested experimentally whether S12 mating type MAT-1 isolates were still able to reproduce sexually (Supp. Table 3). We confirmed outcrossing of isolates of opposite mating type within the S12 lineage by pairing isolates from Molliq (Kosovo) and Nebrodi (southern Italy) (Fig. 7C). Mating pairs from Molliq and Nebrodi grew numerous perithecia, which are the fruiting bodies specific to sexual reproduction (Fig. 7D). Pairings of Bosnian isolates showed no perithecia formation. Using molecular mating type assays, we recovered both mating types among the ascospores produced from successful matings.

Discussion

We analyzed the invasion of the chestnut blight pathogen *C. parasitica* across south-eastern Europe after a first establishment on the continent. Using a global collection of isolates, we found evidence of a demographic expansion consistent with the spread to North America and later Europe over the past century. The Europe population is characterized by the presence of a small number of highly distinct genotypes. Deep sampling of recent invasive genotypes showed that the secondary invasion in south-eastern Europe was caused by a single, highly homogeneous lineage consisting nearly exclusively of a single mating type. The invasive lineage likely transitioned recently from biparental mating populations to a single mating type. The spread across south-eastern Europe left no clear geographic imprint. This lack of genetic differentiation along the invasion route may be due to high levels of gene flow after the initial colonization or a very rapid spread from the center of origin to multiple locations across south-eastern Europe. We show that ongoing TE activity created unexpected levels of insertion polymorphism within the invasive lineage. In addition, the lineage carries a distinct gene deletion and duplication profile compared to the European and North American pool of *C. parasitica.*

The establishment of a European bridgehead population

Central and south-eastern European *C. parasitica* populations likely originated from North American lineages. Clusters of European and North American genotypes were largely overlapping indicating significant gene flow and, possibly, multiple introductions to Europe. These genome-scale analyses are consistent with previous findings documenting multiple North American introductions into central Europe, but also directly from Asia to western Europe (Demené et al., 2019; Dutech et al., 2012; Milgroom et al., 1996). Our results show that central and south-eastern European *C. parasitica* populations were largely established from North American sources alone consistent with previous studies. We were unable to sufficiently sample the North American source populations and, hence, demographic history analyses of North American and European populations remained inconclusive. However, the high diversity at the genetic level with polymorphism at the level of SNPs,

TEs and copy-number variation, as well as at the level of vegetative compatibility types suggest that Europe was repeatedly colonized over the past century. Although genetic diversity could also have accumulated *in situ* in Europe, the establishment of a large set of genotypes and vegetative compatibility types seems difficult to explain with population age alone. Sexual recombination between the three most common vegetative compatibility types in Europe (*i.e.* EU-01, EU-02 and EU-05; (Robin and Heiniger, 2001)) could not account for the observed vegetative compatibility type diversity. Observational records date the first introductions into Europe to the 1930s (Robin and Heiniger, 2001). Hence, based on the observed genetic diversity, a scenario of repeated introductions of different vegetative compatibility types since 1930s seems most plausible.

Within Europe, populations from southern Switzerland, Slovenia, Croatia, and Bosnia are highly diverse and have nearly balanced mating type ratios. We also found that no clear genetic structure according to vegetative compatibility types. This strongly suggests frequent outcrossing and population admixture, consistent with reports of perithecia in the field (Ježić et al., 2012; Prospero et al., 2006; Prospero and Rigling, 2012; Trestic et al., 2001). Low vegetative compatibility type diversity in most European *C. parasitica* populations was thought to have contributed to low population admixture within Europe compared to Asia and North America (Cortesi et al., 1996; Dutech et al., 2012; Prospero and Rigling, 2012). However, our genome-wide analyses revealed frequent and ongoing *in situ* admixture in Europe. Thus, vegetative compatibility type diversity does not necessarily underpin population admixture frequency and genetic diversity in sexually recombining populations. Our findings show that in asexually reproducing populations, such as in the S12 lineage, genotypes tend to cluster according to vegetative compatibility types.

Emergence of an invasive lineage from a European bridgehead

The invasive lineage S12 most likely arose from existing genotypes established in Europe. The closest genotypes to the dominant S12 MAT-1 were S12 MAT-2 isolates found in Bosnia, Kosovo and southern Italy. Analyses based on a coancestry matrix identified populations from Bosnia, Croatia, Switzerland and Georgia, but also Spain sharing the highest ancestry with the S12 lineage. This shows that introductions from outside of Europe are unlikely to explain the emergence of S12. The lineage carries a unique set of copy-number variants compared to other European genotypes underlining the observation of a recombinant S12 genotype. Furthermore, the emergence of S12 was accompanied by a striking evolutionary transition from mixed mating type populations to single mating type outbreak populations. Human activity may have contributed to the shift towards single mating type populations. Shipments of infected chestnut seedlings from Northern Italy and other trading activities could have disseminated the invasive lineage further South. This would have exposed the pathogen to the geographically more fragmented chestnut forests typically found in south-eastern Europe where asexuality or selfing may be advantageous in absence of mating partners. Although *C. parasitica* is able to produce asexual conidia in large quantities, these specific spores are thought to be mainly splash dispersed by rain over short distances (Griffin, 1986). Accounting for occasional dispersal by birds or insects (Heald and

Studhalter, 1914), conidia dispersal is unlikely to contribute substantially to the colonization of new areas. Moreover, human-mediated dispersal could have introduced the lineage to other regions including the Iberian Peninsula and Georgia.

Despite the loss of a mating type in the S12 lineage, we found limited evidence for reticulation indicating at least low levels of recombination. If mating in *C. parasitica* follows the canonical process found in many ascomycetes, isolates of opposite mating type are required. Hence, S12 isolates of mating type MAT-1 may sporadically mate with rare S12 isolates of mating type MAT-2, which are comparatively more diverse. The emergence of the opposite mating type at low frequency could be the result of recombination with other genotypes and subsequent backcrossing. Combined with experimental evidence, we show that the dominant S12 mating type MAT-1 has retained the ability for sexual reproduction. Furthermore, in Bosnia, Croatia, Italy (Sicily) and Turkey the S12 lineage co-exists with other genotypes (*i.e.* vegetative compatibility types EU-01 and EU-02) of both mating types, potentially enabling sexual recombination and diversification *in situ*. The invasive S12 lineage was potentially pre-adapted to the south-eastern European niche as we traced the origins to a likely bridgehead population located in southern Switzerland, northern Italy or the northern Balkans. Niche availability and benefits associated with asexual reproduction to colonize new areas may have pre-disposed the European *C. parasitica* bridgehead population to produce a highly invasive lineage. Such potential pre-adaptation to the south-eastern European environmental conditions, as well as reduced niche availability may also explain why genotypes belonging to the S12 lineage have rarely been found elsewhere in Europe.

Expansion and mutation accumulation within the invasive lineage

The S12 lineage diversified largely through mutation accumulation as nearly all high-confidence SNPs were identified as singletons. Mutation accumulation in absence of substantial recombination resulted in star-like phylogenetic relationships. We found a surprisingly high degree of differentiation among S12 genomes at the level of TE insertion polymorphism. Hence, active transposition of TEs is an important factor in diversifying the invasive lineage and possibly underpin future adaptive evolution. We found no evidence for a preferential association of candidate effector genes with TEs. Even though genome compartmentalization is a feature in many fungal genomes, the association of effector genes with repeat-rich regions is not consistent among pathogens (Torres et al., 2020). Necrotrophic pathogens like *C. parasitica* often rely less on effector-based gene-for-gene interactions and tend to deploy toxins or cell-wall degrading enzymes (Friesen et al., 2008; Torres et al., 2020). Interestingly, *C. parasitica* has recently experienced a loss of genes associated with carbohydrate metabolism and cell-wall degradation, which may have reduced exposure to the host immunity system and increased virulence (Stauber et al., 2020). The lack of genes with confirmed pathogenicity functions largely prevents thorough investigations on factors facilitating virulence evolution (*e.g.* epigenetic silencing or sequence rearrangements).

Analyses of allele frequency spectra suggested that the broader European *C. parasitica* populations effectively removed the most deleterious mutations through purifying selection. In contrast, the S12 lineage shows strong skews towards very low minor allele frequencies of all mutation categories. Interestingly, we found a broader spread in allele frequencies for nearly neutral mutations in the S12 lineages. This suggests that despite the largely clonal population structure, deleterious mutations can still be removed through low levels of recombination and purifying selection. Using accumulated mutations as markers to retrace the spatial expansion of the invasive S12 lineage, we found no indication for a step-wise geographic expansion along potential invasion routes. A lack of genetic clustering across south-eastern Europe may be a consequence of high levels of potentially human-mediated gene flow frequently introducing new genotypes over large distances. However, the lack of geographic structure could also have its origins from substantial population bottlenecks during the spread of S12 across south-eastern Europe. Finally, the largely clonal lineage may also become exposed to processes such as Muller's Ratchet fixing deleterious mutations over time (Felsenstein, 1974).

We show how a highly invasive fungal pathogen lineage in Europe can emerge from an intermediate, genetically diverse bridgehead population in Europe. This is in line with the self-reinforcement invasion model where initial introductions promote secondary spread (Bertelsmeier et al., 2018; Garnas et al., 2016). However, empirical evidence for adaptation in bridgehead populations is often elusive. Additionally, human activity (Banks et al., 2015) and host naivety of the European chestnut could have contributed substantially to the rapid spread without the need for diversification and adaptation in the bridgehead population. To date, there is no evidence for the emergence of tolerance or resistance against *C. parasitica* in European chestnut. However, fungal virulence may be severely reduced by the naturally occurring or artificially introduced parasitic mycovirus CHV1 (Rigling and Prospero, 2018). Interestingly, mycovirus spread is most efficient in asexual *C. parasitica* populations such as the invasive S12 lineage. This is because isolates of the same lineage lack vegetative incompatibility barriers among each other. The mycovirus has been reported in south-eastern Europe where and shows population structure in contrast to the fungal host (Bryner and Rigling, 2012). Our study included only CHV1-free S12 isolates, hence potential associations of viral load, genetic identity and geographic structure remain to be investigated. Outcrossing populations often harbor many different vegetative compatibility groups slowing viral transmission (Robin and Heiniger, 2001). Hence, the presence of the mycovirus may confer a selective advantage to retain sexual cycles and diversify vegetative compatibility types. In turn, the diversification may reduce the evolutionary advantage of the invasive lineage over time.

Materials and Methods

Samples of *C. parasitica*

A total of 230 *C. parasitica* were sequenced. A majority of 182 originated from Albania, Bosnia, Bulgaria, Croatia, Greece, Italy, Kosovo, Macedonia, Serbia, Slovenia, Switzerland, and Turkey (Fig. 2B, Supp. Table 1). All 182 isolates originating from central and south-eastern Europe tested to be virulent and mycovirus-free. The 48 other isolates were from China (*n*=7), South Korea (*n*=8), Japan (*n*=8), North America (*n*=8), Azerbaijan (*n*=3), Georgia (*n*=4), Portugal (*n*=3), Spain (*n*=3) and France (*n*=5) (Supp. Table 1). A total of 125 European isolates belonged to the vegetative compatibility type EU-12, whereas 57 isolates represented other vegetative compatibility types (EU-types) occurring in central and south-eastern Europe. We additionally selected 45 isolates from Bulgaria, Greece, Italy and Macedonia, which were already included in a previous population-wide study on south-eastern European *C. parasitica* diversity by (Milgroom et al., 2008). All samples were collected between 1951–2018 and are stored as glycerol stocks at –80°C in the culture collection of the Swiss Federal Research Institute WSL.

DNA extraction and genotyping

All isolates were first inoculated onto cellophane-covered potato dextrose agar plates (PDA, 39 g/L; BD Becton, Dickinson & Company; Franklin Lakes, USA) (Hoegger et al., 2000) and incubated for a minimum of one week at 24°C, at a 14 h light and 10 h darkness cycle. After a sufficient amount of mycelium and spores had grown, the isolates were harvested by scratching the mycelial mass off the cellophane, transferring it into 2 ml tubes and freeze-drying it for 24 h. For DNA extraction, 15–20 mg of dried material was weighted and single tube extraction was performed using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). DNA quantity was measured using the Invitrogen Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham (MA), USA) and DNA quality was assessed using the Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham (MA), USA). Prior to sequencing, all isolates were screened for their genotype at 10 microsatellite markers (Prospero and Rigling, 2012). Additionally, the isolates were screened for their vegetative compatibility and mating type alleles in two multiplex PCRs, as described in Cornejo et al. (2019). Allele sizes for the genotyping and assays were scored with GeneMapper 5 (Thermo Fisher Scientific, Waltham (MA), USA).

Illumina whole-genome sequencing, variant calling and filtration

Isolates were prepared for sequencing using the TrueSeq Nano DNA HT Library Preparation kit (Illumina, San Diego, USA). The libraries were sequenced at the Functional Genomics Centre Zurich (ETH Zurich and University of Zurich) on Illumina platforms (Illumina, San Diego, USA). The obtained sequences were trimmed with Trimmomatic v0.36 (Bolger et al., 2014) and aligned with Bowtie 2 v2.3.5.1 (Langmead and Salzberg, 2012) and SAMtools v1.9 (Li et al., 2009) to the *C. parasitica* reference genome (43.9 Mb) EP155 v2.0 (Crouch et al., 2020) available at the Joint Genome Institute (http://jgi.doe.gov/). Variant calling, selection and filtration

were conducted with the Genome Analysis Toolkit GATK v3.8 and v4.0.2.0 (McKenna et al., 2010). We retained variants satisfying the following filtration parameters: QUAL: ≥ 100, MQRankSum (lower): ≥ -2.0, QD: ≥ 20.0, MQRankSum (upper): ≤ 2.0, MQ: ≥ 20.0, BaseQRankSum (lower): ≥ - 2.0, ReadPosRankSum (lower): ≥ -2.0, ReadPosRankSum (upper): ≤ 2.0, BaseQRankSum(upper): ≤ 2.0 (Supp. Fig. 1). Furthermore, we used BCFtools v1.9 (Narasimhan et al., 2016) and the R package vcfR (Knaus and Grünwald, 2017) for querying and VCFtools v0.1.16 (Danecek et al., 2011) for downstream variant filtering. Variants were additionally filtered for minor allele count (MAC) ≥ 1, excluding all missing data. Sites were also filtered per genotype, only keeping biallelic SNPs with a minimum depth of 3 and a genotyping quality (GQ) of 99. To exclude SNPs associated with the mating type, we ran an association study with TASSEL 5 (Bradbury et al., 2007) and retrieved *p*-values for each SNP across the genome. We set a *p*-value threshold to remove all SNPs with $p \leq 1x10^{-10}$ with the mating type for further analysis. The SNPs showing strong association with the mating type were located on scaffold 2 between positions 100'853–1'997'710 bp (Supp. Fig. 2).

Phylogenetic reconstruction

The filtered whole-genome SNP dataset was used to build unrooted phylogenetic networks using SplitsTree v4.14.6. (Huson and Bryant, 2006). SplitsTree was also used for calculating the PHI test (Bruen et al., 2006) to test for recombination. The required file conversions for and SplitsTree (*i.e.* from VCF to FASTA format) were done with PGDSpider v2.1.1.5 (Lischer and Excoffier, 2011). Principal component analysis (PCA) were performed as implemented in the R package ade4 (Bougeard and Dray, 2018).

Inference of S12 donor populations

We generated an averaged co-ancestry matrix as inferred by fineSTRUCTURE v2.1.3 (Lawson et al., 2012). The software uses a Markov-Chain-Monte-Carlo (MCMC) based algorithm to infer ancestral contributions based on patterns of haplotype similarity. For analysis we selected the unlinked model and generated the required input files ('phase' files) in R. We ran the fineSTRUCTURE pipeline in 'automatic mode', with -s3iters 500000, -s4iters 100000, -maxretained 10000, -s2chunksperregion 20 and ploidy set to 1. Convergence was assessed through MCMC traces (Supp. Fig. 5A) and population assignments (*i.e.* pairwise coincidence matrix; Supp. Fig. 5B).

Population genetic analyses of demography and selection signatures

We computed allele frequencies and estimated the allele spectrum using VCFtools. We used RStudio (RStudio Team, 2015) and ggplot2 (Wickham, 2016) for visualizations. Synonymous and non-synonymous sites were identified and annotated using SnpEff v4.3t (Cingolani et al., 2012). Variants which were classified by SnpEff as having a "high", "moderate", "low" or "modifying" impact on the encoded protein sequences. The data was summarized in R using the packages dplyr (Wickham et al., 2018), reshape2 (Wickham, 2007), tidyr (Wickham and Henry, 2018), as well as and ggplot2. Furthermore, we performed a genome-wide association study

(GWAS) with TASSEL 5 (Bradbury et al., 2007) for identifying highly associated SNPs with genetic groups such as S12. Nucleotide diversity was calculated with vcftools --site-pi function using vcf-files that were converted to diploid.

The demographic history of the populations was investigated based on the folded site frequency spectrum (SFS), using dadi (Gutenkunst et al., 2009). The vcf file was filtered to remove variants predicted to have a "high" to "moderate" effect by SnpEff and thinned with a 5kb distance in order to obtain a set of most likely neutral SNPs. We grouped the isolates of the two smaller clusters of the PCA together (isolates from China, South Korea, Japan and Georgia) in order to obtain a sufficiently large number of samples. For the European cluster, we kept only one randomly selected S12 isolate to avoid a bias from the large number of S12 clones. We compared the observed SFS against the SFS generated under several demographic models: a neutral model (function snm), population contraction and population expansion (function two_epoch).

Identification of putative selective sweeps was performed using RAiSD v2.5 software that is based on three different genetic signatures of positive selection (with -y 1 -M 0 -w 50 -c 1 parameters) (Alachiotis and Pavlidis, 2018). The Manhattan plot shows the distribution of the computed composite μ statistic of positive selection. To investigate potential false positive selection signals due to the demographic history of the populations, we performed 10'000 ms simulations to model three demographic scenarios in the European population: neutral, expansion and bottleneck. We used the average theta value obtained from the demographic simulations with *dadi* on the two populations for which we could estimate the best fitting demographic history model. The recombination rate was estimated with LDhat v2.2 (McVean and Auton, 2007), with -missfreqcut 0, -samp 2000, -its 5000000, -bpen 20 and -burn 50 (Supp. Fig. 8). The average values of phi and of pairwise SNP distance across the 11 scaffolds were used for ms simulations (Hudson, 2002). The bottleneck scenario was simulated given the parameters "ms 10 10000 -t 558 -r 0.0032457 12980 -eN .10000000000000000000 0.01". For the expansion we considered population doubling "ms 10 10000 -t 558 -r 0.0032457 12980 -eN .10000000000000000000 2" and the neutral model was performed given identical theta and phi parameters "ms 10 10000 -t 558 -r 0.0032457 12980". We performed RAiSD selection scans on each of the three simulated datasets of different demographic scenarios and extracted the highest value of the μ statistic.

***De novo* identification of transposable elements and effector prediction**

We performed a *de novo* identification of transposable elements in the *C. parasitica* reference genome EP155 v2.0 using RepeatModeler v2.0.1 (Flynn et al., 2019). The consensus sequences were merged with the RepBase library (RepBaseRepeatMaskerEdition-20181026) and then used for repeat annotation using RepeatMasker v4.0.7 with a blast cutoff of 250 (Smit et al., 2015). Repeats were then filtered out for low complexity and simple repeats, and parsed using the parseRM_merge_interrupted.pl script from https://github.com/4ureliek/Parsing-RepeatMasker-Outputs. We only retained TEs longer than 100 bp and overlapping identical TEs were merged into single elements for the final annotation. In addition, TEs of the same family separated by less than 200 bp

were considered as part of the same TE and merged into a single element. We analyzed population-level presence/absence variation of transposable elements using the R-based tool ngs_te_mapper, using trimmed reads as above and bwa version 0.7.17-r1188 (Bergman, 2012; Li and Durbin, 2009). Overlapping genes and TEs were identified using the intersect function from the bedtools suite v2.29 (Quinlan and Hall, 2010). Effector prediction was performed based on a machine-learning approach. First, putatively secreted proteins were identified for the *C. parasitica* reference genome with SignalP and TMHMM as implemented in InterProScan v5.31-70.0 (Jones et al., 2014). Then, protein sequences with a predicted secretion signal predicted by both tools were extracted with samtools v1.9 (Li et al., 2009) and used as input for effector prediction with EffectorP v2.0 (Sperschneider et al., 2018).

Genome compartmentalization and copy-number variation analyses

We investigated the genome architecture of *C. parasitica* following the protocol described in (Saunders et al., 2014). Briefly, we computed intergenic distances with the Calculate_FIR_length.pl script using the gene prediction for *C. parasitica* reference genome EP155 v2.0. We defined 40 bins given the range of 3' and 5' intergenic distances and calculated the number of genes that fell within each bin. To infer copy-number variation, we computed the normalized read depth for all isolates using the CNVcaller pipeline (Wang et al., 2017). For this, the reference genome was first split into 800 bp overlapping kmers that were re-aligned to the reference using blasr (-m 5 --noSplitSubreads --minMatch 15 --maxMatch 20 --advanceHalf --advanceExactMatches 10 --fastMaxInterval --fastSDP --aggressiveIntervalCut --bestn 10) to identify duplicated windows (python3 0.2.Kmer_Link.py ref.genome.kmer.aln 800 ref.genome.800.window.link). The resulting windows were then used to calculate the normalized read depth for each isolate from the aligned reads (bam format) in 800 bp windows (Individual.Process.sh -b $bam -h ${i% .bam} -d ref.genome.800.window.link -s scaffold_1). For all further analyses, we used the read depth normalized for the absolute copy number and GC content of each sample (RD_normalized output). Given the observed distribution of the normalized read depth (NRD) across all isolates, we considered windows with NRD > 1.6 to be duplicated and windows with NRD < 0.4 as deleted. To investigate the relative diversity within the S12 and non-S12 groups, we resampled 80 random isolates from each lineage 100 times and investigated SNP, TE and CNV diversity. We quantified the minor allele frequency spectrum in each of the two groups and tested for the group effect (S12 vs. non-S12) on the genetic diversity using Fisher's exact tests in R. We simulated datasets using 2'000 resampling replicates to generate expectations for random distributions (Supp. Fig. 6)

Mating experiments

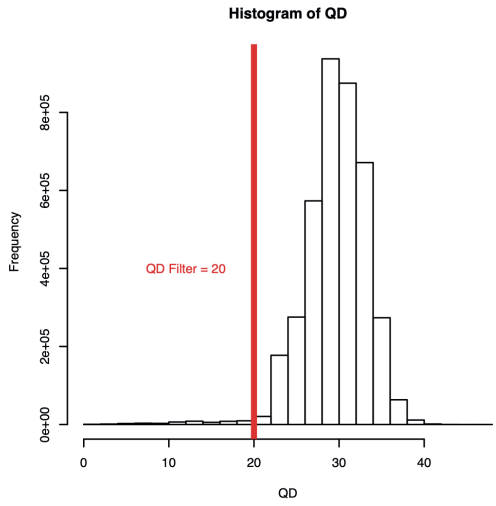
The ability of *C. parasitica* isolates belonging to S12 with mating type MAT-1 to sexually outcross with isolates of opposite mating type, was assessed in an inoculation experiment. For this, we randomly selected four S12 and one non-S12 isolate with mating type MAT-1, as well as five isolates of opposite mating type MAT-2 from populations in Italy (Nebrodi), Kosovo (Molliq) and Bosnia (Konic, Vrnograč, Projsa) (Supp. Table 3). All

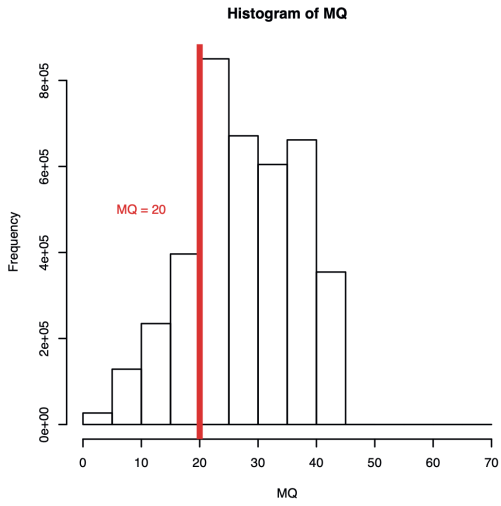
isolates of both mating types belonged to the vegetative compatibility type EU-12 and mating pairs were only formed between isolates from the same geographic source populations. As substrate for the pairings, 40 mm long segments of dormant chestnut (*C. sativa*) stems (15–20 mm in diameter) were split longitudinally in half. The wood pieces were autoclaved twice, placed onto sterile petri dishes (90mm diameter), which were filled with 1.5% water agar to enclose the pieces. Mating pairs were inoculated on opposite sides of each halved stem with three replicates per pairing. The inoculated plates were then incubated at 25°C under a 16 h photoperiod (2500 Lux) for 14 days. After two weeks, mating was stimulated by adding 5 ml sterile water to the plates to suspend and distribute the conidia produced by both isolates over the stem segment. Any excess water was subsequently removed and the plates were incubated at 18°C under an 8 h photoperiod (2500 Lux). After 5 months of incubation, perithecia formation was assessed under a dissecting microscope. To confirm successful outcrossing, single perithecia were carefully extracted from the stromata and crushed in a drop of sterile distilled water. The resulting ascospore suspensions were plated on PDA and incubated at 25°C for 24–36 h. Afterwards, single germinating ascospores were transferred to PDA and incubated in the dark for 3–5 days at 25°C. DNA was then extracted from 10 mg of lyophilized mycelium using the kit and instructions by KingFisher (Thermo Fisher Scientific, Waltham MA, USA). All single ascospore cultures were screened for mating types by performing a multiplex PCR following the protocol described in Cornejo et al. (2019).

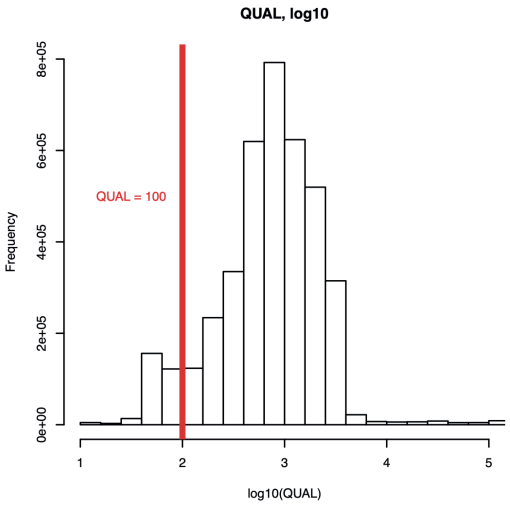
Acknowledgements

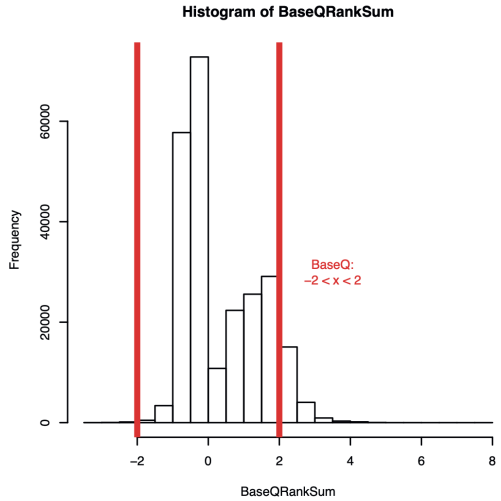
We are grateful to Ursula Oggenfuss for helpful comments on a previous manuscript version. We thank Sandrine Fattore, Hélène Blauenstein, Quirin Kupper, Eva Augustiny, Dario Rüegg and Silvia Kobel for laboratory assistance. We acknowledge the Genetic Diversity Centre (GDC), ETH Zurich, for technical support and facility access. Ludwig Beenken and Martin Wrann helped with documenting mating experiments. Paolo Cortesi, Michael Milgroom, Kiril Sotirovski, Mihajlo Risteski, Marin Ježić and Seçil Akilli kindly provided samples. We thank Pierre Gladieux and Nikhil Kumar Singh for discussions and sharing scripts for data analyses. We are grateful for insightful discussions with Daniel Rigling. Funding was awarded to SP by the Swiss National Science Foundation (grant 170188) and to DC by the Fondation Pierre Mercier pour la science.

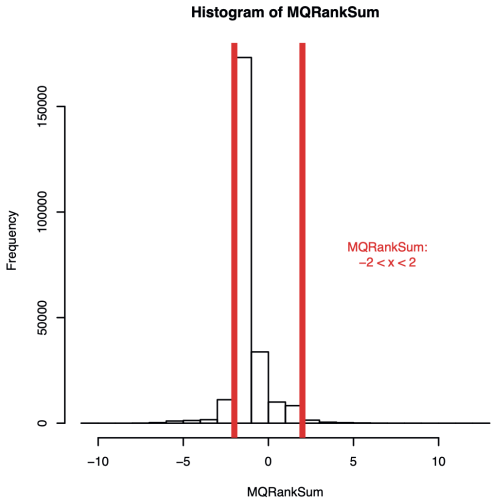
Supplementary Information



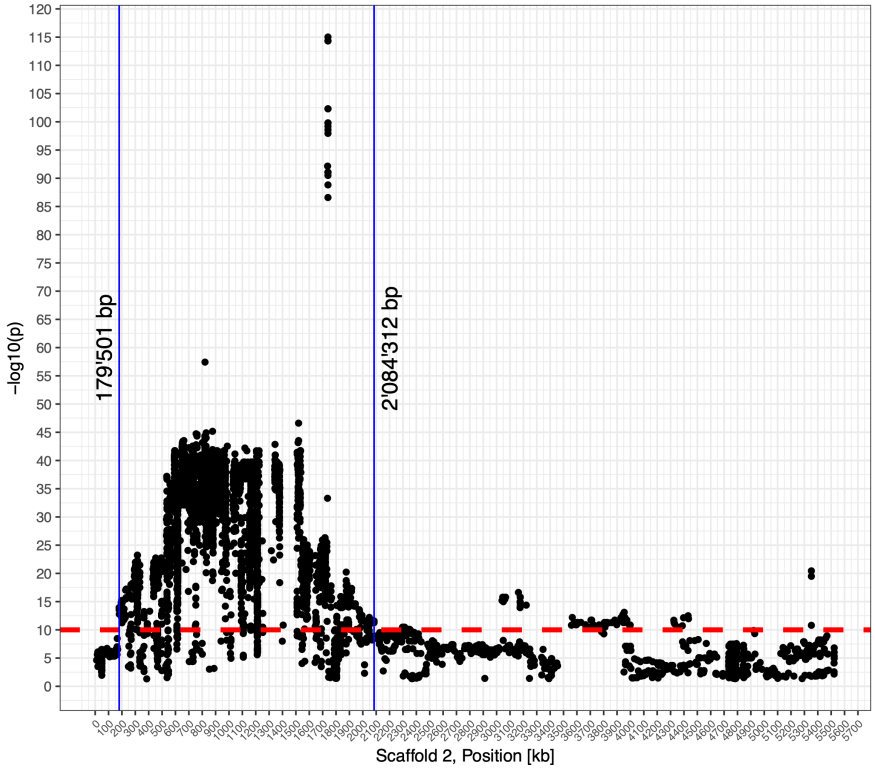




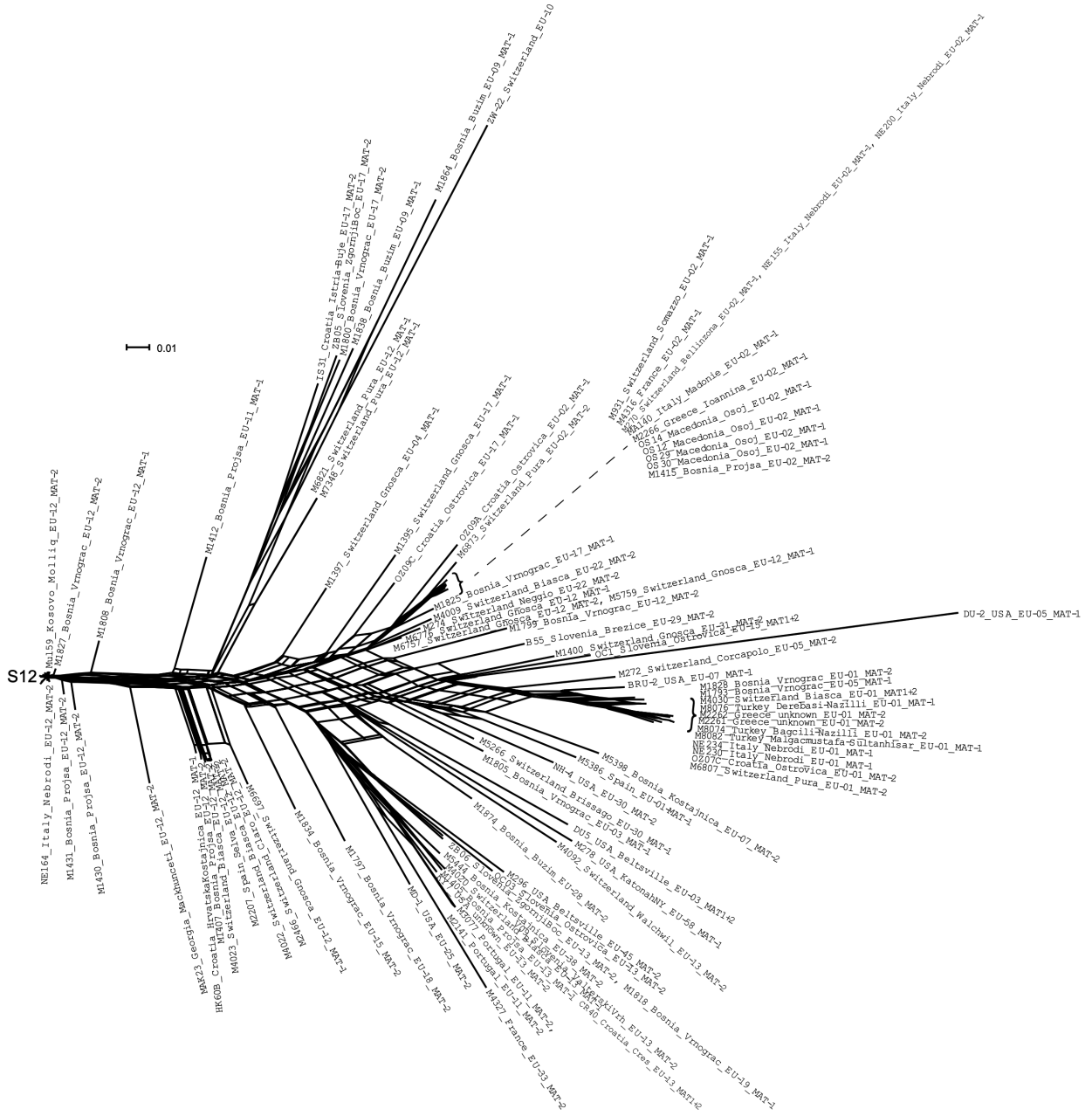




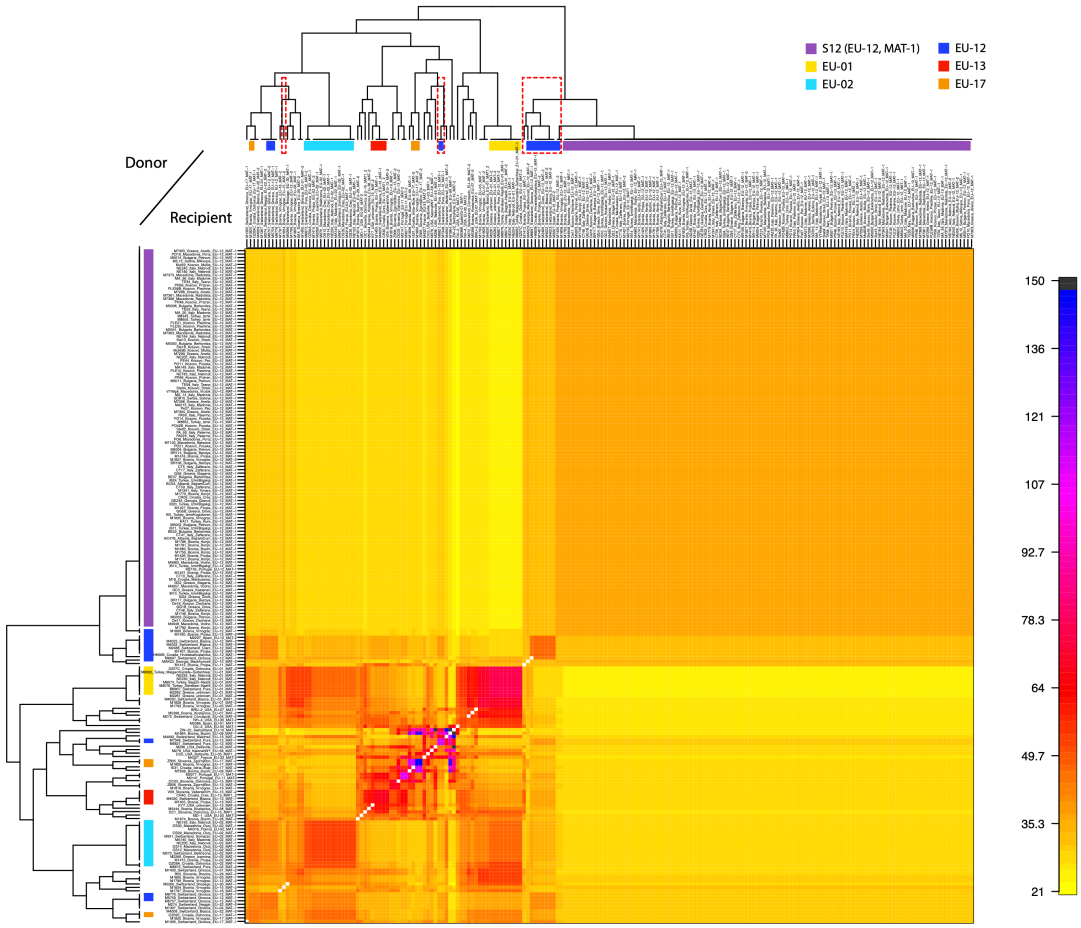
Supplementary Figure 1: Distribution of observed SNP filtration quality values.



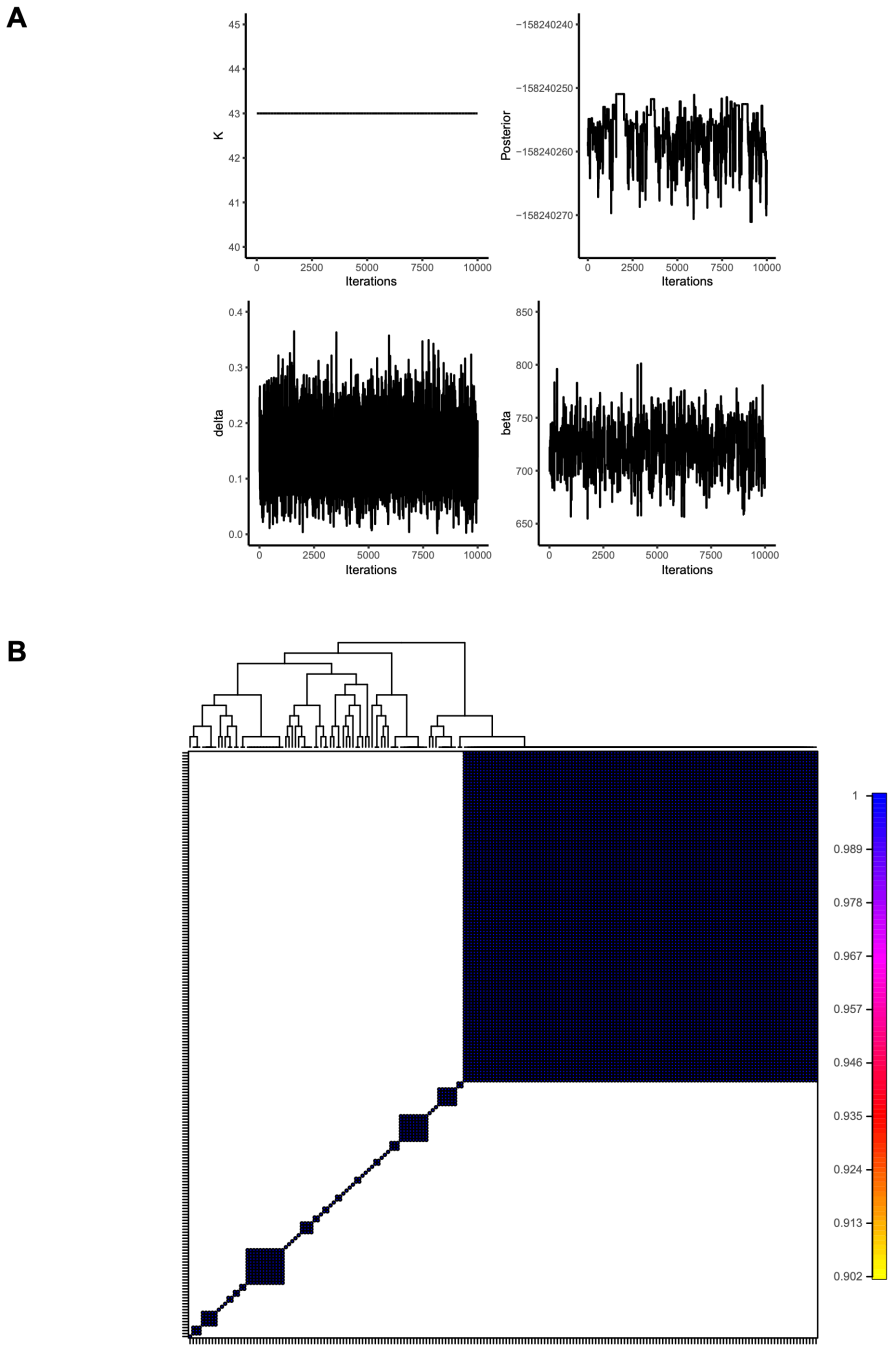
Supplementary Figure 2: Genome-wide association mapping (GWAS) outcome for the mating type locus. The red line shows the p-value based threshold to exclude MAT-locus associated SNPs for population structure analyses.



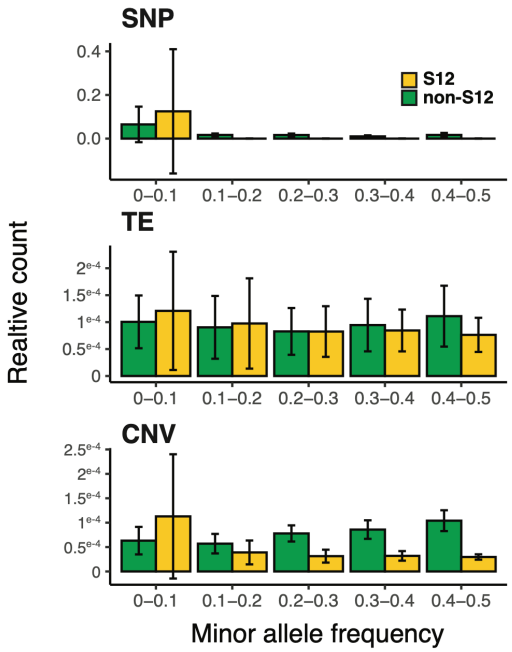
Supplementary Figure 3: SplitsTree of all non-S12 isolates showing full isolate identifiers. Isolates identified as S12 are listed in Supplementary Table 1.



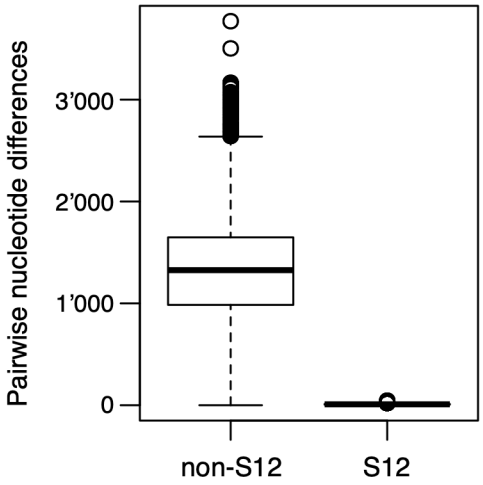
Supplementary Figure 4: Full averaged co-ancestry matrix showing isolate identifiers.



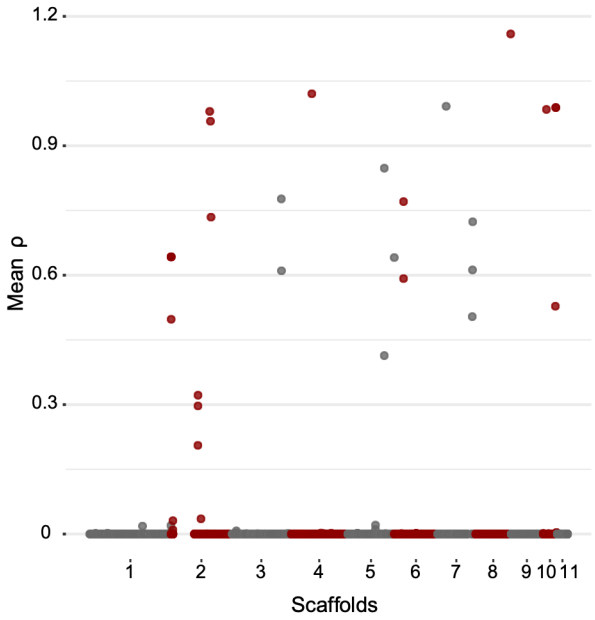
Supplementary Figure 5: fineSTRUCTURE convergence plots. A) fineSTRUCTURE MCMC traces. B) Pairwise coincidence matrix of population assignments.



Supplementary Figure 6: Genetic diversity within the S12 lineage. Minor allele frequency (MAF) distribution spectrum for three classes of variants. The number of polymorphic positions were normalized to the total number of polymorphic positions in S12 and non-S12 isolates, respectively. Error bars represent the standard deviation estimated from a resampling (n = 100) of 80 isolates in each group. We evaluated the difference between the S12 and non-S12 populations using two-way analysis of variance (number of sites ~ minor allele frequency*population). Both SNPs and copy-number variants (CNVs) frequencies show strong contrast ($p < 1e-16$) while we find only minor differences for transposable elements (TEs) between S12 and non-S12 isolates ($p < 0.05$).



Supplementary Figure 7: Pairwise genome-wide nucleotide differences between S12 and non-S12 isolates.



Supplementary Figure 8: Per site recombination rate ρ of European/North American *Cryphonectria parasitica* populations as inferred with LDhat

Supplementary Tables:

See: https://doi.org/10.1101/2020.02.15.950170

Supplementary Table 1: Information on all analyzed isolates including sampling location, vegetative compatibility type (EU-type), SSR genotyping outcome, collector information, collection years and NCBI accessions.

Supplementary Table 2: List of the genes in regions with signatures of recent positive selection (based on RAiSD). Putative functions are shown by PFAM and Gene Ontology (GO) annotations. SNPs and the corresponding composite μ statistic are shown.

Supplementary Table 3: Summary of MAT-1 and MAT-2 isolate pairings used for testing sexual recombination in S12 populations. *) Isolates from the S12 MAT-1 lineage.

References

Alachiotis N, Pavlidis P. 2018. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun Biol* **1**:1–11.

Allendorf FW, Lundquist LL. 2003. Introduction: Population biology, evolution, and control of invasive species. *Conserv Biol* **17**:24–30.

Avolio S. 1978. La castanicoltura in Calabria. Aspetti selvicolturali e linee di intervento. *L'Italia For e Mont* **2**:278–296.

Banks NC, Paini DR, Bayliss KL, Hodda M. 2015. The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecol Lett* **18**:188–199.

Bergman CM. 2012. A proposal for the reference-based annotation of de novo transposable element insertions. *Mob Genet Elements* **2**:51–54.

Bertelsmeier C, Keller L. 2018. Bridgehead effects and role of adaptive evolution in invasive populations. *Trends Ecol Evol*.

Bertelsmeier C, Ollier S, Liebhold AM, Brockerhoff EG, Ward D, Keller L. 2018. Recurrent bridgehead effects accelerate global alien ant spread. *Proc Natl Acad Sci* **115**:5486–5491.

Billiard S, López-Villavicencio M, Hood ME, Giraud T. 2012. Sex, outcrossing and mating types: Unsolved questions in fungi and beyond. *J Evol Biol* **25**:1020–1038. doi:10.1111/j.1420-9101.2012.02495.x

Biraghi A. 1946. Cancro della corteccia del castagno. Ramo editoriale degli agricoltori.

Bock DG, Caseys C, Cousens RD, Hahn MA, Heredia SM, Hübner S, Turner KG, Whitney KD, Rieseberg LH. 2015. What we still don't know about invasion genetics. *Mol Ecol* **24**:2277–2297.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

Bougeard S, Dray S. 2018. Supervised multiblock analysis in R with the ade4 package. *J Stat Softw* **86**:1–17.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633–2635.

Brasier CM, Webber JF. 2019. Is there evidence for post-epidemic attenuation in the Dutch elm disease pathogen *Ophiostoma novo-ulmi*? *Plant Pathol* **68**:921–929.

Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**:2665–2681.

Bryner SF, Rigling D. 2012. Hypovirus virulence and vegetative incompatibility in populations of the chestnut blight fungus. *Phytopathology* **102**:1161–1167.

Buccianti M, Feliciani A. 1966. LA SITUAZIONE ATTUALE DEI CASTAGNETI IN ITALIA (The present situation of chestnut stands in Italy). *AttiConvegno Internazionale sul Castagno, Cuneo* 61–85.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**:80–92.

Cornejo C, Šever B, Kupper Q, Prospero S, Rigling D. 2019. A multiplexed genotyping assay to determine vegetative incompatibility and mating type in *Cryphonectria parasitica*. *Eur J Plant Pathol* 1–11.

Cortesi P, McCulloch CE, Song H, Lin H, Milgroom MG. 2001. Genetic control of horizontal virus transmission in the chestnut blight fungus, *Cryphonectria parasitica*. *Genetics* **159**:107–118.

Cortesi P, Milgroom MG, Bisiach M. 1996. Distribution and diversity of vegetative compatibility types in subpopulations of *Cryphonectria parasitica* in Italy. *Mycol Res* **100**:1087–1093.

Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, Hillman B, Milgroom MG, Pangilinan J, Smith M, others. 2020. Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: A fundamental resource for an archetypical invasive plant pathogen. *Phytopathology*.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, others. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.

Demené A, Legrand L, Gouzy J, Debuchy R, Saint-Jean G, Fabreguettes O, Dutech C. 2019. Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of *Cryphonectria parasitica* in western Europe. *Fungal Genet Biol*.

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

Drenth A, McTaggart AR, Wingfield BD. 2019. Fungal clones win the battle, but recombination wins the war. *IMA Fungus* **10**:18.

Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V. 2012. The chestnut blight fungus world tour: Successive introduction events from diverse origins in an invasive plant fungal pathogen. *Mol Ecol* **21**:3931–3946.
Dutech C, Fabreguettes O, Capdevielle X, Robin C. 2010. Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity (Edinb)* **105**:220–228.
Elliott KJ, Swank WT. 2008. Long-term changes in forest composition and diversity following early logging (1919-1923) and the decline of American chestnut (*Castanea dentata*). *Plant Ecol* **197**:155–172.
Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**:737–756.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AFA. 2019. RepeatModeler2: automated genomic discovery of transposable element families. *BioRxiv* 856591.
Fontaine MC, Gladieux P, Hood ME, Giraud T. 2013. History of the invasion of the anther smut pathogen on *Silene latifolia* in North America. *New Phytol* **198**:946–956.
Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr Opin Microbiol* **46**:34–42.
Frantzeskakis L, Kusch S, Panstruga R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol* **20**:3–7.
Friesen TL, Faris JD, Solomon PS, Oliver RP. 2008. Host-specific toxins: effectors of necrotrophic pathogenicity. *Cell Microbiol* **10**:1421–1428.
Garbelotto M, Rocca G Della, Osmundson T, di Lonardo V, Danti R. 2015. An increase in transmission-related traits and in phenotypic plasticity is documented during a fungal invasion. *Ecosphere* **6**:1–16.
Garnas JR, Auger-Rozenberg M-A, Roques A, Bertelsmeier C, Wingfield MJ, Saccaggi DL, Roy HE, Slippers B. 2016. Complex patterns of global spread in invasive insects: Eco-evolutionary and management consequences. *Biol Invasions* **18**:935–952.
Gau RD, Merz U, Falloon RE, Brunner PC. 2013. Global genetics and invasion history of the potato powdery scab pathogen, *Spongospora subterranea* f. sp. *subterranea*. *PLoS One* **8**:e67944.
Gladieux P, Byrnes III EJ, Aguileta G, Fisher MC, Heitman J, Giraud T. 2011. Epidemiology and evolution of fungal pathogens in plants and animalsGenetics and Evolution of Infectious Disease. Elsevier. pp. 59–132.
Gladieux P, Feurtey A, Hood ME, Snirc A, Clavel J, Dutech C, Roy M, Giraud T. 2015. The population biology of fungal invasions. *Mol Ecol* **24**:1969–1986.
González-Varela G, González AJ, Milgroom MG. 2011. Clonal population structure and introductions of the chestnut blight fungus, *Cryphonectria parasitica*, in Asturias, northern Spain. *Eur J Plant Pathol* **131**:67–79.
Griffin GJ. 1986. Chestnut blight and its control. *Hortic Rev (Am Soc Hortic Sci)* **8**:291–336.
Grünwald NJ, Garbelotto M, Goss EM, Heungens K, Prospero S. 2012. Emergence of the sudden oak death pathogen *Phytophthora ramorum*. *Trends Microbiol* **20**:131–138.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5**:e1000695.
Hayes KR, Barry SC. 2008. Are there any consistent predictors of invasion success? *Biol Invasions* **10**:483–506.
Heald FD, Studhalter RA. 1914. Birds as carriers of the chestnut blight fungus. *J Agric Res* **2**:405–422.
Heiniger U, Rigling D. 1994. Biological control of chestnut blight in Europe. *Annu Rev Phytopathol* **32**:581–599.
Heitman J, Sun S, James TY. 2013. Evolution of fungal sexual reproduction. *Mycologia* **105**:1–27.
Hessenauer P, Fijarczyk A, Martin H, Prunier J, Charron G, Chapuis J, Bernier L, Tanguay P, Hamelin RC, Landry CR. 2020. Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. *Nat Ecol Evol* **4**:626–638.
Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. 2000. Genetic structure of newly established populations of *Cryphonectria parasitica*. *Mycol Res* **104**:1108–1116.
Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**:337–338. doi:10.1093/bioinformatics/18.2.337
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**:254–267. doi:10.1093/molbev/msj030
Idnurm A, Hood ME, Johannesson H, Giraud T. 2015. Contrasted patterns in mating-type chromosomes in fungi: Hotspots versus coldspots of recombination. *Fungal Biol Rev* **29**:220–229.
Ježić M, Krstin L, Rigling D, ĆURKOVIĆ-PERICA M. 2012. High diversity in populations of the introduced plant pathogen, *Cryphonectria parasitica*, due to encounters between genetically divergent genotypes. *Mol Ecol* **21**:87–99.
Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5:

Genome-scale protein function classification. *Bioinformatics* **30**:1236–1240. doi:10.1093/bioinformatics/btu031
Karadžić D, Radulović Z, Sikora K, Stanivuković Z, Ćurguz VG, Oszako T, Milenković I. 2019. Characterisation and pathogenicity of *Cryphonectria parasitica* on sweet chestnut and sessile oak trees in Serbia. *Plant Prot Sci* **55**:191–201.
Knaus BJ, Grünwald NJ. 2017. vcfr: A package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* **17**:44–53.
Knaus BJ, Tabima JF, Shakya SK, Judelson HS, Grünwald NJ. 2020. Genome-Wide Increased Copy Number is Associated with Emergence of Dominant Clones of the Irish Potato Famine Pathogen *Phytophthora infestans*. *MBio* **11**.
Krasnov BR, Shenbrot GI, Khokhlova IS, Stanko M, Morand S, Mouillot D. 2015. Assembly rules of ectoparasite communities across scales: combining patterns of abiotic factors, host composition, geographic space, phylogeny and traits. *Ecography (Cop)* **38**:184–197.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–360. doi:10.1038/nmeth.1923
Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet* **8**:e1002453.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *bioinformatics* **25**:1754–1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352
Lischer HEL, Excoffier L. 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**:298–299.
Lombaert E, Guillemaud T, Cornuet JM, Malausa T, Facon B, Estoup A. 2010. Bridgehead Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. *PLoS One* **5**. doi:10.1371/journal.pone.0009743
Marbuah G, Gren IM, McKie B. 2014. Economics of Harmful Invasive Species: A Review. *Diversity* **6**:500–523. doi:10.3390/d6030500
Mariette N, Mabon R, Corbière R, Boulard F, Glais I, Marquer B, Pasco C, Montarry J, Andrivon D. 2016. Phenotypic and genotypic changes in French populations of *Phytophthora infestans*: Are invasive clones the most aggressive? *Plant Pathol* **65**:577–586.
Marra RE, Cortesi P, Bissegger M, Milgroom MG. 2004. Mixed mating in natural populations of the chestnut blight fungus, *Cryphonectria parasitica*. *Heredity (Edinb)* **93**:189–195.
Marra RE, Milgroom MG. 2001. The mating system of the fungus *Cryphonectria parasitica*: Selfing and self-incompatibility. *Heredity (Edinb)* **86**:134–143.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, others. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 1297–1303. doi:10.1101/gr.107524.110
McVean G, Auton A. 2007. LDhat 2.1: A package for the population genetic analysis of recombination. *Dep Stat Oxford, OX1 3TG, UK*.
Mehrotra RS. 2013. Fundamentals of Plant Pathology. Tata McGraw-Hill Education.
Milgroom MG, Cortesi P. 1999. Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *Proc Natl Acad Sci* **96**:10518–10523.
Milgroom MG, Sotirovski K, Spica D, Davis JE, Brewer MT, Milev M, Cortesi P. 2008. Clonal population structure of the chestnut blight fungus in expanding ranges in southeastern Europe. *Mol Ecol* **17**:4446–4458.
Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneko S. 1996. Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* **88**:179–190.
Myteberi IF, Lushaj Arvjen B, Keča N, Lushaj Arnisa B, Lushaj BM. 2013. Diversity of *Cryphonectria parasitica*, hypovirulence, and possibilities for biocontrol of chestnut canker in Albania. *Int J Microb Res Rev* **1**:11–21.
Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**:1749–1751. doi:/10.1093/bioinformatics/btw044
Oggenfuss U, Badet T, Wicker T, Hartmann FE, Singh NK, Abraham LN, Karisto P, Vonlanthen T, Mundt CC, McDonald BA, others. 2020. A population-level invasion by transposable elements in a fungal pathogen. *BioRxiv*.
Paoletti M, Buck KW, Brasier CM. 2006. Selective acquisition of novel mating type and vegetative incompatibility

genes via interspecies gene transfer in the globally invading eukaryote *Ophiostoma novo-ulmi*. *Mol Ecol* **15**:249–262.
Philibert A, Desprez-Loustau M-L, Fabre B, Frey P, Halkett F, Husson C, Lung-Escarmant B, Marçais B, Robin C, Vacher C, others. 2011. Predicting invasion success of forest pathogenic fungi from species traits. *J Appl Ecol* **48**:1381–1390.
Prospero S, Cleary M. 2017. Effects of host variability on the spread of invasive forest diseases. *Forests* **8**:80.
Prospero S, Conedera M, Heiniger U, Rigling D. 2006. Saprophytic activity and sporulation of *Cryphonectria parasitica* on dead chestnut wood in forests with naturally established hypovirulence. *Phytopathology* **96**:1337–1344.
Prospero S, Rigling D. 2012. Invasion genetics of the chestnut blight fungus *Cryphonectria parasitica* in Switzerland. *Phytopathology* **102**:73–82.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.
Raboin L-M, Selvi A, Oliveira KM, Paulet F, Calatayud C, Zapater M-F, Brottier P, Luzaran R, Garsmeur O, Carlier J, others. 2007. Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen Ustilago scitaminea. *Fungal Genet Biol* **44**:64–76.
Rigling D, Prospero S. 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: Invasion history, population biology and disease control. *Mol Plant Pathol* **19**:7–20. doi:10.1111/mpp.12542
Robin C, Heiniger U. 2001. Chestnut blight in Europe: Diversity of *Cryphonectria parasitica*, hypovirulence and biocontrol. *For Snow Landsc Res* **76**:361–367.
Rossman AY. 2001. A special issue on global movement of invasive plants and fungi. *AIBS Bull* **51**:93–94.
RStudio Team. 2015. RStudio: Integrated Development Environment for R.
Santini A, Ghelardini L, De Pace C, Desprez-Loustau M-L, Capretti P, Chandelier A, Cech T, Chira D, Diamandis S, Gaitniekis T, others. 2013. Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytol* **197**:238–250.
Saunders DGO, Win J, Kamoun S, Raffaele S. 2014. Two-dimensional data binning for the analysis of genome architecture in filamentous plant pathogens and other eukaryotes plant-pathogen interactions. Springer. pp. 29–51.
Schoustra S, Rundle HD, Dali R, Kassen R. 2010. Fitness-associated sexual reproduction in a filamentous fungus. *Curr Biol* **20**:1350–1355.
Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013-2015.
Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *Proc Natl Acad Sci* **90**:4384–4388.
Sotirovski K, Papazova-Anakieva I, Grünwald NJ, Milgroom MG. 2004. Low diversity of vegetative compatibility types and mating type of *Cryphonectria parasitica* in the southern Balkans. *Plant Pathol* **53**:325–333.
Stauber L, Prospero S, Croll D. 2020. Comparative genomics analyses of lifestyle transitions at the origin of an invasive fungal pathogen in the genus *Cryphonectria*. *mSphere* **5**:e00737-20.
Steenwyk JL, Soghigian JS, Perfect JR, Gibbons JG. 2016. Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. *BMC Genomics* **17**:700.
Suehs CM, Affre L, Médail F. 2004. Invasion dynamics of two alien Carpobrotus (Aizoaceae) taxa on a Mediterranean island: II. Reproductive strategies. *Heredity (Edinb)* **92**:550.
Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE. 2015. Clonal reproduction in fungi. *Proc Natl Acad Sci* **112**:8901–8908.
Torres DE, Oggenfuss U, Croll D, Seidl MF. 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol Rev*.
Trestic T, Uscuplic M, Colinas C, Rolland G, Giraud A, Robin C. 2001. Vegetative compatibility type diversity of *Cryphonectria parasitica* populations in Bosnia-Herzegovina, Spain and France. *For Snow Landsc Res* **76**:391–396.
van Boheemen LA, Lombaert E, Nurkowski KA, Gauffre B, Rieseberg LH, Hodgins KA. 2017. Multiple introductions, admixture and bridgehead invasion characterize the introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Mol Ecol* **26**:5421–5434.
Vuković R, Liber Z, Ježić M, Sotirovski K, Ćurković-Perica M. 2019. Link between epigenetic diversity and invasive status of south-eastern European populations of phytopathogenic fungus *Cryphonectria parasitica*. *Environ Microbiol*.
Wang X, Zheng Z, Cai Y, Chen T, Li C, Fu W, Jiang Y. 2017. CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *Gigascience* **6**:gix115.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Wickham H. 2007. Reshaping Data with the reshape Package. *J Stat Softw* **21**:1–20.
Wickham H, François R, Henry L, Müller K. 2018. dplyr: A Grammar of Data Manipulation.
Wickham H, Henry L. 2018. tidyr: Easily Tidy Data with "spread()" and "gather()" Functions.
Wingfield MJ, Slippers B, Roux J, Wingfield BD. 2010. Fifty years of tree pest and pathogen invasions, increasingly threatening world forests. *Fifty years invasion Ecol Leg charles Elt* 89–99.
Wuest CE, Harrington TC, Fraedrich SW, Yun H-Y, Lu S-S. 2017. Genetic variation in native populations of the laurel wilt pathogen, *Raffaelea lauricola*, in Taiwan and Japan and the introduced population in the United States. *Plant Dis* **101**:619–628.
Yang X-M, Sun J-T, Xue X-F, Li J-B, Hong X-Y. 2012. Invasion genetics of the western flower thrips in China: Evidence for genetic bottleneck, hybridization and bridgehead effect. *PLoS One* **7**:e34567.

Chapter 3

Analyses of the temporal and spatial genetic structure of the chestnut blight pathosystem

Lea Stauber[1,2], Daniel Croll[2], Simone Prospero[1]

[1] Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

[2] Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel

Unpublished manuscript

Abstract

Antagonistic pathogen-hyperparasite interactions shape co-evolutionary trajectories and can influence stability of their host populations. The invasive chestnut blight fungus *Cryphonectria parasitica* causes lethal bark lesions on non-Asian *Castanea* spp. species, but can itself be infected with the deleterious mycovirus Cryphonectria hypovirus 1 (CHV1), resulting in reduced fungal virulence and healing of infected chestnut cankers. In Europe, CHV1 constitutes an effective natural biocontrol agent against chestnut blight. However, the impact of CHV1 on fungal population dynamics in recently established chestnut blight populations remains largely unknown. In this study, we analyzed genome-wide diversity of multiple *C. parasitica* populations in southern Switzerland. We investigated changes in fungal diversity and CHV1 prevalence over two time points separated by ~30 years. Our results indicate increased mating among related fungal individuals, resulting in high genetic similarity of genotypes, which is expected to facilitate CHV1 transmission. The fungal population structure showed a high degree of stability over the 30-year period and no detectable impact of CHV1 presence on fungal genetic diversity. Although our results show stable CHV1 incidence in fungal populations over ~30 years, the short-term interaction dynamics are likely highly volatile, making it a suitable system to study antagonistic co-evolutionary dynamics in non-native environments.

Introduction

Host-pathogen interactions drive co-evolutionary trajectories, determining disease dynamics through selection in pathogen virulence and host resistance. At the genetic level, such interactions are commonly described by two scenarios: the "arms race" where selective sweeps repeatedly drive pathogenicity and resistance alleles to fixation, or the "trench warfare" in which allele frequencies are cycling leading to signatures of balancing selection (Tellier et al., 2014). Evidence for both scenarios has been found in natural pathosystems. For instance, signatures of co-evolutionary arms race have been detected in anther smut disease causing fungi which infect Caryophyllaceae species (Hartmann et al., 2019). Contrary, maintenance of genetic variation and antagonistic cycling of virulence and resistance (i.e. trench warfare) were observed in the *Linum-Melampsora* pathosystem (Thrall et al., 2012). New pathosystems may also emerge as a consequence of human activity by introducing pathogens into naive host populations (Rogalski et al., 2017). Host-pathogen interaction dynamics are strongly influenced by factors such as host population size. Demographic events (*e.g.* genetic drift due to bottlenecks) impacts co-evolutionary dynamics and makes predictions on future trajectories challenging (Papkou et al., 2016; Tellier et al., 2014).

The evolution of pathosystems can be shaped by hyperparasites, which infect the pathogens themselves, thereby influencing stability of both host and pathogen populations (Parratt and Laine, 2016). Pathogen-hyperparasite interactions may underlie similar evolutionary forces as the primary host-pathogen interactions. For instance, evidence for co-evolutionary dynamics was found in bacterial populations on horse chestnut trees (*Aesculus hippocastanum*) and hyperparasitic viral phages (Koskella, 2013). Hyperparasites have been successfully used in forest and agricultural systems as biocontrol agents against pathogens and pests, including insect pathogenic fungi or nematode infecting hyperparasitic bacteria (Branine et al., 2019; Mei et al., 2020; Tian et al., 2007). A prime example for a hyperparasite is the Cryphonectria hypovirus 1 (CHV1), which induces hypovirulence in the invasive chestnut blight fungus *Cryphonectria parasitica* (Murr). Barr. CHV1 (family Hypoviridae) is a positive-strand RNA virus located in the cytoplasm of *C. parasitica* (Bryner et al., 2012). In infected *C. parasitica* strains the virus significantly reduces fungal virulence, growth and sporulation, leading to recovery of infected chestnut trees (Anagnostakis, 1982; Rigling and Prospero, 2018). On a molecular level, CHV1 has been found to significantly alter gene expression in *C. parasitica*, affecting genes associated with virulence, as well as primary and secondary metabolism, including genes involved in carbohydrate metabolism (Chun et al., 2020). However, whether CHV1 presence shapes the genetic diversity in fungal host populations remains largely unknown.

The chestnut blight fungus *C. parasitica* has emerged as an invasive pathogen in North America and Europe, causing lethal bark lesions on non-Asian chestnut species (*Castanea* spp.) (Rigling and Prospero, 2018). Interestingly, a striking loss of genes associated with carbohydrate metabolism was recently detected in *C.*

parasitica, potentially underpinning a recent evolutionary lifestyle transition towards higher pathogenicity (Stauber et al., 2020b). During the colonization process of North America and Europe, *C. parasitica* underwent multiple genetic bottlenecks that strongly shaped its population structure and diversity in the invasive range (Dutech et al., 2012). Accordingly, European *C. parasitica* populations are characterized by lower genetic diversity compared to native Asian, but also relative to invasive North American populations (Dutech et al., 2012; Milgroom and Cortesi, 1999). In Europe, *C. parasitica* was first reported in 1938 in northern Italy and successively spread to all major chestnut growing areas (Rigling and Prospero, 2018). Low genetic diversity of European *C. parasitica* populations have partially been attributed to higher rates of clonality, facilitating colonization of new areas in absence of mating partners (Demené et al., 2019; Milgroom et al., 1996; Sotirovski et al., 2004; Stauber et al., 2020a). For instance, south-eastern Europe was predominantly colonized by a single clonal *C. parasitica* lineage, which likely emerged from a sexually reproducing European bridgehead population (Stauber et al., 2020a). Sexual reproduction has predominantly been observed in long established and genetically more diverse European *C. parasitica* populations, including those in southern Switzerland, northern Italy or Croatia (Ježić et al., 2018; Prospero and Rigling, 2012; Stauber et al., 2020a). In *C. parasitica*, sexual reproduction is controlled by a single mating (MAT) locus with two alleles (MAT-1 or MAT-2) (McGuire et al., 2001). For successful mating, opposite mating types from two individuals are required, but fertile individuals carrying both mating type alleles (so called heterokaryons) are occasionally observed (McGuire et al., 2004). Sexual recombination is considered as a major driver for population diversification, including diversification at vegetative incompatibility (*vic*) loci, which can subsequently lead to reduced horizontal transmission rates of CHV1 in *C. parasitica* populations.

As an obligate hyperparasite, CHV1 can only be transmitted horizontally via hyphal anastomosis between fungal individuals, or vertically through asexual spores (conidia) (Rigling and Prospero, 2018). In contrast, sexual spores (ascospores) have consistently been found to be virus-free (Prospero et al., 2006). The virus induces female sterility, thereby restricting fungal sexual reproduction among virus-infected fungal strains (Nuss, 2005). However, virus-infected conidia can act as male partners (spermatia) in sexual crosses (Anagnostakis, 1988). Contrary, horizontal CHV1 transmission is determined by fungal vegetative incompatibility barriers. Vegetative incompatibility in *C. parasitica* is controlled by at least six unlinked biallelic *vic* loci (Cortesi et al., 2001). Successful CHV1 transmission occurs when fungal individuals share the same alleles at all *vic* loci (i.e. belong to the same vegetative compatibility (vc) type), whereas allelic polymorphism at the *vic* loci strongly reduces transmission rates (Cortesi et al., 2001). However, incompatibility barriers might be incomplete and virus transmission is possible even if two individuals are polymorphic at a locus such as *vic4* (Cortesi et al., 2001). Moreover, virus transmission can be asymmetric if individuals are heteroallelic at other *vic* loci (Cortesi et al., 2001). Hence, maintaining balanced and high *vic* allele diversity, as well as selection at loci involved in asymmetric virus transmission are likely important evolutionary forces in *C. parasitica* to limit prevalence of the hyperparasite within populations (Milgroom et al., 2018). However,

despite presence of CHV1 in European *C. parasitica* populations, vc type diversity has remained lower than expected under random mating, facilitating the natural spread of CHV1 (Krstin et al., 2011; Prospero and Rigling, 2012). Concurrently, spatial and temporal variation of CHV1 prevalence has been observed in European *C. parasitica* populations, suggesting fluctuating dynamics between CHV1 and its fungal host (Bryner et al., 2014; Ježić et al., 2018). However, how these dynamics shape the evolution of European *C. parasitica* populations and why sexual recombination has not resulted in significantly higher vc type diversity reducing CHV1 spread in Europe, remains largely unknown.

In this study, we investigated temporal changes in genomic diversity and the influence of CHV1 presence in *C. parasitica* populations in southern Switzerland (Ticino). In this region, European chestnut is the dominant tree species, forming a largely continuous forest belt which ranges up to 900 m above sea level. Chestnut blight was first officially reported in 1948, most likely after spontaneous spread from neighboring northern Italy (Heiniger and Rigling, 1994). CHV1 infected chestnut blight cankers were first observed in 1975 and natural hypovirulence is currently widespread. *C. parasitica* populations in southern Switzerland are considered to be sexually reproducing, as both mating types are present and sexual fruiting bodies are frequently observed in the field (Bissegger et al., 1997; Prospero et al., 2006). For our analyses, we sequenced whole genomes of 142 *C. parasitica* isolates, which were obtained from three populations sampled twice in the 1990s and in 2019. We compared genome-wide diversity in southern Switzerland with the broader European *C. parasitica* diversity and assessed temporal differences of genomic diversity using single nucleotide polymorphisms (SNP). Additionally, we assessed temporal changes in mating type and vc type diversity, as well as CHV1 prevalence. Moreover, we investigated the spatial distribution of *C. parasitica* genotypes and CHV1 presence in the 2019 sampling using GPS data. Overall, we found evidence for non-random mating within fungal populations and stable virus infection rates at ~45% over the sampling period. The results highlight that *C. parasitica* populations can retain low vc type diversity and stable CHV1 infection rates over several decades.

Results

Population structure in southern Switzerland and across Europe

To analyze the structure and diversity of *C. parasitica* populations in Ticino in a broader European and US context, we constructed a population dendrogram based on patterns of haplotype similarity with fineSTRUCTURE (Lawson et al., 2012). The fineSTRUCTURE algorithm identified K=45 populations and no clear geographic clustering (Fig. 1A). Most Ticino samples from both sampling periods (1990s and 2019) grouped in two main clusters grouped with other European genotypes. The first cluster exclusively consisted of 55 EU-02 and 8 EU-06 isolates, whereas the second cluster included 38 EU-01, 32 EU-05 and one EU-21 isolate. Ticino samples belonging to other vc types were found also in other populations as identified by fineSTRUCTURE (Fig. 1A). The PCA showed that most Ticino isolates from both sampling periods were found in one central cluster, with few distant genotypes partially reflecting the broader European and North American diversity (Fig. 1B). Nucleotide diversity within European and North American populations was π=0.27 compared to the Ticino populations of the 1990s sampling with π=0.078 and the 2019 sampling with π=0.069.

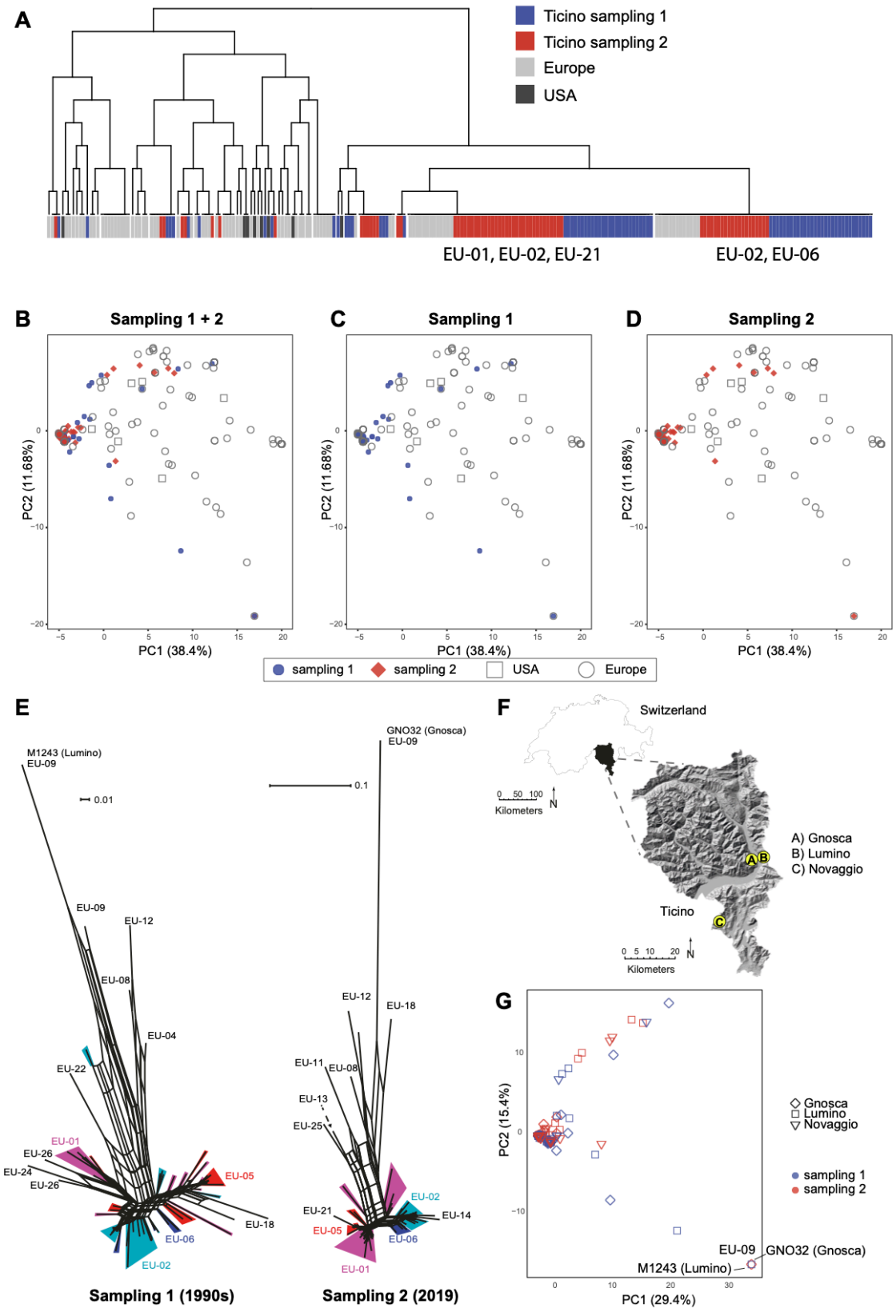


Figure 1. **Genome-wide analyses of population structure and diversity.** A) Dendrogram showing population structure (K=45) of Ticino isolates from both samplings and the broader European and North American diversity, as inferred with fineSTRUCTURE. Isolates from the first Ticino sampling (1990s) are marked in blue,

isolates from the second Ticino sampling (2019) are shown in red. Isolates representing the broader European and North American diversity are shown in light and dark grey. B) Principal component analysis (PCA) of both Ticino samplings and the broader US-European diversity. C) PCA of the first Ticino sampling (1990s) and the broader US-European diversity. D) PCA of the second Ticino sampling (2019) and the broader US-European diversity. E) SplitsTree phylogenetic network comparison of Ticino populations from the first (left) and second (right) sampling. Dominant vc types are highlighted in colors. F) Geographic map showing the sampling locations Gnosca (A), Lumino (B) and Novaggio (C) in Ticino, southern Switzerland. G) PCA of Ticino population structure. Colors mark samplings (blue = first sampling, red = second sampling), shapes indicate sampling locations (*i.e.* Gnosca, Lumino, Novaggio).

Temporal changes in population structure in Ticino

We investigated temporal changes in the Ticino population structure through phylogenetic network analyses. Overall, we detected no substantial changes in population structure over the period of approx. 30 years between the two samplings (Fig. 1E). In both samplings we found genetically closely related, as well as distant genotypes. The closely related genotypes were predominantly of vc-types EU-01, EU-02, EU-05 and EU-06. These genotypes showed clustering in both samplings, but appeared to be slightly more aggregated in the 2019 sampling (Fig. 1E). The most distant genotypes were of vc type EU-09 in both samplings (Fig. 1E). Interestingly, the EU-09 isolate (GNO32) found in 2019 in Gnosca was a clone of the EU-09 isolate (M1243) found in 1990 in Lumino (Fig. 1E and 1G). The phylogenetic networks in both samplings showed clear reticulations, indicating recombination (PHI: $p<0.001$, Fig. 1E). This is consistent with sexual fruiting bodies being frequently observed in the field.

Genotypic similarity and relatedness

We analyzed allele frequencies to gain insight into the genetic structure of the populations across sampling timepoints. Compared to the first sampling (1990s), the second sampling (2019) showed an increase of rare alleles (Fig. 2A). To quantify genetic similarity and relatedness among genotypes at each sampling, we analyzed identity by state (IBS) and identity by descent (IBD). Individuals were found to be highly genetically similar in both samplings, with a mean IBS of 0.92 (sd=0.06) in the 1990s sampling and a mean IBS of 0.94 (sd=0.06) in the 2019 sampling. However, similarity among individuals increased in 2019, where 71.12% of individuals had an IBS ≥ 0.95, compared to 33.76% in the 1990s (Fig. 2B). We additionally analyzed relatedness with IBD, using an estimated recombination rate of $r=1.97x10^{-4}$ for *C. parasitica*. In accordance with IBS, IBD values increased in the 2019 sampling, i.e. more individuals shared genomic regions by descent, suggesting increased mating among relatives (Fig. 2C). The number of deleterious mutations remained at low levels in both samplings (Fig. 2D).

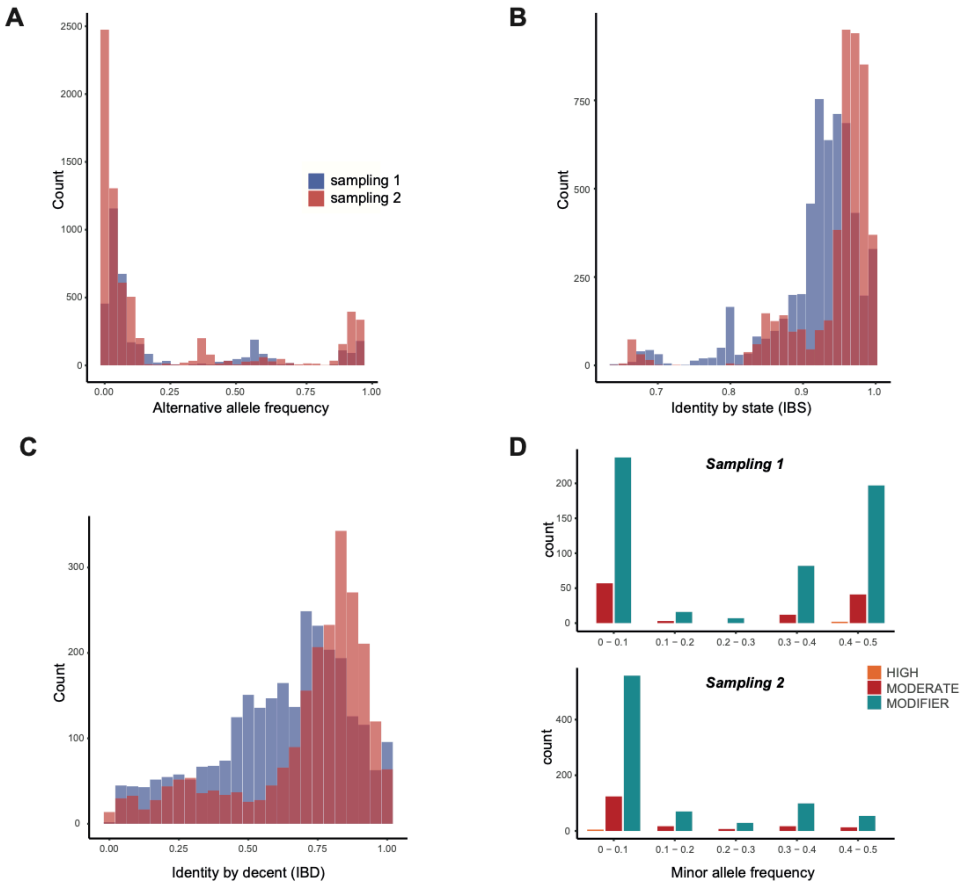


Figure 2. **Polymorphism spectra, identity by state (IBS) and identity by descent (IBD).** A) Alternative allele frequencies of the first (blue) and second (red) Ticino sampling. B) IBS histogram of the first and second Ticino sampling, colors are as in A). C) IBD histogram of the first and second Ticino sampling, colors are as in A) and B). D) Minor allele frequency spectra of high, moderate and modifier impact mutations in both Ticino samplings.

Vegetative compatibility and mating type diversity

Given the frequent observation of perithecia in the field (Bissegger et al., 1997; Prospero et al., 2006), populations in southern Switzerland are considered as sexually reproducing. Therefore, we investigated mating type frequencies were indeed balanced as expected. The mating types in the 1990s sampling were at a balanced ratio of 35 MAT-1 isolates, 36 MAT-2 isolates and three heterokaryons (i.e. with both mating type idiomorphs) (Fig. 3A). In contrast, the 2019 sampling was skewed towards MAT-2, with 42 MAT-2 isolates (61.8%), 20 MAT-1 isolates (29.4%) and 6 heterokaryons (8.8%). The strongest decrease in MAT-1 incidence was observed in Gnosca (Fig. 3A). The dominance of MAT-2 in the 2019 sampling appears not to be due to increased clonal reproduction, as we did not observe more clones in this sampling and changes of mating type ratios were found across all major vc types (Fig. 3B). We indeed observed relatively stable frequencies across both samplings, with EU-01, EU-02, EU-05 and EU-06 being the most abundant vc types (Fig. 3B-C, Table 1). Overall, the observed vc type diversity was lower than the potential diversity under random mating (Table 2). Differences

in vc type diversity between samplings were only marginally significant in Gnosca. (p=0.0396, Table 2). Association studies of genome-wide SNPs confidently identified significant SNPs at all five polymorphic *vic* loci ($p < 1x10^{-50}$ Supp. Fig. 3). Allele frequencies for known *vic* loci, significantly deviated from 0.5 in both samplings with the exception of *vic2* (exact binomial test p<0.05, Fig. 3D).

Table 2. **Expected and observed vc type diversity in relation to the number of observed polymorphic *vic* loci in all sampling sites and at both time points**.

Site	polymorphic *vic* loci	potential vc types	vc types 1990s	vc types 2019	X2	df	P-value
Lumino	5	32	9	8	12.4	11	0.415
Gnosca	5	32	7	5	14.7	7	0.0396
Novaggio	4	16	6	8	12.2	9	0.204

CHV1 prevalence and contributors to genome-wide diversity

Virus incidence appeared to be stable over time, with ~45 % of *C. parasitica* isolates being infected in both samplings. Virus infection equally affected both mating types, but showed some variation among sampling sites (Fig. 3A). We additionally investigated the CHV1 prevalence among the different vc types in both samplings. Virus incidence was higher in abundant vc types, but was also detected in some rare vc types (Fig. 3C). Although mating type ratios, vc type diversity and CHV1 prevalence varied among sampling sites, we did not observe spatial structuring of vc types and CHV1-infected isolates in 2019 (Fig. 3E). To test for the association of vc and mating types, presence/absence of CHV1-infection, as well as the sampling site and year with genetic diversity as inferred by IBS we performed a PERMANOVA. The analysis revealed that only vc types significantly explained variation in the observed genetic diversity (p<0.001, Supp. Table 1). The corresponding R^2 values and effect size suggests that the association with vc types increased over the sampling period (R^2=0.82 and 0.98 in the first and second sampling, respectively).

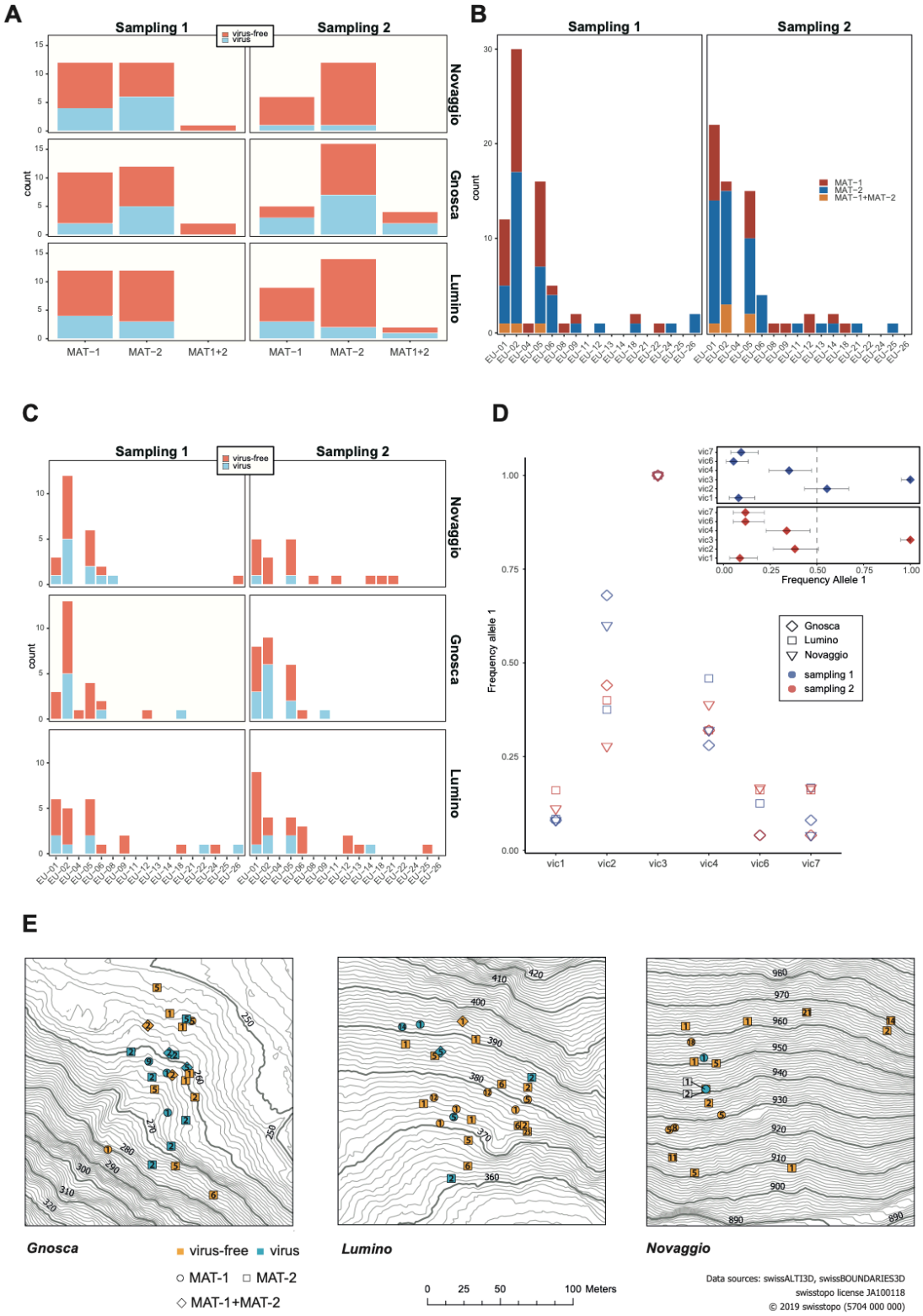


Figure 3. **Mating type and vegetative compatibility (vc) type ratios, virus incidence and spatial distribution.** A) Number of mating types and virus infection in both samplings in Novaggio, Gnosca and

Lumino. B) Ratio of mating types found in individual vc types in both samplings. C) Virus infection among vc types in both samplings in Novaggio, Gnosca and Lumino. D) Large panel: Allele frequencies of allele 1 at all six *vic* loci in all three populations at both samplings. *vic5* is commonly not shown, as the locus has no effect in virus transmission (Cortesi and Milgroom, 1998). Top panel: Allele frequencies of allele 1 at all *vic* loci in both samplings (merged populations) and corresponding confidence intervals showing deviations from 0.5 as inferred by exact binomial tests. E) Spatial distribution of sampled trees in Gnosca, Lumino and Novaggio in 2019. Colors indicate virus infection (orange: virus-free, blue: virus infected), shapes indicate mating type and number show vc type (EU-type) of sampled isolates.

Discussion

We investigated the genetic diversity of *C. parasitica* populations and prevalence of CHV1 in southern Switzerland (Ticino) at two samplings (1990s and 2019), to assess changes in population structure and potential impacts of CHV1 on fungal population after ~30 years. Overall, we detected no substantial changes in the pathogen population structure over the study period, but populations showed signals of increased mating among related individuals. Allele frequencies at known *vic* loci were not found to be at equilibrium, likely explained by the recent migratory bottleneck of *C. parasitica* during the invasion into Europe. Similarly, the number of observed vc types was lower than expected from allele frequencies alone. However, *vic* loci associated with asymmetric virus transmission might potentially be under positive frequency-dependent selection, possibly resulting in selection against *vic* alleles which would further enhance CHV1 spread in *C. parasitica* populations in southern Switzerland. Notably, despite low vegetative compatibility barriers, CHV1 infection was found in only ~45% of isolates in both samplings, indicating long-term infection stability.

Retention of pathogen diversity facilitates mycovirus spread

The mycovirus CHV1 has been shown to have a significant negative impact on the fitness (in particular parasitic growth and sporulation capacity) of its fungal host *C. parasitica* (Choi and Nuss, 1992). Hence, CHV1 may influence fungal population dynamics and act as a selective force at *vic* loci to limit virus transmission within the population. In our study, we found no evidence that CHV1 presence/absence significantly impacts genetic diversity in *C. parasitica* populations in southern Switzerland. Moreover, spatial analyses did not reveal patterns of fungal genotype or CHV1 clustering. However, *C. parasitica* populations likely attain a state of high vegetative compatibility (vc) type diversity to prevent virus transmission, as *vic* alleles have previously been found to be under balancing selection (Milgroom et al., 2018). Interestingly, European *C. parasitica* populations have consistently shown lower vc type diversity than expected under random mating (Ježić et al., 2018). Consistent with previous findings, we found imbalanced *vic* allele frequencies in *C. parasitica* populations in southern Switzerland (Milgroom and Cortesi, 1999). Moreover, the populations were largely dominated by a few vc types (EU-01, EU-02, EU-05 and EU-06) and occurrence of vc types did not substantially change over the span of three decades. The observed diversity of *C. parasitica* populations in southern Switzerland is likely linked to the relatively recent introduction of the pathogen into Europe, where *vic* alleles have not yet reached

a state of equilibrium (Milgroom and Cortesi, 1999). However, our results indicate that in absence of substantial rates of clonality, low vc type diversity is likely maintained by increased mating among related individuals of dominant vc types, as well as genetic drift.

Non-random mating of fungal populations

C. parasitica populations in southern Switzerland show strong signatures of non-random mating as identified by high levels of identity by state (IBS) and descent (IBD). Moreover, the presence of both mating types in all dominant vc types, further point to high levels of mating among those vc types, rather than clonal propagation. In fact, as all dominant vc types (i.e. EU-01, EU-02, EU-05 and EU-06) are only polymorphic at the same two *vic* loci, recombination does not result in the creation of additional vc types (Cortesi and Milgroom, 1998). Subsequently, segregation of genotypes is increasingly linked to conserved markers such as vc types, resulting in high relative importance of those markers to overall observed genome-wide diversity. Strikingly, we found skewed mating type ratios in 2019, with an overrepresentation of mating type MAT-2 at all sampled sites. In absence of strong signals for clonal reproduction, this could indicate unisexual reproduction (i.e. sexual reproduction in absence of the opposite mating type (Feretzaki and Heitman, 2013)). Unisexual reproduction has been reported in several fungal species, including *Cryptococcus neoformans*, *Neurospora* spp., or *Huntiella moniliformis* (Glass and Smith, 1994; Lin et al., 2005; Sun et al., 2014; Wilson et al., 2015). In *C. parasitica*, McGuire et al. (2004) reported no mating type segregation among progeny from perithecia that resulted from selfing. The molecular mechanisms and factors triggering such events in *C. parasitica* remain poorly resolved. However, unisexual reproduction might constitute a plausible mechanism to explain skewed mating type ratios in sexually reproducing *C. parasitica* populations.

Effects of asymmetric virus transmission on vc type diversity

Selective forces acting on *vic* loci associated with asymmetric CHV1 transmission might act as an additional constraint preventing vc type diversification in southern Switzerland. The four dominant vc types only differ at loci *vic2* and *vic4*, the latter locus having no effect on CHV1 transmission (Cortesi et al., 2001). While both alleles at *vic2* show relatively strong virus inhibition effects, asymmetric virus transmission is very pronounced at loci such as *vic1* and *vic7* (Cortesi et al., 2001). For instance, CHV1 transmission is almost 100% successful when the virus donor strain carries allele *vic1-2* and the virus-free recipient carries *vic1-1*. Conversely, CHV1 transmission is highly reduced with a *vic1-1* donor and *vic1-2* recipient (Cortesi et al., 2001). In *C. parasitica* populations with dominance of a few *vic* genotypes, polymorphism at *vic1* might lead to selection of allele *vic1-2* over *vic1-1* (Milgroom et al., 2018). We found all dominant vc types in southern Switzerland to carry *vic1-2,* while allele frequencies for *vic1-1* were ≤ 0.1 in both samplings. This could indicate selection against rare vc types with allele *vic1-1*, which can act as CHV1 donors for dominant vc types with allele *vic1-2*. Contrary, at *vic2* with an overall strong virus inhibition effect (Cortesi et al., 2001), allele frequencies did not significantly deviate from 0.5 over ~30 years, suggesting signals of balancing selection. Hence, in populations with low vc

type diversity, CHV1 transmission among dominant vc types might partially be reduced through maintaining balanced allele frequencies at *vic2*, as well as selection against asymmetric *vic* alleles. However, further analyses are required to disentangle selection at *vic* loci from recent demographic events in European *C. parasitica* populations.

CHV1 infection dynamics in *C. parasitica* populations

Despite retained low vc type diversity and high prevalence of a few dominant vc types in Ticino, we found stable CHV1 infection rates of ~45% among isolates over a period of ~30 years. This CHV1 incidence is within the previously reported range of 40-75% for populations in the same region (Bissegger et al., 1997; Bryner et al., 2012, Ježić et al., submitted). Interestingly, despite the overall CHV1 incidence remaining stable over time, repeated sampling of individual *C. parasitica* cankers revealed a highly dynamic system, where cankers may change between a virus-infected and virus-free state over the course of a few years (Bryner et al., 2014; Ježić et al., 2018). In fact, by drastically slowing fungal metabolism, CHV1 might reduce *C. parasitica* survival on trees and possibly paving the way for the reactivation of virulent, virus-free cankers (Bryner et al., 2014). As a consequence, highly virulent CHV1 strains might be less likely to persist and spread within a fungal population due to the strong negative fitness impact they assert on their fungal host (Morozov et al., 2007). Virus infection dynamics in *C. parasitica* populations could further be determined by the chestnut blight canker microbiome. Recent findings on the American chestnut (*Ca. dentata*) suggest that CHV1 persistence might be influenced by other fungi residing within cankers, potentially promoting the resurgence of virulent *C. parasitica* by outcompeting hypovirulent *C. parasitica* strains (Kolp et al., 2020). Although our results suggest long-term stability of CHV1 infection in *C. parasitica* populations with naturally occurring hypovirulence, the underlying fungal-viral dynamics may not be captured well by overall infection rates. CHV1 genome sequence-based analyses and infection assays would allow to assess temporal persistence and virulence of individual viral genotypes in *C. parasitica* populations.

Trajectories in host-pathogen-parasite interactions

To date, CHV1 is the most effective biological control agent against chestnut blight in Europe, as there is no evidence for resistance in the *Ca. sativa* host (Bryner et al., 2012). Compared to annual plants, perennials such as *Ca. sativa* might be prone to extended inoculum build up and slow selection for resistance due to the lower generation turnover (Dumartinet et al., 2020). Moreover, in southern Switzerland, few new chestnut trees grow from seeds, while most trees regrow as vegetative sprouts from existing stumps or dead trees. Consequently, changes in host genotypic diversity are likely minimal and resistance against chestnut blight is unlikely to rise quickly. Hence, co-evolutionary interactions in the invasive range of *C. parasitica* are likely more pronounced between pathogens and hyperparasites. Our analyses revealed no major changes in the fungal population structure and CHV1 infection rates over three decades while retaining imbalanced *vic* allele ratios and the dominance of single vc types. However, the introduction of novel genotypes with a selective advantage at *vic*

loci could result in higher vc type diversity, subsequently reducing CHV1 presence in southern Switzerland. Further studies focusing on genetic interactions will be required to investigate co-evolutionary dynamics of *C. parasitica* and CHV1. Potential candidate genes include CHV1 virulence-associated proteases (Xiong et al., 2019), as well as genes involved in fungal antiviral RNA silencing (Chun et al., 2020; Zhang et al., 2014). Environmental factors such as temperature have also been identified to impact interactions between CHV1 and *C. parasitica* (Bryner and Rigling, 2011). Hence, studies considering environmental variables in pathogen-hyperparasite systems will further elucidate how changing environmental conditions shape pathogen-hyperparasite co-evolutionary dynamics and the resulting effects on the plant host.

Methods

Origin of *Cryphonectria parasitica* isolates

The 142 *C. parasitica* isolates analyzed in this study originated from three populations (Gnosca, Lumino and Novaggio) in the canton Ticino in southern Switzerland. About half of the isolates (n=74) were obtained from the WSL fungal collection, representing the first sampling in 1990 (Lumino and Gnosca, Bissegger et al., 1997; Prospero and Rigling, 2012) and in 1996 (Novaggio, Prospero and Rigling, 2012) (Table 1). Additional 68 isolates originated from a resampling of the three populations in 2019 (Table 1). These isolates were recovered from randomly selected bark cankers, as described in Prospero and Rigling (2012), which likely differed from those in the first sampling in the 1990s. All isolates from southern Switzerland were screened molecularly for mating and vegetative compatibility (vc) types, following the protocol of Cornejo et al. (2019). Ambiguous molecular vc types results were additionally confirmed by pairing tester strains of known vc type with the unknown isolate and assessing the merging/barrage response (Bissegger et al., 1997). Isolates were prepared for whole-genome sequencing as described in Stauber et al. (2020a). Sequencing was performed on the NovaSeq6000 platform (Illumina, San Diego, USA) at the Functional Genomics Center Zurich (FGCZ). We additionally included 95 sequenced isolates from Stauber et al. (2020), covering the broader European and North American *C. parasitica* diversity (NCBI accessions see Supp. Table 1). All sequences were trimmed, aligned to the *C. parasitica* reference genome EP155 (Crouch et al., 2020) and variants were called and filtered as described in Stauber et al. (2020a). For removal of SNPs associated with the mating type region we identified highly associated SNPs in an association study with TASSEL 5 (Bradbury et al., 2007; Stauber et al., 2020a). We subsequently removed all SNPs in the mating type region with a p-value threshold of $p \leq 1 \times 10^{-5}$ (Supp. Fig. 1).

Table 2. **Sampling locations, sampling years and number of isolates obtained from localities in southern Switzerland**

Location	Sampling Year	n	Publication
Lumino	1990	24	(Bissegger et al., 1997)
	2019	25	
Gnosca	1990	25	(Prospero and Rigling, 2012)
	2019	25	
Novaggio	1996	25	(Prospero and Rigling, 2012)
	2019	18	

CHV1 prevalence in Ticino populations

To quantify CHV1 presence in *C. parasitica* isolates from Gnosca, Lumino and Novaggio, we performed a rapid one step reverse transcriptase polymerase chain reaction (RT-PCR). For this, the isolates were cultured on potato dextrose agar (PDA, 39g/L; BD Becton, Dickinson and Company) and incubated for five days at room temperature and natural day light. Hyphae of growing cultures were then transferred into PCR plate wells, by vertically piercing through the culture with a sterile toothpick and smearing the adhered residues into the PCR wells. Then, 7 µl of reaction mix from the PrimeScript One Step RT-PCR kit Ver.2 (Takara Bio) were added to each well on ice together with the CHV-1 forward and reverse primers hvep1 and hvep2 (Gobbin et al., 2003). The RT-PCR was run following an adjusted protocol by Urayama et al. (2015), with 50 °C for 30 min for reverse transcription, followed by 94°C for 2 min for DNA polymerase activation, then 33 cycles of denaturation at 94°C for 30 s, annealing at 58 °C for 30s and extension at 72 °C for 1 min. Virus presence was then confirmed through gel electrophoresis (1.2% agarose gel, 50 min, 90 VDC) with bands at 350 bp.

Analysis of population structure

For genome-wide analyses of the European and North American population structure we used fineSTRUCTURE, selecting the unlinked model (Lawson et al., 2012). We removed all singletons and generated the required input phasefiles in R. We then ran the second stage of fineSTRUCTURE in "automatic mode" to obtain the combined coancestry matrix, which was used as input for finestructuregreedy.sh bash script. The chunkcounts.out and fineSTRUCTURE tree file were then used to generate a population dendrogram, using functions from the R fineSTRUCTURE library (FinestructureLibrary.R). Additionally, we investigated the European and Ticino diversity by performing a principal component analysis (PCA) as implemented in the R package ade4 v1.7.15 (Bougeard and Dray, 2018). For analysis of the three Ticino populations we generated unrooted phylogenetic networks with SplitsTree v4.14.6 (Huson et al., 2008).

Population genetic and relatedness analyses

We calculated nucleotide diversity for European isolates, as well as both Ticino samplings separately, using the nuc.div function from the R package pegas (Paradis, 2010). Allele frequencies were calculated for each Ticino

sampling with VCFtools v0.1.15 (Danecek et al., 2011). To investigate identity among individuals of the same sampling, we conducted an identity-by-state (IBS) analysis as implemented in the R package SNPrelate v1.22.0 (Zheng et al., 2012). To further investigate relatedness among individuals of the same sampling, an identity-by-descent (IBD) analysis was performed using the hmmIBD software (Schaffner et al., 2018). For this, we changed the number of analyzed scaffolds to nchrom=13. Moreover, the default recombination rate (r) was changed to estimates for *C. parasitica* with $r = \frac{\rho}{2Ne}$. The mean effective population size (2Ne) and the mean population recombination rate ρ were inferred with LDhat (McVean and Auton, 2007; Stauber et al., 2020a). We additionally annotated synonymous and non-synonymous mutations with SNPeff v5.0 (Cingolani et al., 2012). Moreover, genome-wide association study (GWAS) was performed with TASSEL 5 (Bradbury et al., 2007) to identify SNPs associated with known *vic* loci. To assess the relative importance of vc types, mating types, sampling location and CHV1 presence/absence on the observed genetic diversity as inferred by IBS in each sampling, we ran a permutational multivariate analysis of variance (PERMANOVA, (Anderson, 2001)) using the R package vegan v2.5.6 (Oksanen et al., 2019). The potential number of vc types in each population under random mating was calculated using the formula 2n, where n = number of polymorphic *vic* loci at the first sampling. Differences in vc type diversity between the two samplings were tested for significance by performing a chi-square test. We additionally tested if allele frequencies of individual *vic* loci significantly deviated from 0.5 by performing an exact binomial test.

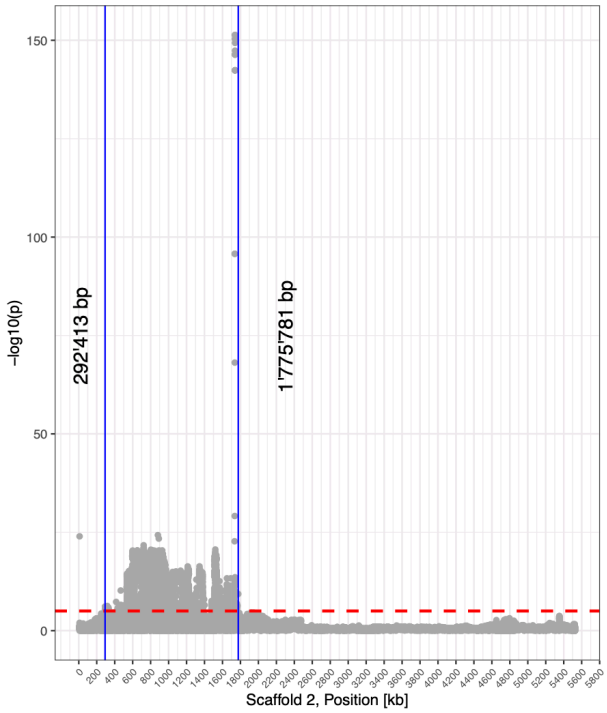
Spatial analysis of Ticino populations

For spatial visualization of genotypes as well as CHV1 distribution at the three sampling sites, GPS coordinates were collected with the app My GPS Location for each sampled tree in 2019 (second sampling). Topographic maps were generated in ArcGIS Pro v2.3.0 (ESRI Inc., USA), using the swissALTI3D elevation model and the swissBOUNDARIES3D for all administrative units and national boundaries (swisstopo license: JA100118).

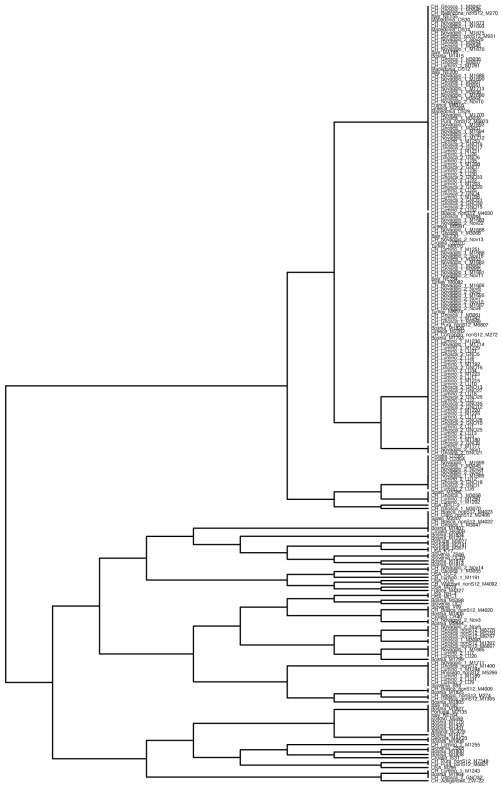
Acknowledgements

We thank Eva Augustiny, Silvia Kobel, Aria Minder, Quirin Kupper, and Hélène Blauenstein for sampling and laboratory assistance. We acknowledge the Genetic Diversity Centre (GDC), ETH Zurich, and the Functional Genomics Center Zurich (FGCZ) for technical support and facility access. Franziska Schmid greatly helped with spatial visualizations. Thanks to Alice Fleurtey for insightful discussions. L.S. was supported by the Swiss National Science Foundation (grant 170188 to S.P.).

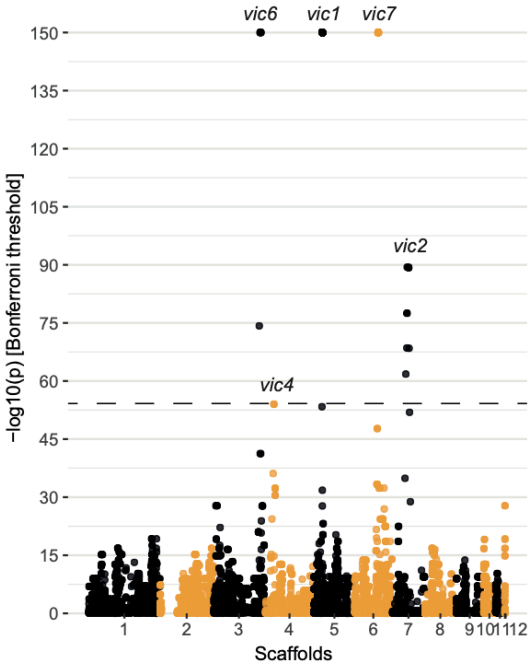
Supplementary Information



Supplementary Figure 1: Association mapping (GWAS) for the mating type locus. *p-value* based threshold for removal of mating type locus (MAT) associated SNPs.



Supplementary Figure 2: Population assignment dendrogram as inferred by fineSTRUCUTRE, with isolate identifiers.



Supplementary Figure 3: Genome-wide association mapping showing significantly SNPs associated with known *vic* loci ($p < 1x10^{-50}$). Associated SNPs with locus *vic3* remain undetected, as it was not found to be heteroallelic in *C. parasitica* populations in southern Switzerland. *p-values* for loci vic1, vic6 and vic7 are $p < 1x10^{-150}$.

Supplementary Table 1: PERMANOVA result tables, testing the impact of vc and mating types, sampling location and virus infection on genome-wide diversity (IBS).

Results sampling 1:

Sampling 1 Diversity	Df	SumsOfSqs	MeanSqs	F.Model	R^2	Pr(>F)
vctype	11	9.9531	0.90483	26.1623	0.8192	0.001***
MAT	2	0.0866	0.0433	1.2518	0.00713	0.244
Sampling location	2	0.0989	0.04947	1.4303	0.00814	0.206
Virus	1	0.0398	0.03984	1.1519	0.00328	0.276
Residuals	57	1.9713	0.03459		0.16225	
Total	73	12.1498			1	

Result sampling 2:

Sampling 2 Diversity	Df	SumsOfSqs	MeanSqs	F.Model	R^2	Pr(>F)
vctype	12	9.855	0.82125	208.372	0.97952	0.001***
MAT	2	0.0039	0.00193	0.489	0.00038	0.631
Sampling location	2	0.0027	0.00135	0.343	0.00027	0.804
Virus	1	0.0044	0.00436	1.107	0.00043	0.318
Tree Diameter	1	0.002	0.00204	0.517	0.0002	0.468
Residuals	49	0.1931	0.00394		0.01919	
Total	67	10.061			1	

References

Anagnostakis SL. 1988. Cryphonectria parasitica, Cause of Chestnut Blight, Advances in Plant Pathology. Academic Press Limited. doi:10.1016/B978-0-12-033706-4.50011-6

Anagnostakis SL. 1982. Biological control of chestnut blight. *Science (80-)* **215**:466–471.

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**:32–46. doi:10.1046/j.1442-9993.2001.01070.x

Bissegger M, Rigling D, Heiniger U. 1997. Population structure and disease development of *Cryphonectria parasitica* in European chestnut forests in the presence of natural hypovirulence. *Phytopathology* **87**:50–59. doi:10.1094/PHYTO.1997.87.1.50

Bougeard S, Dray S. 2018. Supervised multiblock analysis in R with the ade4 package. *J Stat Softw* **86**:1–17.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633–2635.

Branine M, Bazzicalupo A, Branco S. 2019. Biology and applications of endophytic insect-pathogenic fungi. *PLoS Pathog* **15**:e1007831.

Bryner SF, Prospero S, Rigling D. 2014. Dynamics of *Cryphonectria* hypovirus infection in chestnut blight cankers. *Phytopathology* **104**:918–925.

Bryner SF, Rigling D. 2011. Temperature-Dependent Genotype-by-Genotype Interaction between a Pathogenic Fungus and Its Hyperparasitic Virus. *Am Nat* **177**:65–74. doi:10.1086/657620

Bryner SF, Rigling D, Brunner PC. 2012. Invasion history and demographic pattern of Cryphonectria hypovirus 1 across European populations of the chestnut blight fungus. *Ecol Evol* **2**:3227–3241.

Choi GH, Nuss DL. 1992. Hypovirulence of chestnut blight fungus conferred by an infectious viral cDNA. *Science (80-)* **257**:800–803.

Chun J, Ko Y-H, Kim D-H. 2020. Transcriptome Analysis of *Cryphonectria parasitica* Infected With Cryphonectria hypovirus 1 (CHV1) Reveals Distinct Genes Related to Fungal Metabolites, Virulence, Antiviral RNA-Silencing, and Their Regulation. *Front Microbiol* **11**:1711.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**:80–92.

Cornejo C, Šever B, Kupper Q, Prospero S, Rigling D. 2019. A multiplexed genotyping assay to determine vegetative incompatibility and mating type in *Cryphonectria parasitica*. *Eur J Plant Pathol* 1–11.

Cortesi P, McCulloch CE, Song H, Lin H, Milgroom MG. 2001. Genetic control of horizontal virus transmission in the chestnut blight fungus, *Cryphonectria parasitica*. *Genetics* **159**:107–118.

Cortesi P, Milgroom MG. 1998. Genetics of Vegetative Incompatibility in *Cryphonectria parasitica*. *Appl Environ Microbiol* **64**:2988–2994.

Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, Hillman B, Milgroom MG, Pangilinan J, Smith M, others. 2020. Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: A fundamental resource for an archetypical invasive plant pathogen. *Phytopathology*.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, others. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.

Demené A, Legrand L, Gouzy J, Debuchy R, Saint-Jean G, Fabreguettes O, Dutech C. 2019. Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of *Cryphonectria parasitica* in western Europe. *Fungal Genet Biol*.

Dumartinet T, Abadie C, Bonnot F, Carreel F, Roussel V, Habas R, Martinez RT, Perez-Vicente L, Carlier J. 2020. Pattern of local adaptation to quantitative host resistance in a major pathogen of a perennial crop. *Evol Appl* **13**:824–836.

Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V. 2012. The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Mol Ecol* **21**:3931–3946.

Feretzaki M, Heitman J. 2013. Unisexual reproduction drives evolution of eukaryotic microbial pathogens. *PLoS Pathog* **9**:e1003674.

Glass NL, Smith ML. 1994. Structure and function of a mating-type gene from the homothallic species *Neurospora africana*. *Mol Gen Genet MGG* **244**:401–409.

Gobbin D, Hoegger PJ, Heiniger U, Rigling D. 2003. Sequence variation and evolution of Cryphonectria hypovirus 1 (CHV-1) in Europe. *Virus Res* **97**:39–46.

Hartmann FE, de la Vega RC, Carpentier F, Gladieux P, Cornille A, Hood ME, Giraud T. 2019. Understanding adaptation, coevolution, host specialization, and mating system in castrating anther-smut fungi by combining population and comparative genomics. *Annu Rev Phytopathol* **57**:431–457.
Heiniger U, Rigling D. 1994. Biological control of chestnut blight in Europe. *Annu Rev Phytopathol* **32**:581–599.
Huson DH, Kloepper T, Bryant D. 2008. SplitsTree 4.0-Computation of phylogenetic trees and networks. *Bioinformatics* **14**:68–73.
Ježić M, Mlinarec J, Vuković R, Katanić Z, Krstin L, Nuskern L, Poljak I, Idžojtić M, Tkalec M, Ćurković-Perica M. 2018. Changes in cryphonectria parasitica populations affect natural biological control of chestnut blight. *Phytopathology* **108**:870–877. doi:10.1094/PHYTO-07-17-0252-R
Kolp M, Double ML, Fulbright DW, MacDonald WL, Jarosz AM. 2020. Spatial and temporal dynamics of the fungal community of chestnut blight cankers on American chestnut (*Castanea dentata*) in Michigan and Wisconsin. *Fungal Ecol* **45**:100925.
Koskella B. 2013. Phage-mediated selection on microbiota of a long-lived host. *Curr Biol* **23**:1256–1260. doi:10.1016/j.cub.2013.05.038
Krstin L, Novak-Agbaba S, Rigling D, Ćurković-Perica M. 2011. Diversity of vegetative compatibility types and mating types of *Cryphonectria parasitica* in Slovenia and occurrence of associated Cryphonectria hypovirus 1. *Plant Pathol* **60**:752–761.
Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet* **8**:e1002453.
Lin X, Hull CM, Heitman J. 2005. Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. *Nature* **434**:1017–1021.
McGuire IC, Marra RE, Milgroom MG. 2004. Mating-type heterokaryosis and selfing in *Cryphonectria parasitica*. *Fungal Genet Biol* **41**:521–533.
McGuire IC, Marra RE, Turgeon BG, Milgroom MG. 2001. Analysis of mating-type genes in the chestnut blight fungus, *Cryphonectria parasitica*. *Fungal Genetics Biol* **34**:131–144.
McVean G, Auton A. 2007. LDhat 2.1: A package for the population genetic analysis of recombination. *Dep Stat Oxford, OX1 3TG, UK*.
Mei L, Chen M, Shang Y, Tang G, Tao Y, Zeng L, Huang B, Li Z, Zhan S, Wang C. 2020. Population genomics and evolution of a fungal pathogen after releasing exotic strains to control insect pests for 20 years. *ISME J* 1–13.
Milgroom MG, Cortesi P. 1999. Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *Proc Natl Acad Sci U S A* **96**:10518–10523. doi:10.1073/pnas.96.18.10518
Milgroom MG, Smith ML, Drott MT, Nuss DL. 2018. Balancing selection at nonself recognition loci in the chestnut blight fungus, *Cryphonectria parasitica*, demonstrated by trans-species polymorphisms, positive selection, and even allele frequencies. *Heredity (Edinb)* **121**:511–523. doi:10.1038/s41437-018-0060-7
Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneko S. 1996. Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* **88**:179–190.
Morozov AY, Robin C, Franc A. 2007. A simple model for the dynamics of a host-parasite-hyperparasite interaction. *J Theor Biol* **249**:246–253. doi:10.1016/j.jtbi.2007.05.041
Nuss DL. 2005. Hypovirulence: Mycoviruses at the fungal-plant interface. *Nat Rev Microbiol* **3**:632–642. doi:10.1038/nrmicro1206
Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019. vegan: Community Ecology Package.
Papkou A, Gokhale CS, Traulsen A, Schulenburg H. 2016. Host-parasite coevolution: Why changing population size matters. *Zoology* **119**:330–338.
Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**:419–420.
Parratt SR, Laine AL. 2016. The role of hyperparasitism in microbial pathogen ecology and evolution. *ISME J* **10**:1815–1822. doi:10.1038/ismej.2015.247
Prospero S, Conedera M, Heiniger U, Rigling D. 2006. Saprophytic activity and sporulation of *Cryphonectria parasitica* on dead chestnut wood in forests with naturally established hypovirulence. *Phytopathology* **96**:1337–1344.
Prospero S, Rigling D. 2012. Invasion genetics of the chestnut blight fungus *Cryphonectria parasitica* in Switzerland. *Phytopathology* **102**:73–82. doi:10.1094/PHYTO-02-11-0055
Rigling D, Prospero S. 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: Invasion history, population biology and disease control. *Mol Plant Pathol* **19**:7–20. doi:10.1111/mpp.12542

Rogalski MA, Gowler CD, Shaw CL, Hufbauer RA, Duffy MA. 2017. Human drivers of ecological and evolutionary dynamics in emerging and disappearing infectious disease systems. *Philos Trans R Soc B Biol Sci* **372**:20160043.
Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. 2018. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. *Malar J* **17**:196.
Sotirovski K, Papazova-Anakieva I, Grünwald NJ, Milgroom MG. 2004. Low diversity of vegetative compatibility types and mating type of *Cryphonectria parasitica* in the southern Balkans. *Plant Pathol* **53**:325–333.
Stauber L, Badet T, Prospero S, Croll D. 2020a. Emergence and diversification of a highly invasive chestnut pathogen lineage across south-eastern Europe. *bioRxiv*.
Stauber L, Prospero S, Croll D. 2020b. Comparative Genomics Analyses of Lifestyle Transitions at the Origin of an Invasive Fungal Pathogen in the Genus *Cryphonectria* . *mSphere* **5**:1–17. doi:10.1128/msphere.00737-20
Sun S, Billmyre RB, Mieczkowski PA, Heitman J. 2014. Unisexual reproduction drives meiotic recombination and phenotypic and karyotypic plasticity in *Cryptococcus neoformans*. *PLoS Genet* **10**:e1004849.
Tellier A, Moreno-Gámez S, Stephan W. 2014. Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution (N Y)* **68**:2211–2224. doi:10.1111/evo.12427
Thrall PH, Laine AL, Ravensdale M, Nemri A, Dodds PN, Barrett LG, Burdon JJ. 2012. Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecol Lett* **15**:425–435. doi:10.1111/j.1461-0248.2012.01749.x
Tian B, Yang J, Zhang K-Q. 2007. Bacteria used in the biological control of plant-parasitic nematodes: populations, mechanisms of action, and future prospects. *FEMS Microbiol Ecol* **61**:197–213.
Urayama S, Katoh Y, Fukuhara T, Arie T, Moriyama H, Teraoka T. 2015. Rapid detection of *Magnaporthe oryzae* chrysovirus 1-A from fungal colonies on agar plates and lesions of rice blast. *J Gen plant Pathol* **81**:97–102.
Wilson AM, Godlonton T, der Nest MA, Wilken PM, Wingfield MJ, Wingfield BD. 2015. Unisexual reproduction in *Huntiella moniliformis*. *Fungal Genet Biol* **80**:1–9.
Xiong Q, Zhang Linqiao, Waletich J, Zhang Linlin, Zhang C, Zheng X, Qian Y, Zhang Z, Wang Y, Cheng Q. 2019. Characterization of the Papain-Like Protease p29 of the Hypovirus CHV1-CN280 in Its Natural Host Fungus *Cryphonectria parasitica* and Nonhost Fungus *Magnaporthe oryzae*. *Phytopathology* **109**:736–747. doi:10.1094/PHYTO-08-18-0318-R
Zhang DX, Spiering MJ, Nuss DL. 2014. Characterizing the roles of *Cryphonectria parasitica* RNA-dependent RNA polymerase-like genes in antiviral defense, viral recombination and transposon transcript accumulation. *PLoS One* **9**. doi:10.1371/journal.pone.0108653
Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B. 2012. A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* **28**:3326–3328. doi:10.1093/bioinformatics/bts606

General discussion

This doctoral thesis investigated the emergence and population dynamics of the invasive chestnut blight pathogen *C. parasitica* through comparative and population genomic analyses of a large set of *C. parasitica* and *Cryphonectria* spp. genomes. Gene content analyses of bark-inhabiting species in the family of Cryphonectriaceae, revealed a striking loss of genes associated with carbohydrate metabolization (CAZyme) in the invasive chestnut blight pathogen *C. parasitica.* The results suggest CAZyme gene loss might have promoted higher pathogenicity towards *Castanea* species. Further population genomic analyses within the species *C. parasitica* revealed that a highly successful clonal outbreak lineage in south-eastern Europe likely emerged through an intermediary, sexually recombining European bridgehead population. The expansion of *C. parasitica* across south-eastern Europe was accompanied by an evolutionary transition from a sexually reproducing mixed mating type population to a predominantly clonal single mating type population, potentially facilitating colonization of new areas. Lastly, this thesis explored host-pathogen-hypovirus interactions by studying pathogen population structure and hypovirus presence over ~30 years in local *C. parasitica* populations in southern Switzerland. Overall, the results revealed that *C. parasitica* populations in southern Switzerland predominantly consist of genetically highly similar genotypes. The observed genetic diversity is likely explained by the recent introduction of *C. parasitica* into Europe, nonrandom mating, as well as genetic drift. Retention of low vegetative incompatibility barriers of *C. parasitica* populations in southern Switzerland facilitated hypovirus transmission among individuals. However, despite low incompatibility barriers, stable hypovirus infection rates at ~45% can nonetheless be maintained for several decades.

Gene content-based pathogen prediction

Cryphonectria parasitica has recently emerged as a necrotrophic pathogen on non-Asian *Castanea* species and shows distinct CAZyme gene loss, compared to its largely non-pathogenic sister species in the genus *Cryphonectria*. Gene loss in *C. parasitica* not only affected CAZymes associated with saprotrophic lifestyles, but also CAZyme genes with a potential role in virulence. Loss of virulence related CAZymes in *C. parasitica* could constitute a mechanism to avoid triggering host defenses through molecular pattern recognition. Interestingly, recent comparative transcriptomic analyses of hypovirus-infected and hypovirus-free *C. parasitica* strains *in vitro* indicated that the hypovirus impacts carbohydrate metabolism, resulting in lower virulence (Chun et al., 2020). The results underpin the importance of CAZymes as virulence factors in *C. parasitica.* Genes underlaying key nutritional modes such as CAZymes, might play an important role for determining genetic factors of nutritional niche adaptation and pathogen emergence. Necrotrophic pathogens often deploy cell wall-degrading enzymes for host damage and colonization (Rodriguez-Moreno et al., 2018). In contrast, biotrophic pathogens tend to have fewer enzymes involved in cell wall degradation (Lorrain et al.,

2019; van de Vossenberg et al., 2019). Saprotrophic fungi often have an overall reduced CAZyme complement, but can have lifestyle-associated CAZyme expansion, e.g. for cellulose degradation (Ohm et al., 2012). Hence, CAZyme composition likely constitutes an important predictor for the evolution of pathogenic lifestyles. Human activities have highly contributed to the global spread of numerous species, including species which have emerged as invasive pathogens outside their native range. This raises the question if particular species traits facilitate invasiveness and pathogenicity and whether these traits are predictive. Host and geographic range expansion have been identified as a major factor facilitating the emergence of novel pathogens (Giraud et al., 2010; Smith and Guégan, 2010). Moreover, previous studies focusing on biological traits identified factors such as sexual reproduction, dispersal abilities, climate and habitat match, as well as host range as predictors of invasiveness (Philibert et al., 2011). However, whether genomic determinants facilitate the emergence of novel invasive pathogens still remains largely unexplored. Recent advances in genomic lifestyle prediction such as the CAZyme-Assisted Training And Sorting of -trophy (CATAStrophy) tool (Hane et al., 2020) might thus present useful tool gene content based pathogen prediction. Moreover, future comparative genomic or transcriptomic analyses of invasive pathogens and related sister species could further elucidate whether genomic determinants enable species to emerge as invasive pathogens.

The genomics of pathogen invasions

Population genomic analyses are highly useful to uncover the genetic basis of successful pathogen invasions, as well as identifying evolutionary forces and mechanisms shaping pathogen invasions (Welles and Dlugosch, 2018). Pathogen invasions are often characterized by low genetic diversity within founding populations, which reduces adaptive genetic variation (Barnes et al., 2014; Dutech et al., 2012; Havenga et al., 2020). Yet, many pathogen invasions were successful despite low genetic diversity within founding populations. Our population genomic analyses reconstructing the emergence and diversification of a highly successful clonal *C. parasitica* lineage in south-eastern Europe, uncovered that the lineage likely emerged from a European bridgehead population. However, whether bridgehead populations are the source for adaptive evolution in introduced populations is still debated (Bertelsmeier and Keller, 2018). Combining population genomic analyses and identification of candidate loci for adaptation through selection scans, as well as associating loci with environmentally dependent phenotypic traits could elucidate the origins of adaptive evolution of invasive populations (Welles and Dlugosch, 2018). Recent studies also point to genomic architecture and genome plasticity as a source for adaptive potential in invasive pathogens. For instance, in the asexually reproducing invasive forest pathogen *Phytophthora ramorum* genotypic diversity is created through mitotic recombination (Dale et al., 2019). Moreover, accessory genomic regions enriched in effectors and transposable elements are likely important for adaptive evolution in *P. ramorum* lineages (Dale et al., 2019). *C. parasitica* did not display a preferential association of candidate effector genes with transposable elements. However, we detected specific

copy-number variation patterns and evidence for ongoing transposable element activity in the invasive clonal *C. parasitica* lineage in south-eastern Europe, which potentially facilitates rapid adaptation (Lauer et al., 2018; Schrader and Schmitz, 2019; Stapley et al., 2015; Todd and Selmecki, 2020). As of today, few studies have investigated genome evolution of invasive pathogens (e.g. Dale et al., 2019; Hessenauer et al., 2020; Stauber et al., 2020). Hence, complementing population genomic analyses, as well as phenotyping with investigations into the genomic architecture of invasive species will give valuable insight into mechanisms underlaying invasive and adaptive processes.

Evolution of pathogen-hyperparasite systems and implications for biocontrol

The devastating effects of chestnut blight in Europe were largely mitigated by the hyperparasitic mycovirus Cryphonectria hypovirus 1 (CHV1), which has a strong negative fitness impact on *C. parasitica* and is effectively used as a biocontrol agent. Although hyperparasitic mycoviruses are likely ubiquitous, few mycoviruses are known to cause such severe effects on their hosts as CHV1 (Rigling et al., 2020). For instance, a recently discovered DNA mycovirus is suggested to convert the necrotrophic pathogen *Sclerotinia sclerotiorum* to a beneficial endophytic fungus (Zhang et al., 2020). Moreover, deleterious mycoviruses were detected in the invasive pathogen *Ophiostoma novo-ulmi*, which reduce fungal growth and asexual spore viability (Brasier et al., 2004). However, due to the vast diversification of vegetative compatibility types in *O. novo-ulmi*, mycovirus incidence declined rapidly and remain infrequent today (Brasier and Webber, 2019). Similarly, an increase of vegetative compatibility types in European *C. parasitica* populations could threaten mycovirus-mediated biocontrol success. This raises important questions on co-evolutionary dynamics of introduced pathogens and hyperparasites, as well as the long-term effectiveness of biocontrol. To date, no substantial changes in vegetative compatibility type diversity and CHV1 infection rates have thus far been observed in European *C. parasitica* populations. Yet, differences in CHV1 virulence have been found (Milgroom and Cortesi, 2004; Sotirovski et al., 2011), raising the question whether virulent hyperparasites that are suitable for biocontrol are able to persist within pathogen populations and if hyperparasite presence selects for higher pathogen virulence (Parratt and Laine, 2016). While gene-for-gene interactions have extensively been studied in plant-pathogen systems (Barrett et al., 2009; Dodds and Rathjen, 2010), little is known whether similar dynamics determine pathogen-hyperparasite interactions. Similar to plant-pathogen systems, the identification of virulence and resistance associated genes in both pathogens and hyperparasites, as well as selection analyses could elucidate co-evolutionary trajectories of pathogen-hyperparasite systems and the potential implications for biocontrol (Möller and Stukenbrock, 2017).

Conclusions and outlook

With the worldwide emergence of novel invasive pathogens, a fundamental understanding of the genetics of pathogen emergence and invasions is required for successful pest risk assessment and for developing effective control measures. *C. parasitica* presents relevant model to study pathogen invasions, due to its wide distribution range across the northern hemisphere, as well as a well-known disease history dating back over 100 years (e.g. Biraghi, 1946; Heald and Studhalter, 1914; Merkel, 1906). Moreover, the chestnut blight pathosystem provides an interesting model to study mycovirus-mediated biocontrol and the potential impacts on co-evolutionary dynamics of pathogens and hyperparasites on the plant host. Lastly, as the Cryphonectriaceae family harbors several recently emerged bark pathogens, future comparative genomic and transcriptomic analyses of multiple species within this family will enhance our understanding of the genetic foundation in the evolution of bark pathogens. As genome sequencing technologies become cheaper and more broadly available, genomic data offers the possibility to gain new insights and perspectives on the biology of invasive pathogens.

References

Barnes I, Wingfield MJ, Carbone I, Kirisits T, Wingfield BD. 2014. Population structure and diversity of an invasive pine needle pathogen reflects anthropogenic activity. *Ecol Evol* **4**:3642–3661.

Barrett LG, Kniskern JM, Bodenhausen N, Zhang W, Bergelson J. 2009. Continua of specificity and virulence in plant host–pathogen interactions: causes and consequences. *New Phytol* **183**:513–529.

Bertelsmeier C, Keller L. 2018. Bridgehead effects and role of adaptive evolution in invasive populations. *Trends Ecol Evol*.

Biraghi A. 1946. Cancro della corteccia del castagno. Ramo editoriale degli agricoltori.

Brasier C, Buck K, Paoletti M, Crawford L, Kirk S. 2004. Molecular analysis of evolutionary changes in populations of *Ophiostma novo-ulmi*. *For Syst* **13**:93–103.

Brasier CM, Webber JF. 2019. Is there evidence for post-epidemic attenuation in the Dutch elm disease pathogen *Ophiostoma novo-ulmi*? *Plant Pathol* **68**:921–929.

Chun J, Ko Y-H, Kim D-H. 2020. Transcriptome Analysis of *Cryphonectria parasitica* Infected With Cryphonectria hypovirus 1 (CHV1) Reveals Distinct Genes Related to Fungal Metabolites, Virulence, Antiviral RNA-Silencing, and Their Regulation. *Front Microbiol* **11**:1711.

Dale AL, Feau N, Everhart SE, Dhillon B, Wong B, Sheppard J, Bilodeau GJ, Brar A, Tabima JF, Shen D, Brasier CM, Tyler BM, Grünwald NJ, Hamelin RC. 2019. Mitotic recombination and rapid genome evolution in the invasive forest pathogen phytophthora ramorum. *MBio* **10**:1–19. doi:10.1128/mBio.02452-18

Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat Rev Genet* **11**:539–548.

Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V. 2012. The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Mol Ecol* **21**:3931–3946.

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

Hane JK, Paxman J, Jones DAB, Oliver RP, de Wit P. 2020. "CATAStrophy," a Genome-Informed Trophic Classification of Filamentous Plant Pathogens-How Many Different Types of Filamentous Plant Pathogens Are There? *Front Microbiol* **10**:3088.

Havenga M, Wingfield BD, Wingfield MJ, Dreyer LL, Roets F, Chen S, Aylward J. 2020. Low genetic diversity and strong geographic structure in introduced populations of the Eucalyptus foliar pathogen *Teratosphaeria destructans*. *Plant Pathol* **69**:1540–1550.

Heald FD, Studhalter RA. 1914. Birds as carriers of the chestnut blight fungus. *J Agric Res* **2**:405–422.

Hessenauer P, Fijarczyk A, Martin H, Prunier J, Charron G, Chapuis J, Bernier L, Tanguay P, Hamelin RC, Landry CR. 2020. Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. *Nat Ecol Evol* **4**:626–638.

Lauer S, Avecilla G, Spealman P, Sethia G, Brandt N, Levy SF, Gresham D. 2018. Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. *PLoS Biol* **16**:e3000069.

Lorrain C, dos Santos KC, Germain H, Hecker A, Duplessis S. 2019. Advances in understanding obligate biotrophy in rust fungi. *New Phytol* **222**:1190–1206.

Merkel HW. 1906. A deadly fungus on the American chestnut. *Annu Rep New York Zool Soc* **10**:97–103.

Milgroom MG, Cortesi P. 2004. Biological control of chestnut blight with hypovirulence: a critical analysis. *Annu Rev Phytopathol* **42**:311–338.

Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol* **15**:756.

Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, others. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog* **8**:e1003037.

Parratt SR, Laine AL. 2016. The role of hyperparasitism in microbial pathogen ecology and evolution. *ISME J* **10**:1815–1822. doi:10.1038/ismej.2015.247

Philibert A, Desprez-Loustau M-L, Fabre B, Frey P, Halkett F, Husson C, Lung-Escarmant B, Marçais B, Robin C, Vacher C, others. 2011. Predicting invasion success of forest pathogenic fungi from species traits. *J Appl Ecol* **48**:1381–1390.

Rigling D, Robin C, Prospero S. 2020. Mycovirus-Mediated Biological Control. *Ref Modul Life Sci*. doi:10.1016/b978-0-12-809633-8.21516-1

Rodriguez-Moreno L, Ebert MK, Bolton MD, Thomma BPHJ. 2018. Tools of the crook-infection strategies of fungal plant pathogens. *Plant J* **93**:664–674.
Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. *Mol Ecol* **28**:1537–1549.
Smith KF, Guégan J-F. 2010. Changing geographic distributions of human pathogens. *Annu Rev Ecol Evol Syst* **41**:231–250.
Sotirovski K, Rigling D, Heiniger U, Milgroom MG. 2011. Variation in virulence of Cryphonectria hypovirus 1 in Macedonia. *For Pathol* **41**:59–65.
Stapley J, Santure AW, Dennis SR. 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol Ecol* **24**:2241–2252.
Stauber L, Prospero S, Croll D. 2020. Comparative Genomics Analyses of Lifestyle Transitions at the Origin of an Invasive Fungal Pathogen in the Genus *Cryphonectria* . *mSphere* **5**:1–17. doi:10.1128/msphere.00737-20
Todd RT, Selmecki A. 2020. Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. *Elife* **9**:e58349.
van de Vossenberg BTLH, Warris S, Nguyen HDT, van Gent-Pelzer MPE, Joly DL, van de Geest HC, Bonants PJM, Smith DS, Lévesque CA, van der Lee TAJ. 2019. Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of *Synchytrium endobioticum*. *Sci Rep* **9**:1–14.
Welles SR, Dlugosch KM. 2018. Population genomics of colonization and invasion. Population Genomics. Springer. pp. 655–683.
Zhang H, Xie J, Fu Y, Cheng J, Qu Z, Zhao Z, Cheng S, Chen T, Li B, Wang Q. 2020. A 2-kb Mycovirus Converts a Pathogenic Fungus into a Beneficial Endophyte for Brassica Protection and Yield Enhancement. *Mol Plant* **13**:1420–1433.

Acknowledgements

I am very grateful to all the people who contributed and supported me doing this PhD project. In particular, I would like to thank my supervisors Dr. Simone Prospero and Prof. Dr. Daniel Croll. Thank you, Simone, for giving me the chance to work on this amazing project. I really appreciated your help, our discussions and the freedom you gave me to explore new ideas. I am also deeply thankful to Daniel Croll for his exceptional mentorship - thank you for teaching me in bioinformatics, population genetics, evolutionary biology and critical thinking. Our meetings and discussions were truly inspiring and always left me motivated and excited about science. I would also like to thank Dr. Pierre Gladieux, Prof. Dr. Niklaus Grünwald and Prof. Dr. Jacob Koella for agreeing to be experts on my PhD committee.

For generating large genomic data sets a lot of laboratory work is required, which I could not have done all on my own. I would like to thank Mihajlo Risteski, Kiril Sotirovski, Sandrine Fattoire, Dario Rüegg, Eva Augustiny and Christopher Schefer for helping with experiments, sampling, isolation, DNA extractions and genotyping. I am also very grateful to Quirin Kupper, Hélène Blauenstein, Silvia Kobel and Aria Minder for all their technical assistance in the laboratory at WSL and the Genetic Diversity Center Zürich. I would also like to thank Daniel Rigling for all his support and sharing his deep knowledge on the pathosystem with me.

I am also very thankful to everyone in the Laboratory of Evolutionary Genetics in Neuchâtel for all the nice and inspiring conversations, for sharing scripts with me and for always making me feel welcome and being part of the group. I really enjoyed my time with you in Neuchâtel – including beers, Indian food, barbecues and cakes. Moreover, I would like to thank all the members of the Phytopathology group at WSL for all the nice conversations we had and for all motivational support.

Last, but not least, a big thank you to Martin and my family for their support and believing in me. It must have been quite astonishing for them to see my transition from a musician, to an agronomist, to finally a scientist who is enthusiastic about population genomics and evolution of pathogenic fungi. Thank you for always being there for me!