



Emergence and diversification of a highly invasive chestnut pathogen lineage across southeastern Europe

Lea Stauber^{1,2}, Thomas Badet², Alice Feurtey^{2,3}, Simone Prospero¹, Daniel Croll^{2*}

¹Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland; ²Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland; ³Plant Pathology, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland

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 European chestnut stands. To unravel the chestnut blight invasion of southeastern 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 or North America. The expansion across southeastern 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.

*For correspondence:
daniel.croll@unine.ch

Competing interests: The authors declare that no competing interests exist.

Funding: See page 21

Received: 22 February 2020

Accepted: 17 February 2021

Published: 05 March 2021

Reviewing editor: Vincent Castric, Université de Lille, France

© Copyright Stauber et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

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 ([Gladiieux 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 through the reshuffling of genetic variants, 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 they 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). 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 *P. 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 (CNV) 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 ([Oggendorf 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 (*Castanea 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 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 southeastern 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 subsequently spread across southeastern 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 southeastern 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 southeastern 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 of global *C. parasitica* isolates

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 ([Figure 1A](#), [Supplementary file 1](#)). Isolates were sequenced at a mean depth of 8–53× 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 ([Figure 1—figure supplement 1](#)). This region encoded known mating type-associated genes and was characterized by a high SNP density

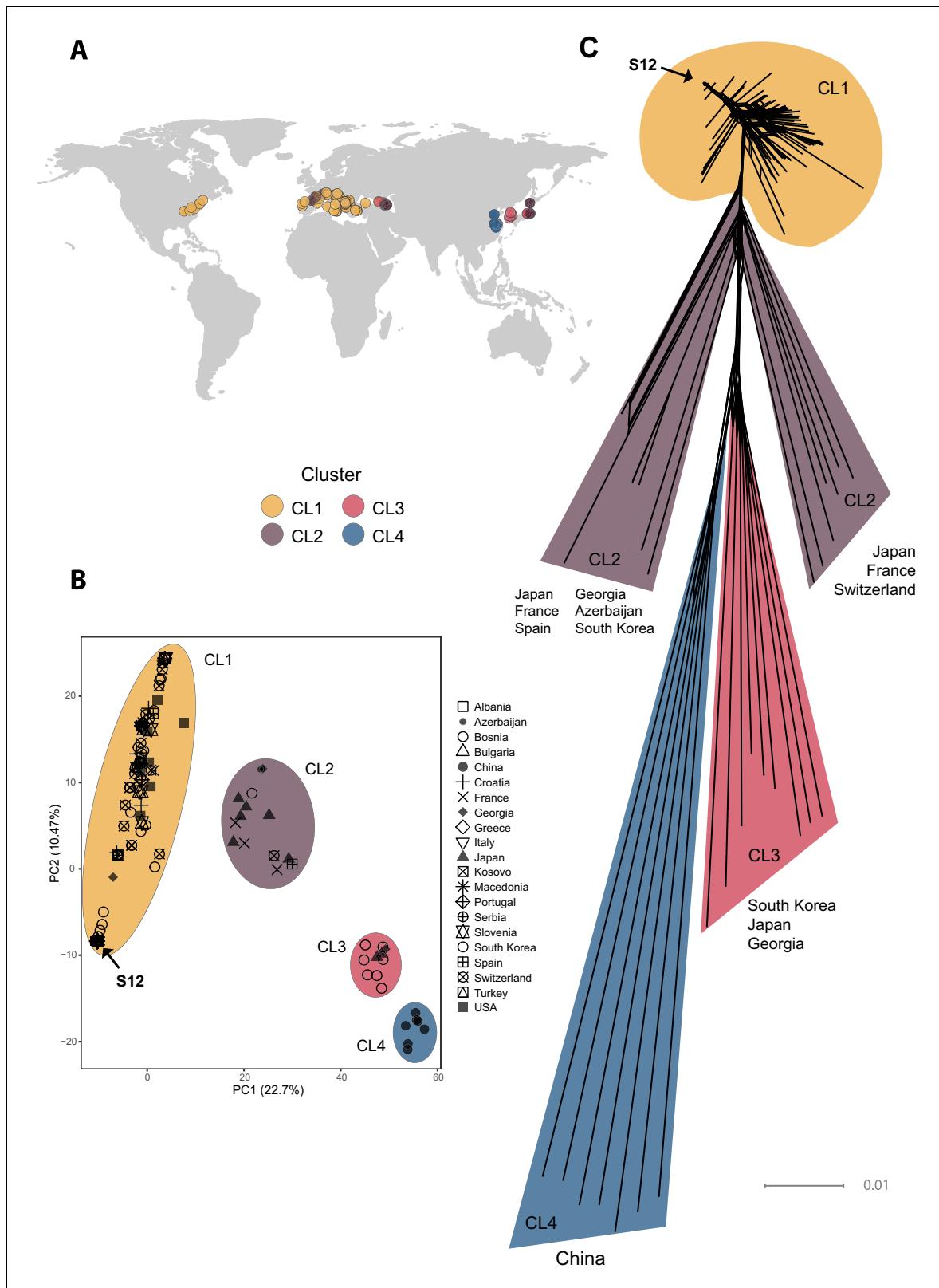


Figure 1. Genome-wide analyses of global *Cyphonectria parasitica* isolates. (A) Map of global sampling locations and (B) principal component analysis (PCA) of the 230 sequenced *C. parasitica* isolates. Colors indicate PCA clusters (CL1–CL4) and are as in (A). (C) SplitsTree phylogenetic network of 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). Figure 1 continued on next page

Figure 1 continued

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Genome-wide association mapping (GWAS) outcome for the mating type (MAT) locus.

Figure supplement 2. Distribution of observed single nucleotide polymorphism (SNP) filtration quality values.

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 (*Figure 1B, C*). The PCA revealed four major clusters (CL1–CL4, *Figure 1B*), showing weak geographic clustering except for isolates of Chinese origin (cluster CL4 in *Figure 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 *Figure 1B*). The phylogenetic network analysis revealed large phylogenetic distances outside the main European/North American CL1 cluster (*Figure 1B, C*).

Reconstruction of the European demographic history

We implemented a demographic model based on each cluster identified by the PCA (*Figure 1B*), merging isolates from the two smaller clusters (i.e., CL3 and CL4, *Figure 1B*) to obtain a sufficient sample size. The folded site frequency spectrum (SFS) for each of the three groups contained a large fraction of singletons, suggesting a population expansion (*Figure 2*). We used the software *dadi* (*Gutenkunst et al., 2009*) to identify the best fitting model considering no population size change or a single instantaneous population size change, as well as more complex models represented in *Figure 2* (*Supplementary file 2*). For the isolates outside the main European/North American group, we found support for a recent population expansion (CL2–CL4 clusters, *Figure 1B, Figure 2*). Indeed, for both clusters, the best Akaike information criterion (AIC) value was obtained for modeling a simple instantaneous expansion (AIC of 70.4 and 115.4 for the CL3+CL4 and CL2, respectively; *Supplementary file 2*). The second-best model was a more complex version of an expansion scenario including an initial bottleneck followed by exponential growth (AIC of 75.0 and 152.7 for the ‘bottlegrowth’ model, for the CL3+CL4 and CL2, respectively). However, for the large European/North American CL1 cluster, despite removing all but a single isolate of the S12 lineage, parameter optimization was not successful for any model or any randomly drawn initial parameters (see *Supplementary file 2*). This is likely explained by the unusual frequency spectra. As shown in *Figure 2*, the observed folded SFS for the European/North American cluster (CL1 *Figure 1B, Figure 2*) is W-shaped (red line in *Figure 2*) instead of a monotonously declining curve as is the case in the empirical spectra for other populations as well as in the simulations (*Figure 2*). The unusual frequency spectrum of the European/North American cluster potentially stems from the presence of multiple small groups of related genotypes due to inbreeding.

Phylogenomic analyses of European and North American isolates

To further investigate the European/North American CL1 cluster containing the S12 lineage (*Figure 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 (*Figure 3A, Figure 3—figure supplement 1*). Genotypes within the cluster show significant evidence for recombination (pairwise homoplasy index [PHI] test; $p < 0.0001$, *Figure 3—figure supplement 2*). 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 (*Figure 3A, Figure 3—figure supplement 1*). Vegetative compatibility-type diversity has been widely used as a proxy for genetic diversity in populations (*Heiniger and Rigling, 1994*). Our findings confirm the clear limitation of these markers as proxies for diversity (*Ježić et al., 2012*). Nevertheless, vegetative compatibility type diversity impacts mycovirus transmission dynamics and remains, hence, a relevant population trait. 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

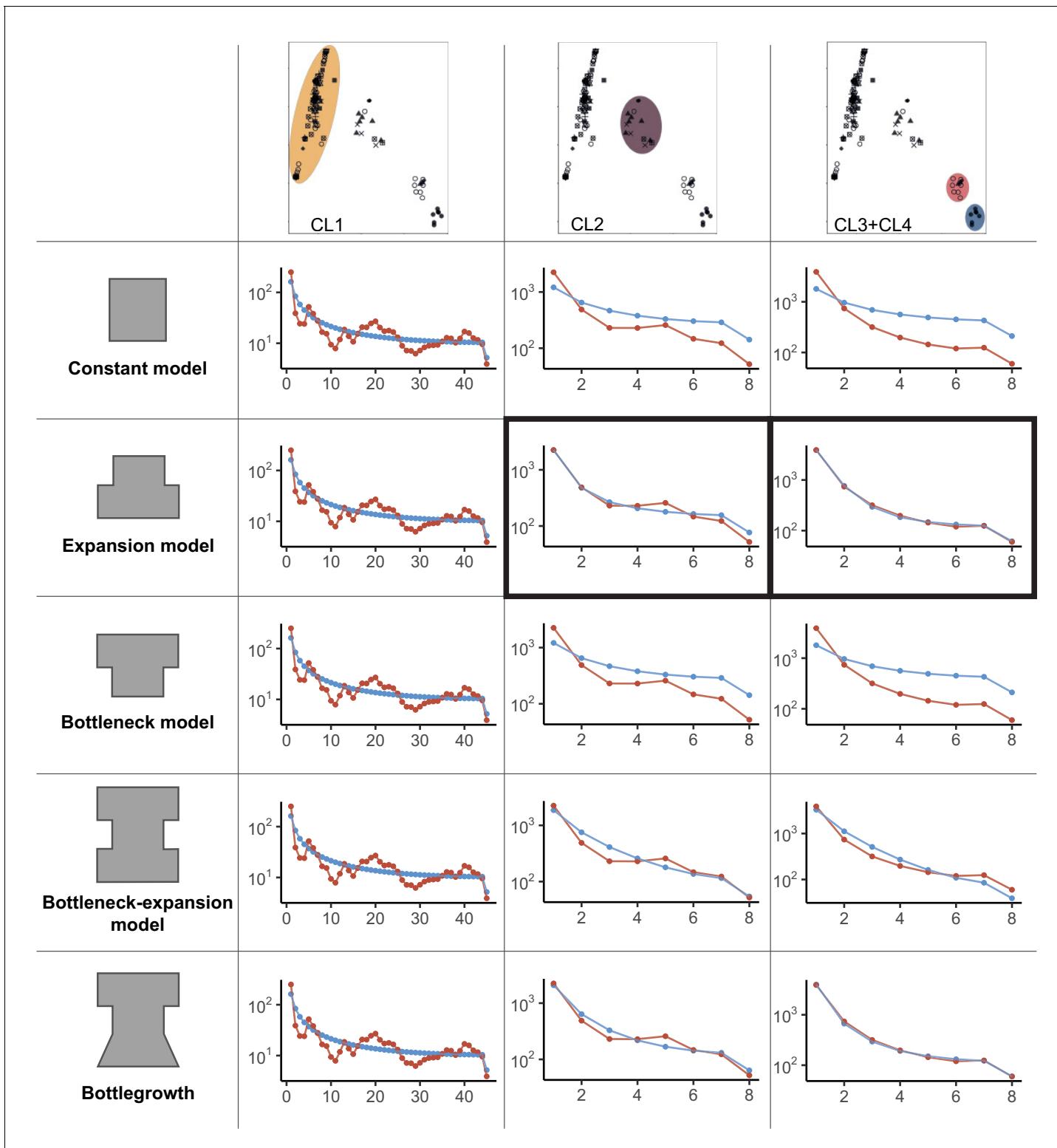


Figure 2. Demographic models for global *Cryphonectria parasitica* populations. Constant, expansion, bottleneck, bottleneck-expansion, and bottlegrowth population models were tested for three clusters (i.e., European/North American CL1, mixed European/Asian CL2, and Asian CL3 and CL4 clusters shown by color highlights). Clusters were inferred by principal component analysis (PCA) (see **Figure 1B**). The two smaller Asian clusters (top-right 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., constant, expansion and bottleneck, bottleneck-expansion and bottlegrowth) as inferred by *dadi*. Blue data points and line represent the model, and red lines show the empirical data. Best fitting models according to the Akaike

Figure 2 continued on next page

Figure 2 continued

information criterion (AIC) model comparisons are marked in black boxes. A detailed list of initial and best-fit parameters, as well as likelihood and AIC values, is shown in **Supplementary file 2**.

analyses by **Milgroom et al., 2008**, this group represents the invasive S12 lineage at the origin of the expansion of *C. parasitica* across southeastern Europe. Additionally, the phylogenetic network revealed closely related but not identical S12 genotypes of mating type MAT-2 ($n = 7$, **Figure 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) (**Figure 3A**, **Figure 3—figure supplement 1**). 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 (**Figure 3A, B**). Additionally, we found the S12 lineage in otherwise more diverse regions of Croatia and Bosnia (**Figure 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 southeastern Europe (**Figure 3B**).

Evidence for selection in European populations

We investigated potential selective sweeps in the focal European/North American cluster. To avoid a bias by the deep sampling of the S12 lineage, we excluded all but one of the S12 MAT-1 isolates (see 'Materials and methods'). We used RAiSD (**Alachiotis and Pavlidis, 2018**) to produce a composite score of selective sweep signals and identified two strong outlier loci. Despite no clear demographic model fitting the genomic data of the focal population, we nevertheless attempted to control for potential demographic history effects. We performed 10,000 simulations of a population under three different models without selection: no population size change (i.e., constant), expansion, and bottleneck. This was followed by a genome-wide selection scan on the simulated datasets using RAiSD. We assessed above what μ values selection signals are most likely exceeding effects by demographic events. The highest values of μ were obtained for the simulations with the expansion scenarios ($\mu = 37$). Consequently, we used this value as a conservative threshold to identify the most robust signatures of selection. Two outlier loci obtained in the selection scan on the observed data have μ values above 37. The strongest sweep locus was located at the boundary of the mating type locus on scaffold 2 (**Figure 3C**). The second sweep locus was detected on scaffold 1 and encompassed an ~120 kb locus at positions 2099–3309 kb. The region contains 335 genes of which 107 encode conserved protein domains (**Figure 3C, Supplementary file 3**). Hence, genetic variation in the two selection sweep regions is potentially underpinning recent adaptation of *C. parasitica* to the European environment.

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, **Figure 1B**), including the S12 lineage ($n = 197$). Within this cluster, $K = 43$ genetic subclusters were identified with fineSTRUCTURE (**Figure 4, Figure 4—figure supplement 1A**). The large number of identified clusters likely stems from the presence of multiple small groups of highly related individuals beyond the S12 lineage. Furthermore, populations only showed weak geographic structuring (**Figure 4—figure supplement 2**). The highest level of coancestry for any given S12 isolate was found within the S12 population (**Figure 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 and 49.7 (**Figure 4, Figure 4—figure supplement 2**). As Central European *C. parasitica* populations were likely established from North American sources alone (**Dutech et al., 2012**), the coancestry analysis points to a European origin of the invasive S12 lineage rather than a novel introduction from North America.

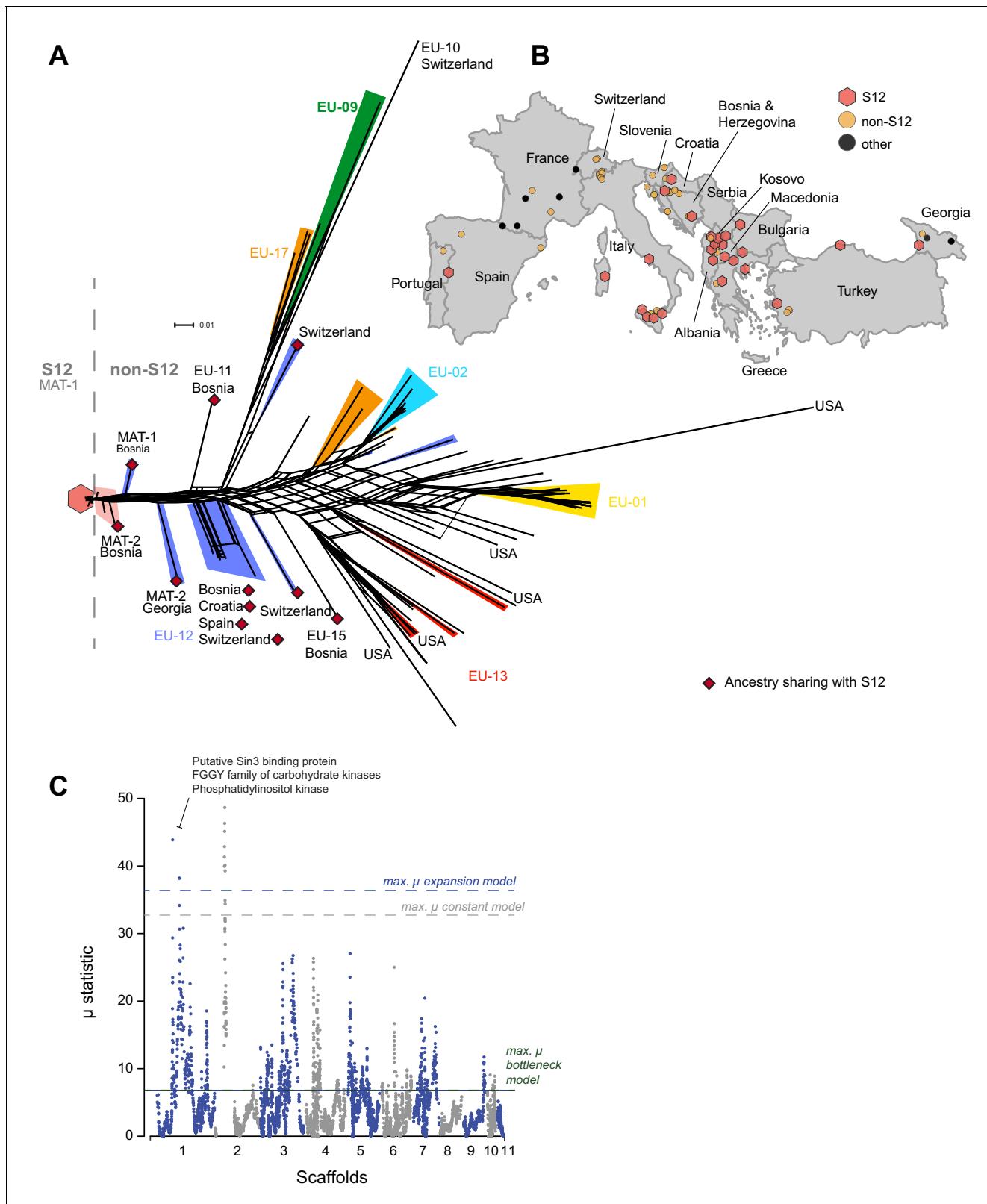


Figure 3. Phylogenetic network structure of the main European/North American subgroup and spatial distribution of S12 across Europe. (A) The highlighted branches represent the most abundant vegetative compatibility types. Isolates belonging to the S12 outbreak lineage (EU-12; mating type MAT-1, $n = 105$) 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 are marked with red diamonds. Figure 3 continued on next page

Figure 3 continued

lineage as inferred by fineSTRUCTURE (**Figure 4**, **Figure 4—figure supplement 2**) are marked with dark red diamonds. (B) Geographic map of European sampling locations. Light red hexagons mark locations where S12 was found, and yellow circles indicate locations containing other genotypes from the main European/North American CL1 cluster (**Figure 1B**). Black circles show the location of highly distinct genotypes outside of the CL1 cluster (**Figure 1A, B**). (C) Genome-wide scan for selective sweeps (RAiSD). Effects due to the unknown demographic history of the European/North American CL1 cluster (**Figure 1B**) were mitigated by implementing population simulations under constant, expansion, and bottleneck scenarios. The three dashed lines show the maximum values observed in the simulated datasets following constant (gray) expansion (blue) and bottleneck (green) demographic models. Encoded protein functions overlapping with the top regions are shown as summaries.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. SplitsTree of all non-S12 isolates showing full isolate identifiers.

Figure supplement 2. Per site population recombination rate ρ of European/North American *Cryphonectria parasitica* populations as inferred with LDhat.

TE landscape and CNV 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 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; **Figure 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 (**Figure 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 (**Figure 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 loci with detectable TE presence/absence polymorphism, we found an over twofold variation in total TE counts across all isolates (**Figure 5C**). The total TE count variation among the genetically diverse non-S12 isolates (North America and Europe only, **Figure 3A**) was larger than the clonal S12. Nevertheless, the count of TE insertions at polymorphic TE insertion sites ranged between 12 and 52 among S12 isolates, which was surprisingly high given their recent emergence and extremely high similarity across the genome (**Figure 5C**). Across TE insertion loci, we found only a marginally significant difference in the TE frequency among S12 compared to non-S12 isolates ($p=0.03$, **Figure 5—figure supplement 1**). This is in strong contrast to segregating SNPs. We found pairwise differences of 0–45 SNPs among S12 isolates and 0–3770 among non-S12 isolates (test statistic t-test mean difference $p<0.001$; **Figure 5—figure supplement 2**). For a direct comparison of TE and SNP diversity, we assessed the coefficient of variation (CV) of normalized TE and SNP counts. For TEs, we found a CV = 0.24 for S12 and CV = 0.33 for non-S12 isolates. For SNPs, we found a CV = 0.007 for S12 and CV = 0.69 for non-S12. Hence, our findings indicate a higher TE diversity among isolates compared to SNP diversity within the S12 lineage. We found that 44% and 49% of the polymorphic TE insertion sites are specific to the S12 and non-S12 group of isolates, respectively. Given the much more recent origin of the S12 lineage, this suggests increased TE activity since the emergence of the S12 lineage contributing to the creation of de novo genetic variation. For SNPs but not for TEs, the minor allele frequency spectrum of the S12 lineage was significantly skewed toward low-frequency variants compared to non-S12 isolates (Fisher’s exact test, $p<0.001$; **Figure 5—figure supplement 2**).

Repetitive sequences such as TEs can trigger non-homologous recombination leading to CNV including deletions. We used normalized read coverage to assess CNV across the species including the S12 lineage (**Figure 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 $p<0.001$, **Figure 5—figure supplement 1**). Genes tend to overlap duplications rather than deletions (odds ratios of 0.37 and 0.13, **Figure 5—figure supplement 2**).

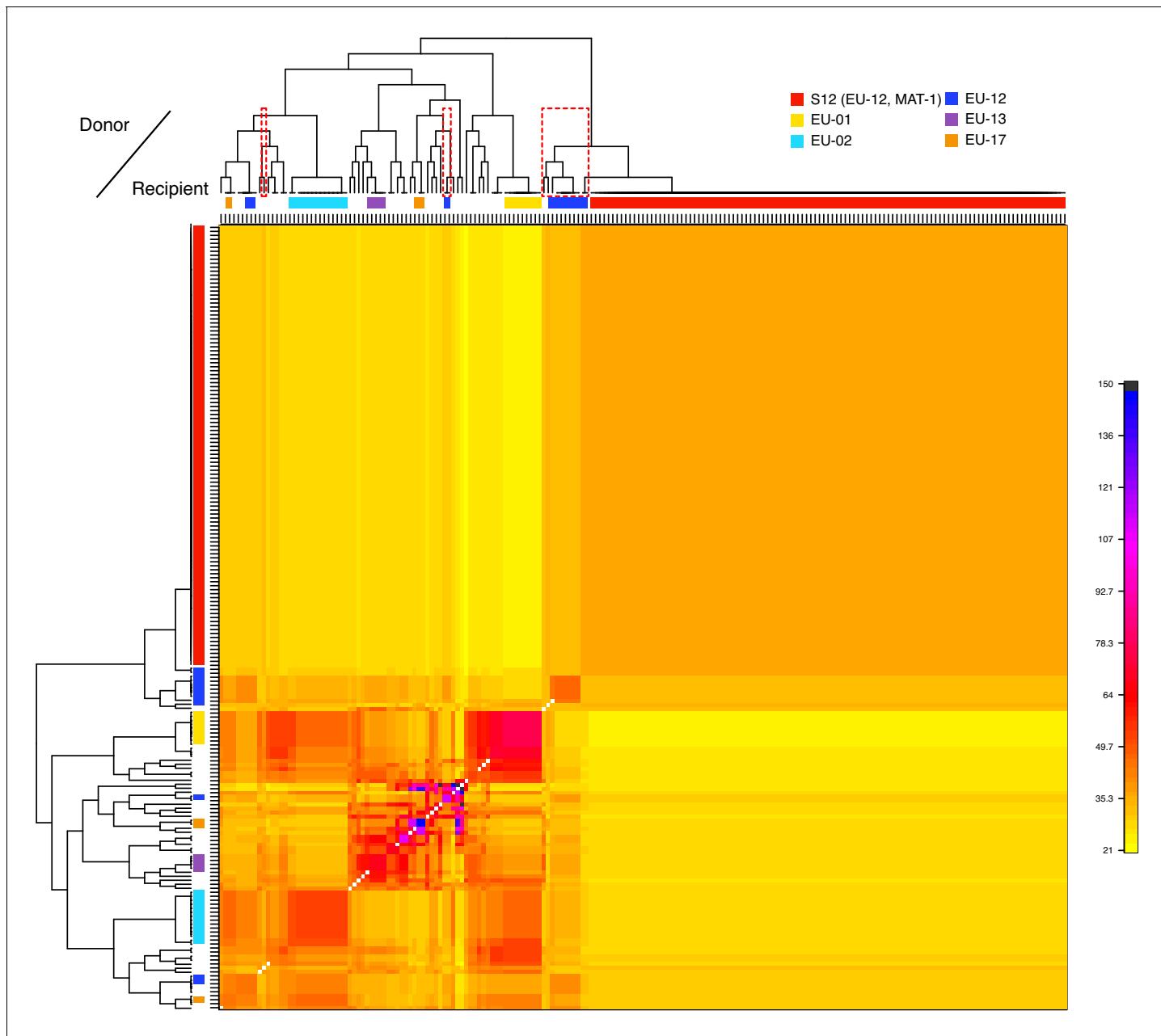


Figure 4. Analysis of donors to the S12 lineage genotypes. Averaged coancestry matrix and population tree of the North American/European subgroup estimated by fineSTRUCTURE. The heatmap indicates averaged coancestry between populations. Populations sharing ancestry with S12 are marked in red-dashed boxes. S12 ancestry-sharing populations are additionally marked in **Figure 3A**. Detailed coancestry matrix with extensive population information is shown in **Figure 4—figure supplement 2**.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. fineSTRUCTURE convergence plots.

Figure supplement 2. Full averaged coancestry matrix showing isolate identifiers.

respectively; two-sided Fisher test, $p\text{-value} < 0.001$). TEs tend to overlap deletions rather than duplications (odds ratios of 5.03 and 1.76, respectively; two-sided Fisher's test, $p\text{-value} < 0.001$) (**Figure 5D**). The mating type region on scaffold 2 and the rDNA locus on scaffold 6 show particularly high levels of CNV (**Figure 5A**). In a joint analysis of all isolates from the main European/North American cluster (CL1, **Figure 1B**), we found that coding sequences overlapping with duplications and deletions are enriched for gene ontology terms associated with protein binding functions and

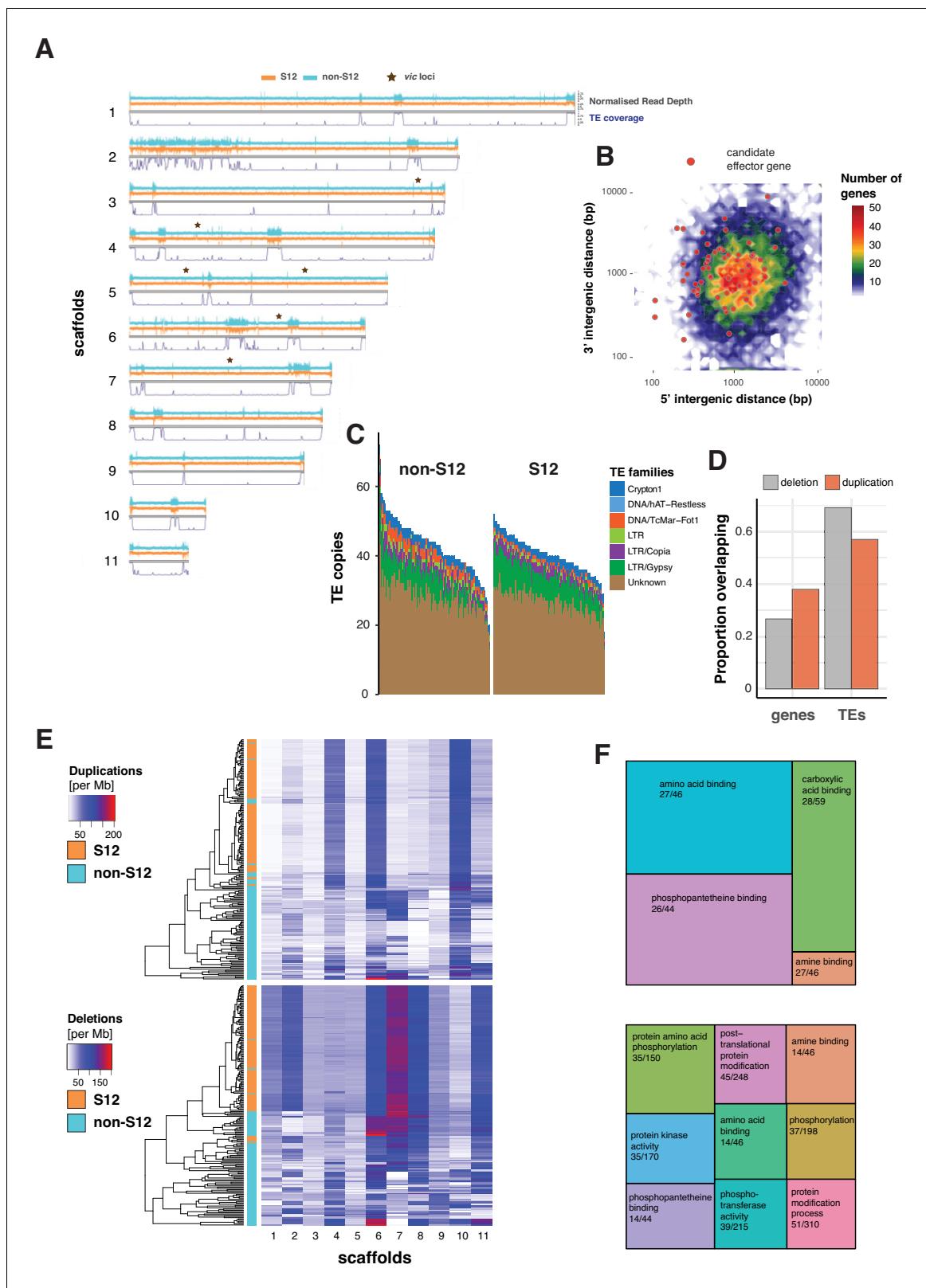


Figure 5. Transposable element (TE) landscape and copy number variation. (A) Genome-wide coverage of transposable elements (in 10 kb windows) matched by normalized read depth for the S12 and non-S12 lineages (North America and Europe only). vic loci: vegetative incompatibility loci. (B) Genome-wide distribution 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) Figure 5 continued on next page

Figure 5 continued

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.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Genetic diversity within the S12 lineage.

Figure supplement 2. Pairwise genome-wide nucleotide differences between S12 and non-S12 isolates.

Figure supplement 3. Normalized read depth distribution across *C. parasitica* isolates.

protein phosphorylation activity, respectively (**Figure 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.

The evolutionary history of 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; **Figure 1B**, **Figure 3A**), we first analyzed allele frequencies across the genome in both groups (**Figure 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 (**Figure 6B**). We analyzed the predicted impact on protein functions of segregating mutations in the S12 lineage and non-S12 populations (**Figure 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 (**Figure 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 $p=0.002028$; **Figure 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, suggesting that deleterious mutations in the S12 lineage can still be removed through low levels of recombination and purifying selection. The extremely low level of polymorphism segregating within the S12 lineage prevents strong inferences of selective sweeps in the lineage.

Retracing S12 invasion routes

To infer potential invasion routes of the S12 outbreak lineage ($n = 105$; **Figure 3A**), we investigated intra-lineage genetic diversity across southeastern Europe. 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 $p>0.05$; **Figure 7A**). The genetic structure assessed by a PCA suggests a differentiation into two to three groups with a weak geographic association (**Figure 7B**). The majority of isolates were found in a cluster along the principal component axis 2 (**Figure 7B**) without apparent geographic structure within the group. The second group consisted of eight isolates from Turkey and one isolate from Georgia consistent with the geographic proximity of the countries (**Figure 7B**). The most diverse group consisted exclusively of isolates from Greece showing reticulation in the SplitsTree (**Figure 7A**). Nucleotide diversity for the S12 outbreak lineage was $\pi = 2.90 \times 10^{-7}$ compared to $\pi = 2.62 \times 10^{-7}$ for the largest subgroup ($n = 89$) and $\pi = 1.80 \times 10^{-7}$ for the Turkish/Georgian subgroup ($n = 8$) excluding the diverse group of Greek isolates. The spread of the Greek isolates suggests that these might have recently undergone sexual reproduction and admixture

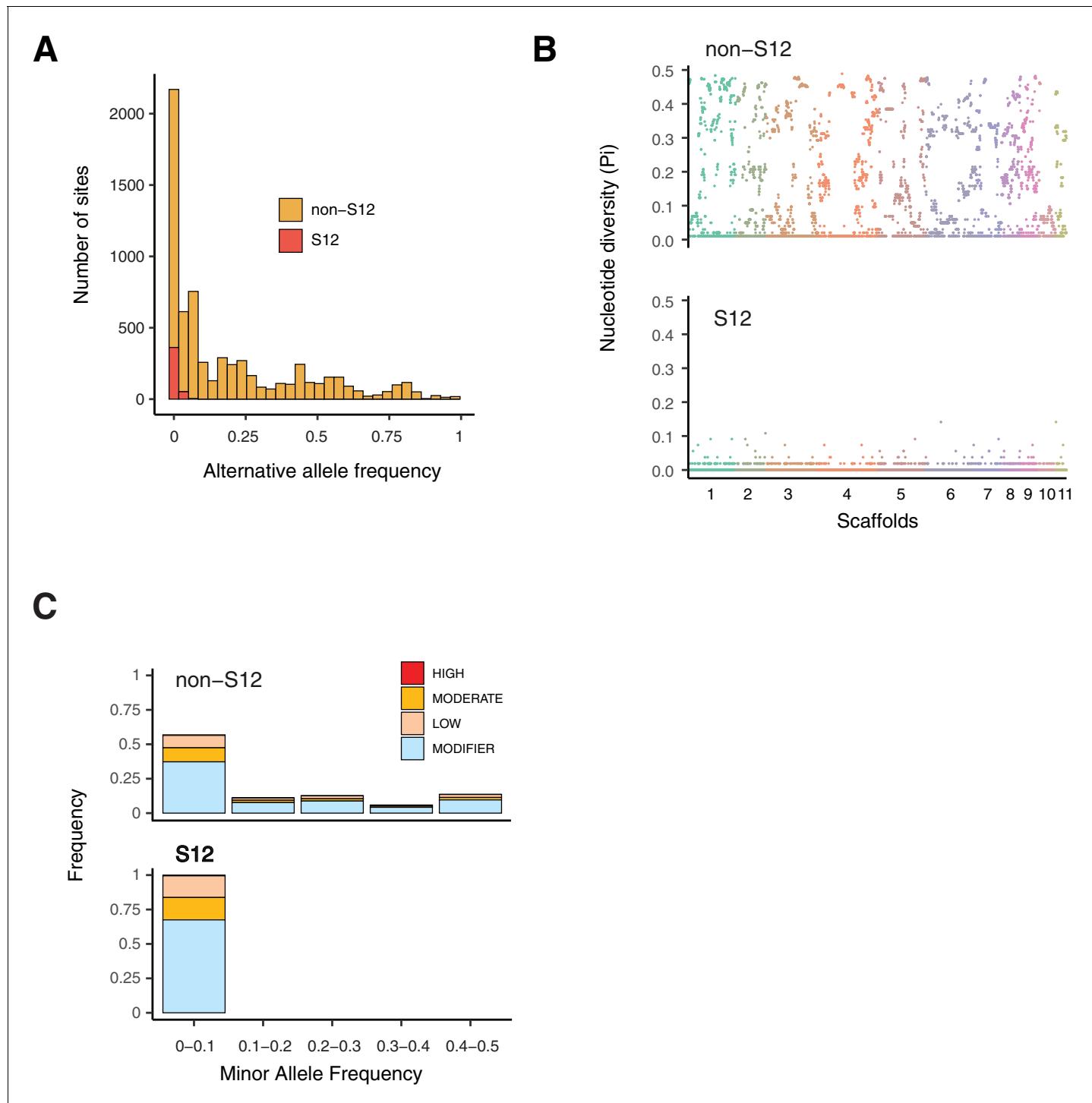


Figure 6. Polymorphism segregating in the S12 lineage. (A) Alternative allele frequencies spectra across the genome for the S12 lineage ($n = 105$) compared to all other analyzed European (non-S12; $n = 92$). (B) Genome-wide nucleotide diversity (π) for the S12 lineage and non-S12 lineages in 10 kb windows. (C) Minor allele frequency spectra of high, moderate, and modifier (i.e., near neutral) impact mutations as identified by SnpEff.

events. As we identified non-S12 genotypes in Greece, either within-lineage recombination or exchange with other lineages may underlie the reticulation and increased diversity.

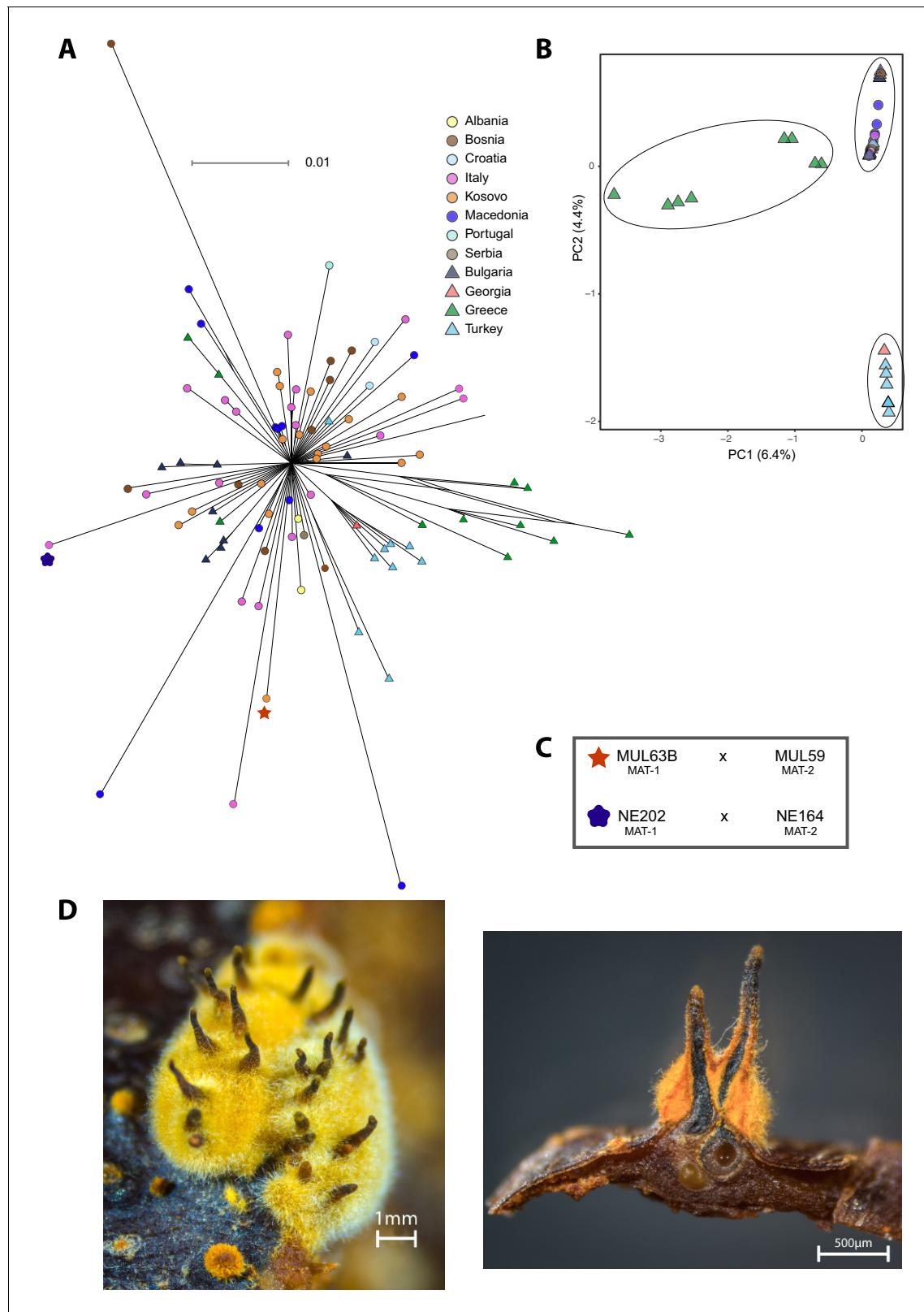


Figure 7. Fine-scale genetic diversity analyses of the S12 lineage. (A) SplitsTree and (B) principal component analysis (PCA) of the S12 mating type MAT-1 outbreak isolates ($n = 105$). Symbols and colors are as in (A). (C) Scheme of successful mating pairs of S12 mating type MAT-1 isolates crossed with 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 5 months of Figure 7 continued on next page

Figure 7 continued

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 peritheciun. When mature, the ascospores are actively ejected into the air through a small opening (ostiole) at the end of the perithecial neck.

Experimental tests of S12 mating competence

We tested experimentally whether S12 mating type MAT-1 isolates were still able to reproduce sexually. We selected S12 isolates from populations in Bosnia, Kosovo, and Italy matching locations where the opposite mating type was also detected (*Supplementary file 4*). After 5 months of incubation, we confirmed outcrossing of isolates of opposite mating type within the S12 lineage by pairing isolates from Molliq (Kosovo) and Nebrodi (southern Italy) (*Figure 7C*). Mating pairs from Molliq and Nebrodi grew numerous perithecia, which are the fruiting bodies specific to sexual reproduction (*Figure 7D*). Pairings of Bosnian isolates showed no perithecia formation. This could have been caused by undetected variation in isolate viability affecting mating competence, mutation accumulation, or environmental sensitivity. 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 southeastern Europe after a first establishment on the continent. The European 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 southeastern Europe was caused by a single, highly homogeneous lineage consisting nearly exclusively of a single mating type. Given the likely origin from the European gene pool with clear evidence for sexual reproduction, the invasive lineage likely transitioned from biparental mating populations to a single mating type. The spread across southeastern 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 southeastern 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 southeastern European *C. parasitica* populations likely originated from North American lineages. Clusters of European and North American genotypes were largely overlapping. Shared genotype clusters between the continents are consistent with historic gene flow from North America 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 southeastern 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 CNV, as well as at the level of vegetative compatibility types, suggests 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 the 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 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. Although we cannot exclude the existence of yet undescribed genetic diversity in North American populations, our results strongly indicate 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 toward 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 southeastern Europe where asexuality or selfing may be advantageous in the 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 coexists 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 southeastern 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 southeastern 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 the 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 toward 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 southeastern 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 southeastern 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 populations. 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 southeastern Europe where it 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, 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 isolates originated from Albania, Bosnia, Bulgaria, Croatia, Greece, Italy, Kosovo, Macedonia, Serbia, Slovenia, Switzerland, and Turkey (Figure 3B, Supplementary file 1). All 182 isolates originating from Central and southeastern 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 ($n = 5$) (*Supplementary file 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 southeastern Europe. We additionally selected 45 isolates from Bulgaria, Greece, Italy, and Macedonia, which were already included in a previous population-wide study on southeastern European *C. parasitica* diversity by **Milgroom et al., 2008**. All samples were collected between 1951 and 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 and Company, Franklin Lakes, USA) (**Hoegger et al., 2000**) and incubated for a minimum of 1 week at 24°C , at a 14 hr light and 10 hr 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 hr. 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), and DNA quality was assessed using the Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific). 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 of vegetative compatibility and mating types were scored with GeneMapper 5 (Thermo Fisher Scientific).

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, CA). The libraries were sequenced at the Functional Genomics Centre Zurich (ETH Zurich and University of Zurich) on Illumina platforms (Illumina). 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: $\text{QUAL} \geq 100$, $\text{MQRankSum} (\text{lower}) \geq -2.0$, $\text{QD} \geq 20.0$, $\text{MQRankSum} (\text{upper}) \leq 2.0$, $\text{MQ} \geq 20.0$, $\text{BaseQRankSum} (\text{lower}) \geq -2.0$, $\text{ReadPosRankSum} (\text{lower}) \geq -2.0$, $\text{ReadPosRankSum} (\text{upper}) \leq 2.0$, $\text{BaseQRankSum} (\text{upper}) \leq 2.0$ (**Figure 1—figure supplement 2**). 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 ≥ 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 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 of $p \leq 1 \times 10^{-10}$ to remove all mating type-associated SNPs for further analysis. The SNPs showing strong association with the mating type were located on scaffold 2 between positions 179,552 – 2,084,312 bp (**Figure 1—figure supplement 1**).

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 SplitsTree (i.e., from VCF to FASTA format) were done with PGDSpider v2.1.1.5 (**Lischer and Excoffier, 2012**). PCA was performed as implemented in the R package ade4 (**Bougeard and Dray, 2018**).

Inference of S12 donor populations

We generated an averaged coancestry matrix as inferred by fineSTRUCTURE v2.1.3 (**Lawson et al., 2012**). The software uses a 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 (**Figure 4—figure supplement 1**) and population assignments (i.e., pairwise coincidence matrix; **Figure 4—figure supplement 1**).

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 that 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**), and ggplot2. Furthermore, we performed a genome-wide association study 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 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 5 kb 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 implemented in dadi (sub-module dadi.Demographics1D). For all models, we ran the optimization 50 times with initial parameters drawn randomly from the bounded distributions (specified for each model below) and the optimization method optimiselog with the maximum number of iterations set to 100 (see **Supplementary file 2** for initial parameters). Models were compared with the AIC. The AIC penalizes models with a higher number of parameters in order to identify the model explaining the largest amount of variation using the fewest possible independent variables. The compared models are described below and include the following parameters: nuF representing the ratio of contemporary to ancient population size, nuB representing the ratio of intermediate population size to ancient population size (relevant for some models only), and T representing the time at which the instantaneous size change happened (in units of 2^*Na generations).

- Standard neutral (i.e., constant) model (snm function) without population size change.
- Sudden expansion model (two_epoch function) with nuF bounded between 1 and 1000 and T between 0 and 2000.
- Sudden contraction model (two_epoch function) with nuF bounded between 1 and 1000 and T between 0 and 2000.
- Model implementing a single instantaneous change in population size at a specific point in time followed by exponential growth (bottlegrowth function) with nuF bounded between 0.01 and 1000, nuB between 0.01 and 100, and T between 0 and 500.
- Model implementing a sudden population contraction followed by a sudden expansion (i.e., a bottleneck with recovery; three-epoch function) with nuF bounded between 1 and 2000, nuB between 0.01 and 1, TB (length of the bottleneck) between 0 and 100, and TF (time since the bottleneck) between 0 and 500.

Datasets and scripts for all demographic analyses are accessible here: <https://github.com/croll-lab/datasets> (**Laboratory of Evolutionary Genetics @ UNINE, 2021**; copy archived at <https://doi.org/10.5281/fa908efdb71d5ba90823c70bf2467e10ce6f12a7>).

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: constant, 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 ([Figure 3—figure supplement 2](#)). 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 constant population 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 TEs and effector prediction

We performed a de novo identification of TEs 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 TEs 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](#)). For effector prediction, we first identified putatively secreted proteins 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](#)). EffectorP comes pre-trained on gene sequences encoding functionally confirmed fungal effectors.

Genome compartmentalization and CNV 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 CNV, we computed the normalized read depth (NRD) 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 NRD 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 NRD across all isolates, we considered windows with NRD >1.6 to be duplicated and windows with NRD <0.4 as deleted considering the 5th and 95th percentiles of the normalized read coverage ([Figure 5—figure supplement 3](#)). 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 2000 resampling replicates to generate expectations for random distributions ([Figure 5—figure supplement 1](#)).

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) (**Supplementary file 4**). 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 and placed onto sterile Petri dishes (90 mm 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 hr photoperiod (2500 lux) for 14 days. After 2 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 hr 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 hr. 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 King-Fisher (Thermo Fisher Scientific). 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 Oggendorf 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 the 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.

Additional information

Funding

Funder	Grant reference number	Author
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung	170188	Simone Prospero
Fondation Pierre Mercier pour la science		Daniel Croll

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Lea Stauber, Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review and editing; Thomas Badet, Formal analysis, Investigation, Visualization, Writing - review and editing; Alice Feurtey, Conceptualization, Formal analysis, Writing - review and editing; Simone Prospero, Conceptualization, Supervision, Funding acquisition, Writing - review

and editing; Daniel Croll, Conceptualization, Supervision, Funding acquisition, Visualization, Writing - original draft, Writing - review and editing

Author ORCIDs

- Lea Stauber  <https://orcid.org/0000-0001-8367-6150>
Thomas Badet  <https://orcid.org/0000-0001-6130-441X>
Simone Prospero  <https://orcid.org/0000-0002-9129-8556>
Daniel Croll  <https://orcid.org/0000-0002-2072-380X>

Decision letter and Author response

- Decision letter <https://doi.org/10.7554/eLife.56279.sa1>
Author response <https://doi.org/10.7554/eLife.56279.sa2>

Additional files

Supplementary files

- Supplementary file 1. Information on all analyzed isolates including sampling location, vegetative compatibility type (EU-type), (SSR) simple sequence repeats genotyping outcome, collector information, collection years, and NCBI accessions.
- Supplementary file 2. Initial and best-fit parameters as well as likelihoods and Akaike information criterion (AIC) values as inferred by *dadi* for demographic analyses. Cluster names correspond to clusters as shown in **Figure 1A**.
- Supplementary file 3. List of the genes in regions with signatures of recent positive selection (based on RAiSD). Putative functions are shown by (PFAM) Protein families (database) and Gene Ontology (GO) annotations. Single nucleotide polymorphisms (SNPs) and the corresponding composite μ statistic are shown.
- Supplementary file 4. Summary of MAT-1 and MAT-2 isolate pairings used for testing sexual recombination in S12 populations. *Isolates from the S12 MAT-1 lineage.
- Transparent reporting form

Data availability

All raw sequencing data is available on the NCBI Short Read Archive (BioProjects PRJNA604575 and PRJNA644891).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Stauber L, Prospero S, Croll D	2020	Population sequencing of <i>Cryphonectria parasitica</i>	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA604575	NCBI BioProject, PRJNA604575
Stauber L, Prospero S, Croll D	2020	Population sequencing of <i>Cryphonectria</i> species	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA644891	NCBI BioProject, PRJNA644891

References

- Alachiotis N, Pavlidis P. 2018. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications Biology* 1:1–11. DOI: <https://doi.org/10.1038/s42003-018-0085-8>, PMID: 30271960
- Allendorf FW, Lundquist LL. 2003. Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* 17:24–30. DOI: <https://doi.org/10.1046/j.1523-1739.2003.02365.x>
- 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. *Ecology Letters* 18:188–199. DOI: <https://doi.org/10.1111/ele.12397>, PMID: 25529499

- Bergman CM. 2012. A proposal for the reference-based annotation of de novo transposable element insertions. *Mobile Genetic Elements* **2**:51–54. DOI: <https://doi.org/10.4161/mge.19479>, PMID: 22754753
- Bertelsmeier C, Ollier S, Liebhold AM, Brockerhoff EG, Ward D, Keller L. 2018. Recurrent bridgehead effects accelerate global alien ant spread. *PNAS* **115**:5486–5491. DOI: <https://doi.org/10.1073/pnas.1801990115>, PMID: 29735696
- Bertelsmeier C, Keller L. 2018. Bridgehead effects and role of adaptive evolution in invasive populations. *Trends in Ecology & Evolution* **33**:527–534. DOI: <https://doi.org/10.1016/j.tree.2018.04.014>, PMID: 29764688
- Billiard S, López-Villavicencio M, Hood ME, Giraud T. 2012. Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *Journal of Evolutionary Biology* **25**:1020–1038. DOI: <https://doi.org/10.1111/j.1420-9101.2012.02495.x>, PMID: 22515640
- Biraghi A. 1946. Cancro Della Cortecchia 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. *Molecular Ecology* **24**:2277–2297. DOI: <https://doi.org/10.1111/mec.13032>, PMID: 25474505
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**:2114–2120. DOI: <https://doi.org/10.1093/bioinformatics/btu170>, PMID: 24695404
- Bougeard S, Dray S. 2018. Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software* **86**:i01. DOI: <https://doi.org/10.1863/jss.v086.i01>
- 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. DOI: <https://doi.org/10.1093/bioinformatics/btm308>, PMID: 17586829
- Brasier CM, Webber JF. 2019. Is there evidence for post-epidemic attenuation in the dutch elm disease pathogen *Ophiostoma novo-ulmi*? *Plant Pathology* **68**:921–929. DOI: <https://doi.org/10.1111/ppa.13022>
- Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**:2665–2681. DOI: <https://doi.org/10.1534/genetics.105.048975>, PMID: 16489234
- Bryner SF, Rigling D. 2012. Hypovirus virulence and vegetative incompatibility in populations of the chestnut blight fungus. *Phytopathology* **102**:1161–1167. DOI: <https://doi.org/10.1094/PHYTO-01-12-0013-R>, PMID: 22857516
- 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 leL, 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* **6**:80–92. DOI: <https://doi.org/10.4161/fly.19695>, PMID: 22728672
- 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*. *European Journal of Plant Pathology* **155**:81–91. DOI: <https://doi.org/10.1007/s10658-019-01751-w>
- Cortesi P, Milgroom MG, Bischiach M. 1996. Distribution and diversity of vegetative compatibility types in subpopulations of *Cryphonectria parasitica* in Italy. *Mycological Research* **100**:1087–1093. DOI: [https://doi.org/10.1016/S0953-7562\(96\)80218-8](https://doi.org/10.1016/S0953-7562(96)80218-8)
- 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. PMID: 11560890
- Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, Hillman BI, Milgroom MG, Pangilinan J, Smith M, Salamov A, Schmutz J, Yadav JS, Grigoriev IV, Nuss DL. 2020. Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: A fundamental resource for an archetypical invasive plant pathogen. *Phytopathology* **110**:1180–1188. DOI: <https://doi.org/10.1094/PHYTO-12-19-0478-A>, PMID: 32207662
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. DOI: <https://doi.org/10.1093/bioinformatics/btr330>, PMID: 21653522
- 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 Genetics and Biology* **130**:122–133. DOI: <https://doi.org/10.1016/j.fgb.2019.06.002>, PMID: 31175938
- Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development* **35**:57–65. DOI: <https://doi.org/10.1016/j.gde.2015.09.001>, PMID: 26451981
- Drenth A, McTaggart AR, Wingfield BD. 2019. Fungal clones win the battle, but recombination wins the war. *IMA Fungus* **10**:18. DOI: <https://doi.org/10.1186/s43008-019-0020-8>, PMID: 32647622
- Dutech C, Fabreguettes O, Capdeville X, Robin C. 2010. Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity* **105**:220–228. DOI: <https://doi.org/10.1038/hdy.2009.164>, PMID: 19997121
- 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. *Molecular Ecology* **21**:3931–3946. DOI: <https://doi.org/10.1111/j.1365-294X.2012.05575.x>, PMID: 22548317
- 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 Ecology* **197**:155–172. DOI: <https://doi.org/10.1007/s11258-007-9352-3>

- Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**:737–756. DOI: <https://doi.org/10.1093/genetics/78.2.737>, PMID: 4448362
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AFA. 2019. RepeatModeler2: automated genomic discovery of transposable element families. *bioRxiv*. DOI: <https://doi.org/10.1101/856591>
- 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 Phytologist* **198**:946–956. DOI: <https://doi.org/10.1111/nph.12177>, PMID: 23406496
- Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology* **46**:34–42. DOI: <https://doi.org/10.1016/j.mib.2018.01.020>, PMID: 29455143
- Frantzeskakis L, Kusch S, Panstruga R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology* **20**:3–7. DOI: <https://doi.org/10.1111/mpp.12738>, PMID: 30557450
- Friesen TL, Faris JD, Solomon PS, Oliver RP. 2008. Host-specific toxins: effectors of necrotrophic pathogenicity. *Cellular Microbiology* **10**:1421–1428. DOI: <https://doi.org/10.1111/j.1462-5822.2008.01153.x>, PMID: 18384660
- Garbelotto M, Rocca GD, 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**:art180. DOI: <https://doi.org/10.1890/ES14-00426.1>
- Garnas RJ, 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. *Biological Invasions* **18**:935–952. DOI: <https://doi.org/10.1007/s10530-016-1082-9>
- 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. DOI: <https://doi.org/10.1371/journal.pone.0067944>, PMID: 23840791
- Gladieux P, Byrnes III EJ, Aguilera G, Fisher MC, Heitman J, Giraud T. 2011. Epidemiology and evolution of fungal pathogens in plants and animals. In: Tibayrenc M, Michel (Ed). *Genetics and Evolution of Infectious Disease*. Elsevier. p. 59–132. DOI: <https://doi.org/10.1016/B978-0-12-799942-5.00004-4>
- Gladieux P, Feurtey A, Hood ME, Snirc A, Clavel J, Dutech C, Roy M, Giraud T. 2015. The population biology of fungal invasions. *Molecular Ecology* **24**:1969–1986. DOI: <https://doi.org/10.1111/mec.13028>, PMID: 25469955
- González-Varela G, González AJ, Milgroom MG. 2011. Clonal population structure and introductions of the chestnut blight fungus, *Cryphonectria parasitica*, in Asturias, northern Spain. *European Journal of Plant Pathology* **131**:67–79. DOI: <https://doi.org/10.1007/s10658-011-9788-0>
- Griffin GJ. 1986. Chestnut blight and its control. *Horticultural Reviews* **8**:291–336. DOI: <https://doi.org/10.1002/9781118060810.ch8>
- Grünwald NJ, Garbelotto M, Goss EM, Heungens K, Prospero S. 2012. Emergence of the sudden oak death pathogen *Phytophthora ramorum*. *Trends in Microbiology* **20**:131–138. DOI: <https://doi.org/10.1016/j.tim.2011.12.006>, PMID: 22326131
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics* **5**:e1000695. DOI: <https://doi.org/10.1371/journal.pgen.1000695>, PMID: 19851460
- Hayes KR, Barry SC. 2008. Are there any consistent predictors of invasion success? *Biological Invasions* **10**:483–506. DOI: <https://doi.org/10.1007/s10530-007-9146-5>
- Heald FD, Studhalter RA. 1914. Birds as carriers of the chestnut blight fungus. *Journal of Agricultural Research* **2**:405–422. DOI: <https://doi.org/10.2307/4071651>
- Heiniger U, Rigling D. 1994. Biological control of chestnut blight in Europe. *Annual Review of Phytopathology* **32**:581–599. DOI: <https://doi.org/10.1146/annurev.py.32.090194.003053>
- Heitman J, Sun S, James TY. 2013. Evolution of fungal sexual reproduction. *Mycologia* **105**:1–27. DOI: <https://doi.org/10.3852/12-253>, PMID: 23099518
- 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. *Nature Ecology & Evolution* **4**:626–638. DOI: <https://doi.org/10.1038/s41559-020-1133-6>, PMID: 32123324
- Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. 2000. Genetic structure of newly established populations of *Cryphonectria parasitica*. *Mycological Research* **104**:1108–1116. DOI: <https://doi.org/10.1017/S0953756299002397>
- Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**:337–338. DOI: <https://doi.org/10.1093/bioinformatics/18.2.337>, PMID: 11847089
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**:254–267. DOI: <https://doi.org/10.1093/molbev/msj030>, PMID: 16221896
- Idnurm A, Hood ME, Johannesson H, Giraud T. 2015. Contrasted patterns in mating-type chromosomes in fungi: hotspots versus coldspots of recombination. *Fungal Biology Reviews* **29**:220–229. DOI: <https://doi.org/10.1016/j.fbr.2015.06.001>, PMID: 2688691
- Ježić M, Krstić L, Rigling D, Čurković-Perica M. 2012. High diversity in populations of the introduced plant pathogen, *Cryphonectria parasitica*, due to encounters between genetically divergent genotypes. *Molecular Ecology* **21**:87–99. DOI: <https://doi.org/10.1111/j.1365-294X.2011.05369.x>, PMID: 22092517
- 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: <https://doi.org/10.1093/bioinformatics/btu031>, PMID: 24451626
- Karadžić D**, Radulović Z, Sikora K, Stanivuković Z, Golubović Ćurguz V, Oszako T, Milenković I. 2019. Characterisation and pathogenicity of *Cryphonectria parasitica* on sweet chestnut and sessile oak trees in Serbia. *Plant Protection Science* **55**:191–201. DOI: <https://doi.org/10.17221/38/2018-PPS>
- 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**:e00326-20. DOI: <https://doi.org/10.1128/mBio.00326-20>, PMID: 3257669
- Knaus BJ**, Grünwald NJ. 2017. Vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17**:44–53. DOI: <https://doi.org/10.1111/1755-0998.12549>, PMID: 27401132
- Krasnov BR**, Shenbrot GI, Khokhlova IS, Stanko M, Morand S, Mouillet D. 2015. Assembly rules of ectoparasite communities across scales: combining patterns of abiotic factors, host composition, geographic space, phylogeny and traits. *EcoGraphy* **38**:184–197. DOI: <https://doi.org/10.1111/ecog.00915>
- Laboratory of Evolutionary Genetics @ UNINE**. 2021. Datasets. Software Heritage. swh:1:rev:fa908efdb71d5ba90823c70bf2467e10ce6f12a7. <https://archive.softwareheritage.org/swh:1:dir:e3c098b0ddcd2bba3e201d3d13293e544c2d58232;origin=https://github.com/crolllab/datasets;visit=swh:1:snp:7e5189913de8400eed7cc2e7b804167ecfb2a;anchor=swh:1:rev:fa908efdb71d5ba90823c70bf2467e10ce6f12a7>
- Langmead B**, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**:357–359. DOI: <https://doi.org/10.1038/nmeth.1923>, PMID: 22388286
- Lawson DJ**, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLOS Genetics* **8**:e1002453. DOI: <https://doi.org/10.1371/journal.pgen.1002453>, PMID: 22291602
- 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: <https://doi.org/10.1093/bioinformatics/btp352>, PMID: 19505943
- Li H**, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI: <https://doi.org/10.1093/bioinformatics/btp324>, PMID: 19451168
- Lischer HE**, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**:298–299. DOI: <https://doi.org/10.1093/bioinformatics/btr642>, PMID: 22110245
- Lombaert E**, Guillemaud T, Cornuet JM, Malusa T, Facon B, Estoup A. 2010. Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLOS ONE* **5**:e9743. DOI: <https://doi.org/10.1371/journal.pone.0009743>, PMID: 20305822
- Marbuah G**, Gren I-M, McKie B. 2014. Economics of harmful invasive species: a review. *Diversity* **6**:500–523. DOI: <https://doi.org/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 Pathology* **65**:577–586. DOI: <https://doi.org/10.1111/ppa.12441>
- Marra RE**, Cortesi P, Bissegger M, Milgroom MG. 2004. Mixed mating in natural populations of the chestnut blight fungus, *Cryphonectria parasitica*. *Heredity* **93**:189–195. DOI: <https://doi.org/10.1038/sj.hdy.6800492>, PMID: 15241462
- Marra RE**, Milgroom MG. 2001. The mating system of the fungus *Cryphonectria parasitica*: selfing and self-incompatibility. *Heredity* **86**:134–143. DOI: <https://doi.org/10.1046/j.1365-2540.2001.00784.x>, PMID: 11380658
- McKenna A**, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernitsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**:1297–1303. DOI: <https://doi.org/10.1101/gr.107524.110>, PMID: 20644199
- McVean G**, Auton A. 2007. LDhat 2.1: A Package for the Population Genetic Analysis of Recombination. 2.1. Dep Stat. <http://www.stats.ox.ac.uk/~mcvean/LDhat/manual.pdf>
- Mehrotra RS**. 2013. *Fundamentals of Plant Pathology*. Tata McGraw-Hill Education.
- 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. DOI: <https://doi.org/10.1080/00275514.1996.12026642>
- 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. *Molecular Ecology* **17**:4446–4458. DOI: <https://doi.org/10.1111/j.1365-294X.2008.03927.x>, PMID: 18803594
- Milgroom MG**, Cortesi P. 1999. Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *PNAS* **96**:10518–10523. DOI: <https://doi.org/10.1073/pnas.96.18.10518>, PMID: 10468641
- Myteberi IF**, Lushaj Aryen B, Keča N, Lushaj Arnisa B, Lushaj BM. 2013. Diversity of *Cryphonectria parasitica*, hypovirulence, and possibilities for biocontrol of chestnut canker in Albania. *International Journal of Microbiology Research and Reviews* **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: <https://doi.org/10.1093/bioinformatics/btw044>, PMID: 26826718
- Oggenfuss U**, Badet T, Wicker T, Hartmann FE, Singh NK, Abraham LN, Karisto P, Vonlanthen T, Mundt CC, McDonald BA. 2020. A population-level invasion by transposable elements in a fungal pathogen. *bioRxiv*. DOI: <https://doi.org/10.1101/2020.02.11.944652>

- 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*. *Molecular Ecology* **15**:249–262. DOI: <https://doi.org/10.1111/j.1365-294X.2005.02728.x>, PMID: 16367844
- Philibert A, Desprez-Loustau M-L, Fabre B, Frey P, Halkett F, Husson C, Lung-Escarmant B, Marçais B, Robin C, Vacher C, Makowski D. 2011. Predicting invasion success of forest pathogenic fungi from species traits. *Journal of Applied Ecology* **48**:1381–1390. DOI: <https://doi.org/10.1111/j.1365-2664.2011.02039.x>
- 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. DOI: <https://doi.org/10.1094/PHYTO-96-1337>, PMID: 18943666
- Prospero S, Cleary M. 2017. Effects of host variability on the spread of invasive forest diseases. *Forests* **8**:80. DOI: <https://doi.org/10.3390/f8030080>
- Prospero S, Rigling D. 2012. Invasion genetics of the chestnut blight fungus *Cryphonectria parasitica* in Switzerland. *Phytopathology* **102**:73–82. DOI: <https://doi.org/10.1094/PHYTO-02-11-0055>, PMID: 21848397
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: <https://doi.org/10.1093/bioinformatics/btq033>, PMID: 2110278
- Raboin LM, Selvi A, Oliveira KM, Paulet F, Calatayud C, Zapater MF, Brottier P, Luzaran R, Garsmeur O, Carlier J, D'Hont A. 2007. Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*. *Fungal Genetics and Biology* **44**:64–76. DOI: <https://doi.org/10.1016/j.fgb.2006.07.004>, PMID: 16979360
- Rigling D, Prospero S. 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: invasion history, population biology and disease control. *Molecular Plant Pathology* **19**:7–20. DOI: <https://doi.org/10.1111/mpp.12542>
- Robin C, Heiniger U. 2001. Chestnut blight in Europe: diversity of *Cryphonectria parasitica*, hypovirulence and biocontrol. *Forest Snow and Landscape Research* **76**:361–367.
- Rossman AMY. 2001. A special issue on global movement of invasive plants and fungi. *BioScience* **51**:93–94. DOI: [https://doi.org/10.1641/0006-3568\(2001\)051\[0093:ASIOGM\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0093:ASIOGM]2.0.CO;2)
- RStudio Team. 2015. RStudio: Integrated Development Environment for R. RStudio. 3.2.1. <http://www.rstudio.org/>
- Santini A, Ghelardini L, De Pace C, Desprez-Loustau ML, Capretti P, Chandelier A, Cech T, Chira D, Diamandis S, Gaitniekis T, Hantula J, Holdenrieder O, Jankovsky L, Jung T, Jurc D, Kirisits T, Kunca A, Lygis V, Malecka M, Marcais B, et al. 2013. Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytologist* **197**:238–250. DOI: <https://doi.org/10.1111/j.1469-8137.2012.04364.x>, PMID: 23057437
- Saunders DG, Win J, Kamoun S, Raffaele S. 2014. Two-dimensional data binning for the analysis of genome architecture in filamentous plant pathogens and other eukaryotes. *Methods in Molecular Biology* **1127**:29–51. DOI: https://doi.org/10.1007/978-1-62703-986-4_3, PMID: 24643550
- Schoustra S, Rundle HD, Dali R, Kassen R. 2010. Fitness-associated sexual reproduction in a filamentous fungus. *Current Biology* **20**:1350–1355. DOI: <https://doi.org/10.1016/j.cub.2010.05.060>, PMID: 20598542
- Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open. RepeatMasker. 4.0. <http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>
- Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *PNAS* **90**:4384–4388. DOI: <https://doi.org/10.1073/pnas.90.10.4384>, PMID: 8506277
- Sotirovski K, Papazova-Anakieva I, Grunwald NJ, Milgroom MG. 2004. Low diversity of vegetative compatibility types and mating type of *Cryphonectria parasitica* in the southern Balkans. *Plant Pathology* **53**:325–333. DOI: <https://doi.org/10.1111/j.0032-0862.2004.01006.x>
- Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology* **19**:2094–2110. DOI: <https://doi.org/10.1111/mpp.12682>, PMID: 29569316
- 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. DOI: <https://doi.org/10.1128/mSphere.00737-20>, PMID: 33055257
- Steenwyk JL, Soghian 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. DOI: <https://doi.org/10.1186/s12864-016-3044-0>, PMID: 27590805
- Suehs CM, Affre L, Médail F. 2004. Invasion dynamics of two alien *Carpobrotus* (Aizoaceae) taxa on a Mediterranean Island: ii. reproductive strategies. *Heredity* **92**:550–556. DOI: <https://doi.org/10.1038/sj.hdy.6800454>, PMID: 15100707
- Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE. 2015. Clonal reproduction in fungi. *PNAS* **112**:8901–8908. DOI: <https://doi.org/10.1073/pnas.1503159112>, PMID: 26195774
- Torres DE, Oggenfuss U, Croll D, Seidl MF. 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews* **34**:136–143. DOI: <https://doi.org/10.1016/j.fbr.2020.07.001>
- 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. *Forest Snow and Landscape Research* **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. *Molecular Ecology* **26**:5421–5434. DOI: <https://doi.org/10.1111/mec.14293>, PMID: 28802079

- 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*. *Environmental Microbiology* **21**:4521–4536. DOI: <https://doi.org/10.1111/1462-2920.14742>, PMID: 31314941
- 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. DOI: <https://doi.org/10.1093/gigascience/gix115>
- Wickham H**. 2007. Reshaping data with the reshape package. *Journal of Statistical Software* **21**:i12. DOI: <https://doi.org/10.18637/jss.v021.i12>
- Wickham H**. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. DOI: <https://doi.org/10.1007/978-0-387-98141-3>
- Wickham H**, François R, Henry L, Müller K. 2018. dplyr: A Grammar of Data Manipulation. Github. 1.0.5. <https://github.com/tidyverse/dplyr>
- Wickham H**, Henry L. 2018. tidyr: Easily Tidy Data with “spread()” and “gather()” Functions. Tidyr. 1.1.2. <https://tidyverse.org/>
- 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. DOI: <https://doi.org/10.1002/9781444329988.ch8>
- Wuest CE**, Harrington TC, Fraedrich SW, Yun HY, Lu SS. 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 Disease* **101**:619–628. DOI: <https://doi.org/10.1094/PDIS-10-16-1517-RE>, PMID: 30677356
- Yang XM**, Sun JT, Xue XF, Li JB, Hong XY. 2012. Invasion genetics of the western flower Thrips in China: evidence for genetic bottleneck, hybridization and bridgehead effect. *PLOS ONE* **7**:e34567. DOI: <https://doi.org/10.1371/journal.pone.0034567>, PMID: 22509325