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

Lea Stauber[1,2], Thomas Badet[2], Simone Prospero[1*], Daniel Croll[2*]

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

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

* Corresponding authors: simone.prospero@wsl.ch, daniel.croll@unine.ch

Data availability:

All raw sequencing data is available on the NCBI Short Read Archive (BioProject PRJNA604575)

## 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 American chestnut and a recent outbreak threatens European chestnut trees. To unravel the mechanisms underpinning the invasion of south-eastern Europe, we sequenced 188 genomes of predominantly European strains. Genotypes outside of the invasion zone showed high levels of diversity with evidence for frequent and ongoing recombination. The invasive lineage emerged from the highly diverse European genotype pool rather than a secondary introduction from Asia. The expansion across south-eastern Europe was mostly clonal and is dominated by a single mating type suggesting a fitness advantage of asexual reproduction. Our findings show how an intermediary, highly diverse bridgehead population gave rise to an invasive, largely clonally expanding pathogen.

# Introduction

Over the past century, a multitude of invasive species have emerged as threat to forest and agricultural ecosystems worldwide (Wingfield et al., 2010; Santini et al., 2013). 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 spread via living plants and plant products (Rossman, 2001). To successfully colonize a new environment, fungal pathogens have to overcome several invasion barriers including effective dispersal abilities, changes in available hosts, competition with other fungi, and niche availability (Gladieux et al., 2015). This may be achieved with plastic phenotypic changes followed by rapid genetic adaptation (Garbelotto et al., 2015). However, because the number of initial founders is often low, invasive populations are frequently of low genetic diversity which reduces adaptive genetic variation (Allendorf & Lundquist, 2003; Yang et al., 2012). Yet, many fungal plant pathogen invasions were successful despite low genetic diversity within founding populations (Raboin et al., 2007; Fontaine et al., 2013; Wuest et al., 2017).

A major hypothesis 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 & Keller, 2018). An alternative explanation suggests that first introductions simply serve as repeated sources of inoculum for additional secondary invasions without creating adaptive genotypes *in situ* (Bertelsmeier & 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). 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). The low availability of mating partners can be a major cost for individuals at the invasion front. To avoid mating costs, invasive fungal pathogens often switch from sexual to predominantly asexual reproduction (Heitman et al., 2013; Suehs et al., 2004). However, the switch to asexual progeny reproduction limits genetic diversification and adaptive potential (Taylor et al., 2015; Drenth et al., 2019). Populations of some highly successful invaders, such as the ascomycete *Ophiostoma novo-ulmi* (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 A1 mating type was present followed by the secondary introduction of the A2 mating type (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 & Webber, 2019; Paoletti et al., 2006). Although switches in reproductive modes can be a key factor for invasion success (Philibert et al., 2011), mechanisms underlying such switches remain poorly understood (Billiard et al., 2012).

The ascomycete *Cryphonectria parasitica* (Murr.) Barr. is the causal agent of chestnut blight, a lethal bark disease of *Castanea* species (Rigling & 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 & Swank, 2008). In 1938, the fungus was first detected on the European chestnut (*C. sativa*) near Genoa (Italy) and is now established in all major chestnut-growing countries in

Europe (Rigling & Prospero, 2018). The damage to the European chestnut may have been reduced by the presence of the *Cryphonectria hypovirus 1* (CHV-1) which acts as a biological control agent of chestnut blight (Rigling & 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 & Rigling, 1994). Hyphal anastomoses are controlled by a vegetative compatibility system and the virus spreads most effectively between strains of the same vegetative compatibility type (Cortesi et al., 2001).

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

127 In this study, we sequenced complete genomes of a comprehensive collection of European *C. parasitica*
128 strains with a fine-scaled sampling throughout the S12 invasion zone. Based on high-confidence
129 genome-wide single nucleotide polymorphisms (SNPs), we identified the most likely origin and
130 recapitulated the invasion process across south-eastern Europe. We show that the invasive *C. parasitica*
131 lineage arose through an intermediary, highly diverse bridgehead population. During the expansion, the
132 lineage became dominated by a single mating type but retained the ability to reproduce sexually.
133
134
135 # Results
136
137 **Genome-wide polymorphism analyses and phylogenomic reconstruction**
138 We analyzed complete genomes of 188 *C. parasitica* isolates covering the European outbreak region of
139 the invasive S12 lineage, as well as reference isolates from South Korea ($n$ = 2) and the United States
140 ($n$ = 4). Isolates were sequenced at a mean depth of 8–27.5X to detect high-confidence genome-wide
141 SNPs. A region of 179'501 -2'084'312 bp on scaffold 2 was associated to the mating type locus based
142 on association mapping $p$-values (Supp. Fig. 2). This region encoded known mating type associated
143 genes (Idnurm et al., 2015) and was characterized by a high SNP density consistent with observations
144 in other fungi (Taylor et al., 2015). We removed SNPs within the mating type associated region to avoid
145 confounding genetic structure with mating type divergence. We retained 17'873 SNPs and constructed
146 a whole-genome maximum likelihood phylogenetic tree (Fig. 1A). The tree revealed two major clades
147 splitting two Asian and one Swiss isolate (vegetative compatibility type EU-65) from the European and
148 North American isolates. The large European/North American clade was subdivided into three clades.
149 The clades a and c showed mixed geographic origins including North American isolates and European
150 isolates of various vegetative compatibility (EU-)types (Fig. 1A, Supp. Fig. 3). In contrast, clade b
151 consisted of genetically highly similar isolates ($n$ = 113; Fig. 1A) of vegetative compatibility type EU-
152 12, which predominantly originated from the Balkans, Italy, Turkey and Georgia. Approximately 93%
153 ($n$ = 105) of all isolates in the EU-12 clade were MAT-1 (Figure 1A) and were found in Albania, Bosnia,
154 Bulgaria, Croatia, Georgia, Greece, Italy, Macedonia, Serbia and Turkey (Fig. 1B; Supp. Table 1).

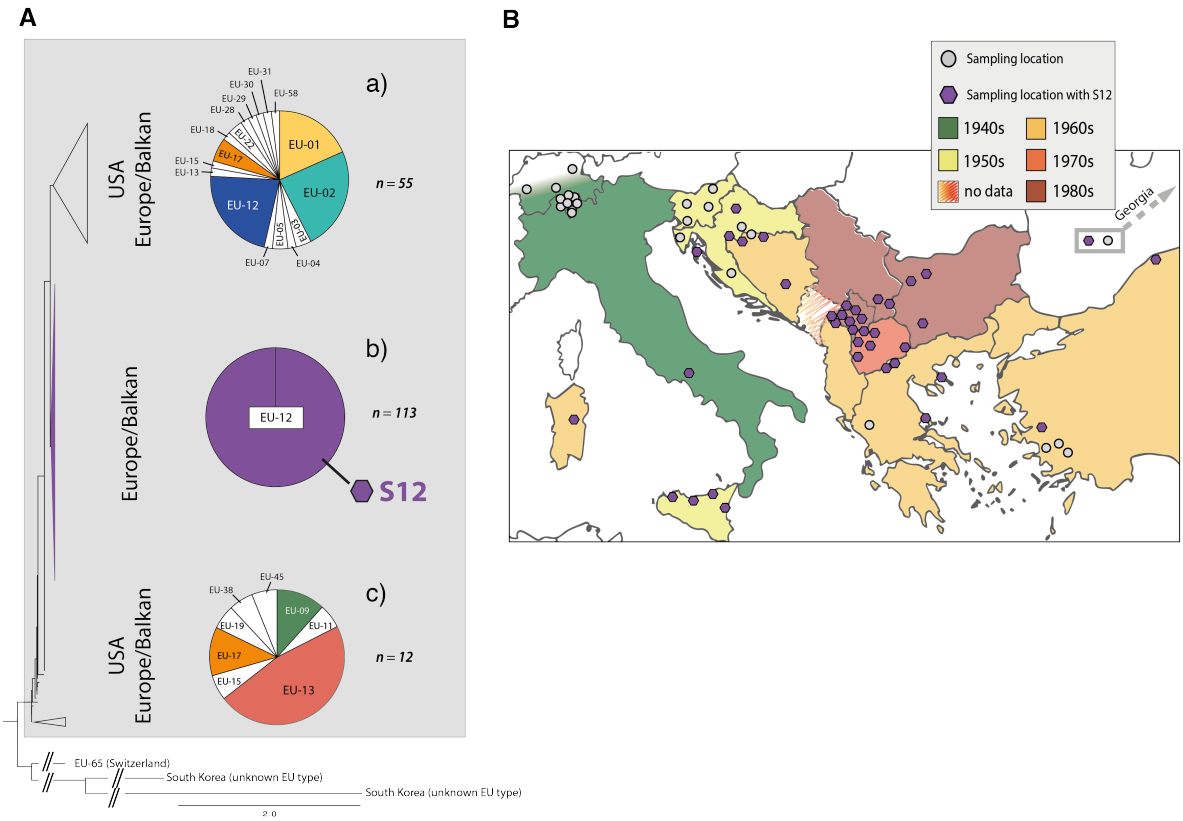
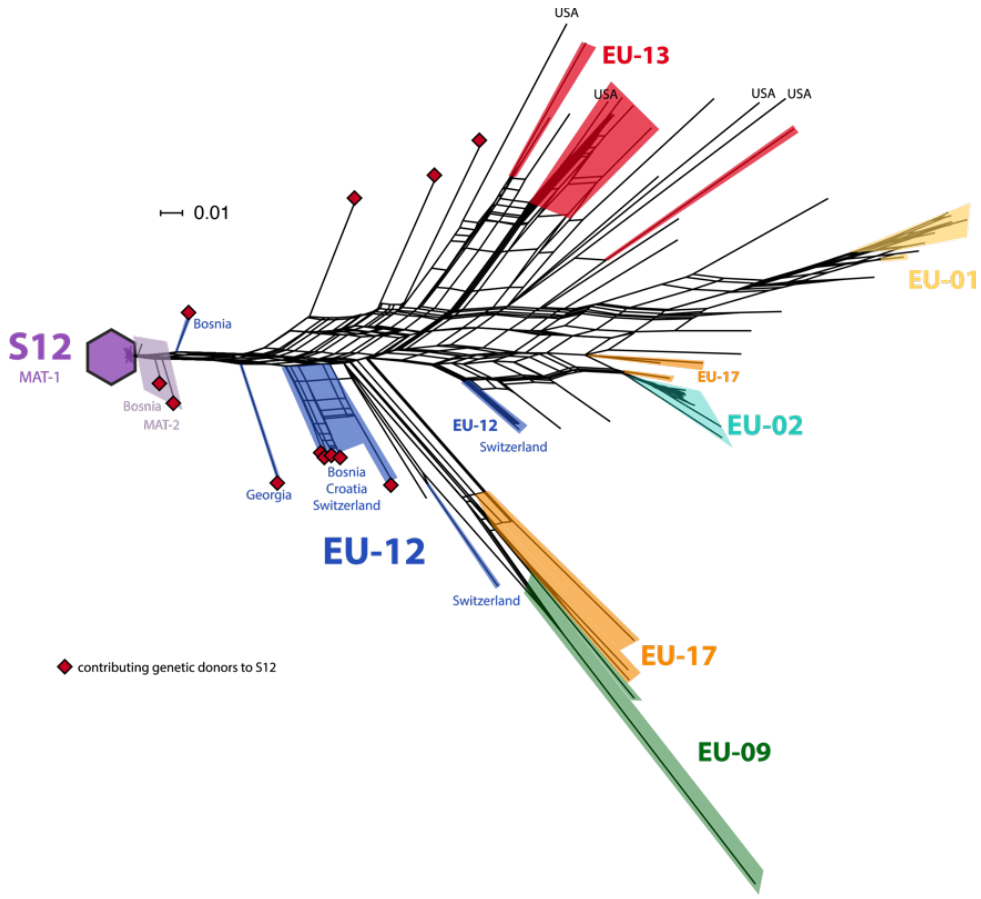



**Figure 1: Genome-wide analyses of *Cryphonectria parasitica* isolates focusing on the invasion in south-eastern Europe**. A) Whole-genome single nucleotide polymorphism (SNP) maximum likelihood tree constructed from 188 sequenced *C. parasitica* isolates. The US/European clade is shown with a grey box. The pie charts represent the proportions of vegetative compatibility types within each of the three clades. The most frequent vegetative compatibility types are highlighted in color (the color scheme matches with Figure 2). B) Map of European *C. parasitica* sampling locations. Grey circles show sampling locations of various vegetative compatibility types. Purple coloured hexagons represent sampling locations where EU-12 (mating type MAT-1) isolates were found (marked as "S12 lineage"). First observations of chestnut blight in the corresponding countries and regions are marked in a colour scheme according to decade. A tree with individual isolate labelling is shown in Supp. Fig. 3.

We performed a SplitsTree phylogenetic network analysis to account for reticulation caused by recombination. The network showed a high diversification, with both long branching and reticulation (Fig. 2, Supp. Fig. 4). The PHI-test revealed significant evidence for recombination ($p < 0.0001$). Despite the high level of genetic diversity, we found no evidence for geographic structure. Moreover, we found no clustering of isolates belonging to the same vegetative compatibility type with the exception of some EU-01, EU-02 and EU-12 (S12) genotypes from the Balkans. 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 ($n$ = 104). Consistent with analyses by (Milgroom et al., 2008), this group represents the invasive S12 lineage at the origin of the expansion of

177 *C. parasitica* across south-eastern Europe. Additionally, the phylogenetic network revealed closely
178 related but not identical S12 genotypes of mating type MAT-2 ($n$ = 7, Fig. 2). Hence, S12 outbreak
179 strains of MAT-2 connect the nearly uniform cluster of S12 MAT-1 strains with the remaining genetic
180 diversity of the major European subgroup of *C. parasitica*. The S12 cluster was furthermore connected
181 with the remaining genotypes of the major clade by two EU-12 isolates from Bosnia (M1808 with MAT-
182 1) and Georgia (MAK23 with MAT-2).




183

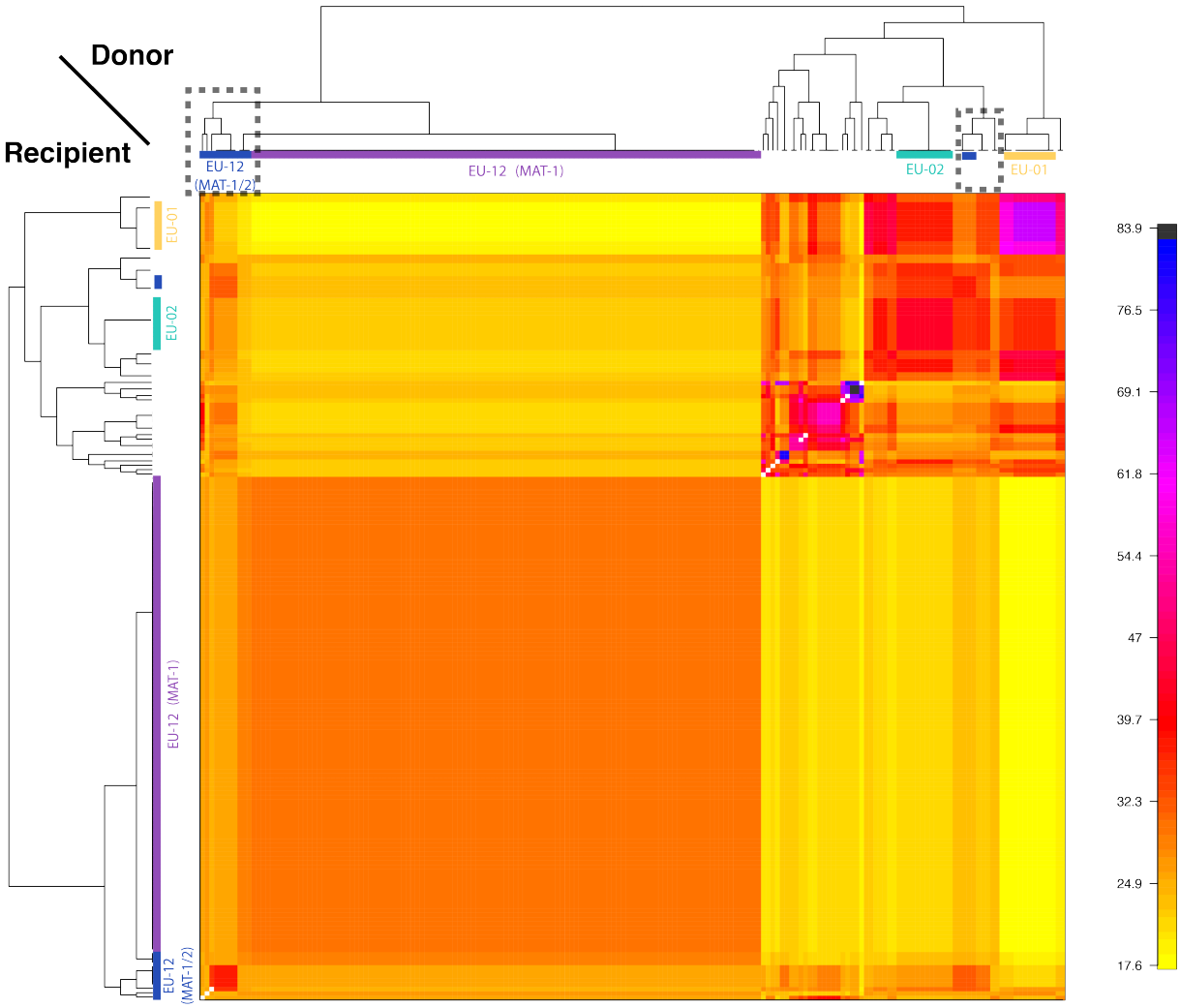
184 **Figure 2: Phylogenetic network structure of the US/European subgroup analyzed by SplitsTree**.
185 The highlighted branches represent the most abundant vegetative compatibility types (color scheme
186 matching Fig. 1A). Isolates belonging to the S12 outbreak lineage (EU-12; mating type MAT-1, $n$ =
187 104) are marked with a purple hexagon. S12 isolates of mating type MAT-2 are highlighted in light-
188 purple. Additional EU-12 isolates not belonging to the S12 lineage are highlighted in blue with
189 information on the country of origin. Genetic donors of the S12 lineage as inferred by fineSTRUCTURE
190 (Fig. 3, Table 1, Supp. Fig. 5) are marked with red squares.
191

192

193 **Identification of the most likely S12 founder populations**

194 The maximum likelihood phylogenetic tree and the SplitsTree network revealed that the invasive S12
195 lineage has closely related genotypes occurring in Europe. Thus, to dissect the genetic origin of S12, we

196 performed a co-ancestry matrix analysis using fineSTRUCTURE considering all isolates of the major
197 North American/European subgroup, including the S12 lineage (*n* = 185; Figure 1A). The averaged co-
198 ancestry matrix revealed no direct ancestors of the invasive S12 lineage among the major clade of
199 different vegetative compatibility types. However, we found an association with a coefficient of 24.9–
200 32.3 between the recipient S12 genotypes and European donors from different locations (Fig. 3, Table
201 1, Supp. Fig. 5). The donors mostly originated from the North Balkans (Bosnia and Croatia), as well as
202 Southern Switzerland, with the exception of isolate MAK23 from Georgia.
203



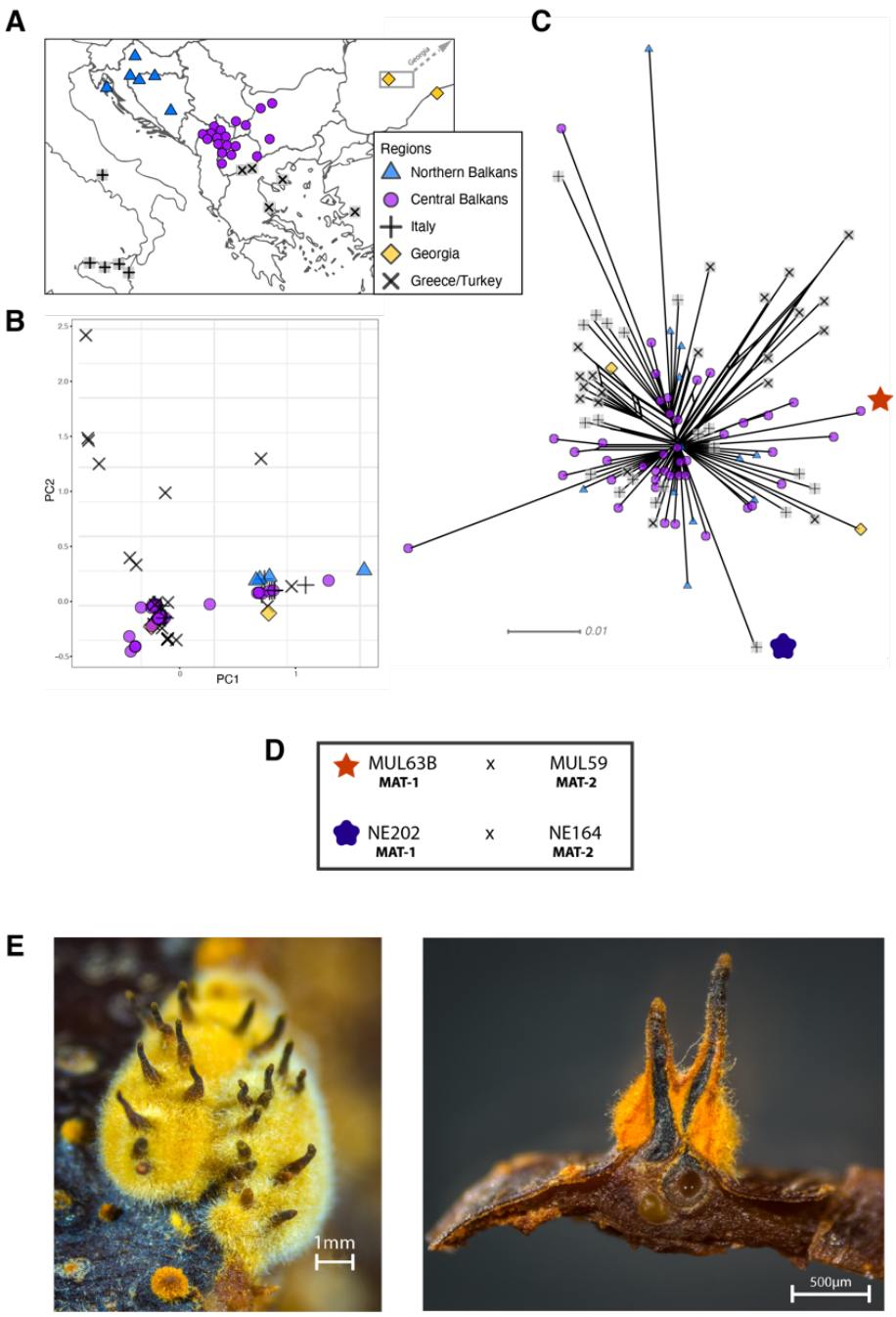

204

205 **Figure 3: Analysis of donors to the S12 lineage genotypes**. The averaged co-ancestry matrix and
206 phylogenetic tree of the North American/European subgroup estimated by fineSTRUCTURE. The
207 matrix shows recipient genotypes as rows and donors as columns. The heatmap indicates averaged
208 coancestry coefficients between donors and recipients (i.e. darker colors indicate stronger genetic
209 relationship between donors and recipients). Contributing S12 donors are marked in grey-dashed boxes.
210 A detailed list of contributing donors is shown in Table 1, Fig. 2 and Supp. Fig. 5.
211

212 **Retracing S12 invasion routes and retention of mating competence**

213 To infer potential invasion routes of the S12 outbreak lineage, we investigated intra-lineage genetic diversity
214 across south-eastern Europe. We focused only on S12 isolates of mating type MAT-1 to delimit the closest
215 genotypes contributing to the outbreak ($n$ = 104; Fig. 2). The closely related genotypes segregated 468 high-
216 confidence SNPs across the genome. The genetic structure assessed by a principal component analysis
217 showed loose clustering of genotypes across south-eastern Europe (Fig. 4A and B, Supp. Fig. 6). We assigned
218 genotypes to five regions: Italy, Northern Balkans, Central Balkans, Greece/Turkey and Georgia (Fig. 4A).
219 Italy, Northern and Central Balkans, as well as Georgia harbored mainly genotypes of two dominant clusters.
220 In contrast, the Greece/Turkey region contained genotypes of the two dominant clusters but also a broad
221 diversity of further genotypes. We analyzed evidence for reticulation in the phylogenetic relationships among
222 genotypes but found a star-like structure. We found minor evidence for reticulation among Central Balkans
223 genotypes (Fig. 4C). Consistent with the phylogenetic network pattern, we found significant evidence for
224 recombination within the S12 lineage (PHI test; $p$ = 0.0035). We tested experimentally whether S12 mating
225 type MAT-1 isolates were still able to reproduce sexually. We confirmed outcrossing of isolates of opposite
226 mating type within the S12 lineage by pairing isolates from Molliq (Kosovo) and Nebrodi (southern Italy)
227 (Fig. 4D). Mating pairs from Molliq and Nebrodi grew numerous perithecia, which are the fruiting bodies
228 specific to sexual reproduction (Fig. 4E). Pairings of Bosnian isolates showed no perithecia formation. Using
229 molecular mating type assays, we recovered both mating types among the ascospores produced from
230 successful matings.

231




232

233 **Figure 4: Fine-scale genetic diversity analyses of the S12 lineage**. A) Geographic map showing the
234 collected S12 isolates with mating type MAT-1 of *Cryphonectria parasitica* (*n*=104) according to five
235 geographic regions. B) Principal component analysis (PCA) and C) SplitsTree of the S12 mating type
236 MAT-1 outbreak isolates. Symbols and colors are as in A). D) Scheme of successful mating pairs of
237 S12 mating type MAT-1 isolates crossed with isolates from the opposite mating type, of the same
238 geographic origin. Symbols are as in C). E) Photographic images of sexual *C. parasitica* fruiting bodies
239 (*i.e.* perithecia) emerging from crosses of S12 mating type MAT-1 isolates with isolates of the opposite
240 mating type after five months of incubation under controlled conditions. Left: Perithecia embedded in a
241 yellow-orange stromatic tissue. Right: Cross-section of perithecia and chestnut bark. Flask-shaped
242 structures with a long cylindrical neck develop in yellow-orange stromatic tissue and are embedded in
243 the bark (except for the upper part). The ascospores are formed in sac-like structures (asci) in the basal
244 part of the perithecium. When mature, the ascospores are actively ejected into the air through a small
245 opening (ostiole) at the end of the perithecial neck.
246

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

248 Invasive pathogen lineages may have undergone crucial genomic rearrangements producing more fit
249 genotypes. Here, we generated a *de novo* identification and annotation of transposable elements (TEs)
250 for the *C. parasitica* genome. We found that 12% of the genome was composed of TEs with striking
251 variation along the assembled scaffolds (*i.e.* quasi-chromosomes; Fig. 5A). In particular, regions on
252 scaffold 2 matching the mating type locus are highly enriched in TEs suggesting that the large non-
253 recombining region has undergone substantial degeneration (Fig. 5A). In contrast, the vegetative
254 incompatibility (*vic*) loci are located in regions devoid of TEs. In fungal pathogens, effector genes and
255 TEs are often co-localized in fast-evolving compartments of the so-called "two-speed genome" (Dong
256 et al., 2015). However, the *C. parasitica* genome shows no apparent compartmentalization into gene-
257 sparse and gene-rich regions. We used machine learning to predict secreted proteins most likely acting
258 as effectors to manipulate the host. In contrast to other plant pathogens, effector gene candidates showed
259 no tendency to localize in gene-sparse regions of the genome (Fig-ure 5B). Non-repressed TEs can
260 potentially create additional copies in the genome leading to intra-species variability in TE content. To
261 detect such TE activity, we performed genome-wide scans of *C. parasitica* isolates for presence or
262 absence of TEs based on split read and target site duplication information. At the loci with detectable
263 TE presence/absence polymorphism, we found an over two-fold variation in total TE counts across the
264 genome (Fig. 5C). The total TE count variation among the genetically diverse non-S12 isolates (North
265 America and Europe only) was larger than the clonal S12 isolates. Nevertheless, the TE count variation
266 among the S12 was surprisingly high given their recent emergence and extremely high similarity across
267 the genome (Fig. 5C). This suggests that TE activity may have continued after the emergence of the S12
268 lineage and has created *de novo* genetic variation.

269

270 Repetitive sequences such as TEs can trigger non-homologous recombination leading to copy-number
271 variation including deletions. We used normalized read coverage to assess copy-number variation across
272 the species including the S12 lineage (Fig. 5A). Genes tend to overlap duplications rather than deletions
273 and TEs tend to overlap deletions rather than duplications (Fig. 5D). The mating type region on scaffold
274 2 and the rDNA locus on scaffold 6 show particularly high levels of copy-number variation. Scaffolds 4

275 and 6 were overall rich in duplications and scaffolds 6 and 7 were rich in deletions (Fig. 5E). On average
276 the S12 lineage bears more deletions per isolate compared to non-S12 lineages (~71.2 versus 64.6
277 respectively). In a joint analysis of all isolates (except from Asia), we found that coding sequences
278 overlapping with duplications and deletions are enriched for gene ontology terms associated with protein
279 binding functions and protein phosphorylation activity, respectively (Fig. 5F). Overall, our results show
280 that the S12 lineage underwent specific gene deletion and duplication patterns compared to the broad
281 diversity of non-S12 isolates.
282

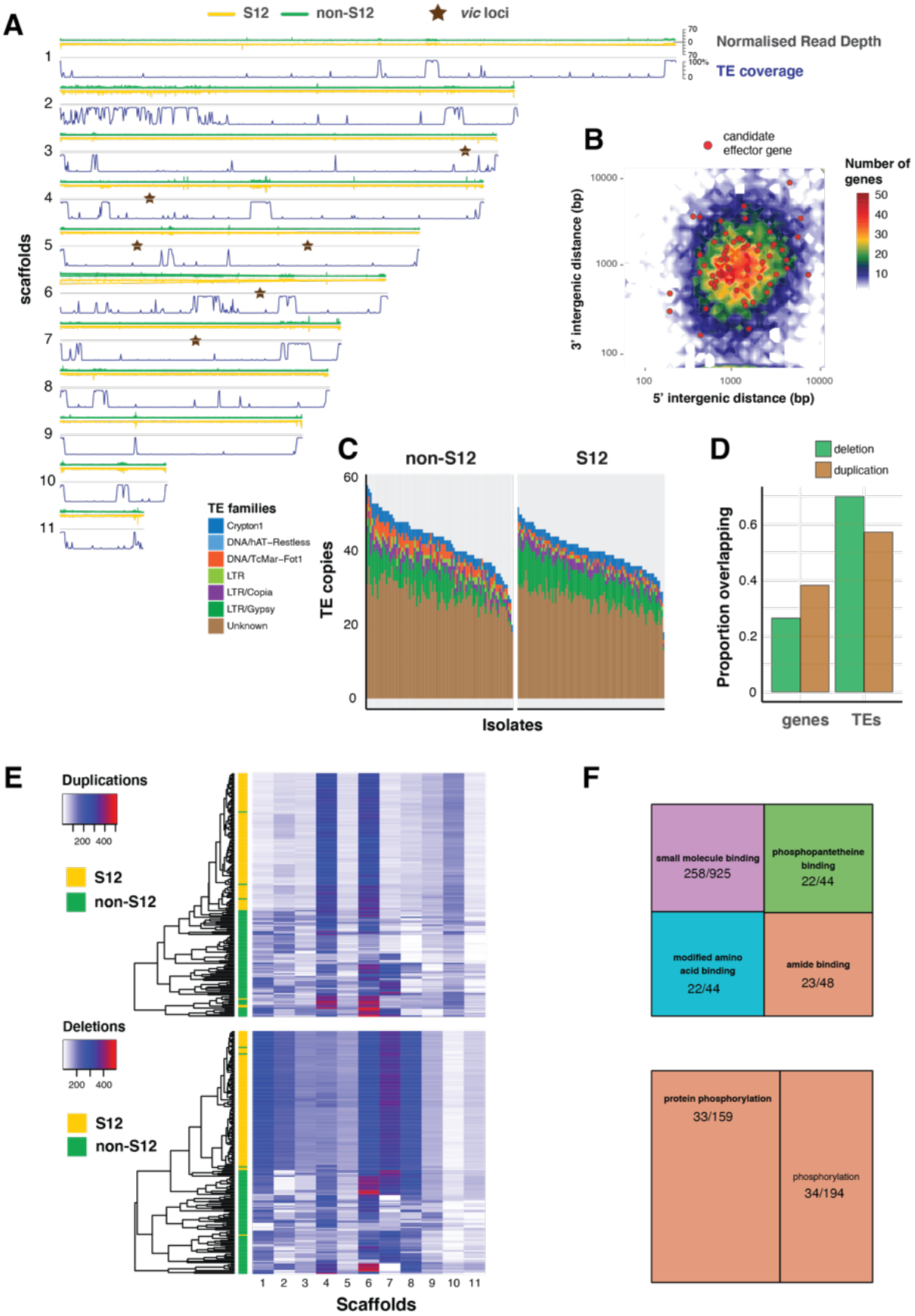



**Figure 5: Transposable element (TE) landscape and copy-number variation.** A) Genome-wide coverage of transposable elements (in 10kb windows) matched by with normalized read depth for the S12 and non-S12 lineages (North America and Europe only). *vic* loci: vegetative incompatibility loci.

287 B) Genome-wide distribution of intergenic distances according to the length of 5' and 3' flanking
288 regions. Red dots represent genes encoding predicted effector proteins. C) Counts of detected TE
289 sequences across S12 and non-S12 isolates using split reads and target site duplication information. D)
290 Proportion of normalized read depth windows (800 bp) with evidence for duplications (normalized read
291 depth > 1.6) or deletions (< 0.4) overlapping with genes and TEs. E) Heatmap showing the number of
292 windows (800 bp) with duplications and deletions. The dendrogram shows the similarity in duplication
293 or deletion profiles for S12 and non-S12 isolates (North America and Europe only). F) Molecular
294 functions (based on gene ontology) enriched in duplicated and deleted regions (upper and lower panel,
295 respectively). Significance of the enrichment was at an alpha = 0.05 Bonferroni threshold. The numbers
296 represent the number of genes with the matching gene ontology term in a duplicated or deleted region,
297 and across the genome, respectively.
298

299

300 **Polymorphism and allele frequency spectra within the outbreak lineage**

301 To gain insights into evolutionary forces shaping polymorphism in the outbreak S12 mating type MAT-
302 1 lineage versus non-S12 populations, we first analyzed allele frequencies across the genome in both
303 groups (Fig. 6A). The S12 lineage segregated virtually no intermediate allele frequencies in the range
304 of 0.05–0.95. In contrast, the non-S12 genotypes showed overall a wide spectrum of allele frequencies
305 across the genome. Genome-wide nucleotide diversity was extremely reduced in the S12 compared to
306 non-S12 populations (Fig. 6B). We analyzed the predicted impact on protein functions of segregating
307 mutations in the S12 lineage and non-S12 populations (Figure 6C). We found 4 highly and 94
308 moderately deleterious SNPs within S12 in contrast to 29 high and 773 moderately deleterious mutations
309 in non-S12 groups (Fig. 6C). Three of the high impact SNPs in the S12 lineage were classified as stop
310 gain mutations, as well as one splice acceptor variant (insertion variant). Two of these high impact
311 mutations affect proteins of the major facilitator superfamily, as well as a protein containing a LCCL
312 domain and an ecdysteroid kinase. Non-S12 populations showed an over-representation of low
313 frequency high-impact mutations (Fig. 6C). This is consistent with purifying selection reducing the
314 frequency of these mutations due to fitness costs. Within the S12 lineage nearly all segregating
315 mutations were at very low frequency. We found only modifier (*i.e.* nearly neutral) mutations rising to
316 higher frequency within the lineage. The extremely low level of polymorphism segregating within the
317 S12 lineage prevents strong inferences of selective sweeps since the origin of the lineage. Hence, we
318 focused on potential selective sweeps in the broader European and North American populations. To
319 avoid a bias by the deep sampling of the S12 lineage, we excluded all but one of the S12 MAT-1 isolates
320 (see Methods). We used RAiSD to produce a composite score of selective sweep signals and identified

321 three strong outlier loci. The first sweep locus is located at the boundary of the mating type locus on
322 scaffold 2 (Fig. 6D). The strongest sweep locus is on scaffold 3 and encompasses a ~471 kb locus at
323 positions 939-1'410 kb. The region contains 154 genes of which 107 encode conserved protein domains
324 (Fig. 6D, Supp. Table 2). Ten genes overlap with SNPs showing the strongest signature of selection ($\mu$
325 statistic >40) of which two genes encode MFS transporters. The third selective sweep region of ~190
326 kb region on scaffold 6 encompasses 60 genes (Fig. 6D, Suppl. Table S2). Three genes overlap with
327 SNPs of $\mu$ > 30 and encode for a RTA1-like protein transporter, an oxidoreductase and an anaphase
328 promoting protein, respectively (Fig. 6D, Suppl. Table S2).

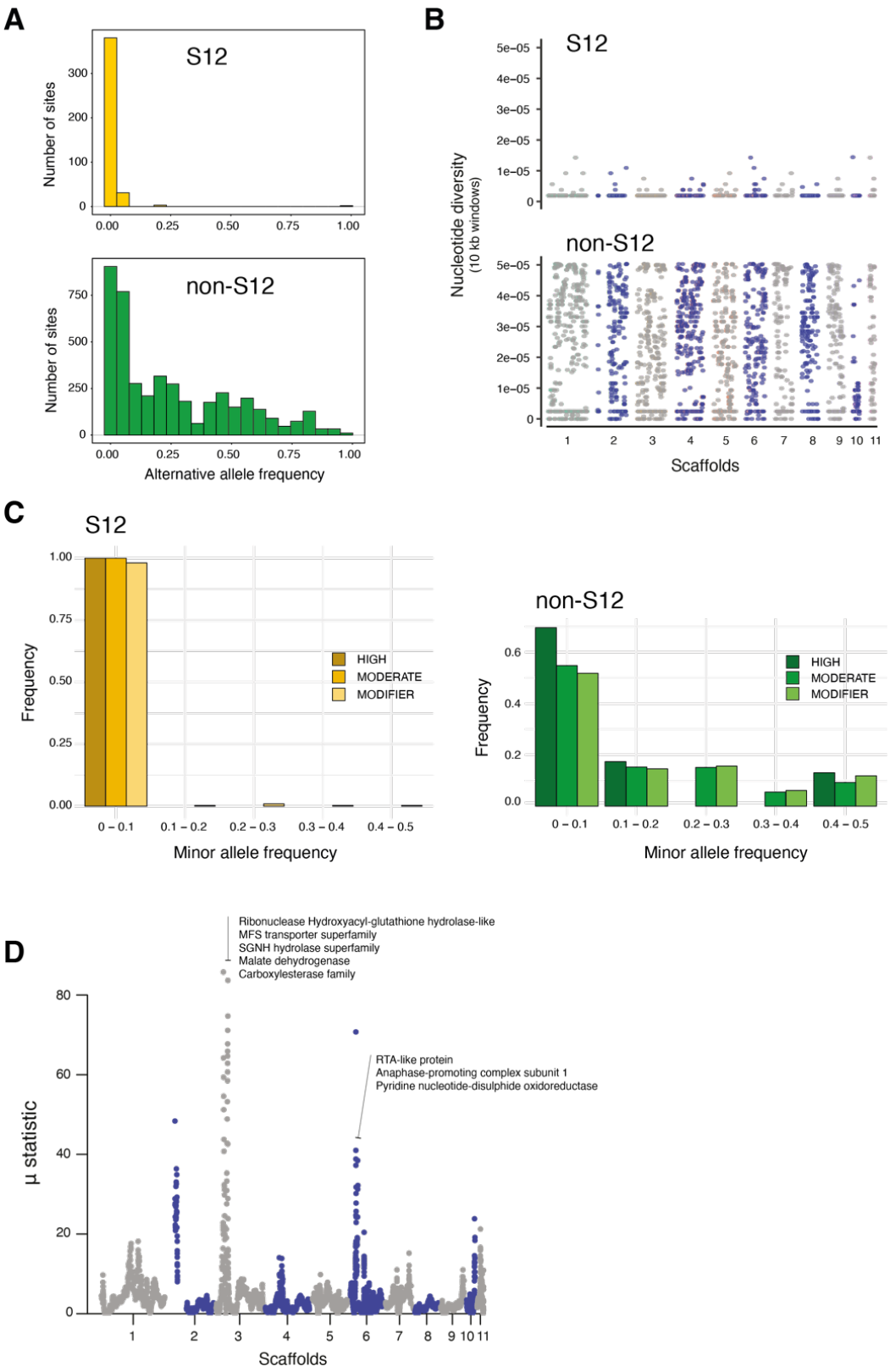



**Figure 6: Polymorphism segregating in the S12 lineage.** A) Alternative allele frequencies spectra across the genome for the S12 lineage ($n$ = 104) compared to all other analyzed European (non-S12; $n$ = 80). B) Genome-wide nucleotide diversity (pi) for the S12 lineage and for the non-S12 lineages in 10 kb windows. C) Minor allele frequency spectra of high, moderate and modifier (*i.e.* near neutral) impact mutations as identified by SnpEff. D) Genome-wide scan for selective sweeps (RAiSD). Encoded protein functions in the two top loci are shown as summaries.

337 # Discussion

338

339 We analyzed at the genome level how the chestnut blight pathogen *C. parasitica* successfully invaded
340 south-eastern Europe after a first establishment on the continent. Deep sampling of the invasive
341 genotypes showed that the invasion was caused by a single, highly homogeneous lineage consisting
342 nearly exclusively of a single mating type. The invasive lineage likely transitioned recently from
343 biparental mating populations to a single mating type. The spread across south-eastern Europe,
344 suspected to originate in northern Italy, left no clear geographic imprint. This lack of genetic
345 differentiation along the invasion route may be due to high levels of gene flow after the initial
346 colonization or a very rapid spread from the center of origin to multiple locations across south-eastern
347 Europe. Ongoing TE activity created unexpected levels of insertion polymorphism within the invasive
348 lineage. The lineage carries a distinct gene deletion and duplication profile compared to the European
349 and North American pool of *C. parasitica.* During the invasion, sexual reproduction was likely
350 sporadically and may have introgressed genetic material from outside of the invasive lineage.

351

352 **The establishment of a European bridgehead population**

353 Central and south-eastern European *C. parasitica* populations likely originated from North American
354 lineages. Clusters of European and North American genotypes were largely overlapping indicating
355 significant gene flow and, possibly, multiple introductions to Europe. These genome-scale analyses are
356 consistent with previous findings documenting multiple North American introductions into France, but
357 also directly from Asia (Milgroom et al., 1996; Dutech et al., 2012; Demené et al., 2019). Our results
358 suggest that central and south-eastern European *C. parasitica* populations were largely established from
359 North American sources alone. The high diversity at the genetic level with polymorphism at the level
360 of SNPs, TEs and copy-number variation as well as at the level of vegetative compatibility types suggest
361 that Europe was repeatedly colonized over the past century. Although genetic diversity could also have
362 accumulated *in situ* in Europe, the establishment of a large set of genotypes and vegetative compatibility
363 types seems difficult to explain with population age alone. Sexual recombination between the three most
364 common vegetative compatibility types in Europe (*i.e.* EU-01, EU-02 and EU-05; Robin & Heiniger

365 (2001)) could not account for the observed vegetative compatibility type diversity. Observational
366 records date the first introductions into Europe to the 1930s (Robin & Heiniger, 2001). Hence, based on
367 the observed genetic diversity, a scenario of repeated introductions of different vegetative compatibility
368 types since 1930s seems most plausible.
369
370 Populations from Southern Switzerland, Slovenia Croatia and Bosnia are highly diverse, have nearly
371 balanced mating type ratios, and a lack clustering according to vegetative compatibility types. This
372 strongly suggests frequent outcrossing and population admixture, consistent with reports of perithecia
373 in the field (Ježić et al., 2012; Prospero et al., 2006; Prospero & Rigling, 2012; Trestic et al., 2001).
374 Low vegetative compatibility type diversity in most European *C. parasitica* populations was thought to
375 have contributed to low population admixture within Europe compared to Asia and North America
376 (Cortesi et al., 1996; Dutech et al., 2012; Prospero & Rigling, 2012). However, our genome-wide
377 analyses revealed frequent and ongoing *in situ* admixture in Europe. Thus, vegetative compatibility type
378 diversity does not necessarily underpin population admixture frequency and genetic diversity in sexually
379 recombining populations. Our findings show that in asexually reproducing populations, such as in the
380 S12 lineage, genotypes tend to cluster according to vegetative compatibility types.
381
382 **Emergence of an invasive lineage from a European bridgehead**
383 The invasive lineage S12 most likely arose from existing genotypes established in Europe. We identified
384 a series of closely related genotypes to the S12 lineage in Bosnia, Croatia, Georgia and southern
385 Switzerland. Strikingly, the closest genotypes to the dominant S12 MAT-1 were S12 MAT-2 isolates
386 found in Bosnia, Kosovo and southern Italy. Analyses based on a coancestry matrix identified a group
387 of more distantly related genotypes from Bosnia, Croatia, Switzerland and Georgia having made the
388 strongest genetic contributions to the S12 lineage. This shows that introductions from outside of Europe
389 are unlikely to explain the emergence of S12. The lineage carries a unique set of copy-number variants
390 compared to other European genotypes underlining the observation of a recombinant S12 genotype.
391 Furthermore, the emergence of S12 was accompanied by a striking evolutionary transition from mixed
392 mating type populations to single mating type outbreak populations. Human activity may have

393 contributed to the shift towards single mating type populations. Shipments of infected chestnut seedlings
394 from Northern Italy and other trading activities could have disseminated the invasive lineage further
395 South. This would have exposed the pathogen to the geographically more fragmented chestnut forests
396 typically found in south-eastern Europe where asexuality or selfing may be advantageous. Although *C.*
397 *parasitica* is able to produce asexual conidia in large quantities, these specific spores are thought to be
398 mainly splash dispersed by rain over short distances (Griffin, 1986). Accounting for occasional dispersal
399 by birds or insects (Heald & Studhalter, 1914), conidia dispersal is unlikely to contribute substantially
400 to the colonization of new areas.
401
402 Despite the loss of a mating type in the S12 lineage, we found genome-wide evidence for reticulation
403 indicating at least low levels of recombination. If mating in *C. parasitica* follows the canonical process
404 found in many ascomycetes, isolates of opposite mating type are required. Hence, S12 isolates of mating
405 type MAT-1 may sporadically mate with rare S12 isolates of mating type MAT-2, which are
406 comparatively more diverse. The emergence of the opposite mating type at low frequency could be the
407 result of recombination with other genotypes and subsequent backcrossing. Combined with
408 experimental evidence, we show that the dominant S12 mating type MAT-1 has retained the ability for
409 sexual reproduction. Furthermore, in Bosnia, Croatia, Italy (Sicily) and Turkey the S12 lineage co-exists
410 with other genotypes (*i.e.* vegetative compatibility types EU-01 and EU-02) of both mating types,
411 potentially enabling sexual recombination and diversification *in situ*. The invasive S12 lineage was
412 likely pre-adapted to the south-eastern European niche as we traced the origins to a likely Italian
413 bridgehead population. Niche availability and benefits associated with asexual reproduction to colonize
414 new areas may have pre-disposed the European *C. parasitica* bridgehead population to produce a highly
415 invasive lineage.
416
417 **Expansion and mutation accumulation within the invasive S12 lineage**
418 The MAT-1 S12 lineage diversified largely through mutation accumulation as nearly all high-
419 confidence SNPs were identified as singletons. Mutation accumulation in absence of substantial
420 recombination resulted in star-like phylogenetic relationships. We found a surprisingly high degree of

421 differentiation among S12 genomes at the level of TE insertion polymorphism. Hence, active
422 transposition of TEs is an important factor in diversifying the invasive lineage and possibly underpin
423 future adaptive evolution. Analyses of allele frequency spectra suggested that the broader European *C.*
424 *parasitica* populations efficiently removed the most deleterious mutations through purifying selection.
425 In contrast, the S12 lineage shows strong skews towards very low minor allele frequencies of all
426 mutation categories. Interestingly, we found a broader spread in allele frequencies for nearly neutral
427 mutations in the S12 lineages. This suggests that despite the largely clonal population structure,
428 deleterious mutations can still be removed through low levels of recombination and purifying selection.
429 Using accumulated mutations as markers to retrace the spatial expansion of the invasive S12 lineage,
430 we found no indication for a step-wise geographic expansion along potential invasion routes. A lack of
431 genetic clustering across south-eastern Europe may be a consequence of high levels of gene flow
432 frequently introducing new genotypes over large distances. However, the lack of geographic structure
433 could also have its origins from substantial population bottlenecks during the spread of S12 across south-
434 eastern Europe. Finally, the largely clonal lineage may also become exposed to processes such as
435 Muller's Ratchet fixing deleterious mutations over time (Felsenstein, 1974).
436
437 We show how a highly invasive fungal pathogen lineage can emerge from an intermediate, genetically
438 diverse bridgehead population. This is in line with the self-reinforcement invasion model where initial
439 introductions promote secondary spread (Bertelsmeier et al., 2018; Garnas et al., 2016). However,
440 empirical evidence for adaptation in bridgehead populations is often elusive. Additionally, human
441 transport (Banks et al. 2015) and host naivety of the European chestnut could have contributed
442 substantially to the rapid spread without the need for diversification and adaptation in the bridgehead
443 population. As a response to fungal pathogen invasions among bats, some host populations evolved
444 resistance (Langwig et al., 2017). There is no evidence for the emergence of tolerance or resistance in
445 the European chestnut, however deployed control measures such as the artificial introduction of the
446 *Cryphonectria hypovirus I* (CHV-1) can severely reduce fungal virulence (Rigling & Prospero, 2018).
447 Interestingly, the mycoviral spread should be facilitated in asexual populations such as the invasive S12
448 lineage due to the lack of vegetative compatibility barriers. Outcrossing populations often harbor many

449 different vegetative compatibility groups slowing transmission (Robin & Heiniger, 2001). Hence, the
450 presence of the mycovirus may favor sexual reproduction and the immigration of additional vegetative
451 compatibility types. In turn, the diversification may reduce the evolutionary advantage of the invasive
452 lineage.
453
454
455 **Materials and Methods**
456
457 **Samples of *C. parasitica***
458 A total of 188 virulent and mycovirus-free isolates of *C. parasitica* were sequenced. A majority of 182
459 originated from Albania, Bosnia, Bulgaria, Croatia, Georgia, Greece, Italy, Kosovo, Macedonia, Serbia,
460 Slovenia, Switzerland, and Turkey (Fig. 1B, Supp. Table 1, Supp. Fig. 7). The six other isolates were
461 from South Korea (*n*=2) and North America (*n*=4) (Supp. Table 1). A total of 125 European isolates
462 belonged to the vegetative compatibility type EU-12, whereas 57 isolates represented other vegetative
463 compatibility types (EU-types) occurring in central and south-eastern Europe. A total of 45 isolates from
464 Bulgaria, Greece, Italy and Macedonia were already included in a previous population-wide study on
465 south-eastern European *C. parasitica* diversity by (Milgroom et al., 2008). All samples were collected
466 between 1951–2018 and are stored as glycerol stocks at –80°C in the culture collection of the Swiss
467 Federal Research Institute WSL.
468
469 **DNA extraction and genotyping**
470 All isolates were first inoculated onto cellophane-covered potato dextrose agar plates (PDA, 39 g/L; BD
471 Becton, Dickinson & Company; Franklin Lakes, USA) (Hoegger et al., 2000) and incubated for a
472 minimum of one week at 24°C, at a 14 h light and 10 h darkness cycle. After a sufficient amount of
473 mycelium and spores had grown, the isolates were harvested by scratching the mycelial mass off the
474 cellophane, transferring it into 2 ml tubes and freeze-drying it for 24 h. For DNA extraction, 15–20 mg
475 of dried material was weighted and single tube extraction was performed using the DNeasy Plant Mini
476 Kit (Qiagen, Hilden, Germany). DNA quantity was measured using the Invitrogen Qubit 3.0

477 Fluorometer (Thermo Fisher Sci-entific, Waltham (MA), USA) and DNA quality was assessed using
478 the Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham (MA), USA). Prior to
479 sequencing, all isolates were screened for their genotype at 10 microsatellite markers (Prospero &
480 Rigling, 2012). Additionally, the isolates were screened for their vegetative compatibility and mating
481 type alleles in two multiplex PCRs, as described in Cornejo et al. (2019). Allele sizes for the genotyping
482 and assays were scored with GeneMapper 5 (Thermo Fisher Scientific, Waltham (MA), USA).
483
484 **Illumina whole-genome sequencing, variant calling and filtration**
485 Isolates were prepared for sequencing using the TrueSeq Nano DNA HT Library Preparation kit
486 (Illumina, San Diego, USA). The libraries were sent to the Functional Genomics Centre Zurich (ETH
487 Zurich and University of Zurich) and sequenced on the Illumina Hiseq 4000 platform (Illumina, San
488 Diego, USA). The obtained sequences were trimmed with Trimmomatic v0.36 (Bolger et al., 2014) and
489 aligned with Bowtie 2 v2.3.5.1 (Langmead & Salzberg, 2012) and SAMtools v1.9 (Li et al., 2009) to
490 the *C. parasitica* reference genome (43.9 Mb) EP155 v2.0 available at the Joint Genome Institute
491 (http://jgi.doe.gov/). Variant calling, selection and filtration were conducted with the Genome Analysis
492 Toolkit GATK v3.8 and v4.0.2.0 (McKenna et al., 2010). We retained variants satisfying the following
493 filtration parameters: QUAL: ≥ 100, MQRankSum (lower): ≥ -2.0, QD: ≥ 20.0, MQRankSum (upper):
494 ≤ 2.0, MQ: ≥ 20.0, BaseQRankSum (lower): ≥ - 2.0, ReadPosRankSum (lower): ≥ -2.0,
495 ReadPosRankSum (upper): ≤ 2.0, BaseQRankSum(upper): ≤ 2.0 (Supp. Fig. 1). Furthermore, we used
496 BCFtools v1.9 (Narasimhan et al., 2016) and the R package vcfR (Knaus & Grünwald, 2017; R Core
497 Team, 2014) for querying and VCFtools v0.1.16 (Danecek et al., 2011) for downstream variant filtering.
498 Variants were additionally filtered for minor allele count (MAC) ≥ 1, excluding all missing data. Sites
499 were also filtered per genotype, only keeping biallelic SNPs with a minimum depth of 3 and a
500 genotyping quality (GQ) of 99. To exclude SNPs associated with the mating type, we ran an association
501 study with TASSEL 5 (Bradbury et al., 2007) and retrieved *p*-values for each SNP across the genome.
502 We set a *p*-value threshold to remove all SNPs with $p \leq 1x10^{-10}$ with the mating type for further analysis.
503 The SNPs showing strong association with the mating type were located on scaffold 2 between positions
504 100’853–1’997’710 bp (Supp. Fig. 2).

505 **Phylogenetic reconstruction**

506 The filtered whole-genome SNP dataset was used to build phylogenetic trees. We generated a
507 maximum likelihood (ML) tree with RAxML v8.2.11 (Stamatakis, 2006), applying rapid bootstrapping
508 generating 100 maximum likelihood (ML) trees. Phylogenetic trees were displayed using FigTree
509 v1.4.3 (Rambaut, 2016). We also generated an unrooted phylogenetic network with using SplitsTree
510 v4.14.6. (Huson & Bryant, 2006). SplitsTree was also used for calculating the PHI test (Bruen et al.,
511 2006) to test for recombination. The required file conversions for using RAxML and SplitsTree (*i.e.*
512 from VCF to FASTA format) were done with PGDSpider v2.1.1.5 (Lischer & Excoffier, 2011). We
513 also performed a principal component analysis (PCA) as implemented in the R package ade4
514 (Bougeard & Dray, 2018).

515

516 **Inference of S12 donor populations**

517 We generated an averaged co-ancestry matrix as inferred by fineSTRUCTURE v2.1.3 (Lawson et al.,
518 2012). The software uses a Markov-Chain-Monte-Carlo (MCMC) based algorithm to infer ancestral
519 contributions based on patterns of haplotype similarity. We ran the fineSTRUCTURE pipeline in
520 'automatic mode', with 500 Expectation-Maximation (EM) and 300'000 MCMC iterations, 400'000
521 maximization steps to infer the best tree and with ploidy set to 1. fineSTRUCTURE input files were
522 created with LDhat (McVean & Auton, 2007) and the R package vcfR.

523

524 **Population genetic analyses and SNP impact assessment**

525 We computed allele frequencies and estimated the allele spectrum using VCFtools. We used RStudio
526 (RStudio Team, 2015) and ggplot2 (Wickham, 2016) for visualizations. Synonymous and non-
527 synonymous sites were identified and annotated using SnpEff v4.3t (Cingolani et al., 2012). Variants
528 which were classified by SnpEff as having a "high" , "moderate" or "modifying" impact on fungal
529 protein sequences, were further processed in R using the packages dplyr (Wickham et al., 2018),
530 reshape2 (Wickham, 2007), tidyr (Wickham & Henry, 2018), as well as and ggplot2. Furthermore, we
531 performed a genome-wide association study (GWAS) with TASSEL 5 for identifying highly associated
532 SNPs with genetic groups such as S12. Nucleotide diversity was calculated with vcftools --site-pi

533 function using vcf-files that were converted to diploid. Identification of putative selective sweeps was
534 performed using RAiSD v2.5 software that is based on three different genetic signatures of positive
535 selection (with -y 1 -M 0 -w 50 -c 1 parameters) (Alachiotis & Pavlidis, 2018). The Manhattan plot
536 shows the distribution of the computed composite $\mu$ statistic of positive selection.
537

538 **Transposable element *de novo* identification and population scans**

539 We performed a *de novo* identification of transposable elements in the *C. parasitica* reference genome
540 EP155 v2.0 using RepeatModeler v2.0.1 (Flynn et al., 2019). The consensus sequences were merged to
541 the RepBase (RepBaseRepeatMaskerEdition-20181026) and then used for repeat annotation using
542 RepeatMasker v4.0.7 with a blast cutoff at 250 (Smit et al., 2015). Repeats were then filtered out for
543 low complexity and simple repeats, and parsed using the parseRM_merge_interrupted.pl script from
544 https://github.com/4ureliek/Parsing-RepeatMasker-Outputs. We only retained elements longer than 100
545 bp and overlapping identical elements were merged into single elements for the annotation. In addition,
546 elements of the same family separated by less than 200 bp were considered as part of the same TE and
547 merged into a single element. We analyzed population-level presence/absence variation of transposable
548 elements using the R-based tool ngs_te_mapper, using the previously trimmed reads and bwa version
549 0.7.17-r1188 (Bergman, 2012; Li & Durbin, 2009; Linheiro & Bergman, 2012). Overlapping genes and
550 TEs were identified using the intersect function from the bedtools suite v2.29 (Quinlan & Hall, 2010).
551

552 **Genome compartmentalization and copy-number variation analyses**

553 We investigated the genome architecture of *C. parasitica* following the protocol described in (Saunders
554 et al., 2014). Briefly, we computed intergenic distances with the Calculate_FIR_length.pl script using
555 the gene prediction for *C. parasitica* reference genome EP155 v2.0. We defined 40 bins given the range
556 of 3' and 5' intergenic distances and calculated the number of genes that fell within each bin. To infer
557 copy-number variation, we computed the normalized read depth for all isolates using the CNVcaller
558 pipeline (Wang et al., 2017). Briefly, the reference genome was first split into 800 bp overlapping kmers
559 that were re-aligned to the reference using blasr (-m 5 --noSplitSubreads --minMatch 15 --maxMatch
560 20 --advanceHalf --advanceExactMatches 10 --fastMaxInterval --fastSDP --aggressiveIntervalCut --

561 bestn 10) to identify duplicated windows (python3 0.2.Kmer_Link.py ref.genome.kmer.aln 800
562 ref.genome.800.window.link). The resulting windows were then used to calculate the normalized read
563 depth for each isolate from the aligned reads (bam format) in 800 bp windows (Individual.Process.sh -
564 b $ bam -h $ { i% .bam} -d ref.genome.800.window.link -s scaffold_1). For all further analyses, we
565 used the read depth normalized for the absolute copy number and GC content of each sample
566 (RD_normalized output). Given the observed distribution of the normalised read depth (NRD) across
567 all isolates, we considered windows with NRD > 1.6 to be duplicated and windows with NRD < 0.4 as
568 deleted.
569
570 **Mating experiments**
571 The ability of *C. parasitica* isolates belonging to S12 with mating type MAT-1 to sexually outcross with
572 isolates of opposite mating type, was assessed in an inoculation experiment. For this, we randomly
573 selected four S12 and one non-S12 isolate with mating type MAT-1, as well as five isolates of opposite
574 mating type MAT-2 from populations in Italy (Nebrodi), Kosovo (Molliq) and Bosnia (Konic,
575 Vrnograč, Projsa) (Supp. Table 2). All isolates of both mating types belonged to the vegetative
576 compatibility type EU-12 and mating pairs were only formed between isolates from the same geographic
577 source populations. As substrate for the pairings, 40 mm long segments of dormant chestnut (*C. sativa*)
578 stems (15–20 mm in diameter) were split longitudinally in half. The wood pieces were autoclaved twice,
579 placed onto sterile petri dishes (90mm diameter), which were filled with 1.5% water agar to enclose the
580 pieces. Mating pairs were inoculated on opposite sides of each halved stem with three replicates per
581 pairing. The inoculated plates were then incubated at 25°C under a 16 h photoperiod (2500 Lux) for 14
582 days. After two weeks, mating was stimulated by adding 5 ml sterile water to the plates to suspend and
583 distribute the conidia produced by both isolates over the stem segment. Any excess water was
584 subsequently removed and the plates were incubated at 18°C under an 8 h photoperiod (2500 Lux).
585 After 5 months of incubation, perithecia formation was assessed under a dissecting microscope. To
586 confirm successful outcrossing, single perithecia were carefully extracted from the stromata and crushed
587 in a drop of sterile distilled water. The resulting ascospore suspensions were plated on PDA and
588 incubated at 25°C for 24–36 h. Afterwards, single germinating ascospores were transferred to PDA and

incubated in the dark for 3–5 days at 25°C. DNA was then extracted from 10 mg of lyophilized mycelium using the kit and instructions by KingFisher (Thermo Fisher Scientific, Waltham MA, USA). All single ascospore cultures were screened for mating types by performing a multiplex PCR following the protocol described in Cornejo et al. (2019).

## Acknowledgements

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

## 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), 1–11.

Allendorf, F. W. & Lundquist, L. L. (2003). Introduction: Population biology, evolution, and control of invasive species. *Conservation Biology*, *17*(1), 24–30.

Avolio, S. (1978). La castanicoltura in Calabria. Aspetti selvicolturali e linee di intervento. *L'Italia Forestale e Montana*, *2*, 278–296.

Bergman, C. M. (2012). A proposal for the reference-based annotation of de novo transposable element insertions. *Mobile genetic elements*, *2*(1), 51–54.

Bertelsmeier, C. & Keller, L. (2018). Bridgehead effects and role of adaptive evolution in invasive populations. *Trends in Ecology & Evolution*.

Bertelsmeier, C., Ollier, S., Liebhold, A. M., Brockerhoff, E. G., Ward, D., & Keller, L. (2018). Recurrent bridgehead effects accelerate global alien ant spread. *Proceedings of the National Academy of Sciences*, *115*(21), 5486–5491.

Billiard, S., López-Villavicencio, M., Hood, M., & Giraud, T. (2012). Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *Journal of Evolutionary Biology*, *25*(6), 1020–1038.

Biraghi, A. (1946). *Cancro della corteccia del castagno*. Ramo editoriale degli agricoltori.

Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., Turner, K. G., Whitney, K. D., & Rieseberg, L. H. (2015). What we still don't know about invasion genetics. *Molecular Ecology*, *24*(9), 2277–2297.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

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

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635.

Brasier, C. & Webber, J. (2019). Is there evidence for post-epidemic attenuation in the Dutch elm disease pathogen *Ophiostoma novo-ulmi*? *Plant Pathology*, *68*(5), 921–929.

Bruen, T. C., Philippe, H., & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics*, *172*(4), 2665–2681.

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, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (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*(2), 80–92.

652 Cornejo, C., Šever, B., Kupper, Q., Prospero, S., & Rigling, D. (2019). A multiplexed genotyping
653 assay to determine vegetative incompatibility and mating type in *Cryphonectria parasitica*. *European*
654 *Journal of Plant Pathology*, 1–11.

655 Cortesi, P., McCulloch, C. E., Song, H., Lin, H., & Milgroom, M. G. (2001). Genetic control of
656 horizontal virus transmission in the chestnut blight fungus, *Cryphonectria parasitica*. *Genetics*,
657 *159*(1), 107–118.

658 Cortesi, P., Milgroom, M. G., & Bisiach, M. (1996). Distribution and diversity of vegetative
659 compatibility types in subpopulations of *Cryphonectria parasitica* in Italy. *Mycological Research*,
660 *100*(9), 1087–1093.

661 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E.,
662 Lunter, G., Marth, G. T., Sherry, S. T., et al. (2011). The variant call format and VCFtools.
663 *Bioinformatics*, *27*(15), 2156–2158.

664 Demené, A., Legrand, L., Gouzy, J., Debuchy, R., Saint-Jean, G., Fabreguettes, O., & Dutech, C.
665 (2019). Whole-genome sequencing reveals recent and frequent genetic recombination between clonal
666 lineages of *Cryphonectria parasitica* in western Europe. *Fungal Genetics and Biology*.

667 Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: waltz
668 with plants. *Current opinion in genetics & development*, *35*, 57–65.

669 Drenth, A., McTaggart, A. R., & Wingfield, B. D. (2019). Fungal clones win the battle, but
670 recombination wins the war. *IMA fungus*, *10*(1), 18.

671 Dutech, C., Barrès, B., Bridier, J., Robin, C., Milgroom, M., & Ravigné, V. (2012). The chestnut
672 blight fungus world tour: Successive introduction events from diverse origins in an invasive plant
673 fungal pathogen. *Molecular Ecology*, *21*(16), 3931–3946.

674 Dutech, C., Fabreguettes, O., Capdevielle, X., & Robin, C. (2010). Multiple introductions of divergent
675 genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity*,
676 *105*(2), 220–228.

677 Elliott, K. J. & Swank, W. T. (2008). Long-term changes in forest composition and diversity following
678 early logging (1919–1923) and the decline of American chestnut (*Castanea dentata*). *Plant Ecology*,
679 *197*(2), 155–172.

680 Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, *78*(2), 737–756.
681 Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2019).

682 Repeatmodeler2: automated genomic discovery of transposable element families. *BioRxiv*, 856591.
683 Fontaine, M. C., Gladieux, P., Hood, M. E., & Giraud, T. (2013). History of the invasion of the anther
684 smut

685 pathogen on *Silene latifolia* in North America. *New Phytologist*, *198*(3), 946–956.

686 Garbelotto, M., Rocca, G. D., Osmundson, T., di Lonardo, V., & Danti, R. (2015). An increase in
687 transmission-related traits and in phenotypic plasticity is documented during a fungal invasion. *Eco-*
688 *sphere*, *6*(10), 1–16.

689 Garnas, J. R., Auger-Rozenberg, M.-A., Roques, A., Bertelsmeier, C., Wingfield, M. J., Saccaggi, D.
690 L., Roy, H. E., & Slippers, B. (2016). Complex patterns of global spread in invasive insects: Eco-
691 evolutionary and management consequences. *Biological Invasions*, *18*(4), 935–952.

692 Gau, R. D., Merz, U., Falloon, R. E., & Brunner, P. C. (2013). Global genetics and invasion history of
693 the potato powdery scab pathogen, *Spongospora subterranea* f. sp. *subterranea*. *PLoS One*, *8*(6),
694 e67944.

695 Gladieux, P., Feurtey, A., Hood, M. E., Snirc, A., Clavel, J., Dutech, C., Roy, M., & Giraud, T.
696 (2015). The population biology of fungal invasions. *Molecular Ecology*, *24*(9), 1969–1986.

697 González-Varela, G., González, A. J., & Milgroom, M. G. (2011). Clonal population structure and
698 intro- ductions of the chestnut blight fungus, *Cryphonectria parasitica*, in Asturias, northern Spain.
699 *European Journal of Plant Pathology*, *131*(1), 67–79.

700 Griffin, G. J. (1986). Chestnut blight and its control. *Horticultural Reviews*, *8*, 291–336.

701 Grünwald, N. J., Garbelotto, M., Goss, E. M., Heungens, K., & Prospero, S. (2012). Emergence of the
702 sudden oak death pathogen *Phytophthora ramorum*. *Trends in Microbiology*, *20*(3), 131–138.

703 Heald, F. & Studhalter, R. A. (1914). Birds as carriers of the chestnut blight fungus. *Journal of*
704 *Agriculture Research.*, *2*, 405–422.

705 Heiniger, U. & Rigling, D. (1994). Biological control of chestnut blight in Europe. *Annual review of*
706 *Phy- topathology*, *32*(1), 581–599.

707 Heitman, J., Sun, S., & James, T. Y. (2013). Evolution of fungal sexual reproduction. *Mycologia*,
708 *105*(1), 1–27.

709 Hoegger, P. J., Rigling, D., Holdenrieder, O., & Heiniger, U. (2000). Genetic structure of newly
710 established populations of *Cryphonectria parasitica*. *Mycological Research*, *104*(9), 1108–1116.

711 Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies.
712 *Molecular Biology and Evolution*, *23*(2), 254–267.

713 Idnurm, A., Hood, M. E., Johannesson, H., & Giraud, T. (2015). Contrasted patterns in mating-type
714 chromosomes in fungi: Hotspots versus coldspots of recombination. *Fungal Biology reviews*, *29*(3-4),
715 220–229.

716 Ježić, M., Krstin, L., Rigling, D., & Ćurković-Perica, M. (2012). High diversity in populations of the
717 introduced plant pathogen, *Cryphonectria parasitica*, due to encounters between genetically divergent
718 genotypes. *Molecular Ecology*, *21*(1), 87–99.

719 Karadžić, D., Radulović, Z., Sikora, K., Stanivuković, Z., Ćurguz, V. G., Oszako, T., & Milenković, I.
720 (2019). Characterisation and pathogenicity of *Cryphonectria parasitica* on sweet chestnut and sessile
721 oak trees in Serbia. *Plant Protection Science*, *55*(3), 191–201.

722 Knaus, B. J. & Grünwald, N. J. (2017). vcfr: A package to manipulate and visualize variant call format
723 data in R. *Molecular Ecology Resources*, *17*(1), 44–53.

724 Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*,
725 *9*, 357–360.

726 Langwig, K. E., Hoyt, J. R., Parise, K. L., Frick, W. F., Foster, J. T., & Kilpatrick, A. M. (2017).
727 Resistance in persisting bat populations after white-nose syndrome invasion. *Philosophical*
728 *Transactions of the Royal Society B: Biological Sciences*, *372*(1712), 20160044.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS genetics*, *8*(1), e1002453.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Linheiro, R. S. & Bergman, C. M. (2012). Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PloS one*, *7*(2).

Lischer, H. E. & Excoffier, L. (2011). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299.

Lombaert, E., Guillemaud, T., Cornuet, J.-M., Malausa, T., Facon, B., & Estoup, A. (2010). Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PloS one*, *5*(3), 1–7.

Marbuah, G., Gren, I.-M., & McKie, B. (2014). Economics of harmful invasive species: a review. *Diversity*, *6*(3), 500–523.

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*(4), 577–586.

Marra, R., Cortesi, P., Bissegger, M., & Milgroom, M. (2004). Mixed mating in natural populations of the chestnut blight fungus, *Cryphonectria parasitica*. *Heredity*, *93*(2), 189–195.

Marra, R. E. & Milgroom, M. G. (2001). The mating system of the fungus *Cryphonectria parasitica*: Selfing and self-incompatibility. *Heredity*, *86*(2), 134–143.

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

McVean, G. & Auton, A. (2007). LDhat 2.1: A package for the population genetic analysis of recombination. *Department of Statistics, Oxford, OX1 3TG, UK*.

Milgroom, M. G. & Cortesi, P. (1999). Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *Proceedings of the National Academy of Sciences*, *96*(18), 10518–10523.

Milgroom, M. G., Sotirovski, K., Spica, D., Davis, J. E., Brewer, M. T., Milev, M., & Cortesi, P. (2008). Clonal population structure of the chestnut blight fungus in expanding ranges in southeastern Europe. *Molecular Ecology*, *17*(20), 4446–4458.

Milgroom, M. G., Wang, K., Zhou, Y., Lipari, S. E., & Kaneko, S. (1996). Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia*, *88*(2), 179–190.

Myteberi, I. F., Lushaj, A. B., Keča, N., Lushaj, A. B., & Lushaj, B. M. (2013). Diversity of *Cryphonectria parasitica*, hypovirulence, and possibilities for biocontrol of chestnut canker in Albania. *Int. J. Microb. Res. Rev*, *1*, 11–21.

767 Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016).
768 BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation
769 sequencing data. *Bioin- formatics*, *32*(11), 1749–1751.

770 Paoletti, M., Buck, K. W., & Brasier, C. M. (2006). Selective acquisition of novel mating type and
771 vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote
772 *Ophiostoma novo- ulmi*. *Molecular Ecology*, *15*(1), 249–262.

773 Philibert, A., Desprez-Loustau, M.-L., Fabre, B., Frey, P., Halkett, F., Husson, C., Lung-Escarmant,
774 B., Marçais, B., Robin, C., Vacher, C., et al. (2011). Predicting invasion success of forest pathogenic
775 fungi from species traits. *Journal of Applied Ecology*, *48*(6), 1381–1390.

776 Prospero, S., Conedera, M., Heiniger, U., & Rigling, D. (2006). Saprophytic activity and sporulation
777 of *Cryphonectria parasitica* on dead chestnut wood in forests with naturally established
778 hypovirulence. *Phy- topathology*, *96*(12), 1337–1344.

779 Prospero, S. & Rigling, D. (2012). Invasion genetics of the chestnut blight fungus *Cryphonectria*
780 *parasitica* in Switzerland. *Phytopathology*, *102*(1), 73–82.

781 Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
782 features. *Bioinformatics*, *26*(6), 841–842.

783 R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R
784 Foundation for Statistical Computing.

785 Raboin, L.-M., Selvi, A., Oliveira, K. M., Paulet, F., Calatayud, C., Zapater, M.-F., Brottier, P.,
786 Luzaran, R., Garsmeur, O., Carlier, J., et al. (2007). Evidence for the dispersal of a unique lineage
787 from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*. *Fungal*
788 *Genetics and Biology*, *44*(1), 64–76.

789 Rambaut, A. (2016). Figtree v1. 4.3 software. Institute of Evolutionary Biology, University of
790 Edinburgh. Rigling, D. & Prospero, S. (2018). *Cryphonectria parasitica*, the causal agent of chestnut
791 blight: Invasion history, population biology and disease control. *Molecular Plant Pathology*, *19*(1), 7–
792 20.
793
794 Robin, C. & Heiniger, U. (2001). Chestnut blight in Europe: Diversity of *Cryphonectria parasitica*,
795 hypovirulence and biocontrol. *Forest Snow and Landscape Research*, *76*(3), 361–367.
796 Rossman, A. Y. (2001). A special issue on global movement of invasive plants and fungi. *AIBS*
797 *Bulletin*, *51*(2), 93–94.

798 RStudio Team (2015). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio,
799 Inc.

800 Santini, A., Ghelardini, L., De Pace, C., Desprez-Loustau, M.-L., Capretti, P., Chandelier, A., Cech,
801 T., Chira, D., Diamandis, S., Gaitniekis, T., et al. (2013). Biogeographical patterns and determinants
802 of invasion by forest pathogens in Europe. *New Phytologist*, *197*(1), 238–250.

803 Saunders, D. G., Win, J., Kamoun, S., & Raffaele, S. (2014). Two-dimensional data binning for the
804 analysis of genome architecture in filamentous plant pathogens and other eukaryotes. In *Plant-*
805 *Pathogen Interactions* (pp. 29–51). Springer.

806 Smit, A., Hubley, R., & Green, P. (2015). Repeatmasker open-4.0. 2013–2015.

Sotirovski, K., Papazova-Anakieva, I., Grünwald, N., & Milgroom, M. (2004). Low diversity of vegetative compatibility types and mating type of *Cryphonectria parasitica* in the southern Balkans. *Plant Pathology*, *53*(3), 325–333.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690.

Suehs, C., Affre, L., & Médail, F. (2004). Invasion dynamics of two alien *Carpobrotus* (Aizoaceae) taxa on a Mediterranean island: II. Reproductive strategies. *Heredity*, *92*(6), 550.

Taylor, J. W., Hann-Soden, C., Branco, S., Sylvain, I., & Ellison, C. E. (2015). Clonal reproduction in fungi. *Proceedings of the National Academy of Sciences*, *112*(29), 8901–8908.

Trestic, T., Uscuplic, M., Colinas, C., Rolland, G., Giraud, A., & Robin, C. (2001). Vegetative compatibility type diversity of *Cryphonectria parasitica* populations in Bosnia-Herzegovina, Spain and France. *For. Snow Landsc. Res.*, *76*(3), 391–396.

van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., & Hodgins, K. A. (2017). Multiple introductions, admixture and bridgehead invasion characterize the introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Molecular Ecology*, *26*(20), 5421–5434.

Vuković, R., Liber, Z., Ježić, M., Sotirovski, K., & Ćurković-Perica, M. (2019). Link between epigenetic diver- sity and invasive status of south-eastern European populations of phytopathogenic fungus *Cryphonectria parasitica*. *Environmental Microbiology*.

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*(12), gix115.

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, *21*(12), 1–20.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A Grammar of Data Manipulation*. R package version 0.7.8.

Wickham, H. & Henry, L. (2018). *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. R package version 0.8.2.

Wingfield, M. J., Slippers, B., Roux, J., & Wingfield, B. D. (2010). Fifty years of tree pest and pathogen invasions, increasingly threatening world forests. *Fifty years of invasion ecology: The legacy of charles elton*, 89–99.

Wuest, C. E., Harrington, T. C., Fraedrich, S. W., Yun, H.-Y., & Lu, S.-S. (2017). Genetic variation in native populations of the laurel wilt pathogen, *Raffaelea lauricola*, in Taiwan and Japan and the introduced population in the United States. *Plant Disease*, *101*(4), 619–628.

Yang, X.-M., Sun, J.-T., Xue, X.-F., Li, J.-B., & Hong, X.-Y. (2012). Invasion genetics of the western flower thrips in China: Evidence for genetic bottleneck, hybridization and bridgehead effect. *PLoS One*, *7*(4), e34567.

846 **Table**

847

848 **Table 1: Donor isolates for S12 identified using FineStructure**. The geographic origin (country and
849 population), vegetative compatibility (vc) type and mating type of donors contributing to genotypes of
850 the invasive S12 lineage are given. See Fig. 3 for the corresponding co-ancestry matrix.

851

852

| Country | Population | vc type | Mating type | Donor isolate |
|---|---|---|---|---|
| Bosnia | Projsa | EU-11 | MAT-1 | M1412 |
| | Projsa | EU-12 | MAT-2 | M1407 |
| | Projsa | EU-12 | MAT-2 | M1430 |
| | Projsa | EU-12 | MAT-2 | M1431 |
| | Vrnograč | EU-12 | MAT-1 | M1808 |
| | Vrnograč | EU-15 | MAT-2 | M1834 |
| | Vrnograč | EU-18 | MAT-2 | M1797 |
| Croatia | Kostajnica | EU-12 | MAT-1 | HK60B |
| Georgia | Mackhunceti | EU-12 | MAT-2 | MAK23 |
| Switzerland | Gnosca | EU-12 | MAT-1 | M6697 |
| | Biasca | EU-12 | MAT-2 | M4023 |
| | Biasca | EU-12 | MAT-2 | M4022 |
| | Claro | EU-12 | MAT-2 | M2466 |

853