Million years old recombination suppression and balancing selection in a region partially linked to the mating-type locus in the invasive chestnut blight fungal pathogen *Cryphonectria parasitica*

Fanny E. Hartmann[1]*, Ricardo C. Rodriguez de la Vega[1], Arthur Demené[2], Thomas Badet[3], Jean-Philippe Vernadet[1], Quentin Rougemont[1], Amandine Labat[1], Alodie Snirc[1], Lea Stauber[5], Daniel Croll[3], Simone Prospero[4], Cyril Dutech[2], Tatiana Giraud[1]

[1] Ecologie Systematique et Evolution, CNRS, Université Paris-Saclay, AgroParisTech, 91198 Gif-sur-Yvette, France

[2] BIOGECO, INRAE, Univ. Bordeaux, 69 route d'Arcachon, Cestas F-33610, France.

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

[4] Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

[5] Institute for Infectious Diseases, University of Bern, Switzerland

*Corresponding author: Fanny E. Hartmann

E-mail: fanny.hartmann@universite-paris-saclay.fr

Key words: recombination suppression, fungi, mating-type chromosome, inversion, deleterious allele sheltering, supergenes, evolutionary strata, sex chromosomes

Abstract

Recombination suppression often evolves in sex chromosomes and around mating-type loci. In the invasive chestnut blight fungus *Cryphonectria parasitica* (Ascomycota), a genomic region was previously suggested to lack recombination and to be partially linked to the mating-type (MAT) locus based on the analysis of a few progenies. Using hundreds of available *C. parasitica* genomes and generating several new high-quality genome assemblies from the native and introduced range of the pathogen, we show that a ca. 1.2 Mb genomic region proximal to the mating-type locus lacks recombination worldwide. In invasive populations, this MAT-proximal region displayed two highly differentiated haplotypes, that were strongly associated to mating types, but not completely. High-quality assemblies revealed an inversion in one of the haplotypes and footprints of degeneration worldwide, the MAT-proximal region being enriched in gene disruptions, non-synonymous substitutions and transposable elements in both haplotypes. The divergence between the two haplotypes was estimated to have occurred at least 1.5 million years ago and two haplotypes segregate in all continents, including the native range. High differentiation between haplotypes, their occurrence on different continents, their balanced frequencies within populations, their genomic rearrangements and degeneration worldwide, altogether suggest an ancient recombination suppression maintained by selection. The MAT-Proximal region carries multiple genes upregulated under virus infection or vegetative incompatibility reaction. This study sheds light on a case of a large non-recombining region partially linked to a mating compatibility locus, and on balancing selection maintaining differentiated haplotypes, possibly involving deleterious mutations and/or host or virus adaptation in a devastating tree pathogen.

Introduction

Recombination increases the efficiency of selection, the purging of deleterious alleles and the generation of potentially beneficial allelic combinations (Otto and Lenormand 2002). However, recombination can also break up beneficial allelic combinations and, therefore, can be selected against, generating supergenes, i.e., large regions without recombination encompassing multiple genes (Schwander et al. 2014). This is often the case in genomic regions controlling mating compatibility, such as regions determining sex, mating type or self-incompatibility in plants, animals, fungi, algae and oomycetes (Charlesworth et al. 2005; Bergero and Charlesworth 2009; Umen 2011; Charlesworth 2016; Dussert et al. 2020; Hartmann et al. 2021). Such recombination suppression keeps alleles linked across different genes that prevent self-compatibility or intermediate sexual phenotypes.

Recombination suppression can sometimes extend away from sex-determining genes or mating-type loci (Bergero and Charlesworth 2009; Furman et al. 2020; Hartmann et al. 2021). Such extension can be due to antagonistic selection, i.e. selection for different alleles in different sexes or mating types (Charlesworth 2023). Extension of recombination suppression can also occur near mating-type loci or sex-determining genes because of selection for inversions carrying fewer deleterious mutations than average in the population, their few deleterious mutations being sheltered when linked to a permanently heterozygous allele, such as on a Y chromosome or a mating-type chromosome (Jay et al. 2022; Jay, D.L. Jeffries, Hartmann, et al. 2024).

On the long term, however, because of less efficient selection, genomic regions without recombination show signs of degeneration, such as transposable element accumulation, rearrangements, non-synonymous substitutions, decreased gene expression, reduced frequency of optimal codons and gene losses (Bachtrog 2013; Carpentier et al. 2022; Duhamel et al. 2023). The non-recombining regions can nevertheless be maintained by balancing selection (Jay et al. 2021; Berdan et al. 2022). Chromosomal rearrangements such as inversions are often considered to be the mechanistic cause of recombination suppression, although they may also be a mere consequence of recombination suppression. The evolutionary mechanisms leading to recombination suppression and allowing their persistence remain however poorly known (Jay, D. Jeffries, Hartmann, et al. 2024). Reporting more

87 diverse constellations of recombination suppression in a variety of organisms with contrasting
88 life-history traits is important for understanding the general patterns of recombination
89 suppression, their evolution and maintenance (Ironside 2010; Charlesworth 2016; Furman et
90 al. 2020; Hartmann et al. 2021).
91
92 In the invasive chestnut blight fungal pathogen *Cryphonectria parasitica*, mating
93 compatibility is determined by a mating-type locus that displays two alleles, called *MAT-1*
94 and *MAT-2* that are both permanently heterozygous in the diploid and dikaryotic stages
95 (McGuire et al. 2001). A large region without recombination was suggested to occur near the
96 mating-type locus, although not completely linked to it (*i.e.,* at 3.9 cM of the mating-type
97 locus), based on segregation of RAPD and RFLP markers in three crosses involving five
98 different parents, from the USA, Japan or Italy (Kubisiak and Milgroom 2006a). Recent
99 genome-wide association studies in a worldwide collection and in local *C. parasitica*
100 populations in southern Switzerland found significant association of SNPs with the mating-
101 type locus across a large region (>1 Mb), further suggesting the existence of reduced
102 recombination in this region (Stauber et al. 2021; Stauber et al. 2022). This genomic region
103 was enriched in SNPs, transposable elements and copy-number variants, such as deletions,
104 which are consistent with sequence degeneration and a lack of recombination (Stauber et al.
105 2021), and which prevented its assembly so far.
106
107 The evolutionary history of the species is well documented as it is an invasive and highly
108 damaging pathogen, having almost caused the extinction of American chestnut (*Castanea*
109 *dentata*) in North America (Anagnostakis 1987). From its center of origin in Asia and its
110 original hosts, the Chinese chestnut (*Ca. mollissima*) and the Japanese chestnut (*Ca. crenata*),
111 *C. parasitica* has first invaded North America, killing millions of American chestnuts (*Ca.*
112 *dentata*). In Europe, at least two distinct introduction events occurred, on the European
113 chestnut (*Ca. sativa*): one introduction from North America to Italy, and one directly from
114 Asia, probably to the Pyrenees Mountains. European strains will be therefore hereafter
115 referred to as invasive European strains introduced from North America or invasive European
116 strains introduced directly from Asia. The chestnut blight symptoms have been less severe in
117 Europe than in North America, due to lower susceptibility of *C. sativa* and a virus infecting *C.*
118 *parasitica* and causing hypovirulence (Dutech et al. 2010; Dutech et al. 2012). The genetic
119 determinants of the adaptation of the pathogen to its new environments and hosts remain
120 largely unknown (Lovat and Donnelly 2019).

121

122 Lineages with an apparently clonal structure have been identified in invasive populations
123 while sexually reproducing populations occur both in the native and introduced ranges
124 (Dutech et al. 2012; Stauber et al. 2021). The lineages with a predominant clonal structure
125 likely still undergo rare sex events, as shown by the introgression of the missing mating-type
126 allele in most of them (Demené et al. 2019). Furthermore, although this ascomycete fungus is
127 mostly found as haploid mycelia, some isolates have been reported to be heterokaryotic at the
128 mating-type locus in several invasive European and North American populations, i.e., with
129 cells carrying different nuclei, of the opposite mating types *MAT-1* and *MAT-2* (McGuire et
130 al. 2004; McGuire et al. 2005; Dutech et al. 2010; Stauber et al. 2021; Stauber et al. 2022).
131 The theoretical model of recombination suppression based on recessive deleterious mutations
132 showed that non-recombining fragments could be selected for even with a substantial fraction
133 of the life cycle being haploid (Jay et al. 2022). Hence, due to previous evidence of
134 recombination suppression, the natural occurrence of a heterokaryotic stage and its
135 archetypical evolutionary history of invasive pathogen, *C. parasitica* seems an excellent
136 biological model to study the evolution of recombination suppression and test recent
137 theoretical predictions about the existence of non-recombining regions only partially linked to
138 sex- or mating-type determining loci.

139

140 Here, we therefore studied the occurrence of recombination suppression in the genomic
141 region proximal to the mating-type locus in *C. parasitica,* using the previously published
142 genomes and further generated six new high-quality long-read based genome assemblies of
143 strains from the native and introduced range of the pathogen. More specifically, we tested
144 whether recombination was completely suppressed in this region partially linked to the
145 mating-type locus in *C. parasitica*. We first analyze sexually reproducing invasive European
146 populations, for which extensive genomic datasets are available in local populations. We
147 assessed, in these populations, the degree of linkage disequilibrium within the MAT-proximal
148 region, the level of differentiation between haplotypes, as well as the frequency of the two
149 non-recombining haplotypes and their association with mating types. We investigated
150 whether other regions of the genome showed reduced recombination rates and whether
151 genomic footprints of degeneration were present in the non-recombining region, such as
152 genomic rearrangements, transposable element accumulation, non-synonymous substitutions
153 and gene disruptions. We also tested whether this non-recombining region has been gradually
154 expanding, estimated the age of recombination suppression and investigated whether the

155 predicted gene functions in the MAT-proximal region could help understanding the
156 evolutionary cause for recombination suppression and partial linkage to the mating-type
157 locus. We then extended the analyses to other populations with less extensive genomic
158 resources, i.e. other invasive populations and native Asian populations, investigating the
159 presence of recombination suppression footprints and the occurrence of the two differentiated
160 haplotypes worldwide, as well as the possibility of introgression, using available genomes of
161 closely related species, pathogenic and non-pathogenic (Stauber et al. 2021). Finally, we used
162 available expression data to investigate whether the MAT-Proximal region carried genes
163 upregulated under infection by the virus responsible for hypovirulence or during vegetative
164 incompatibility reactions, a phenomenon avoiding hyphal fusions between individuals,
165 considered to protect against virus transmission.
166
167 **Results**
168 ***Footprints of recombination suppression in a large region (> 1 Mb) proximal to the***
169 ***mating-type locus in sexually reproducing invasive populations***
170 We first performed population genomic analyses on available genome sequences within local
171 populations sampled in southern Switzerland over two temporal frames (early 1990 and
172 2019), and within the more broadly geographically distributed CL1 cluster in central and
173 southeastern Europe, both having a recombining genetic structure and being invasive
174 European populations introduced from North America (Table S1 (Stauber et al. 2021)). We
175 studied only monokaryotic strains, i.e. having either *MAT-1* or *MAT-2* in their genomes, for
176 phasing haplotypes and performed stringent SNPs filtering by masking repeats on the EP155
177 reference genome (Crouch et al. 2020) and removing missing data and rare variants to ensure
178 robustness of our analyses. We identified 8,900 SNPs segregating among 71 strains of the
179 1990 Swiss population (35 MAT-1; 36 MAT-2; Table S1), 9,646 SNPs segregating among 62
180 strains of the Swiss 2019 population (20 MAT-1; 42 MAT-2; Table S1) and 15,104 SNPs
181 segregating among 88 strains of the CL1 cluster (41 MAT-1; 47 MAT-2; Table S1), which
182 represent an average density of about 20 SNPs/100 kb, in agreement with previous studies in
183 this invasive fungus (Stauber et al. 2021; Stauber et al. 2022). We investigated the genome-
184 wide linkage disequilibrium (LD) landscape and looked for large blocks (>1 Mb) of high LD
185 (r2>0.9) among SNPs within contigs, to look for signatures of reduced recombination along
186 the genome. LD indeed decays by half across a few 100 kb on average in C. parasitica in
187 recombining regions of the genome (Demené et al. 2019). Therefore, high LD beyond >1 Mb

188 represents strong evidence of recombination cessation and the SNP density is sufficient to
189 study such LD variation.
190
191 We found a large block of high LD on the chromosome called scaffold_2 of the EP155
192 genome in the three populations (red arrow in Fig 1 and Fig S1), corresponding to the mating-
193 type chromosome. The high-LD block was located near the mating-type locus but did not
194 encompass the mating-type locus itself and will be called hereafter the "MAT-proximal
195 region". In the 1990 and 2019 Swiss populations, SNPs between 0.493 Mb and 1.720 Mb in
196 the mating-type chromosome were all in high LD (with r^2>0.9 between SNPs distant by up to
197 1 Mb; in this region, mean r^2 of all SNP pairs in the Swiss 1990 population = 0.89; mean r^2 of
198 all SNP pairs in the Swiss 2019 population = 0.83; Fig S1). In the CL1 cluster, the large high-
199 LD block was also present, but was split into two blocks, a large one and a smaller, peripheral
200 one (orange arrows in Fig 1). We indeed detected values of r^2>0.9 between SNPs distant by
201 up to 1 Mb, between 0.532 Mb and 1.535 Mb (mean r^2 of all SNP pairs = 0.94). Between
202 1.535 and 1.717 Mb, SNPs were also in high LD (mean r^2 = 0.84). The LD between the two
203 high-LD blocks was lower (r^2= 0.59) than within blocks but was still higher than elsewhere
204 along the genome (orange rectangle at the right border of the red triangle; Fig 1). A similar
205 pattern also existed in the Swiss populations (orange rectangle at the right border of the red
206 triangle; Fig S1), although less marked. This is likely due to rare recombination events at one
207 specific locus near the edge of the large fully non-recombining region, lowering the LD level
208 in populations between the two parts of the otherwise fully non-recombining region. The
209 mating-type locus was located at 1.737 Mb in the EP155 genome, distant by 17 kb and 20 kb
210 from the LD block in the Swiss populations and CL1 cluster, respectively. SNP density was
211 on average 48 SNPs/100 kb in the high-LD block region, i.e. higher than in the rest of the
212 chromosome (Fig S2C). When sampling SNPs distant of at least 50 kb in this region to have
213 the same SNP density all over the chromosome, we also detected the high-LD block, which
214 indicates that the SNP density did not generated biases in the LD pattern (Fig S2D).
215
216 We found no other regions of the genomes with r^2>0.9 between SNPs across such a large
217 genomic distance in any population. On scaffold_6 in the CL1 cluster, two smaller SNP
218 blocks had a r^2>0.9 between each other despite being distant (Fig S2C; between SNPs blocks
219 at 3-20 kb and 2.151-2.278 kb), which is likely due to the major intra-scaffold translocation
220 described in this region between the EP155 and ESM15 strains (Demené et al. 2022). The
221 maximum size of the blocks with r^2>0.9 on other scaffolds ranged from 135 to 540 kb in the

222 CL1 cluster, as previously described (Demené et al. 2019). The average distance between
223 SNPs in high LD (r^2>0.9) across the genome was higher on the scaffold_2 than on other
224 scaffolds (in the CL1 cluster, pairwise Wilcoxon test p-value < 2e-16 with Bonferroni
225 correction). Furthermore, the proportion of SNPs in high LD (r^2>0.9) across the genome was
226 also much higher on the scaffold_2 than on other scaffolds due to the high LD in the MAT-
227 proximal region (Fig S2D, Table S2). Therefore, the large (1Mb) and localized block with
228 maximal LD values proximal to the mating-type locus both in the Swiss populations and the
229 CL1 genetic cluster stands out as exceptional in the genome and indicates full recombination
230 cessation, which supports previous inferences from progeny segregation analyses (Kubisiak
231 and Milgroom 2006b). Indeed, even low rates of recombination homogenize alleles and
232 prevents LD building (Dufresnes et al. 2015). The MAT-proximal non-recombining region
233 was actually larger (1.23 Mb), but with likely rare recombination events at one precise locus
234 near its edge, lowering LD between its two fully non-recombining parts.
235
236 Consistent with a lack of recombination, the MAT-proximal region formed two genetically
237 highly differentiated haplotypes in these invasive Swiss and CL1 populations (Fig 2A; Fig S3
238 A(1)-C(1)), as shown by the two clusters on the principal component analysis (PCA) using
239 only SNPs located in the MAT-proximal region, while no structure was detected in the rest of
240 the genome (Fig 2B-C; Fig S3A(2)-C(2)). The neighbor-net networks further supported the
241 existence of two differentiated haplotypes in the MAT-Proximal region, contrasting with an
242 otherwise recombining structure genome-wide (Fig 2D-E; Fig S3B-D). In the Swiss 2019
243 population, a few reticulations between haplotypes were found and three strains (LU3,
244 Nov10, Nov4) appeared to have an intermediate sequence in the MAT-Proximal region
245 between the two haplotypes (Fig S3D). These intermediates were rare and may correspond to
246 incomplete lineage sorting, gene conversion, infrequent recombination between haplotypes
247 and/or may represent a third haplotype.
248
249 In agreement with the view of a partial linkage with the mating-type locus, the two genetic
250 clusters in the MAT-proximal region appeared strongly associated with mating types,
251 although the association did not reach 100% (Figs 2A and D; Fig S3, Table 1). Indeed, the
252 distribution of mating types between the two PCA clusters strongly deviated from
253 expectations under random association (Table 1; chi-squared test: χ^2 = 33.3; p-value = 5.876e-
254 08, with r^2 = 0.38 in CL1). We hereafter call the two differentiated haplotypes MAT-Prox1

255 and MAT-Prox2, referring to the mating-type association. The EP155 genome used as
256 reference for SNP calling above carried the MAT-Prox1 haplotype.
257
258 ***Inversion in the non-recombining MAT-proximal region***
259 To investigate whether rearrangements between haplotypes were present in the MAT-
260 proximal region, we sequenced the genome, using PacBio HiFi, of two strains with alternative
261 MAT-proximal haplotypes: M1400 (MAT-2; MAT-Prox2) and M6697 (MAT-1; MAT-
262 Prox1), originating from the Swiss population in Gnosca and belonging to the CL1 cluster, i.e.
263 European invasive populations introduced from North America (Stauber et al. 2021; Stauber
264 et al. 2022). We built high-quality genome assemblies: statistics of the assemblies were in the
265 same range as those of the genome assemblies of the ESM15 and EP155 strains previously
266 sequenced (Table S3; Demené et al 2022; Crouch et al 2020). The mating-type chromosome
267 could be assembled as a single contig for the first time, in the M1400 genome, likely
268 corresponding to a full chromosome, and was assembled into two contigs in the M6697
269 genome. We therefore used the M1400 genome as a reference for subsequent analyses. To
270 identify the location of the MAT-proximal region in the M1400 genome, we computed LD by
271 read mapping in the 1990 Swiss populations. We found the large block (> 1 Mb) of high LD
272 (r^2>0.9) between 7,285,137 and 8,828,934 bp (red arrow; Fig. S4). The MAT-proximal region
273 was also located 20 kb away from the mating-type locus (located at 7.265 Mb) and was 1.540
274 Mb long using the M1400 genome as reference. Consistent with the results using the EP155
275 genome as reference, we found that the high-LD region was divided into two higher-LD
276 blocks near the edge at 7.392 Mb (orange arrows in Fig. S4).
277
278 The two new high-quality genome assemblies from the invasive Swiss population (introduced
279 from North America) revealed an inversion between the two haplotypes in the MAT-proximal
280 region (Fig 3). The two newly sequenced PacBio genomes were indeed collinear except for
281 the mating-type chromosome (Fig S5A-B-C), where we found a large inversion (>1 Mb)
282 between the M1400 and M6697 genomes (Fig 3; Fig S5A-B-C). Breakpoints of the inversion
283 were located at ca. 7.455 and 8.563 Mb of the tig00000001 contig in the M1400 genome and
284 at ca. 3.380 and 4.185 Mb of the tig00000060 contig in the M16697 genome. The region
285 affected by the inversion was ca. 300 kb smaller in the M6697 genome than in the M1400
286 genome, while the whole MAT-proximal region was ca. 100 kb smaller. The split into the two
287 high-LD blocks in natural populations reported above was outside of the inversion, i.e. 70 kb
288 from the inversion breakpoint (orange arrows in Fig S4). LD in natural populations was

289 higher within the inversion (median value of r^2 = 1) than in other regions of the MAT-
290 proximal region (median value of r^2 = 0.64; Wilcoxon rank sum test with continuity
291 correction; W = 4.8714e+10; p-value < 2e-16).
292
293 We looked for centromeres, as the non-recombining regions near the mating-type locus in
294 other ascomycetes, when they occur, either capture the centromere (Menkis et al. 2008; Sun et
295 al. 2017), or are associated to the occurrence of a single crossing-over between the centromere
296 and the mating-type locus (Grognet et al. 2014; Hartmann et al. 2021; Vittorelli et al. 2022).
297 The centromere in *C. parasitica* may be at 4.380-4.536 Mb on the mating-type chromosome
298 in the M1400 genome, as we detected here a peak in TE density and a drop in GC content
299 (Fig 4A-B). A dotplot of repeats in this region also presented a pattern typical of centromeres
300 (Fig S5D). The MAT-proximal region and the inversion thus did not include the putative
301 centromere and was instead located about 2.9 Mb away of the putative centromere.
302
303 ***Higher genetic differentiation and lower genetic diversity within the MAT-proximal region***
304 ***than in recombining regions***
305 When analyzing polymorphism within the invasive 1990 Swiss population using the M1400
306 genome as reference, we found much higher genetic differentiation between strains of the two
307 MAT-proximal haplotypes in this region than elsewhere along the genome (Fig 4C; Fig S6A;
308 Table S4), as expected for a non-recombining region in LD with the mating-type locus. For
309 example, the F_{ST} values between the two non-recombining haplotypes within the MAT-
310 proximal region in the 1990 Swiss population were nearly maximal (median F_{ST}=0.93 per 50
311 kb window), while they were near zero in the rest of the genome (median F_{ST}=0.02; Wilcoxon
312 test p-value < 2e-16 with Bonferroni correction; Fig 4C; Fig S6A). Such F_{ST} values close to 1
313 indicate a lack of shared polymorphism and therefore support the inference of recombination
314 suppression in the MAT-proximal region.
315
316 Within each haplotype, the genetic diversity at the MAT-proximal region was lower than in
317 the rest of the genome (Wilcoxon test p-value < 2e-16 for the MAT-Prox1 haplotype; p-value
318 = 2.3e-07 for the MAT-Prox2 haplotype; Fig 4E-F; Fig S6C; Table S4), as expected for a
319 region without recombination associated with the mating-type locus, as its effective
320 population size is half as in recombining regions. The diversity was especially low in the pool
321 of isolates with the MAT-Prox1 haplotype (see all MAT-Prox1 sequences clustering on a
322 single point on the PCA in Fig 2A and in the neighbor-net network in Fig 2D, in contrast to

323 the more scattered MAT-Prox2 sequences). Such a very low diversity may be due to a recent
324 selective sweep or to particularly strong bottleneck in this haplotype during the invasion.
325
326 We detected signals for long-term balancing selection in the MAT-proximal haplotypes.
327 Computation of the Tajima's D statistics within all strains suggested signatures of balancing
328 selection in the MAT-proximal region as expected in a region without recombination
329 associated with the mating-type locus maintained at a frequency close to 1/2. Tajima's D
330 values in the MAT-proximal region pooling all sequences (median D=3.0 per 50 kb window)
331 were higher than in the rest of the genome (median D=-0.14; Wilcoxon test p-value < 2e-16
332 with Bonferroni correction; Fig 4G; Fig S6D). In contrast, Tajima's D was significantly lower
333 than the rest of the genome within each pool of haplotype, including in the MAT-Prox2
334 haplotype, suggestive of positive selection (Wilcoxon test p-value = 0.0037 for the MAT-
335 Prox1 haplotype; p-value = 0.0022 for the MAT-Prox2 haplotype with Bonferroni correction;
336 Fig 4H-I; Fig S6D; Table S4). We found no difference in population diversity statistics
337 between the recombining part of the mating-type contig and the other contigs (Table S4A).
338
339 The study of synonymous divergence (d_S) between M1400 and M6697 haplotypes suggests
340 that the recombination suppression is at least 1.5 million years old and that there is no pattern
341 of evolutionary strata within the MAT-proximal region, i.e. segments with different levels of
342 differentiation between haplotypes that would indicate stepwise expansion of recombination
343 suppression away from the mating-type locus. Per-gene synonymous divergence (d_S) is
344 typically used for detecting gradual expansion of recombination cessation and for estimating
345 its age, as it is considered a good proxy for the time since recombination suppression. Indeed,
346 when recombination is suppressed, mutations accumulate independently in the two non-
347 recombining haplotypes. We plotted the per-gene synonymous divergence (d_S) between
348 M1400 and M6697 along the M1400 genome, as its mating-type contig is likely being
349 assembled as a full chromosome (Fig S7A&E; Table S4). Consistent with recombination
350 suppression, we found significantly higher d_S values in the MAT-proximal region (computed
351 for 71 genes) than in the other regions of the mating-type contig (pairwise Wilcoxon test p-
352 value <2e-16 with Bonferroni correction) and other contigs (pairwise Wilcoxon test p-value
353 <2e-16 with Bonferroni correction). We found no significant differences between the other
354 regions of the mating-type contig and other contigs (pairwise Wilcoxon test p-value = 0.23
355 with Bonferroni correction). The d_S pattern displayed no indication of gradual expansion of
356 recombination cessation within the MAT-proximal region, as there was no stair-like pattern.

357 Note that the per-gene d_S could not be computed for the genes with a copy not predicted in
358 one of the haplotypes and therefore mostly in the oldest, most degenerated regions (see
359 below). Using synonymous substitution rate estimates in fungi (Kasuga et al. 2002; Taylor
360 and Berbee 2006) and the mean d_S values across genes of the MAT-proximal region, we
361 estimated the age of the divergence between the two haplotypes to be at least 1.5 Million
362 years when considering the mean d_S values of all genes in the MAT-proximal region (d_S
363 mean= 0.0495) (Table S4).
364
365 ***Enrichment in transposable elements in the non-recombining region and inversion***
366 ***breakpoints***
367 When studying the M1400 and M6697 high-qualities assemblies, we detected an enrichment
368 in transposable elements (TEs) in the MAT-proximal region in the two haplotypes compared
369 to the rest of the genome. The percentage of bp occupied by TEs (TE load) was higher than
370 50% in the MAT-proximal region, while it was only 9% on average in other regions. TE load
371 in the MAT proximal region was higher in the M6697 haplotype (76%, MAT-1; MAT-Prox1)
372 than the M1400 haplotype (68% MAT-2; MAT-Prox2). Class I retrotransposons with Gypsy
373 (LTR-Ty3) and LARD elements were the most abundant TEs in all genomic regions
374 (autosomes, MAT-Proximal region and the rest of the mating-type contig), representing >70%
375 of TE annotations (Fig 5A-B), as shown previously (Demené et al. 2022). The TE load
376 however varied significantly among TE families and genomic regions (ANOVA; Table S5).
377 In M6697 for example, TIR elements appeared less abundant in the non-recombining than
378 recombining regions while LTR-Ty3 elements were more frequent in the MAT-proximal
379 region (Fig 5B). In M1400, LARD elements were more frequent in the MAT-proximal region
380 than in recombining regions (Fig 5B).
381
382 The presence of transposable elements at the two inversion breakpoints in the two genomes
383 (Fig 5A) suggests that the inversion may have occurred via non-homologous recombination
384 mediated by these elements. By dating retrotransposon insertions using between-copy
385 divergence and within-copy LTR divergence in both M1400 and M6697 genomes (Fig 5C
386 and D), we found that TEs within the inversion and around the inversion breakpoints were
387 older than TEs farther from the breakpoints. This is expected in non-recombining regions as
388 selection is there less efficient to purge TE insertions so that they remain there longer
389 (Duhamel et al. 2023). The date estimates of LTR element insertions (Fig 5D) further
390 indicated that the TE accumulation in the MAT-proximal region was as old as a few million

391 years, confirming the age estimated based on sequence divergence between MAT-proximal
392 haplotypes.
393
394 ***Gene functions, pseudogenization and degeneration in the non-recombining region***
395 Recombination suppression could allow linking genes with mating-type antagonistic effects to
396 the mating-type locus, or link together co-adapted alleles, for example for parasitizing
397 different host genotypes. Out of 127 genes functionally annotated in the MAT-proximal
398 region of the EP155 genome, 50 genes were supported by RNAseq data acquired *in vitro* for
399 this strain (Chun et al. 2020), suggesting that the region contain functionally active genes.
400 Among these 50 genes, nine were upregulated during barrage allorecognition of the EP155
401 strain with a compatible strain (Belov et al. 2021) and five were upregulated during infection
402 of the EP155 strain with the hypovirus CHV1 (Chun et al. 2020), suggesting that the MAT-
403 Proximal region may have a role in host-parasite interactions and vegetative incompatibility.
404 We found no gene function known to be involved in mating or mating compatibility, nor in
405 pathogenicity, in the MAT-proximal region (Table S6). We nevertheless identified genes
406 potentially coding for proteins with kinase domain or proteins with homeodomains (Table
407 S6), that are often involved in the regulation of developmental pathways.
408
409 By comparing gene predictions in the M6697 and M1400 newly assembled genomes, we
410 identified a higher prevalence of gene disruptions (i.e. genes predicted in a single of the two
411 haplotypes) in the MAT-proximal region than elsewhere in the genomes. Out of 175 groups
412 of single-copy orthologous genes in the MAT-proximal region, 80 (46%) were only present in
413 M1400 and 24 (14%) only present in M6697. In other genomic regions, only 3% of genes
414 were only present in one genome. This suggests that gene disruption occurred in the non-
415 recombining region, in higher proportions than elsewhere in the genome (Fisher's exact test
416 p-value < 2.2e-16), and especially in the M6697 genome (MAT-1; MAT-Prox1). The
417 enrichment in TEs and in gene disruptions in the MAT-Proximal region explains the lower
418 gene density in this region (Fig 4) and supports the inference of an old full recombination
419 suppression. The occurrence of high TE load and gene degeneration in both MAT-Proximal
420 haplotypes further confirms the lack of homozygotes in this region, in which recombination
421 could purge deleterious mutations, or indicates that recombination suppression occurs even in
422 homozygotes, although it remains to be investigated.
423

424 We found overall higher non-synonymous substitution rate compared to the baseline
425 substitution rates (d_N/d_S values) between the M1400 and M6697 genomes in the MAT-
426 proximal region (median per gene d_N/d_S = 0.80) than in other contigs (d_N/d_S = 0.43; pairwise
427 Wilcoxon test p-value = 1.7e-07 with Bonferroni correction; Fig S7A; Table S4). This is
428 consistent with relaxed purifying selection due to recombination suppression in the MAT-
429 proximal region. In addition, 13 genes displayed $d_N/d_S > 1$, suggesting that they may evolve
430 under positive selection.
431
432 ***Recombination suppression in the MAT-Proximal region in other populations assessed***
433 ***from high-quality genome assemblies***
434 We analyzed five additional high-quality genome assemblies for strains from the Asian native
435 range (CL2 and CL4 clusters) and the invasion range (including two North American strains
436 belonging to the CL1 cluster and one European strain introduced directly from Asia and
437 belonging to the CL2 cluster), three of which were generated for the present study (Table S3).
438 Most of the high-quality genomes displayed the same MAT-Proximal chromosomal
439 arrangement as M1400 (MAT-Prox2, Swiss population and belonging to the CL1 cluster),
440 suggesting that it is the ancestral state (Fig S8A-D), although the incomplete assemblies may
441 prevent detecting other rearrangements. The inversion detected in M6697 (MAT-Prox1, Swiss
442 population and belonging to CL1 cluster) was also present in the EP155 strain (MAT-Prox1,
443 North America and belonging to CL1 cluster, Crouch et al, 2020; Fig S8D), indicating that
444 this inversion was already segregating in early invasive populations in North America. In
445 other genomic regions outside of the MAT-Proximal region, all analyzed high-quality
446 genomes were colinear to the ESM15 reference genome, except the highly rearranged EP155
447 genome, as previously reported (Demené et al. 2022).
448
449 The MAT-Proximal region in all the high-quality genomes, from Asian native strains and
450 invasive strains, displayed signs of degeneration, with a high TE load and gene disruptions, as
451 seen from lower gene densities and missing orthologs (Fig6; Fig S8), confirming the ancient
452 age of recombination suppression and its occurrence in the native range. Breaks of synteny
453 between genome assemblies were associated with presence or absence of transposable
454 elements and genes (Fig S8). Study of d_S and d_N/d_S between the MRC10 strain, belonging to
455 the European clonal lineages of direct Asian origin (CL2 cluster), the M6697, M1400 and the
456 two Asian native strains further confirmed recombination suppression in the native range and
457 in invasive strains from different origins (North America or directly from Asia). The d_S and

458 d_N/d_S values were indeed significantly higher in the MAT-proximal region than other
459 genomic regions (Fig S7 B-E). Mean d_S values in the MAT-proximal region were in the same
460 range as d_S values between the M6697 and M1400 genomes, suggesting a similar date of
461 recombination suppression and, therefore, that recombination suppression was ancestral. The
462 MAT-proximal region appeared smaller in the two Asian strains XIM9508 and ESM15 (<1
463 Mb) than in invasive strains (Table S3), but this may be due to the incomplete genome
464 assemblies.
465
466 ***Presence of the two MAT-Proximal haplotypes in other populations***
467 By analyzing available additional Illumina genomes in the native range and in invasive
468 populations, with less dense sampling within populations, we recovered the two haplotypes in
469 the MAT-proximal region, i.e., high differentiation between two clusters in the MAT-
470 Proximal region, contrasting with a lack of genetic subdivision in other genomic regions or a
471 completely different genetic subdivision, corresponding to previously described population
472 genetic structures (Figs 7 and FigS9). In particular, the neighbor-net network with multiple
473 strains sequenced previously from the CL2 and CL3 clusters (from the introduced range and
474 the native range, South Korean and Japan) and from the CL4 cluster (from the native range,
475 China;) reinforces the view of balancing selection on the MAT-Proximal haplotypes.
476 Sequences from the CL1 and CL2 clusters were indeed intermingled in the MAT-proximal
477 region (Fig 7A), while the structure in the rest of the genome instead corresponded to the
478 different population clusters, i.e., CL1, CL2, CL3 and CL4 (Fig 7B).
479
480 However, the two haplotypes were less differentiated in the native range and in invasive
481 populations originating directly from Asia than in invasive populations from the first invasion
482 wave via North America. For example, branch lengths on the neighbor-net indicate a lower
483 differentiation between haplotypes within CL2 (Fig 7A) than within the European CL1 or
484 Swiss invasive populations (Fig 2D; Fig S3). Some reticulations were present between MAT-
485 proximal haplotypes in CL2, which may again correspond to incomplete lineage sorting, gene
486 conversion or rare recombination events.
487
488 We found no association of the MAT-Proximal haplotypes with mating types at the world
489 scale (Figs 7 and S9), which may be due to the lack of extensive sampling within local
490 populations. As the MAT-Proximal region is only partially linked to the mating-type locus,
491 founder effects may yield different associations between MAT-Proximal haplotypes and

492 mating types in different populations. Only in the S12 European invasive clonal lineage (only
493 MAT-1), we found a single MAT-Proximal haplotype (MAT-Prox1; Fig S9B). In two
494 invasive European lineages of American origin with a predominantly clonal structure (re092
495 and re103), the strains with a mating type different from their clone-mates due to localized
496 introgression from other clonal lineages (Demené et al. 2019) also had their MAT-Proximal
497 region introgressed (Fig S9C); this finding reinforces the view of the strong linkage between
498 the mating-type locus and the MAT-Proximal region, and of an advantage of maintaining an
499 association between MAT-proximal haplotypes and mating types within populations. The
500 three 2019 Swiss strains that exhibited an intermediate sequence in the MAT-Proximal region
501 clustered with strains from Europe with a direct Asian origin (Fig S9C). This may represent a
502 third MAT-Proximal haplotype, associated with the second invasion in Europe, directly from
503 Asia.
504
505 ***No evidence of introgression in the MAT-proximal region from other species***
506 The stronger differentiation in the MAT-Proximal region in the Swiss and CL1 invasive
507 populations than in other populations could also be due to an introgression event from a
508 closely related species; however, the small genetic distance and the reticulations observed in
509 the neighbor-net network between the M6697 Swiss reference genome and the Asian
510 genomes M8444 and XA19 rather indicate that the MAT-Prox1 haplotype in Switzerland
511 originated from Asian *C. parasitica* populations, likely a population different from those at
512 the origin of the CL2 invasive cluster (Fig 7).
513
514 Gene genealogies also provided no evidence of introgression in the MAT-proximal region
515 from closely related *Cryphonectria* species. We performed a gene orthology analysis
516 including three genomes of the closely related species *C. japonica*, C. *carpinicola*, and *C.*
517 *naterciae* that were previously sequenced with short read technologies (Stauber et al. 2020).
518 We retrieved only nine genes from the MAT proximal region that were shared between the
519 M6697 and M1400 genomes that had orthologs in one of the related species. This may be due
520 to low quality assemblies in the outgroups (Stauber et al. 2020) and/or a region without
521 recombination also in the outgroups with gene losses and many repeats. We found no gene
522 with a topology consistent with introgression from another species into the MAT-Proximal
523 region, i.e. with a MAT-Prox1 or MAT-Prox2 haplotype that would branch with an outgroup
524 allele rather than with the alternative haplotype of *C. parasitica*. The high load of TEs and
525 gene disruptions in the two MAT-Proximal haplotypes does not fit with the introgression

526 hypothesis either, unless the introgression occurred from a species with recombination
527 suppression in the MAT-Proximal region.
528
529 **Discussion**
530 ***A large and million-year-old non-recombining region partially linked to the mating-type***
531 ***locus in* Cryphonectria parasitica**
532 We found strong evidence for the complete cessation of recombination in a region of more
533 than 1 Mb (1.2 Mb to 1.5 Mb depending on the haplotype), strongly but not completely linked
534 to the mating-type locus in *C. parasitica*. Indeed, we detected maximal levels of linkage
535 disequilibrium in otherwise recombining populations and the existence of two highly
536 differentiated haplotypes, without shared polymorphism, while even low rates of
537 recombination can homogenize alleles and prevent LD build-up (Dufresnes et al. 2015). The
538 full cessation of recombination was further supported by the existence of a previously
539 unknown, large rearrangement (an inversion) in the center of the non-recombining region in
540 invasive strains. The higher TE load in this region constitutes additional evidence for full
541 recombination cessation, as the lack of recombination renders selection against TE
542 accumulation less efficient, in particular due to Muller's ratchet: two chromosomes with
543 different TE insertions cannot recombine to produce a chromosome free of TE insertions. TE
544 takes hundreds of thousands of years to accumulate in non-recombining regions (Duhamel et
545 al. 2023), so that the high TE load in the MAT-Proximal region in *C. parasitica* indicates old
546 recombination cessation. The particularly high frequency of missing genes and non-
547 synonymous substitutions in the MAT-Proximal region worldwide further indicates
548 degeneration and ancient recombination suppression. The estimates for the age of
549 recombination suppression based on the divergence between haplotypes and on TE insertions
550 were in fact at least a few million years. However, in contrast to other known non-
551 recombining regions on sex or mating-type chromosomes (Bergero and Charlesworth 2009;
552 Hartmann et al. 2021), the region without recombination was not completely linked to the
553 mating-type locus. Indeed, the occurrence of rare recombination events between the non-
554 recombining MAT-Proximal region and the mating-type locus are supported by previous
555 segregation analyses (Kubisiak and Milgroom 2006b) and the lack of association found here
556 between the MAT-Proximal haplotypes and the mating types at the worldwide scale, as well
557 as their incomplete association within populations.
558
559 ***Large inversion in the MAT-proximal non-recombining region***

560 We found a large inversion between the two highly differentiated MAT-proximal haplotypes
561 in invasive populations originating from North America (CL1 genetic cluster, Swiss 1990 and
562 2019 populations and US populations). The chromosomal arrangement of the M1400 MAT-
563 Prox2 haplotype is likely the ancestral state as it is the most frequent worldwide, although a
564 more complete sampling in the native range is required to obtain definitive conclusion. The
565 inversion would therefore be the M6697 MAT-Prox1 haplotype and may contribute to
566 recombination suppression. The non-recombining region was defined based on LD pattern
567 and was larger than the inversion. This may indicate that additional proximal mechanisms
568 suppress recombination. The inversion may even be a consequence rather than a cause of
569 recombination cessation, as previously found in other fungal mating-type chromosomes
570 (Grognet et al. 2014; Sun et al. 2017; Branco et al. 2018; Vittorelli et al. 2022). Other
571 proximal causes of recombination cessation may for example be genetic recombination
572 modifiers, transposable elements, epigenetic marks, histone modifications or lack of
573 recruitment of proteins responsible for double-strand breaks (Maloisel and Rossignol 1998;
574 Boideau et al. 2022; Legrand et al. 2024). The lack of rearrangements at the edge of the
575 MAT-proximal region may allow rare recombination or gene conversion events, which may
576 explain the presence of two distinct LD blocks with lower LD between them and the lower
577 differentiation at the edges of the MAT-proximal region. Alternatively, the LD around the
578 inversion may be formed by less frequent recombination as recombination is often modified
579 around inversion breakpoints (Pegueroles et al. 2010; Stevison et al. 2011) and the actual size
580 of the non-recombining region may be smaller than suggested from the LD pattern.
581
582 ***The MAT-proximal non-recombining region is million-year-old and degenerated***
583 Following recombination suppression, selection efficacy decreases. We detected footprints of
584 degeneration, with particularly high occurrence of gene disruptions and of TE insertions in the
585 MAT-proximal region, both in Europe, North America and in the native Asian range of the
586 fungus. A previous study (Demené et al. 2022) had already reported an enrichment of LTR-
587 Ty3 (*Gypsy*) element in the region near the mating-type locus in *C. parasitica*. We found
588 LTR-Ty3 elements at the inversion breakpoints and an enrichment in the MAT-Prox1
589 haplotype. Interestingly, the same TE family (LTR-Ty3) seems to have expanded in the non-
590 recombining regions in other fungi with recombination suppression on their mating-type
591 chromosomes (Duhamel et al. 2023), suggesting they may be more prone to accumulate in
592 non-recombining regions. Accumulations of transposable elements have also been detected in
593 other mating-type chromosomes, sex chromosomes and supergenes (Bachtrog 2013; Fontana

594 et al. 2020; Carpentier et al. 2022; Duhamel et al. 2023); some other large supergenes do not
595 seem to exhibit TE load, likely due to recent introgression, gene conversion or double
596 crossovers (Harringmeyer and Hoekstra 2022; Huang et al. 2022; Matschiner et al. 2022; Hill
597 et al. 2023).
598
599 The transposable element accumulation in the MAT-proximal region and the enrichment in
600 gene disruptions support the estimated age of recombination suppression at a few million
601 years ago based on mean pairwise gene sequence divergence between haplotypes. The finding
602 of the two haplotypes and of the inversion in the two invaded continents, Europe and North
603 America, as well as the occurrence of degeneration footprints in the MAT-Proximal region in
604 the native Asian range, further support the inference of an ancient recombination suppression
605 maintained by balancing selection. Indeed, even rare recombination events allow purging
606 deleterious mutations and TE insertions. This is in agreement with previous genetic maps
607 reporting lack of recombination near the mating-type locus in crosses involving Japanese,
608 North American and Italian isolates (Kubisiak and Milgroom 2006a). The high TE load and
609 degeneration level in terms of gene disruption in strains from Asian and North American
610 origins indicate a worldwide recombination cessation, so likely ancestral to all these native
611 and invasive populations. The strong degeneration, and in both haplotypes, further indicates
612 that even crosses between strains carrying the same MAT-Proximal haplotype does not allow
613 recombination in this region, which may be due to recombination suppression in the MAT-
614 Proximal region regardless of the haplotype, i.e., not being due to the inversion.
615
616 The differentiation between the MAT-Proximal haplotypes appeared greater in the CL1
617 cluster and Swiss populations (from the first invasion event, from North America) than in the
618 other clusters, which may be due to a founder effect or the lack of sampling in the source
619 population in Asia. The MAT-Prox1 haplotype in the populations from the first invasion
620 event from North America shows an inversion, a very low diversity, and the strongest
621 footprints of degeneration. Neighbornet networks show some reticulations between and
622 within haplotypes in some populations (e.g. CL2 that include the invasive strains from Asia,
623 but also the Swiss 2019 populations), and a few, rare intermediates between the two MAT-
624 Proximal haplotypes, which is likely due to incomplete lineage sorting, or, alternatively, to
625 rare recombination or gene conversion events, without leading to haplotype homogenization,
626 likely due to lower fitness of intermediates. These findings point to a strong balancing
627 selection maintaining the two haplotypes differentiated and non-recombining, and perhaps

628 different selection regimes or successive selection events during the various invasion waves
629 of *C. parasitica.*
630
631 The higher differentiation between the MAT-Proximal haplotypes in the first invasion wave
632 in Europe could also have resulted from introgression from a closely related species. The
633 evolution of non-recombining regions by introgression has been reported for example in ants
634 and butterflies (Jay et al. 2018; Helleu et al. 2022; Stolle et al. 2022). However, we found
635 little evidence for an introgression: i) the two haplotypes are present in the CL2 cluster of
636 invasive European populations directly originating from Asia, ii) the MAT-Prox1 haplotype
637 in the CL1 cluster more differentiated from the MAT-Prox2 haplotype than in the CL2 cluster
638 is genetically close to an Asian haplotype, iii) footprints of recombination suppression are
639 present in the native range in terms of degeneration and particular genomic structure in the
640 MAT-Proximal region, and iv) we found no signatures of introgression from closely related
641 species in gene genealogies.
642
643 ***The selective forces potentially maintaining the two haplotypes in the MAT-proximal***
644 ***region***
645 The genes found missing in the non-recombining region in one or the other haplotype may
646 help maintaining the two haplotypes by selection due to the sheltering of combinations of
647 deleterious mutations in repulsion, i.e. thanks to a heterozygote advantage called pseudo-
648 overdominance (Abu-Awad and Waller 2023). Such pseudo-overdominance advantage may
649 only be a consequence of recombination suppression, and contribute to its maintenance, or
650 may even be its initial cause (Branco et al. 2017; Jay et al. 2021; Jay et al. 2022; Charlesworth
651 2023; Jay, D.L. Jeffries, Hartmann, et al. 2024). Indeed, this situation of a non-recombining
652 region partially linked to the mating-type locus and with two haplotypes maintained by
653 selection fits predictions from a recent theoretical model showing that recombination
654 suppression can be selected for near permanently heterozygous alleles for protecting
655 haplotypes with fewer deleterious mutations than average in the population, while sheltering
656 their few recessive deleterious alleles. Importantly, the model predicts that selection can also
657 occur for regions partially linked to heterozygous alleles (Jay et al. 2022). Similar pseudo-
658 overdominance effect due to the sheltering of deleterious mutations can lead to balancing
659 selection in the context of local adaptation (Jay, Aubier, et al. 2024). These models require the
660 existence of a substantial diploid or dikaryotic phase for sheltering recessive deleterious
661 alleles in a heterozygous state, although the mechanism also works with a substantial haploid

662 phase (Jay et al. 2022). In fact, although *C. parasitica* has a predominant haploid phase,
663 strains can be found as dikaryons heterozygous for the mating-type locus in some natural
664 populations, and not only as monokaryons (McGuire et al. 2004; McGuire et al. 2005;
665 Stauber et al. 2021; Stauber et al. 2022). The importance and frequency of heterokaryons in
666 *C. parasitica* in nature would deserve further investigations. It may be sufficient if recessive
667 deleterious alleles are sheltered in ca 20% of individuals for selecting recombination
668 suppression (Jay et al. 2022), and/or if the genes involved are preferentially expressed during
669 the diploid or dikaryotic phases. However, it is challenging to assess whether the degree of
670 linkage between the mating-type locus and the MAT-Proximal region and the frequency of
671 dikaryotic stages are sufficient to fit the model. This hypothesis of balancing selection, and in
672 particular the hypothesis of overdominance (i.e. heterozygote advantage), does not require the
673 same association between the MAT-Proximal haplotypes and mating types across all
674 populations worldwide, it is sufficient that the MAT-Proximal haplotypes are strongly
675 associated to mating types within populations. Further sampling will be required to test
676 whether the MAT-Proximal haplotypes are associated to mating types within local
677 populations in the CL3 and CL4 clusters and in the native Asian range.
678
679 It could indeed also be that genes evolve under overdominance in this region for other reasons
680 than deleterious mutations, so that a partial linkage to a permanently heterozygous locus is
681 beneficial, perhaps in relationship with the pathogenic lifestyle of the fungus. One could
682 imagine, for example, that heterozygosity could be advantageous at genes involved in
683 virulence against the tree host, and especially the novel hosts colonized in invasive ranges, or
684 in resistance against a parasitic virus known to negatively affect fitness in *C. parasitica* (Choi
685 and Nuss 1992; Brusini et al. 2017). As a matter of fact, multiple genes present in the MAT-
686 Proximal region are up-regulated under infection by the hypovirus CHV1 or during the
687 vegetative incompatibility reaction considered to play a role in the prevention of virus
688 transmission (Choi et al. 2012; Rigling and Prospero 2018). As an alternative to a
689 heterozygote advantage, the selection maintaining the recombination suppression and the two
690 haplotypes in the MAT-proximal region could be some kind of negative-frequency dependent
691 selection of beneficial allelic combinations, possibly linked to a trench-warfare-like arms race
692 with the host tree, the virus or the microbial community, where partial linkage to the mating-
693 type locus would help maintenance in balanced proportions and therefore would prevent allele
694 loss (Tellier et al. 2014; Jay, Aubier, et al. 2024). Such a role in host-pathogen interactions
695 would also be consistent with the selective sweep footprints detected in the MAT-proximal

696 region in a previous study (Stauber et al. 2021), if there is recurrent positive selection for
697 improving the efficiency of the pathogen weapons within each of one of the two MAT-
698 proximal haplotypes. The detection of the selective sweep may have, however, been impacted
699 by particular demographic effect during the invasion (Moinet et al. 2022). Balancing selection
700 and location adjacent to a non-recombining region was recently found for loci involved in
701 host resistance in the *Daphnia–Pasteuria* system (Fredericksen et al. 2023). The MAT-
702 Proximal region may in this case even include genes involved different traits under balancing
703 selection, and would then constitute a supergene (Schwander et al. 2014). The transposable
704 elements insertion polymorphism observed from the high quality assemblies may also be
705 adaptive (Casacuberta and González 2013; Orteu et al. 2024).
706
707 Another hypothesis for explaining such extension of recombination suppression beyond the
708 mating-type locus is antagonistic selection, i.e. linkage of alleles that would improve fitness
709 of a MAT-1 gamete while being deleterious in a MAT-2 gamete, or vice-versa. However, full
710 recombination suppression with the mating-type locus would be expected under such
711 antagonistic selection, as well as the same association between the MAT-Proximal alleles and
712 the mating-type in all populations, in contrast to our findings. In addition, we found no
713 particular predicted function in the MAT-proximal region that could be related to mating
714 compatibility and there are very little functions, if any, with possible antagonistic functions
715 between mating types in fungi (Bazzicalupo et al. 2019; Hartmann et al. 2021).
716
717 **Conclusion**
718 In conclusion, we provide strong evidence for the existence of a non-recombining region
719 partially linked to the mating-type locus and maintained by selection in the chestnut blight
720 fungus *C. parasitica*. The non-recombining region in *C. parasitica* displayed contrasting
721 differentiation levels between haplotypes and different chromosomal arrangements in native
722 and invasive populations. We found footprints of balancing selection in the MAT-proximal
723 region in the first introduction of the pathogen in Europe from North America and a
724 chromosomal inversion. The non-recombining region displayed lower differentiation between
725 haplotypes in Asia and in the second invasion wave directly from Asia. This non-recombining
726 region may thus underlie important adaptive traits and thereby provide important applications
727 for the control of a devastating tree pathogen. In addition, the high-quality genome assemblies
728 provided here, from the native and introduced ranges, will more generally be useful for
729 studies aiming at understanding the evolution of this invasive and damaging pathogen.

730

731 **Material and Methods**

732

733 **Strains and genomic data**

734 For population genomic analyses, we studied a collection of 386 monokaryotic *C. parasitica*
735 strains sampled worldwide, from the native and introduced ranges of the pathogen, and
736 sequenced previously with the short-read Illumina technology (strain information are
737 presented in Table S1; (Demené et al. 2019; Stauber et al. 2021; Stauber et al. 2022)). Data
738 were downloaded from NCBI Bioproject numbers PRJNA604575, PRJNA644891 and
739 PRJNA706885. We focused our analyses first on European invasive strains originating from
740 North America. We studied 88 strains belonging to the CL1 genetic cluster, in central and
741 southeastern Europe (Stauber et al. 2021). We excluded the putative clonal genotypes
742 previously identified (Stauber et al. 2021). We also studied 71 strains sampled in the 1990 and
743 62 strains sampled in 2019 in southern Switzerland (Stauber et al. 2022). Mating-type ratios
744 close to ½ and population structure analyses of the CL1 genetic cluster and the Swiss 1990
745 population suggest frequent recombination in these populations (Stauber et al. 2021; Stauber
746 et al. 2022). The mating-type ratio was 1/3 *MAT-1* - 2/3 *MAT-2* in the 2019 Swiss population
747 and population structure analyses suggested regular sexual reproduction and recent population
748 bottleneck. To study more broadly the presence of the two haplotypes in the MAT-proximal
749 region, we analyzed monokaryotic strains belonging to the CL2, CL3, CL4 genetic clusters,
750 as well as the S12 European invasive lineage (Stauber et al. 2021) and additional
751 monokaryotic strains from the US and Europe (with an Asian or North American origin
752 (Demené et al. 2019)). We excluded heterokaryotic strains, i.e. strains having both MAT-1
753 and MAT-2 alleles, as the phase was challenging to infer.

754

755 For mapping and SNP calling, we first used as reference the 43.9 Mb genome sequence
756 EP155 v2.0 of the strain EP155 (*MAT-2,* North America, CL1 cluster) available at the Joint
757 Genome Institute (http://jgi.doe.gov/)(Crouch et al. 2020). For comparative genomics, we
758 used the published genome of the strain ESM15 (MAT-2, Japan, CL2 cluster; (Demené et al.
759 2022)) available at DDBJ/ENA/GenBank on the bioproject PRJNA700518 under the
760 accession JAGDFO000000000. We additionally sequenced *de novo,* with long-read
761 technologies, five strains from the native and introduced ranges of the pathogen. We
762 sequenced with PacBio Hifi the genomes of the strains M1400 (MAT-2) and M6697 (MAT-
763 1) sampled in Gnosca, southern Switzerland (Stauber et al. 2021). Mycelia were stored as

764 glycerol stocks at -80 C after strain isolation. To produce mycelium for DNA extraction,
765 isolates were inoculated onto cellophane-covered potato dextrose agar plates (PDA, 39 g/L;
766 BD Becton, Dickinson and Company, Franklin Lakes, USA) (Hoegger et al. 2000) and
767 incubated for a minimum of 1 week at 24°C, at a 14 hr light and 10 hr darkness cycle.
768 Mycelium and spores were harvested by scratching the mycelial mass off the cellophane,
769 transferring it into 2 mL tubes and freeze-drying it for 24 hr (Stauber et al. 2021). DNA
770 extraction was performed with the NucleoBond High Molecular Weight DNA kit from
771 Macherey-Nagel, with the mechanical disruption of about 30 mg of lyophilized mycelium
772 with two beads of 3 mm for 5 min at 30 Hz. Sequencing was outsourced to Novogene, the
773 Netherlands. We additionally sequenced with Oxford Nanopore MinION technology the
774 strains XIM9508 (China, CL4 cluster, MAT-1), MRC10 (South Western France introduced
775 directly from Asia, CL2 cluster, MAT-2) and DUM-005 (USA, CL1 cluster, MAT-2). These
776 isolates had been collected for previous studies (Milgroom et al. 1996; Liu et al. 2003; Dutech
777 et al. 2012). The protocols for mycelium production, DNA extraction and sequencing for
778 these three strains were the same as in (Demené et al. 2022).
779
780 **Short-read data processing and SNP calling**
781 We used SNP calling datasets of the genomes from monokaryotic strains of the CL1, CL2,
782 CL3, CL4 genetic clusters, the S12 invasive lineage and the Swiss populations against the
783 EP155 reference v2.0 genome obtained in (Stauber et al. 2021; Stauber et al. 2022). We
784 performed mapping and raw SNP calling of short-read data using the M1400 new genome
785 assembly as a reference as described in (Stauber et al. 2021; Stauber et al. 2022). Briefly, we
786 trimmed reads with Trimmomatic v0.36 (Bolger et al. 2014) and aligned them with Bowtie 2
787 v2.3.5.1 (Langmead et al. 2009) and SAMtools v1.9 (Li et al. 2009) to the EP155 v2.0
788 genome. Raw SNP calling and filtration for quality were conducted with the genome analysis
789 toolkit GATK v3.8 and v4.0.2.0 (McKenna et al. 2010). We used the filtration parameters
790 described in (Stauber et al. 2021): QUAL>=100, MQRankSum (lower) >= -2.0, QD:_20.0,
791 MQRankSum (upper) <=2.0, MQ:_20.0, BaseQRankSum (lower) >=-2.0, ReadPosRankSum
792 (lower) >=-2.0, ReadPosRankSum (upper) <=2.0, BaseQRankSum (upper) <=2.0. We further
793 removed SNPs overlapping with transposable elements predicted *de novo* in the EP155 v2.0
794 genome by (Stauber et al. 2021).
795
796 **Population genomics analyses**

797 For all population genomics analyses, we excluded SNPs with missing data in more than 10%
798 of the strains and kept only polymorphic strains with vcftools v0.1.16 (Danecek et al. 2011).
799 To study linkage disequilibrium, we further excluded rare variants (minor allele frequency
800 <0.1) with the vcftools (Danecek et al. 2011) option --maf 0.1. We computed LD with the --
801 hap-r2 option of vcftools (Danecek et al. 2011) for each scaffold and each population
802 separately. We used the --thin 50000 option of vcftools (Danecek et al. 2011) to sample SNPs
803 distant of at least 50 kb. We used the R package LDheatmap v1.0-6 (Shin et al. 2006) to plot
804 LD r^2 values among SNP pairs. To perform principal component analyses (PCAs), we first
805 used vcftools (Danecek et al. 2011) to convert VCF format files in Plink format. We then used
806 the Plink v1.90b5.3 (Purcell et al. 2007) –pca command to run PCA analysis. We used the R
807 package PopGenome v2.7.5 (Pfeifer et al. 2014) to compute nucleotide diversity, Tajima's D
808 values and the F_{ST} index in 50 kb window overlapping over 10 kbp. Windows containing
809 fewer than 5 SNPs were removed from the analysis. We used the R package ggplot2 (2_3.5.0)
810 to plot results. Pairwise Wilcoxon tests were performed in R with Bonferroni or false
811 discovery rate correction.
812
813 We built neighbornet networks using SplitsTree4 v 4.19.2 (Huson 1998). VCF file were
814 converted to nexus format using PGDSpider v1.0 tool (Lischer and Excoffier 2012). For the
815 study of other populations, we included only one strain of each haplotype of the invasive
816 European population but not all CL1 and Swiss strains for network readability; we
817 nevertheless checked that the results were the same with all CL1 and Swiss strains.
818
819 **Long-read based assemblies**
820 PacBio Hifi reads of strains M1400 and M6697 were both assembled using canu v1.8 (Koren
821 et al. 2017) program with a set genome size of 44 Mb. Multiple assembly pipelines were used
822 for the other strains. Oxford Nanopore MinION reads and Illumina reads of the genomes of
823 both XIM9508 and MRC10 strains were assembled using HybridSPAdes (Antipov et al.
824 2016). Assemblies were manually curated and scaffolds were cut when an evidence of a
825 chimeric connection was detected (i.e. mis-paired short reads) as previously described
826 (Demené et al. 2022). For DUM005, the assembly was generated by Ra with basic parameters
827 that uses Oxford Nanopore MinION reads and corrects the assembly with Illumina reads
828 (https://github.com/lbcb-sci/ra) as it outperformed the HybridSPAdes assembly. As the
829 HybridSpades assembly of the MRC10 strain suggested absence of collinearity with the
830 MAT-proximal region M1400 and M6697, we further checked the assembly of the MAT-

proximal region of MRC010 by generating a meta-assembly of this strain. We used the assembler tool canu v2.2 with an estimated genome size of 42 Mb. We also used Flye v2.9.3-b1797 (Kolmogorov et al. 2019) with an estimated genome size of 42 Mb, --nano-raw for reads with an error rate below 20 and a coverage for initial disjointig assembly of 50. Then we used ragtag (Alonge et al. 2019) to patch the canu assembly with the Flye assembly as query in a first loop. In a second loop, the canu assembly was patched with the first loop ragtag assembly. Finally, we polished this second output assembly of ragtag with short reads and the consensus part of medaka v1.11.3 with no change in the parameters (https://github.com/nanoporetech/medaka). Assemblies statistics were obtained with quast v5.1 (Gurevich et al. 2013). We assessed the completeness of each assembly using the Benchmarking of Universal Single-Copy Orthologue (BUSCO) tools with the Sordariomyceta ortholog set (sordariomycetes_odb10, 2020-08-05, n = 3817 orthologous groups searched) (Manni et al. 2021). Gene models were predicted with the Helixer v0.3.1 pipeline (Holst et al. 2023). We also run Helixer pipeline for ESM15 and EP155 strains for the gene orthology analysis. Statistics of the obtained gene annotation was obtained with the Agat v1.0.0 tool (Dainat et al. 2020). Transposable elements of all genomes were annotated using Repeatmasker v4-0-7 (Smit et al. 2013) and the customized library built for *C. parasitica* in (Demené et al. 2022) contained in the "Curated_TEs_database_ESM015_EP155.fa" file (available on the "Portail Data INRAE: Chromosomal rearrangements but no change of genes and transposable elements repertoires in an invasive forest-pathogenic fungus" at https://doi.org/10.15454/UTIB8U). The class Gipsy invader was renamed LTR-Ty3. We filtered out TE copies shorter than 100bp. We used the HybridSPAdes preliminary assembly of MRC10 for gene model annotations. To study gene functions of the EP155 genome, we used the gene annotation available at http://jgi.doe.gov/ (Crouch et al. 2020). To study support from RNAseq data, we used RNAseq data from the EP155 strain cultivated *in vitro* available at Genebank under Project ID number PRJNA588887 and accessions numbers SRR10428542, SRR10428543, SRR10428544 (Chun et al. 2020). Raw reads were mapped using STAR v2.7.10a (Dobin et al. 2013) with the settings. --alignIntronMax 1000 --limitBAMsortRAM 1629219965--quantMode GeneCounts. We used the program featureCounts (Liao et al. 2013) v2.0.6 with the options -p --countReadPairs -M - -B -O --largestOverlap. We considered a gene to be supported for read count >10.

Estimation of retrotranposon insertion time and nucleotide divergence time

865 To get an estimation of the insertion date of the transposable elements present in the MAT-
866 proximal non-recombining region, we applied two complementary methods. We first used the
867 divergence between the LTR sequences in retrotransposons, as these LTR sequences at their
868 edges are identical at the time of TE insertion and then diverge with time. For this, we used a
869 *de-novo* prediction of LTR retrotransposons using LTRharvest GenomeTools 1.6.2
870 (Ellinghaus et al. 2008; Gremme et al. 2013) and LTR_Finder v1.07 (Xu and Wang 2007). To
871 prepare data for LTRharvest, we first created enhanced suffix for the M1400 and M6697
872 genome assemblies using the GenomeTools Suffixerator (-dna -suf -lcp -bwt -bck -mirrored
873 parameters). We ran LTRharvest using two pipelines, designed to identify LTR
874 retrotransposons with a minimum and maximum LTR length of 100 bp and 7000 bp
875 respectively and at least 85% identity between two LTR regions, with and without canonical
876 TGCA motifs, respectively: i) -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -similar
877 85 -vic 10 -seed 20 -motif TGCA -motifmis 1; ii) -minlentltr 100, -maxlenltr 7000, -mintsd 4,
878 -maxtsd 6, -similar 85, -vic 10, -seed 20. Similarly, we ran LTR_Finder on the M1400 and
879 M6697 genome assemblies to retrotransposons with both TGCA and non-TGCA motifs and a
880 minimum and maximum LTR length of 100 bp and 7000 bp respectively and at least 85%
881 identity between two LTR regions (-D 15000, -d 1000, -l 100, -L 7000, -p 20, -C, -M 0.85).
882 Finally, we used LTR_retriever v2.8.5 (Ou and Jiang 2018) with default parameters to filter
883 out false positive LTR candidates identified by LTRharvest and LTR_Finder and get an
884 estimation of each element insertion date.
885
886 As a second method to estimate the insertion date of the transposable elements present in the
887 MAT-proximal non-recombining region, we used the set of curated consensus sequences from
888 (Demené et al. 2022) to annotate the inversion sequence or its surroundings. We first used
889 samtools faidx v1.9 (Li et al. 2009) to extract the sequence corresponding to inversion and
890 their 50 kb surroundings in M1400 and M6697 genome assemblies. We annotated the
891 sequences corresponding to the inversions and the concatenated 50 kb surroundings in both
892 isolates using RepeatMasker version 4.1.5 and rmblast as search engine (v2.10.0) with -no_is
893 -pa 20 -cutoff 250 -a parameters. Finally, we parsed the RepeatMasker .out file using the
894 helped script parseRM_merge_interrupted.pl and omitting Simple_repeat and
895 Low_complexity regions (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs). We
896 then built a summary of the alignments using the RepeatMasker helper script
897 buildSummary.pl and calculated sequence divergence using the calcDivergenceFromAlign.pl
898 script to finally render results with the createRepeatLandscape.pl from the same helper suite.

899

900

901 **Comparative genomics analyses**

902 Genome synteny between long reads assemblies were studied using the nucmer v3.1 program
903 (Marçais et al. 2018). Outputs were plotted with the R programs ggplot2 (Wickham 2009),
904 genoPlotR (Guy et al. 2010) and RIdeogram (Hao et al. 2020). The dotplot in the putative
905 centromere region of the M1400 mating-type contig was performed using the online
906 megablast alignment tool available at https://blast.ncbi.nlm.nih.gov/(last accessed 13th May
907 2024).

908

909 To study gene disruption, synonymous and non-synonymous divergence (d_S and d_N) and
910 testing introgression in the MAT-proximal region, we build orthology relationships for genes
911 of the genome assemblies of *C. parasitica* strains M1400, M6697, XIM9508, MRC10, DUM-
912 005, ESM15 and EP155 and included three genomes of the closely related species
913 *Cryphonectria japonica* (IF-6), *Cryphonectria carpinicola* (CS3), and *Cryphonectria*
914 *naterciae* (M3664). Genome, gene annotation and species tree of these closely related species
915 were previously published by (Stauber et al. 2020). Genome data were retrieved from NCBI
916 bioproject number PRJNA644891 and accession IDs JACWRX000000000 for IF-6,
917 JACWRQ000000000 for CS3 and JACWST000000000 for M3664. We run OrthoFinder
918 v2.3.7 (Emms and Kelly 2019) analysis on protein sequences. We used the translatorX v1.1
919 program (Abascal et al. 2010) with default parameters that use a codon-based approach to
920 align orthologous gene coding sequences. To compute d_S and d_N vaues of one-to-one
921 orthologs between pairs of genome assemblies, we use the yn00 program of PAML (Yang
922 2007). Estimation of divergence time between haplotypes was performed using computed
923 gene d_S values and the formula Tgenerations = dS/2μ. We used previous estimates of
924 substitution rates in fungi (Kasuga et al. 2002; Taylor and Berbee 2006) and considered that
925 *C. parasitica* undergoes one generation a year (Guerin et al. 2001). We build gene coding
926 sequences trees with the outgroup genomes in the MAT-proximal region using iqtree2
927 v2.2.2.6 (Minh et al. 2020) with 1000 bootstraps and used the Newick Utilities
928 (https://github.com/tjunier/newick_utils) for displaying phylogenetics tree.

929

930 **References**

931 Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide
932 sequences guided by amino acid translations. *Nucleic Acids Res.* 38:W7-13.

933 Abu-Awad D, Waller D. 2023. Conditions for maintaining and eroding pseudo-
934 overdominance and its contribution to inbreeding depression. *Peer Community J.*
935 [Internet] 3. Available from:
936 https://peercommunityjournal.org/articles/10.24072/pcjournal.224/

937 Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB,
938 Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft
939 genomes. *Genome Biol.* 20:224.

940 Anagnostakis SL. 1987. Chestnut blight: The classical problem of an introduced pathogen.
941 *Mycologia* [Internet]. Available from:
942 https://www.tandfonline.com/doi/abs/10.1080/00275514.1987.12025367

943 Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for
944 hybrid assembly of short and long reads. *Bioinformatics* 32:1009–1015.

945 Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-
946 chromosome degeneration. *Nat. Rev. Genet.* 14:113–124.

947 Bazzicalupo AL, Carpentier F, Otto SP, Giraud T. 2019. Little evidence of antagonistic
948 selection in the evolutionary strata of fungal mating-type chromosomes
949 (*Microbotryum lychnidis-dioicae*). *G3 Genes Genomes Genet.* 9:1987–1998.

950 Belov AA, Witte TE, Overy DP, Smith ML. 2021. Transcriptome analysis implicates
951 secondary metabolite production, redox reactions, and programmed cell death during
952 allorecognition in Cryphonectria parasitica. *G3 GenesGenomesGenetics* 11:jkaa021.

953 Berdan EL, Blanckaert A, Butlin RK, Flatt T, Slotte T, Wielstra B. 2022. Mutation
954 accumulation opposes polymorphism: supergenes and the curious case of balanced
955 lethals. *Philos. Trans. R. Soc. B Biol. Sci.* 377:20210199.

956 Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex
957 chromosomes. *Trends Ecol. Evol.* 24:94–102.

958 Boideau F, Richard G, Coriton O, Huteau V, Belser C, Deniot G, Eber F, Falentin C, Ferreira
959 de Carvalho J, Gilet M, et al. 2022. Epigenomic and structural events preclude
960 recombination in Brassica napus. *New Phytol.* 234:545–559.

961 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
962 sequence data. *Bioinformatics* 30:2114–2120.

963 Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza
964 S, Brandenburg J-T, Coelho MA, Hood ME, et al. 2017. Evolutionary strata on young
965 mating-type chromosomes despite the lack of sexual antagonism. *Proc. Natl. Acad.*
966 *Sci.* 114:7067–7072.

967 Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, Prieur SL, Coelho
968 MA, Vienne DM de, Hartmann FE, Begerow D, et al. 2018. Multiple convergent
969 supergene evolution events in mating-type chromosomes. *Nat. Commun.* 9:2000.

970 Brusini J, Wayne ML, Franc A, Robin C. 2017. The impact of parasitism on resource
971 allocation in a fungal host: the case of *Cryphonectria parasitica* and its mycovirus,
972 Cryphonectria Hypovirus 1. *Ecol. Evol.* 7:5967–5976.

973 Carpentier F, Rodríguez de la Vega RC, Jay P, Duhamel M, Shykoff JA, Perlin MH, Wallen
974 RM, Hood ME, Giraud T. 2022. Tempo of degeneration across independently evolved
975 nonrecombining regions. *Mol. Biol. Evol.* 39:msac060.

976 Casacuberta E, González J. 2013. The impact of transposable elements in environmental
977 adaptation. *Mol. Ecol.* 22:1503–1517.

978 Charlesworth B. 2023. The fitness consequences of genetic divergence between polymorphic
979 gene arrangements. *Genetics*:iyad218.

980 Charlesworth D. 2016. The status of supergenes in the 21st century: recombination
981 suppression in Batesian mimicry and sex chromosomes and other complex
982 adaptations. *Evol. Appl.* 9:74–90.

983 Charlesworth D, Vekemans X, Castric V, Glémin S. 2005. Plant self-incompatibility systems:
984 a molecular evolutionary perspective. *New Phytol.* 168:61–69.

985 Choi GH, Dawe AL, Churbanov A, Smith ML, Milgroom MG, Nuss DL. 2012. Molecular
986 characterization of vegetative incompatibility genes that restrict hypovirus
987 transmission in the chestnut blight fungus *Cryphonectria parasitica*. *Genetics*
988 190:113–127.

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

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

995 Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, Hillman BI, Milgroom
996 MG, Pangilinan J, Smith M, et al. 2020. Genome sequence of the chestnut blight
997 fungus Cryphonectria parasitica EP155: A fundamental resource for an archetypical
998 invasive plant pathogen. *Phytopathology®* 110:1180–1188.

999 Dainat J, Hereñú D, Pucholt P. 2020. AGAT:Another Gff Analysis Toolkit to handle
1000 annotations in any GTF. *GFF Format*.

1001 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter
1002 G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools.
1003 *Bioinformatics* 27:2156–2158.

1004 Demené A, Laurent B, Cros-Arteil S, Boury C, Dutech C. 2022. Chromosomal
1005 rearrangements with stable repertoires of genes and transposable elements in an
1006 invasive forest-pathogenic fungus. *Peer Community J.* [Internet] 2. Available from:
1007 https://peercommunityjournal.org/articles/10.24072/pcjournal.127/

1008 Demené A, Legrand L, Gouzy J, Debuchy R, Saint-Jean G, Fabreguettes O, Dutech C. 2019.
1009 Whole-genome sequencing reveals recent and frequent genetic recombination between
1010 clonal lineages of Cryphonectria parasitica in western Europe. *Fungal Genet. Biol.*
1011 130:122–133.

1012 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
1013 Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*
1014 29:15–21.

1015 Dufresnes C, Borzée A, Horn A, Stöck M, Ostini M, Sermier R, Wassef J, Litvinchuck SN,
1016 Kosch TA, Waldman B, et al. 2015. Sex-chromosome homomorphy in palearctic tree
1017 frogs results from both turnovers and X-Y recombination. *Mol. Biol. Evol.* 32:2328–
1018 2337.

1019 Duhamel M, Hood ME, Vega RCR de la, Giraud T. 2023. Dynamics of transposable element
1020 accumulation in the non-recombining regions of mating-type chromosomes in anther-
1021 smut fungi. :2022.08.03.502670. Available from:
1022 https://www.biorxiv.org/content/10.1101/2022.08.03.502670v2

1023 Dussert Y, Legrand L, Mazet ID, Couture C, Piron M-C, Serre R-F, Bouchez O, Mestre P,
1024 Toffolatti SL, Giraud T, et al. 2020. Identification of the first oomycete mating-type
1025 locus sequence in the grapevine downy mildew pathogen, *Plasmopara viticola*. *Curr.*
1026 *Biol.* 30:3897-3907.e4.

1027 Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V. 2012. The chestnut blight
1028 fungus world tour: successive introduction events from diverse origins in an invasive
1029 plant fungal pathogen. *Mol. Ecol.* 21:3931–3946.

1030 Dutech C, Fabreguettes O, Capdevielle X, Robin C. 2010. Multiple introductions of divergent
1031 genetic lineages in an invasive fungal pathogen, Cryphonectria parasitica, in France.
1032 *Heredity* 105:220–228.

1033 Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for
1034 de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.

1035 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
1036 genomics. *Genome Biol.* 20:238.

1037 Fontana S, Chang N-C, Chang T, Lee C-C, Dang V-D, Wang J. 2020. The fire ant social
1038 supergene is characterized by extensive gene and transposable element copy number
1039 variation. *Mol. Ecol.* 29:105–120.

1040 Fredericksen M, Fields PD, Pasquier LD, Ricci V, Ebert D. 2023. QTL study reveals
1041 candidate genes underlying host resistance in a Red Queen model system. *PLOS*
1042 *Genet.* 19:e1010570.

1043 Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank
1044 JE. 2020. Sex chromosome evolution: So many exceptions to the rules. *Genome Biol.*
1045 *Evol.*

1046 Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A Comprehensive Software Library
1047 for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans.*
1048 *Comput. Biol. Bioinform.* 10:645–656.

1049 Grognet P, Bidard F, Kuchly C, Tong LCH, Coppin E, Benkhali JA, Couloux A, Wincker P,
1050 Debuchy R, Silar P. 2014. Maintaining two mating types: structure of the mating type
1051 locus and its role in heterokaryosis in *Podospora anserina*. *Genetics* 197:421–432.

1052 Guerin L, Froidefond G, Xu X-M. 2001. Seasonal patterns of dispersal of ascospores of
1053 Cryphonectria parasitica (chestnut blight). *Plant Pathol.* 50:717–724.

1054 Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for
1055 genome assemblies. *Bioinformatics* 29:1072–1075.

1056 Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome
1057 visualization in R. *Bioinformatics* 26:2334–2335.

1058 Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. RIdeogram: drawing SVG
1059 graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.*
1060 6:e251.

1061 Harringmeyer OS, Hoekstra HE. 2022. Chromosomal inversion polymorphisms shape the
1062 genomic landscape of deer mice. *Nat. Ecol. Evol.* 6:1965–1979.

1063 Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, Malagnac F,
1064 Grognet P, Giraud T. 2021. Recombination suppression and evolutionary strata around
1065 mating-type loci in fungi: documenting patterns and understanding evolutionary and
1066 mechanistic causes. *New Phytol.* 229:2470–2491.

1067 Helleu Q, Roux C, Ross KG, Keller L. 2022. Radiation and hybridization underpin the spread
1068 of the fire ant social supergene. *Proc. Natl. Acad. Sci.* 119:e2201040119.

1069 Hill J, Enbody ED, Bi H, Lamichhaney S, Lei W, Chen J, Wei C, Liu Y, Schwochow D,
1070 Younis S, et al. 2023. Low mutation load in a supergene underpinning alternative male
1071 mating strategies in ruff (*Calidris pugnax*). *Mol. Biol. Evol.* 40:msad224.

1072 Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. 2000. Genetic structure of newly
1073 established populations of *Cryphonectria parasitica*. *Mycol. Res.* 104:1108–1116.

1074 Holst F, Bolger A, Günther C, Maß J, Triesch S, Kindel F, Kiel N, Saadat N, Ebenhöh O,
1075 Usadel B, et al. 2023. Helixer–de novo prediction of primary Eukaryotic gene models
1076 combining deep learning and a hidden markov model. :2023.02.06.527280. Available
1077 from: https://www.biorxiv.org/content/10.1101/2023.02.06.527280v2

1078 Huang K, Ostevik KL, Elphinstone C, Todesco M, Bercovich N, Owens GL, Rieseberg LH.
1079 2022. Mutation load in sunflower inversions is negatively correlated with inversion
1080 heterozygosity.Takahashi A, editor. *Mol. Biol. Evol.* 39:msac101.

1081 Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinforma. Oxf.*
1082 *Engl.* 14:68–73.

1083 Ironside JE. 2010. No amicable divorce? Challenging the notion that sexual antagonism
1084 drives sex chromosome evolution. *BioEssays* 32:718–726.

1085 Jay P, Aubier TG, Joron M. 2024. The interplay of local adaptation and gene flow may lead to
1086 the formation of supergenes. *Mol. Ecol.*:e17297.

1087 Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M. 2021.
1088 Mutation load at a mimicry supergene sheds new light on the evolution of inversion
1089 polymorphisms. *Nat. Genet.* 53:288–293.

1090 Jay P, Jeffries D, Hartmann FE, Véber A, Giraud T. 2024. Why do sex chromosomes
1091 progressively lose recombination? *Trends Genet.* [Internet]. Available from:
1092 https://www.sciencedirect.com/science/article/pii/S0168952524000672

1093 Jay P, Jeffries DL, Hartmann FE, Véber A, Giraud T. 2024. Why do sex chromosomes
1094 progressively lose recombination? *Trends in Genetics*:in press.

1095 Jay P, Tezenas E, Véber A, Giraud T. 2022. Sheltering of deleterious mutations explains the
1096 stepwise extension of recombination suppression on sex chromosomes and other
1097 supergenes. *PLOS Biol.* 20:e3001698.

1098 Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J, Dasmahapatra
1099 KK, Joron M. 2018. Supergene evolution triggered by the introgression of a
1100 chromosomal inversion. *Curr. Biol.* 28:1839-1845.e3.

1101 Kasuga T, White TJ, Taylor JW. 2002. Estimation of nucleotide substitution rates in
1102 eurotiomycete fungi. *Mol. Biol. Evol.* 19:2318–2324.

1103 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using
1104 repeat graphs. *Nat. Biotechnol.* 37:540–546.

1105 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable
1106 and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
1107 *Genome Res.* 27:722–736.

1108 Kubisiak TL, Milgroom MG. 2006a. Markers linked to vegetative incompatibility (vic) genes
1109 and a region of high heterogeneity and reduced recombination near the mating type
1110 locus (MAT) in *Cryphonectria parasitica*. *Fungal Genet. Biol.* 43:453–463.

1111 Kubisiak TL, Milgroom MG. 2006b. Markers linked to vegetative incompatibility (*vic*) genes
1112 and a region of high heterogeneity and reduced recombination near the mating type
1113 locus (*MAT*) in *Cryphonectria parasitica*. *Fungal Genet. Biol. FG B* 43:453–463.

1114 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
1115 alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

1116 Legrand S, Saifudeen A, Bordelet H, Vernerey J, Guille A, Bignaud A, Thierry A, Acquaviva
1117 L, Gaudin M, Sanchez A, et al. 2024. Absence of chromosome axis protein
1118 recruitment prevents meiotic recombination chromosome-wide in the budding yeast
1119 Lachancea kluyveri. *Proc. Natl. Acad. Sci.* 121:e2312820121.

1120 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
1121 R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
1122 25:2078–2079.

1123 Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read
1124 mapping by seed-and-vote. *Nucleic Acids Res.* 41:e108–e108.

1125 Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for
1126 connecting population genetics and genomics programs. *Bioinformatics* 28:298–299.

1127 Liu YC, Linder-Basso D, Hillman BI, Kaneko S, Milgroom MG. 2003. Evidence for
1128 interspecies transmission of viruses in natural populations of filamentous fungi in the
1129 genus Cryphonectria. *Mol. Ecol.* 12:1619–1628.

1130 Lovat C-A, Donnelly DJ. 2019. Mechanisms and metabolomics of the host–pathogen
1131 interactions between Chestnut (Castanea species) and Chestnut blight (Cryphonectria
1132 parasitica). *For. Pathol.* 49:e12562.

1133 Maloisel L, Rossignol JL. 1998. Suppression of crossing-over by DNA methylation in
1134 *Ascobolus*. *Genes Dev.* 12:1381–1389.

1135 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: Novel
1136 and streamlined workflows along with broader and deeper phylogenetic coverage for
1137 scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38:4647–4654.

1138 Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4:
1139 A fast and versatile genome alignment system. *PLOS Comput. Biol.* 14:e1005944.

1140 Matschiner M, Barth JMI, Tørresen OK, Star B, Baalsrud HT, Brieuc MSO, Pampoulie C,
1141 Bradbury I, Jakobsen KS, Jentoft S. 2022. Supergene origin and maintenance in
1142 Atlantic cod. *Nat. Ecol. Evol.* 6:469–481.

1143 McGuire IC, Davis JE, Double ML, MacDonald WL, Rauscher JT, McCawley S, Milgroom
1144 MG. 2005. Heterokaryon formation and parasexual recombination between
1145 vegetatively incompatible lineages in a population of the chestnut blight fungus,
1146 *Cryphonectria parasitica*. *Mol. Ecol.* 14:3657–3669.

1147 McGuire IC, Marra RE, Milgroom MG. 2004. Mating-type heterokaryosis and selfing in
1148 *Cryphonectria parasitica*. *Fungal Genet. Biol.* 41:521–533.

1149 McGuire IC, Marra RE, Turgeon BG, Milgroom MG. 2001. Analysis of mating-Type genes
1150 in the chestnut blight fungus, *Cryphonectria parasitica*. *Fungal Genet. Biol.* 34:131–
1151 144.

1152 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
1153 Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a
1154 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
1155 *Res.* 20:1297–1303.

1156 Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. 2008. The mating-type chromosome
1157 in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early
1158 evolution of sex chromosomes. *PLOS Genet.* 4:e1000030.

1159 Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneko S. 1996. Intercontinental population
1160 structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* 88:179–
1161 190.

1162 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A,
1163 Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic
1164 inference in the genomic era. *Mol. Biol. Evol.* 37:1530–1534.

1165 Moinet A, Schlichta F, Peischl S, Excoffier L. 2022. Strong neutral sweeps occurring during a
1166 population contraction. *Genetics* 220:iyac021.

1167 Orteu A, Kucka M, Gordon IJ, Ng'iru I, van der Heijden ESM, Talavera G, Warren IA,
1168 Collins S, ffrench-Constant RH, Martins DJ, et al. 2024. Transposable element
1169 insertions are associated with Batesian mimicry in the pantropical butterfly
1170 *Hypolimnas misippus*. *Mol. Biol. Evol.* 41:msae041.

1171 Otto SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. *Nat. Rev.*
1172 *Genet.* 3:252–261.

1173 Ou S, Jiang N. 2018. LTR_retriever: A highly accurate and sensitive program for
1174 identification of long terminal repeat retrotransposons. *Plant Physiol.* 176:1410–1422.

1175 Pegueroles C, Ordóñez V, Mestres F, Pascual M. 2010. Recombination and selection in the
1176 maintenance of the adaptive value of inversions. *J. Evol. Biol.* 23:2709–2717.

1177 Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient
1178 Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31:1929–
1179 1936.

1180 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P,
1181 de Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome
1182 association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.

1183 Rigling D, Prospero S. 2018. Cryphonectria parasitica, the causal agent of chestnut blight:
1184 invasion history, population biology and disease control. *Mol. Plant Pathol.* 19:7–20.

1185 Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr. Biol.*
1186 24:R288–R294.

1187 Shin J-H, Blay S, McNeney B, Graham J, others. 2006. LDheatmap: an R function for
1188 graphical display of pairwise linkage disequilibria between single nucleotide
1189 polymorphisms. *J. Stat. Softw.* 16:1–10.

1190 Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available at:
1191 http://www.repeatmasker.org Accessed: November 30, 2017.

1192 Stauber L, Badet T, Feurtey A, Prospero S, Croll D. 2021. Emergence and diversification of a
1193 highly invasive chestnut pathogen lineage across southeastern Europe.Castric V,
1194 Weigel D, Laine A-L, Castric V, editors. *eLife* 10:e56279.

1195 Stauber L, Croll D, Prospero S. 2022. Temporal changes in pathogen diversity in a perennial
1196 plant–pathogen–hyperparasite system. *Mol. Ecol.* 31:2073–2088.

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* 5:10.1128/msphere.00737-20.

Stevison LS, Hoehn KB, Noor MAF. 2011. Effects of Inversions on Within- and Between-Species Recombination and Divergence. *Genome Biol. Evol.* 3:830–841.

Stolle E, Pracana R, López-Osorio F, Priebe MK, Hernández GL, Castillo-Carrillo C, Arias MC, Paris CI, Bollazzi M, Priyam A, et al. 2022. Recurring adaptive introgression of a supergene variant that determines social organization. *Nat. Commun.* 13:1180.

Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. 2017. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat. Commun.* 8:1–8.

Taylor JW, Berbee ML. 2006. Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* 98:838–849.

Tellier A, Moreno-Gámez S, Stephan W. 2014. Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. *Evolution* 68:2211–2224.

Umen JG. 2011. Evolution of sex and mating loci: An expanded view from Volvocine algae. *Curr. Opin. Microbiol.* 14:634–641.

Vittorelli N, Hartmann, Giraud T. 2022. Stepwise recombination suppression around the mating-type locus in the fungus *Schizothecium tetrasporum* (Ascomycota, Sordariales).

Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer Science & Business Media

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35:W265–W268.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Acknowledgments

This work was supported by the European Research Council (ERC) EvolSexChrom (832352) grant and a Fondation Louis D grant from the Institut de France to TG, and by the ANR PIA grant # ANR-20-IDEES-0002 to F.E.H. We acknowledge the GenOuest bioinformatics core facility (https://www.genouest.org) for providing the computing infrastructure, (GPU) for genome annotation". The authors declare that they have no competing interests.

1233 TG and FEH conceptualized the study and acquired funding. TG supervised the study. AL, AS, CD,
1234 AD, SP, DC, LS and TB contributed to data and their availability. AL and AS performed *in vitro*
1235 culture and DNA extraction. FEH, RRDLV, QR, JPV, LS and TB analyzed genomes. FEH and TG
1236 wrote the original draft. All authors edited the manuscript.
1237
1238 The raw data and the new assemblies produced in this study will be published pending scientific
1239 review.
1240
1241 **Tables**
1242 **Table 1: Distribution of mating types (*MAT-1* and *MAT-2*) among non-recombining**
1243 **haplotypes of the MAT-proximal region in the European invasive *Cryphonectria***
1244 ***parasitica* CL1 genetic cluster and the Swiss populations.** MAT-Prox1 and MAT-Prox2
1245 haplotypes were defined based on the clusters of the principal component analysis.
1246
1247 **Figures**
1248
1249 **Figure 1: Linkage disequilibrium (LD) analysis along the contig carrying the mating-**
1250 **type locus in a *Cryphonectria parasitica* European invasive population.** LD heatmaps
1251 using single nucleotide polymorphisms (SNPs; n=3815) located on the contig carrying the
1252 mating-type locus (scaffold_2 of the EP155 genome) in the CL1 genetic cluster (European
1253 invasive population introduced from North America); pairs of SNPs with high LD, i.e. r^2
1254 >0.9, correspond to the red triangle. The mating-type locus location is shown with a green
1255 triangle and the MAT-proximal region lacking recombination is shown with a red arrow. The
1256 two high-LD blocks within the MAT-proximal region are shown with orange arrows. SNPs at
1257 the limit of the MAT-proximal region and the two high LD blocks were manually highlighted
1258 with red dotted lines and an orange line.
1259
1260 **Figure 2: Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-**
1261 **proximal region lacking recombination and other genomic regions in a *Cryphonectria***
1262 ***parasitica* European invasive population. A-C.** Principal component analysis (PCA). Two
1263 principal components are presented. Percentage of variance explained by each PC is indicated
1264 into brackets. Strains are colored according to their mating type (*MAT-1* or *MAT-2*). **D-E.**
1265 Neighbor-net network from a SplitsTree analysis. These analyses were performed in the CL1
1266 genetic cluster (European invasive populations introduced from North America) based on: **A**

1267 **and D.** SNPs (n=2,220) located within the MAT-proximal region along the contig carrying
1268 the mating-type locus (scaffold_2 of the EP155 genome).; **B.** SNPs (n=1,595) located in other
1269 regions along the contig carrying the mating-type locus; **C and E** SNPs (n=11,289) located on
1270 other contigs of the EP155 genome. On panels A and D, the two clusters corresponding to the
1271 MAT-Prox1 and MAT-Prox2 haplotypes are shown with red circles. The identified haplotype
1272 of each strain is indicated in Table S1. The number of strains within the MAT-Prox1
1273 haplotype is indicated on panel A by the letter n. The newly sequenced M1400 and M6697
1274 strains are highlighted with a black rectangle
1275
1276 **Figure 3: Synteny and rearrangements between the two newly sequenced genomes of**
1277 **European invasive strains introduced from North America (M1400 and M6697) in the**
1278 **chromosome carrying the MAT-proximal region lacking recombination in**
1279 ***Cryphonectria parasitica*.** The mating-type locus is located with a green diamond. The MAT-
1280 proximal region defined from LD analyses is indicated with red arrows.
1281
1282 **Figure 4: Genetic diversity and divergence between non-recombining haplotypes of the**
1283 **MAT-proximal region along the contig carrying the mating-type locus in a**
1284 ***Cryphonectria parasitica* European invasive population.** A-B. GC content (A) and
1285 transposable element densities (B) along the contig carrying the mating-type locus
1286 (tig_0000001) of the M1400 *C. parasitica* genome. C. Relative divergence (F_{ST}) between
1287 strains of the MAT-Prox1 and MAT-Prox2 haplotypes: D-E. Nucleotide diversity within
1288 pools of strains for each MAT-Proximal haplotype; F-G-H Tajima's D for all strains pooled
1289 and within pools of strains for each MAT-Proximal haplotype; The MAT-proximal region
1290 defined from LD analyses and the inversion between M1400 and M6697 genomes are
1291 indicated with red and blue arrows respectively. The mating-type locus location is shown with
1292 a green triangle. The location of the putative centromere is indicated by a yellow circle. All
1293 population statistics were computed for the 1990 Swiss population along the mating-type
1294 contig (tig_0000001) of the M1400 genome per 50-kb window overlapping over 10 kbp.
1295 Windows containing fewer than 5 SNPs were removed from the analysis.
1296
1297 **Figure 5: Annotation of transposable elements (TEs) and estimates of their insertion**
1298 **time in the MAT-proximal region lacking recombination and other genomic regions in**
1299 ***Cryphonectria parasitica*.** A. Annotation of transposable elements (TEs) and gene density
1300 along the MAT-proximal region in M1400 and M6697 genomes. The figure only shows the

1301 genomic region from 7 Mb on tig_0000001 of the M1400 genome and from 2.8 Mb on
1302 tig_0000060 of the M6697 genome. The mating-type locus is located with a green triangle.
1303 The two high-LD blocks within the MAT-proximal region are shown with orange arrows.
1304 Synteny for 10 kb segments with identity > 90 % is shown in red and inversion in blue.
1305 Transposons larger than 5 kb are shown in purple for the transposons annotated as LTR-Ty3
1306 by and in green for other transposons. Orthologous genes shared between genomes are shown
1307 in black and unique to each each genome in grey. B. Relative percentage of each TE
1308 annotation (in %) in the non-recombining MAT-proximal region and other recombining
1309 regions of M6697 and M1400 genomes. C. Pairwise genetic distance between TE copies
1310 within the inversion and around (50 kb) in both M1400 and M6697 genomes. Distribution of
1311 the Kimura substitution levels computed using the consensus sequence for the TEs. D.
1312 Estimates of the age of intact copies of LTR within and around the inversion.
1313
1314 **Figure 6: Transposable elements content and gene content variation in seven high**
1315 **quality genome assemblies from other *Cryphonectria parasitica* populations from the**
1316 **native and the introduced range. A.** TE load (percentage of base pairs occupied by
1317 transposable elements) in the seven high quality genome assemblies in the MAT proximal
1318 region and other recombining regions. **B.** Number of orthologous genes present in the seven
1319 high quality genome assemblies. Only single copy orthologous genes shared between M6697
1320 and M1400 or present only in one this genome are shown.
1321
1322 **Figure 7: Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-**
1323 **proximal region lacking recombination (A) and other genomic regions (B) in other**
1324 **resequenced *Cryphonectria parasitica* populations from the native and the introduced**
1325 **range.** Neighbor-net network from a SplitsTree analysis based on: **A** SNPs (n=4,120) located
1326 within the MAT-proximal region along the contig carrying the mating-type locus (scaffold_2
1327 of the EP155 genome); **B** SNPs (n= 103,058) located on other contigs of the EP155 genome.
1328 Color of empty circles around strain ID indicate the genetic clusters the strains belong to
1329 (CL1, CL2, CL3, CL4; n= 33 strains**)**. Dotted circles indicate strains of the native range and
1330 plain circles indicate strains of the introduced range. Red and blue dots near strain ID indicate
1331 the mating type (*MAT-1* or *MAT-2*, respectively).
1332
1333

1334 Table 1

1335

	Population	CL1 genetic cluster		Swiss 1990 population		Swiss 2019 population	
	Haplotype	MAT-Prox1	MAT-Prox2	MAT-Prox1	MAT-Prox2	MAT-Prox1	MAT-Prox2
Number of strains in each PCA cluster	total	36	52	20	51	8	54
	MAT-1	30	11	18	17	5	15
	MAT-2	6	41	2	34	3	39

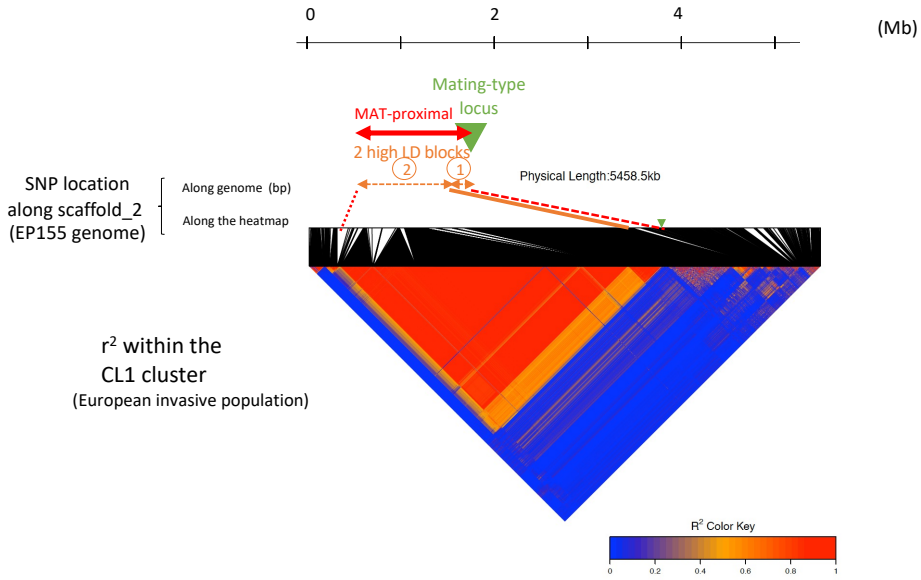
1336

1337

1338

1339 Fig 1

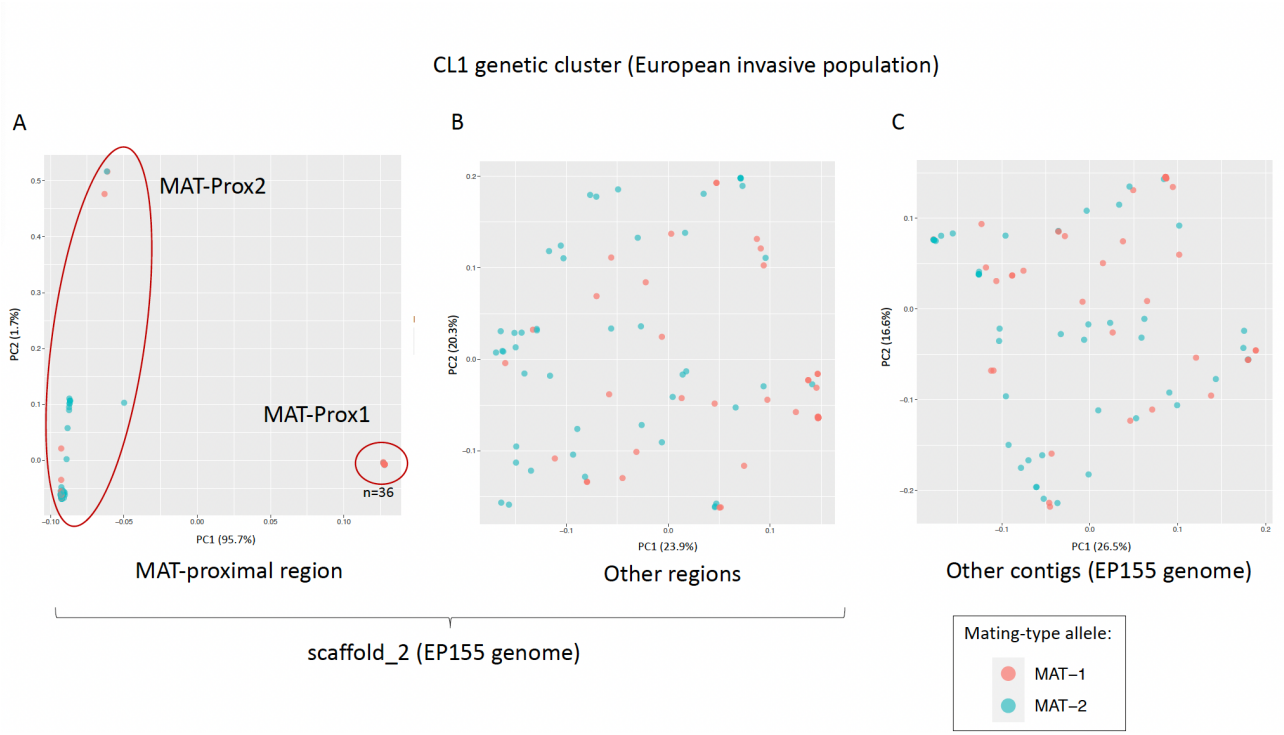
1340



1341

1342

Fig 2


CL1 genetic cluster (European invasive population)
A
B
C
MAT-Prox2
MAT-Prox1
n=36
PC1 (95.7%)
PC2 (1.7%)
PC1 (23.9%)
PC2 (20.3%)
PC1 (26.5%)
PC2 (16.6%)
MAT-proximal region
Other regions
Other contigs (EP155 genome)
scaffold_2 (EP155 genome)
Mating-type allele:
MAT−1
MAT−2

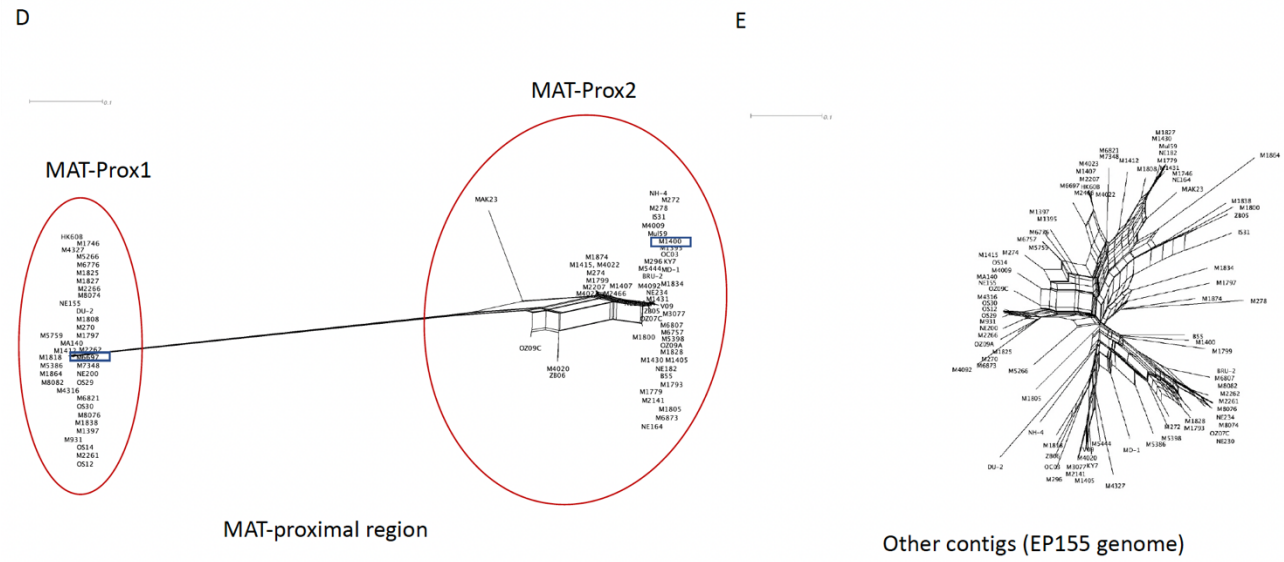

D
E
MAT-Prox2
MAT-Prox1
MAT-proximal region
Other contigs (EP155 genome)

Fig 3

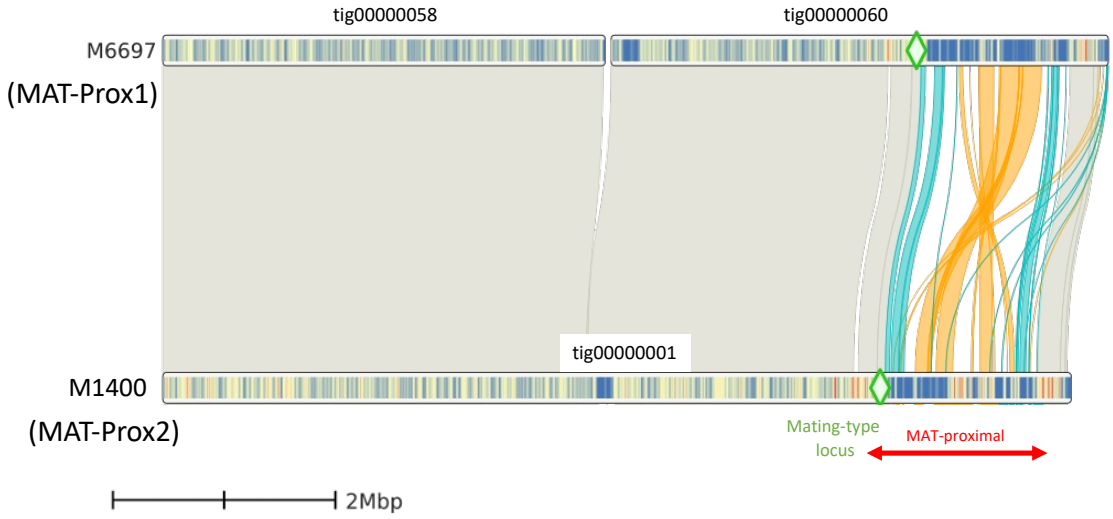

tig00000058
tig00000060
M6697
(MAT-Prox1)
tig00000001
M1400
(MAT-Prox2)
Mating-type
locus
MAT-proximal
2Mbp

Fig 4

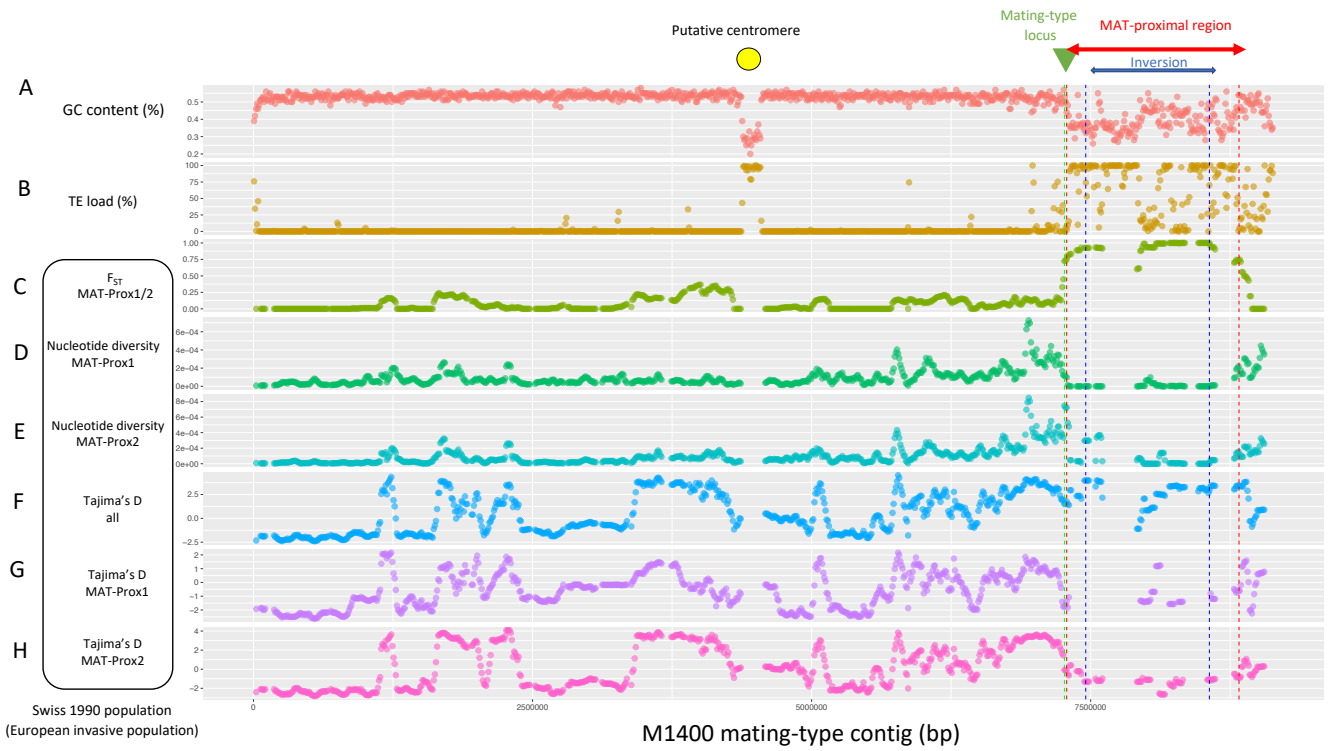

Putative centromere
Mating-type locus
MAT-proximal region
Inversion
A
GC content (%)
B
TE load (%)
C
F_{ST} MAT-Prox1/2
D
Nucleotide diversity MAT-Prox1
E
Nucleotide diversity MAT-Prox2
F
Tajima's D all
G
Tajima's D MAT-Prox1
H
Tajima's D MAT-Prox2
Swiss 1990 population (European invasive population)
M1400 mating-type contig (bp)

Fig5

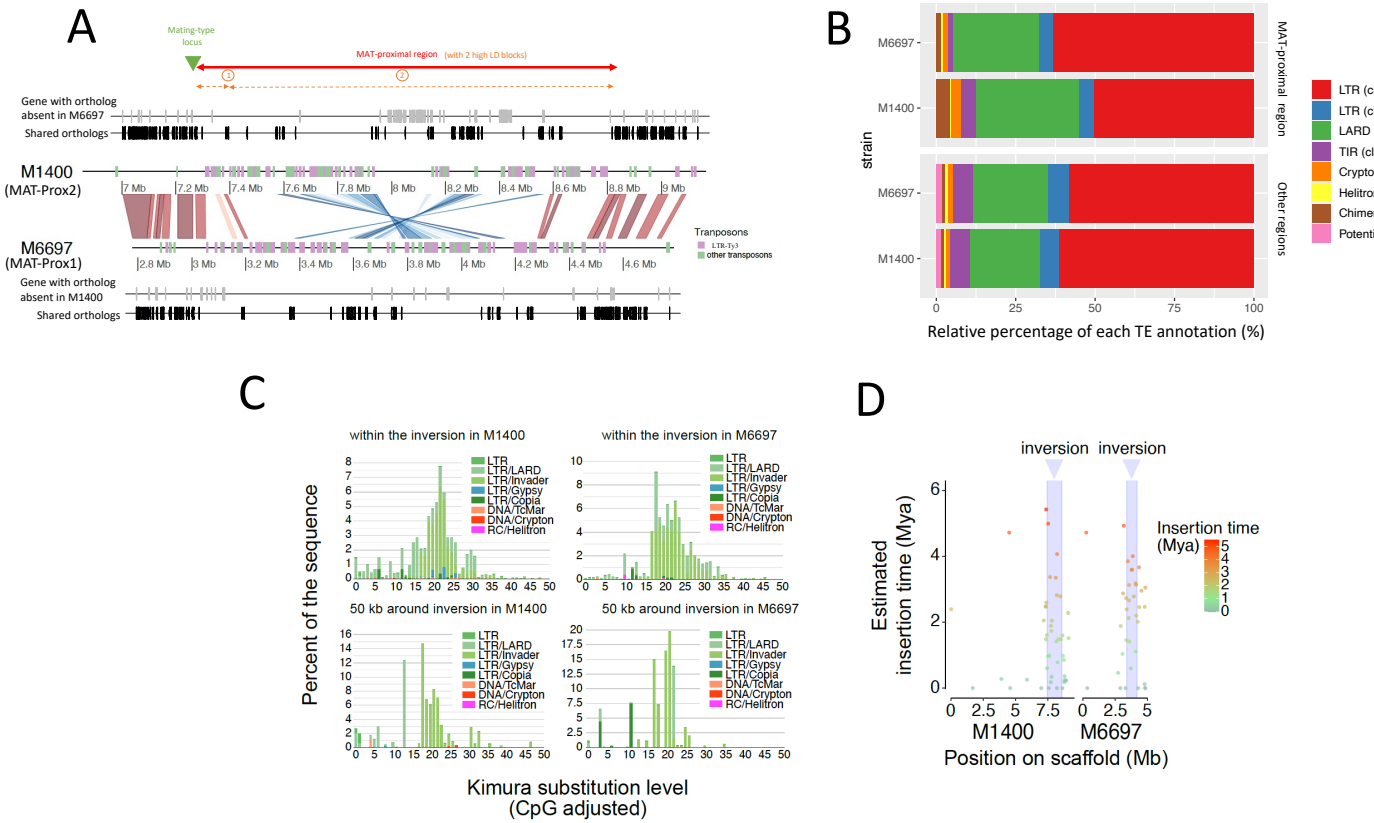

A
Mating-type locus
MAT-proximal region (with 2 high LD blocks)
Gene with ortholog absent in M6697
Shared orthologs
M1400 (MAT-Prox2)
M6697 (MAT-Prox1)
Tranposons
other transposons
Gene with ortholog absent in M1400
Shared orthologs
B
M6697
M1400
strain
MAT-proximal region
Other regions
Relative percentage of each TE annotation (%)
C
within the inversion in M1400
within the inversion in M6697
50 kb around inversion in M1400
50 kb around inversion in M6697
Percent of the sequence
LTR
LTR/LARD
LTR/Invader
LTR/Gypsy
LTR/Copia
DNA/TcMar
DNA/Crypton
RC/Helitron
Kimura substitution level (CpG adjusted)
D
inversion
inversion
Insertion time (Mya)
Estimated insertion time (Mya)
M1400
M6697
Position on scaffold (Mb)

Fig 6

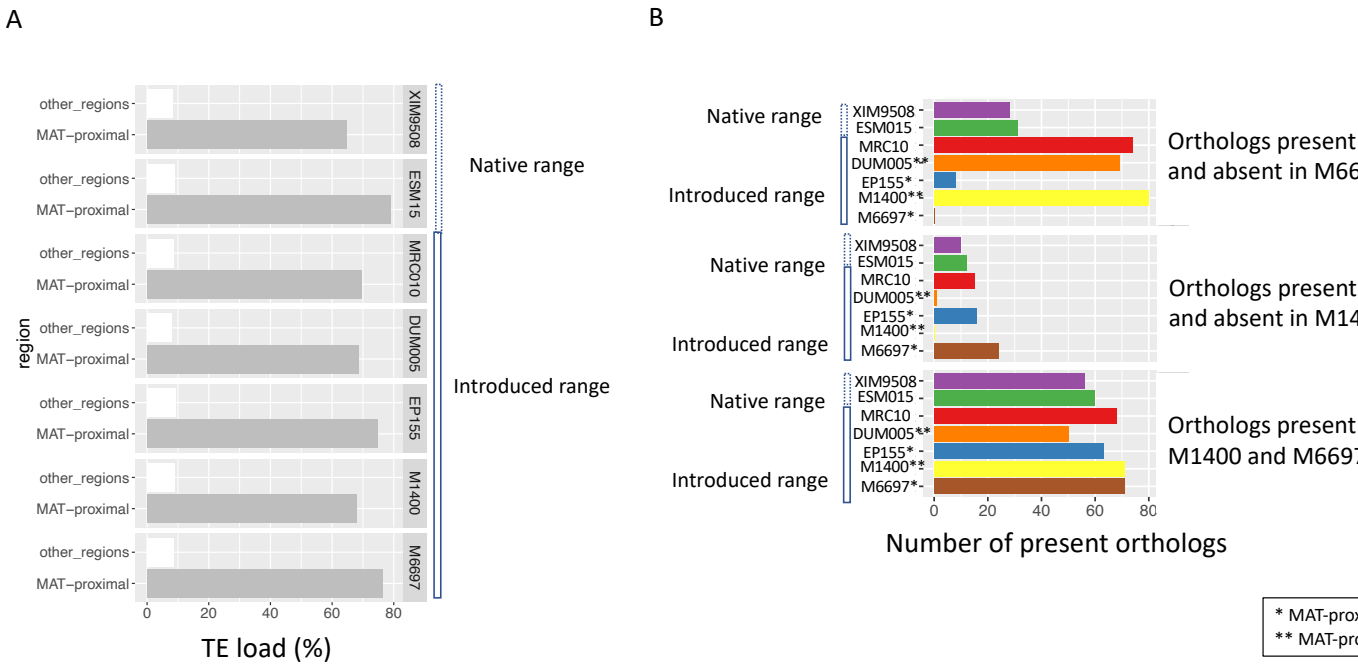
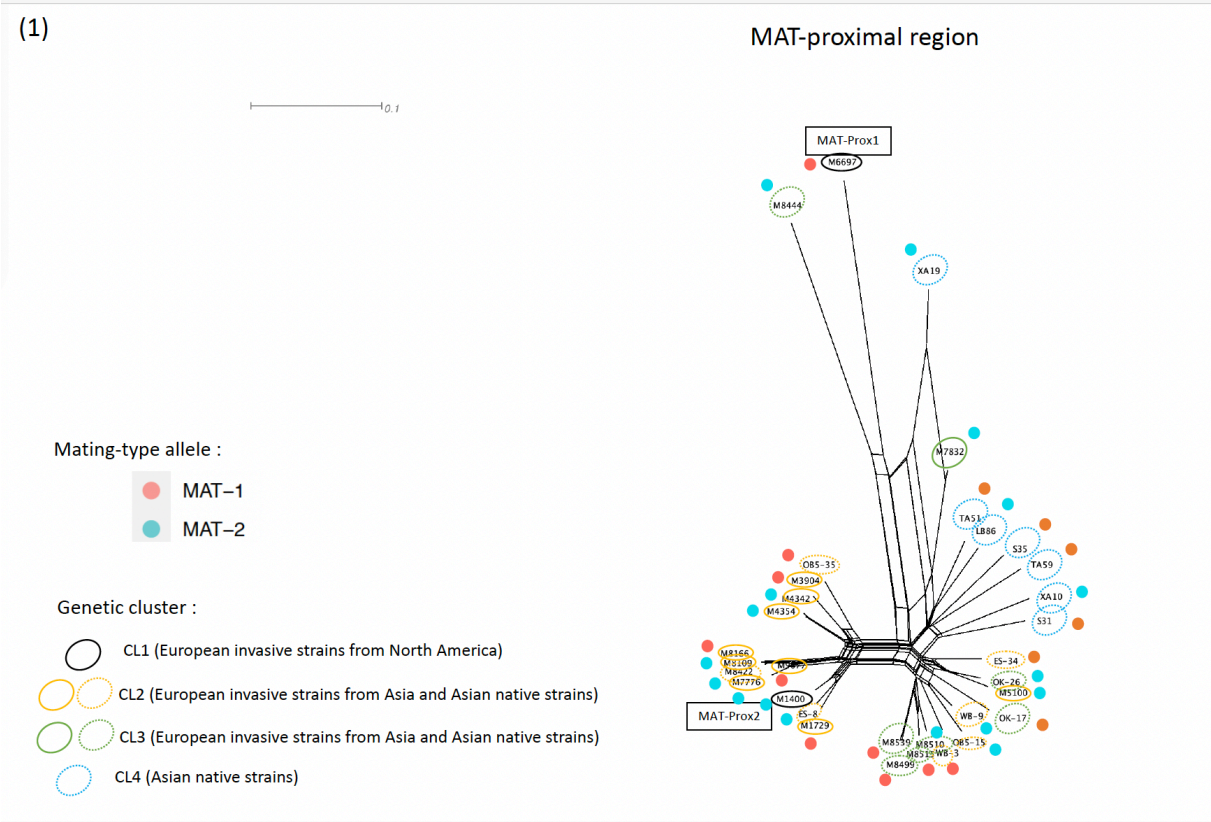
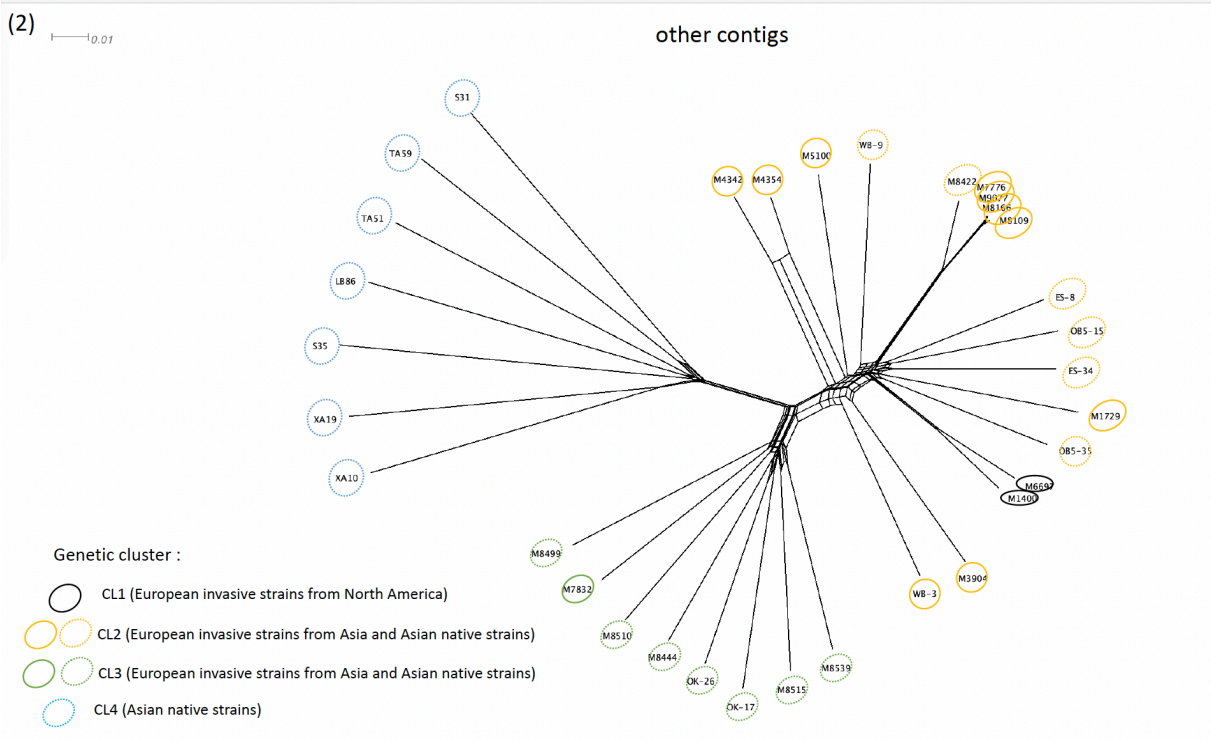

A
B
other_regions
MAT−proximal
region
XIM9508
ESM15
MRC010
DUM005
EP155
M1400
M6697
Native range
Introduced range
0
20
40
60
80
TE load (%)
Native range
Introduced range
XIM9508
ESM015
MRC10
DUM005**
EP155*
M1400**
M6697*
Orthologs present
and absent in M66
Orthologs present
and absent in M14
Orthologs present
M1400 and M6697
Number of present orthologs
* MAT-prox
** MAT-pro

Fig 7


(1)
MAT-proximal region
0.1
MAT-Prox1
M6697
M8444
XA19
M7832
TA51
LB86
S35
TA59
XA10
S31
OB5-35
M3904
M4342
M4354
M8166
M5109
M8422
M7776
M1400
ES-8
M1729
MAT-Prox2
ES-34
OK-26
M5100
W8-9
OK-17
OB5-15
M8539
M8510
M8515
W8-3
M8499
Mating-type allele :
MAT–1
MAT–2
Genetic cluster :
CL1 (European invasive strains from North America)
CL2 (European invasive strains from Asia and Asian native strains)
CL3 (European invasive strains from Asia and Asian native strains)
CL4 (Asian native strains)


(2)
0.01
other contigs
S31
TA59
TA51
LB86
S35
XA19
XA10
M4342
M4354
M5100
W8-9
M8422
M7776
M8166
M8109
ES-8
OB5-15
ES-34
M1729
OB5-35
M1400
M3904
W8-3
M8499
M7832
M8510
M8444
OK-26
OK-17
M8515
M8539
Genetic cluster :
CL1 (European invasive strains from North America)
CL2 (European invasive strains from Asia and Asian native strains)
CL3 (European invasive strains from Asia and Asian native strains)
CL4 (Asian native strains)