Million years old recombination suppression and balancing selection in a region 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], Quentin Rougemont[1], Arthur Demené[2], Thomas Badet[3], Daniel Croll[3], Amandine Labat[1], Simone Prospero[4], Alodie Snirc[1], Lea Stauber[5], 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 (short version)

Recombination suppression often evolves around mating compatibility loci, generating supergenes. Using hundreds of available genomes and generating high-quality genome assemblies from the invasive chestnut blight fungus *Cryphonectria parasitica*, we show that a 1.2 Mb genomic region proximal to the mating-type locus lacks recombination worldwide, thus constituting a supergene, but only partially linked to the mating-type locus. In invasive populations, this MAT-proximal region displayed two highly differentiated haplotypes, strongly associated to mating types, but not completely. High-quality genome assemblies revealed an inversion in the MAT-proximal region, and footprints of degeneration worldwide, with gene disruptions, non-synonymous substitutions and transposable elements. The two haplotypes segregate in all continents and have diverged at least 1.5 million years ago. This study reveals a large supergene partially linked to a mating compatibility locus, with balancing selection maintaining differentiated haplotypes, possibly involving deleterious mutations and/or host adaptation in a devastating tree pathogen.

Teaser:

A new supergene discovered near the mating-type locus in the devastating tree pathogen responsible for the chestnut blight disease

Abstract (longer version)

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, thus constituting a supergene. In invasive populations with extensive local genomic data and a recombining genome-wide structure, this MAT-proximal region displayed two highly differentiated haplotypes, that were strongly associated to mating types, but not completely, and with footprints of selective sweeps. High-quality assemblies of the two haplotypes in invasive and

native populations revealed an inversion in one of the two MAT-proximal 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 in the supergene was estimated to have occurred at least 1.5 million years ago and two haplotypes segregate in all continents, including the native range. Footprints of selective sweep, 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. This study sheds light on an interesting case of a large supergene partially linked to a mating compatibility locus, and on balancing selection maintaining differentiated haplotypes, possibly involving deleterious mutations and/or host 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 (*1*). However, recombination can also break up beneficial allelic combinations and, therefore, be selected against, generating supergenes, i.e., large regions without recombination encompassing multiple genes (*2*). 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 (*3–8*). Such recombination suppression keeps alleles linked across different genes that prevent self-compatibility or intermediate sexual phenotypes. Even beyond sex and mating-type chromosomes, there is an increasing number of supergene reports (*2*). Striking examples include supergenes controlling social structure in ants, wing color patterns in butterflies, reproductive morphs in birds or host-parasite interactions (*9–12*).

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 (*13–15*). Recombination suppression in sex chromosomes and other supergenes can nevertheless be maintained by balancing selection or heterozygote advantage (*11*, *16*). The mechanisms leading to recombination suppression and allowing their persistence remain poorly. Reporting more diverse supergene constellations in a variety of organisms with contrasting life-history traits is important for understanding the general patterns of their evolution and maintenance (*6*, *8*, *17*, *18*).

Recombination suppression can extend progressively away from sex-determining or mating-type loci, and around other types of supergenes (*4*, *8*, *18*). This is usually detected by plotting the synonymous differentiation (d_S) between alleles along the chromosome and detecting "evolutionary strata" of differentiation (i.e. with different d_S levels), as differentiation between alleles increases with time since recombination suppression (*4*). A recent theoretical model showed that, even when only partially linked to a mating-type or sex-determining allele, a recombination suppression event trapping a region with fewer deleterious alleles than average could be selected for and be maintained as heterozygous in linkage disequilibrium with the mating-type or sex-determining locus (*19–21*).

108 A situation that may correspond to this theoretically predicted case has been reported in the
109 invasive chestnut blight fungal pathogen *Cryphonectria parasitica*, where a large region
110 without recombination was suggested to occur near the mating-type locus, although not
111 completely linked to it (*i.e.,* at 3.9 cM of the mating-type locus), based on segregation of
112 RAPD and RFLP markers in three crosses involving five different parents, from the USA,
113 Japan or Italy (*22*). Recent genome-wide association studies in a worldwide collection and in
114 local *C. parasitica* populations in southern Switzerland found significant association of SNPs
115 with the mating-type locus across a large region (>1 Mb), further suggesting the existence of
116 reduced recombination in this region (*23*, *24*). This genomic region was enriched in SNPs,
117 transposable elements and copy-number variants, such as deletions, which are consistent with
118 sequence degeneration and a lack of recombination (*23*), and which prevented its assembly so
119 far. The strongest footprint of selective sweeps detected genome-wide in *C. parasitica* was
120 found matching this region (*23*), indicating recent positive selection, and thus possibly
121 selection for recombination suppression. In *C. parasitica*, the mating-type locus displays two
122 alleles (*MAT-1* and *MAT-2*) that are both permanently heterozygous in the diploid and
123 dikaryotic stages (*25*). Although this ascomycete fungus is mostly found as haploid mycelia,
124 some isolates have been reported to be heterokaryotic at the mating-type locus in several
125 invasive European and North American populations, i.e., carrying different nuclei, of the
126 opposite mating types *MAT-1* and *MAT-2* (*23*, *24*, *26–28*). The theoretical model based on
127 recessive deleterious mutations showed that non-recombining fragments could be selected for
128 even with a substantial fraction of the life cycle being haploid (*19*). Hence, *C. parasitica*
129 seems an excellent biological model to test recent theoretical predictions about the existence
130 of non-recombining regions partially linked to sex- or mating-type determining loci.
131
132 The evolutionary history of the species is well documented as it is an invasive and highly
133 damaging pathogen, having almost caused the extinction of American chestnut (*Castanea*
134 *dentata*) in North America (*29*). From its center of origin in Asia and its original hosts, the
135 Chinese chestnut (*Ca. mollissima*) and the Japanese chestnut (*Ca. crenata*), *C. parasitica* has
136 first invaded North America, killing millions of American chestnuts (*Ca. dentata*). In Europe,
137 at least two distinct introduction events occurred, on the European chestnut (*Ca. sativa*): one
138 introduction from North America to Italy, and one directly from Asia, probably to the
139 Pyrenees Mountains. European strains will be therefore hereafter referred to as invasive
140 European strains introduced from North America or invasive European strains introduced
141 directly from Asia. The chestnut blight symptoms have been less severe in Europe than in

142 North America, due to lower susceptibility of *C. sativa* and a virus infecting *C. parasitica* and
143 causing hypovirulence (*28*, *30*). Clonal lineages have been identified in invasive populations
144 while sexually reproducing populations occur both in the native and introduced ranges (*23*,
145 *30*). The clonal lineages likely still undergo rare sex events, as shown by the introgression of
146 the missing mating-type allele in most of them (*31*).
147
148 Here, we studied the occurrence of recombination suppression in the genomic region
149 proximal to the mating-type locus in *C. parasitica,* using the previously published genomes
150 and further generated six new high-quality long-read based genome assemblies of strains from
151 the native and introduced range of the pathogen. More specifically, we tested whether
152 recombination was completely suppressed in this region partially linked to the mating-type
153 locus in *C. parasitica*. We first analyze sexually reproducing invasive European populations,
154 for which extensive genomic datasets are available in local populations. We assessed, in these
155 populations, the degree of linkage disequilibrium within the MAT-proximal region, the level
156 of differentiation between haplotypes, as well as the frequency of the two non-recombining
157 haplotypes and their association with mating types. We investigated whether other regions of
158 the genome showed reduced recombination rates and whether genomic footprints of
159 degeneration were present in the non-recombining region, such as genomic rearrangements,
160 transposable element accumulation, non-synonymous substitutions and gene disruptions. We
161 also tested whether this non-recombining region has been gradually expanding, estimated the
162 age of recombination suppression and investigated whether the predicted gene functions in the
163 MAT-proximal region could help understanding the evolutionary cause for recombination
164 suppression and partial linkage to the mating-type locus. We then extended the analyses to
165 other populations with less extensive genomic resources, i.e. other invasive populations and
166 native Asian populations, investigating the presence of recombination suppression footprints
167 and the occurrence of the two differentiated haplotypes worldwide, as well as the possibility
168 of introgression, using available genomes of closely related species, pathogenic and non-
169 pathogenic (*23*).
170
171 **Results**
172 **Footprints of recombination suppression in a large region (> 1 Mb) proximal to the**
173 **mating-type locus in sexually reproducing invasive populations**
174 We first performed population genomic analyses on available genome sequences within local
175 populations sampled in southern Switzerland over two temporal frames (early 1990 and

176 2019), and within the more broadly geographically distributed CL1 cluster in central and
177 southeastern Europe, both having a recombining genetic structure and being invasive
178 European populations introduced from North America (Table S1 (*23*)). We studied only
179 monokaryotic strains, i.e. having either *MAT-1* or *MAT-2* in their genomes, for phasing
180 haplotypes. After masking repeats on the EP155 reference genome (*32*) and removing missing
181 data and rare variants, we identified 8,900 SNPs segregating among 71 strains of the 1990
182 Swiss population (35 *MAT-1*; 36 *MAT-2*; Table S1), 9,646 SNPs segregating among 62 strains
183 of the Swiss 2019 population (20 *MAT-1*; 42 *MAT-2*; Table S1) and 15,104 SNPs segregating
184 among 88 strains of the CL1 cluster (41 *MAT-1*; 47 *MAT-2*; Table S1) (*23*, *24*). Such a low
185 number of SNPs is consistent with expectations and previous studies in this invasive clonal
186 fungus. We investigated the genome-wide linkage disequilibrium (LD) landscape and looked
187 for large blocks (>1 Mb) of high LD (r^2>0.9) among SNPs within contigs, to look for
188 signatures of reduced recombination along the genome. LD indeed decays by half across a
189 few 100 kb on average in *C. parasitica* in recombining regions of the genome (*31*), so that
190 high LD beyond this distance is indicative of recombination cessation.
191
192 We found a large block of high LD on the chromosome called scaffold_2 of the EP155
193 genome in the three populations (red arrow in Fig 1 and Fig S1), corresponding to the mating-
194 type chromosome. The high-LD block was located near the mating-type locus but did not
195 encompass the mating-type locus itself and will be called hereafter the "MAT-proximal
196 region". In the 1990 and 2019 Swiss populations, SNPs between 0.493 Mb and 1.720 Mb in
197 the mating-type chromosome were all in high LD (with r^2>0.9 between SNPs distant by up to
198 1 Mb; in this region, mean r^2 of all SNP pairs in the Swiss 1990 population = 0.89; mean r^2 of
199 all SNP pairs in the Swiss 2019 population = 0.83; Fig S1). In the CL1 cluster, the large high-
200 LD block was also present, but was split into two blocks, a large one and a smaller, peripheral
201 one (orange arrows in Fig 1). We indeed detected values of r^2>0.9 between SNPs distant by
202 up to 1 Mb, between 0.532 Mb and 1.535 Mb (mean r^2 of all SNP pairs = 0.94). Between
203 1.535 and 1.717 Mb, SNPs were also in high LD (mean r^2 = 0.84). The LD between the two
204 high-LD blocks was lower (r^2= 0.59) than within blocks but was still higher than elsewhere
205 along the genome (orange rectangle at the right border of the red triangle; Fig 1). A similar
206 pattern also existed in the Swiss populations (orange rectangle at the right border of the red
207 triangle; Fig S1), although less marked. This is likely due to rare recombination events at one
208 specific locus near the edge of the large fully non-recombining region, lowering the LD level
209 in populations between the two parts of the otherwise fully non-recombining region. The

210 mating-type locus was located at 1.737 Mb in the EP155 genome, distant 17 kb and 20 kb
211 from the LD block in the Swiss populations and CL1 cluster, respectively.
212
213 We found no other regions of the genomes with r^2>0.9 between SNPs across such a large
214 genomic distance in any population. On scaffold_6 in the CL1 cluster, two smaller SNP
215 blocks had a r^2>0.9 between each other despite being distant (Fig S2A; between SNPs blocks
216 at 3-20 kb and 2.151-2.278 kb), which is likely due to the major intra-scaffold translocation
217 described in this region between the EP155 and ESM15 strains (*33*). The maximum size of
218 the blocks with r^2>0.9 on other scaffolds ranged from 135 to 540 kb in the CL1 cluster, as
219 previously described (*31*). The average distance between SNPs in high LD (r^2>0.9) across the
220 genome was higher on the scaffold_2 than on other scaffolds (in the CL1 cluster, pairwise
221 Wilcoxon test p-value < 2e-16 with Bonferroni correction). Furthermore, the proportion of
222 SNPs in high LD (r^2>0.9) across the genome was also much higher on the scaffold_2 than on
223 other scaffolds due to the high LD in of the MAT-proximal region (Fig S2B, Table S2).
224 Therefore, the large (1Mb) and localized block with maximal LD values proximal to the
225 mating-type locus both in the Swiss populations and the CL1 genetic cluster stands out as
226 exceptional in the genome and indicates full recombination cessation, which supports
227 previous inferences from progeny segregation analyses (*34*). Indeed, even low rates of
228 recombination homogenize alleles and prevents LD building (*35*). The MAT-proximal non-
229 recombining region was actually larger (1.23 Mb), but with likely rare recombination events
230 at one precise locus near its edge, lowering LD between its two fully non-recombining parts.
231
232 Noteworthy, we also found blocks of high LD (r^2>0.9) around three *vic* loci (i.e., *vic2*, *vic6*
233 and *vic7* on scaffold_3, scaffold_6, and scaffold_7; Fig S2C), known to be involved in
234 vegetative incompatibility (vic) and under balancing selection in *C. parasitica* (*24*, *36*). These
235 high-LD blocks were smaller than the MAT-proximal region but nevertheless larger than
236 elsewhere in the genome (across more than 100 kb, Fig S2C).
237
238 Consistent with a lack of recombination, the MAT-proximal region formed two genetically
239 highly differentiated haplotypes in these invasive Swiss populations (Fig 2A; Fig S3), as
240 shown by the two clusters on the principal component analysis (PCA) using only SNPs
241 located in the MAT-proximal region (Fig 2A), while no structure was detected in the rest of
242 the genome (Figs 2B and C; Fig S3). The neighbor-net trees further supported the existence of
243 two differentiated haplotypes in the MAT-Proximal region, contrasting with an otherwise

recombining structure genome-wide (Fig 2D and E). Three strains from the Swiss 2019 population appeared to have an intermediate sequence in the MAT-Proximal region between the two haplotypes (Fig S3D).-These intermediates were rare and may correspond to ancestral polymorphism and/or a third haplotype.

In agreement with the view of a partial linkage with the mating-type locus, the two genetic clusters in the MAT-proximal region appeared strongly associated with mating types, although the association did not reach 100% (Figs 2A and D; Fig S3, Table 1). Indeed, the distribution of mating types between the two PCA clusters strongly deviated from expectations under random association (Table 1; chi-squared test: $\chi^2 = 33.3$; p-value = 5.876e-08 in CL1). We hereafter call the two differentiated haplotypes MAT-Prox1 and MAT-Prox2, referring to the mating-type association. The EP155 genome used as reference for SNP calling above carried the MAT-Prox1 haplotype.

Inversion in the non-recombining MAT-proximal region

To investigate whether rearrangements between haplotypes were present in the MAT-proximal region, we sequenced the genome, using PacBio HiFi, of two strains with alternative MAT-proximal haplotypes: M1400 (MAT-2; MAT-Prox2) and M6697 (MAT-1; MAT-Prox1), originating from the Swiss population in Gnosca and belonging to the CL1 cluster, i.e. European invasive populations introduced from North America (*23*, *24*). We build high-quality genome assemblies: statistics of the assemblies were in the same range as those of the genome assemblies of the ESM15 and EP155 strains previously sequenced (Table S3; Demené et al 2022; Crouch et al 2020). The mating-type chromosome could be assembled as a single contig for the first time, in the M1400 genome, likely corresponding to a full chromosome, and was assembled into two contigs in the M6697 genome. We therefore used the M1400 genome as a reference for subsequent analyses. To identify the location of the MAT-proximal region in the M1400 genome, we computed LD by read mapping in the 1990 Swiss populations. We found the large block (> 1 Mb) of high LD (r^2>0.9) between 7,285,137 and 8,828,934 bp (red arrow; Fig. S4). The MAT-proximal region was also located 20 kb away from the mating-type locus (located at 7.265 Mb) and was 1.540 Mb long using the M1400 genome as reference. Consistent with the results using the EP155 genome as reference, we found that the high-LD region was divided into two higher-LD blocks near the edge at 7.392 Mb (orange arrows in Fig. S4).

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 S5), where we found a large inversion (>1 Mb) between
282 the M1400 and M6697 genomes (Fig 3; Fig S5). Breakpoints of the inversion were located at
283 ca. 7.455 and 8.563 Mb of the tig00000001 contig in the M1400 genome and at ca. 3.380 and
284 4.185 Mb of the tig00000060 contig in the M16697 genome. The region affected by the
285 inversion was ca. 300 kb smaller in the M6697 genome than in the M1400 genome, while the
286 whole MAT-proximal region was ca. 100 kb smaller. The split into the two high-LD blocks in
287 natural populations reported above was outside of the inversion, i.e. 70 kb from the inversion
288 breakpoint (orange arrows in Fig S4). LD in natural populations was higher within the
289 inversion (median value of $r^2 = 1$) than in other regions of the MAT-proximal region (median
290 value of $r^2 = 0.64$; Wilcoxon rank sum test with continuity correction; W = 4.8714e+10; p-
291 value < 2e-16).
292
293 We looked for centromeres, as the non-recombining regions near mating-type loci in other
294 ascomycetes, when they occur, either capture the centromere (*37*, *38*), or are associated to the
295 occurrence of a single crossing-over between the centromere and the mating-type locus (*8*, *39*,
296 *40*). The centromere in *C. parasitica* was likely at 4.380-4.536 Mb on the mating-type
297 chromosome in the M1400 genome, as we detected here a peak in TE density and a drop in
298 GC content (Fig 4A-B). The MAT-proximal region and the inversion thus did not include the
299 putative centromere and was instead located about 2.9 Mb away of the putative centromere.
300
301 **Higher genetic differentiation and lower genetic diversity within the MAT-proximal**
302 **region than in recombining regions**
303 When analyzing polymorphism within the invasive 1990 Swiss population using the M1400
304 genome as reference, we found higher genetic differentiation between strains of the two
305 MAT-proximal haplotypes in this region than elsewhere along the genome (Fig 4C-D; Fig
306 S6A-B; Table S4), as expected for a non-recombining region in LD with the mating-type
307 locus. For example, the F_{ST} values between the two non-recombining haplotypes within the
308 MAT-proximal region in the 1990 Swiss population were nearly maximal (median F_{ST}=0.93
309 per 50 kb window), while they were near zero in the rest of the genome (median F_{ST}=0.02;
310 Wilcoxon test p-value < 2e-16 with Bonferroni correction; Fig 4C; Fig S6A). Such F_{ST} values
311 close to 1 indicate a lack of shared polymorphism and therefore support the inference of

312 recombination suppression in the MAT-proximal region. We also found higher absolute
313 genetic divergence (d_{XY}) in the MAT-proximal region compared to other regions (Wilcoxon
314 test p-value < 2e-16 with Bonferroni correction), with a peak of number of fixed differences
315 in the middle of the MAT-Proximal region and of the inversion (Fig 4D; Fig S6A; Table S4),
316 suggesting a pattern of gradual expansion of recombination cessation within the MAT-
317 proximal region, from its center to the edges. The 95th percentile of highest d_{XY} values was for
318 the windows encompassing the genomic region between 8.190 à 8.280 Mb.
319
320 Within each haplotype, the genetic diversity at the MAT-proximal region was lower than in
321 the rest of the genome (Wilcoxon test p-value < 2e-16 for the MAT-Prox1 haplotype; p-value
322 = 2.3e-07 for the MAT-Prox2 haplotype; Fig 4E-F; Fig S6C; Table S4), as expected for a
323 region without recombination associated with the mating-type locus, as its effective
324 population size is half as in recombining regions. The diversity was especially low in the pool
325 of isolates with the MAT-Prox1 haplotype (see all MAT-Prox1 sequences clustering on a
326 single point on the PCA in Fig 2A and in the neighbor-net tree in Fig 2D, in contrast to the
327 more scattered MAT-Prox2 sequences); such a lack of diversity contrasting with other
328 genomic regions and the other haplotype suggests that a selective sweep occurred in this
329 MAT-Prox1 haplotype. This supports previous inferences of a selective sweep in this region
330 based on site-frequency spectrum analyses (*23*).
331
332 In addition to footprints of selective sweeps at short time scales, we detected signals for long-
333 term balancing selection in the MAT-proximal haplotypes. Computation of the Tajima's D
334 statistics within all strains suggested signatures of balancing selection in the MAT-proximal
335 region as expected in a region without recombination associated with the mating-type locus
336 maintained at a frequency close to 1/2. Tajima's D values in the MAT-proximal region
337 pooling all sequences (median D=3.0 per 50 kb window) were higher than in the rest of the
338 genome (median D=-0.14; Wilcoxon test p-value < 2e-16 with Bonferroni correction; Fig 4G;
339 Fig S6D). In contrast, Tajima's D was significantly lower than the rest of the genome within
340 each pool of haplotype, including in the MAT-Prox2 haplotype, further supporting the
341 inference of positive selection (Wilcoxon test p-value = 0.0037 for the MAT-Prox1
342 haplotype; p-value = 0.0022 for the MAT-Prox2 haplotype with Bonferroni correction; Fig
343 4H-I; Fig S6D; Table S4). We found no difference in population diversity statistics between
344 the recombining part of the mating-type contig and the other contigs (Table S4A).
345

346 The study of synonymous divergence (d_S) between M1400 and M6697 haplotypes suggests
347 that the recombination suppression is at least 1.5 million years old and that there may be a
348 pattern of evolutionary strata within the MAT-proximal region, i.e. segments with different
349 levels of differentiation between haplotypes that could indicate stepwise expansion of
350 recombination suppression away from the mating-type locus. Per-gene synonymous
351 divergence (d_S) is typically used for detecting gradual expansion of recombination cessation
352 and for estimating its age, as it is considered a good proxy for the time since recombination
353 suppression. Indeed, when recombination is suppressed, mutations accumulate independently
354 in the two non-recombining haplotypes. We plotted the per-gene synonymous divergence (d_S)
355 between M1400 and M6697 along the M1400 genome, as its mating-type contig is likely
356 being assembled as a full chromosome (Fig S7A&E; Table S4). Consistent with
357 recombination suppression, we found significantly higher d_S values in the MAT-proximal
358 region (computed for 71 genes) than in the other regions of the mating-type contig (pairwise
359 Wilcoxon test p-value <2e-16 with Bonferroni correction) and other contigs (pairwise
360 Wilcoxon test p-value <2e-16 with Bonferroni correction). We found no significant
361 differences between the other regions of the mating-type contig and other contigs (pairwise
362 Wilcoxon test p-value = 0.23 with Bonferroni correction). The region with higher d_{XY} values
363 within the MAT-proximal region (between 8.190 à 8.280 Mb) showed significantly higher d_S
364 values than the rest of the MAT-Proximal region (Wilcoxon test p-value = 0.014) supporting
365 the view of a gradual expansion of recombination cessation within the MAT-proximal region,
366 from its center to the edges. Note that the per-gene d_S could not be computed for the genes
367 with a copy not predicted in one of the haplotypes and therefore mostly in the oldest, most
368 degenerated regions (see below), which may prevent detecting the oldest stratum using d_S
369 values, while the d_{XY} values were computed in sliding windows all along the MAT-proximal
370 region. Using synonymous substitution rate estimates in fungi (*41*, *42*) and the mean d_S values
371 across genes of the MAT-proximal region, we estimated the age of the divergence between
372 the two haplotypes to be at least 1.5 Million years when considering the mean d_S values of all
373 genes in the MAT-proximal region (dS mean= 0.0495) and at least 7.2 Million years when
374 considering only the 15% genes with highest d_S values dS mean= 0.24), under the hypothesis
375 of gradual expansion of the non-recombining region (Table S4).
376
377 **Enrichment of transposable elements in the non-recombining region and inversion**
378 **breakpoints**

379 When studying the M1400 and M6697 high-qualities assemblies, we detected an enrichment
380 of transposable elements (TEs) in the MAT-proximal region in the two haplotypes compared
381 to the rest of the genome. The percentage of bp occupied by TEs (TE load) was higher than
382 50% in the MAT-proximal region, while it was only 9% on average in other regions. TE load
383 in the MAT proximal region was higher in the M6697 haplotype (76%, MAT-1; MAT-Prox1)
384 than the M1400 haplotype (68% MAT-2; MAT-Prox2). Class I retrotransposons with Gypsy
385 (LTR-Ty3) and LARD elements were the most abundant TEs in all genomic regions
386 (autosomes, MAT-Proximal region and the rest of the mating-type contig), representing >70%
387 of TE annotations (Fig 5A-B), as shown previously (*33*). The TE load however varied
388 significantly among TE families and genomic regions (ANOVA; Table S5). In M6697 for
389 example, TIR elements appeared less abundant in the non-recombining than recombining
390 regions while LTR-Ty3 elements were more frequent in the MAT-proximal region (Fig 5B).
391 In M1400, LARD elements were more frequent in the MAT-proximal region than in
392 recombining regions (Fig 5B).
393
394 The presence of transposable elements at the two inversion breakpoints in the two genomes
395 (Fig 5A) suggests that the inversion may have occurred via non-homologous recombination
396 mediated by these elements. By dating retrotransposon insertions using between-copy
397 divergence and within-copy LTR divergence in both M1400 and M6697 genomes (Fig 5C
398 and D), we found that TEs within the inversion and around the inversion breakpoints were
399 older than TEs farther from the breakpoints. This is expected in non-recombining regions as
400 selection is there less efficient to purge TE insertions so that they remain there longer
401 (Duhamel et al. 2023). The date estimates of LTR element insertions (Fig 5D) further
402 indicated that the TE accumulation in the MAT-proximal region was as old as a few million
403 years, confirming the age estimated based on sequence divergence between MAT-proximal
404 haplotypes.
405
406 **Gene functions, pseudogenization and degeneration in the non-recombining region**
407 Recombination suppression could allow linking genes with mating-type antagonistic effects to
408 the mating-type locus, or link together co-adapted alleles, for example for parasitizing
409 different host genotypes. We however found no gene function known to be involved in
410 mating or mating compatibility, nor in pathogenicity, in the MAT-proximal region (Table S6).
411 We nevertheless identified genes potentially coding for proteins with kinase domain or
412 proteins with homeodomains (Table S6), that are often involved in the regulation of

413 developmental pathways. Gene functions in the d_{XY} peak (95th percentile) included general
414 functions such as protein kinase-like domain, SET domain (i.e. with methyltransferase
415 activity) and RNA-binding domain families (Table S6).
416
417 By comparing gene predictions in the M6697 and M1400 newly assembled genomes, we
418 identified a higher prevalence of gene disruptions (i.e. genes predicted in a single of the two
419 haplotypes) in the MAT-proximal region than elsewhere in the genomes. Out of 175 groups
420 of single-copy orthologous genes in the MAT-proximal region, 80 (46%) were only present in
421 M1400 and 24 (14%) only present in M6697. In other genomic regions, 3% of genes were
422 only present in one genome. This suggests that gene disruption occurred in the non-
423 recombining region, in higher proportions than elsewhere in the genome (Fisher's exact test
424 p-value < 2.2e-16), and especially in the M6697 genome (MAT-1; MAT-Prox1). These gene
425 disruptions are mainly in the middle of the MAT-proximal region and of the inversion (Fig
426 5A), i.e. the region corresponding to the d_{XY} peak in the Swiss population and to the high d_S
427 values between the M1400 and M6697 genomes (Figs 4D and S7). In the M1400 genome,
428 gene disruptions are more scattered all along in the MAT-proximal region (Fig 5A). The
429 enrichment in TEs and in gene disruptions in the MAT-Proximal region explains the lower
430 gene density in this region (Fig 4) and supports the inference of an old full recombination
431 suppression. The occurrence of high TE load and gene degeneration in both MAT-Proximal
432 haplotypes further confirms the lack of homozygotes in this region, in which recombination
433 could purge deleterious mutations, or indicates that recombination suppression occurs even in
434 homozygotes, although it remains to be investigated.
435
436 We found higher non-synonymous substitution rate between the M1400 and M6697 genomes
437 compared to the baseline substitution rates (d_N/d_S values) in the MAT-proximal region
438 (median per gene d_N/d_S = 0.80) than in other contigs (d_N/d_S = 0.43; pairwise Wilcoxon test p-
439 value = 1.7e-07 with Bonferroni correction), consistent with relaxed purifying selection due
440 to recombination suppression in the MAT-proximal region (Fig S7A; Table S4).
441
442 **Recombination suppression in the MAT-Proximal region in other populations assessed**
443 **from high-quality genome assemblies**
444 We analyzed five additional high-quality genome assemblies for strains from the Asian native
445 range (CL2 and CL4 clusters) and the invasion range (including two North American strains
446 belonging to CL1 cluster and one European strain introduced directly from Asia and

447 belonging to CL2 cluster), three of which were generated for the present study (Table S3).
448 Most of the high-quality genomes displayed the same MAT-Proximal chromosomal
449 arrangement as M1400 (MAT-Prox2, Swiss population and belonging to CL1 cluster),
450 suggesting that it is the ancestral state (Fig S8A-C), although the incomplete assemblies may
451 prevent detecting other rearrangements. The inversion detected in M6697 (MAT-Prox1, Swiss
452 population and belonging to CL1 cluster) was also present in the EP155 strain (MAT-Prox1,
453 North America and belonging to CL1 cluster, Crouch et al, 2020; Fig S8D), indicating that
454 this inversion was already segregating in early invasive populations in North America. In
455 other genomic regions outside of the MAT-Proximal region, all analyzed high-quality
456 genomes were colinear to the ESM15 reference genome, except the highly rearranged EP155
457 genome, as previously reported (*33*).
458
459 The MAT-Proximal region in all the high-quality genomes, from Asian native strains and
460 invasive strains, displayed signs of degeneration, with a high TE load (Fig 6B) and gene
461 disruptions, as seen from lower gene densities (Fig S8A) and missing orthologs (Fig 6C),
462 confirming the ancient age of recombination suppression and its occurrence in the native
463 range. Study of d_S and d_N/d_S between the MRC10 strain, belonging to the European clonal
464 lineages of direct Asian origin (CL2 cluster), the M6697, M1400 and the two Asian native
465 strains further confirmed recombination suppression in the native range and in invasive strains
466 from different origins (North America or directly from Asia). The d_S and d_N/d_S values were
467 indeed significantly higher in the MAT-proximal region than other genomic regions (Fig S7
468 B-E). Mean d_S values in the MAT-proximal region were in the same range as d_S values
469 between the M6697 and M1400 genomes, suggesting a similar date of recombination
470 suppression and, therefore, that recombination suppression was ancestral. The MAT-proximal
471 region appeared smaller in the two Asian strains XIM9508 and ESM15 (<1 Mb) than in
472 invasive strains (Table S3), but this may be due to the incomplete genome assemblies.
473
474 **Presence of the two MAT-Proximal haplotypes in other populations**
475 By analyzing available additional Illumina genomes in the native range and in invasive
476 populations, with less dense sampling within populations, we recovered the two haplotypes in
477 the MAT-proximal region, i.e., high differentiation between two clusters in the MAT-
478 Proximal region, contrasting with a lack of genetic subdivision in other genomic regions or a
479 completely different genetic subdivision, corresponding to previously described population
480 genetic structures (Figs 7 and FigS9). In particular, the neighbor-net tree with multiple strains

481 sequenced previously from the clusters CL2 and CL3 (from the introduced range and the
482 native range, South Korean and Japan) and from the cluster CL4 (from the native range,
483 China;) reinforces the view of balancing selection on the MAT-Proximal haplotypes.
484 Sequences from the CL1 and CL2 clusters were indeed intermingled in the MAT-proximal
485 region (Fig 7A), while the structure in the rest of the genome instead corresponded to the
486 different population clusters, i.e., CL1, CL2, CL3 and CL4 (Fig 7B).
487
488 However, the two haplotypes were less differentiated in the native range and in invasive
489 populations originating directly from Asia than in invasive populations from the first invasion
490 wave via North America. For example, branches length on the neighbor-net indicates a lower
491 differentiation between haplotypes within CL2 (Fig 7A) than within the European CL1 or
492 Swiss invasive populations (Fig 2D; Fig S3). Within CL2, F_{ST} and d_{xy} values between the two
493 haplotypes in the MAT-proximal region were lower than in the invasive population coming
494 from North America (Table S7).
495
496 We found no association of the MAT-Proximal haplotypes with mating types at the world
497 scale (Figs 7 and S9), which may be due to the lack of extensive sampling within local
498 populations. As the MAT-Proximal region is only partially linked to the mating-type locus,
499 founder effects may yield different associations between MAT-Proximal haplotypes and
500 mating types in different populations. Only in the S12 European invasive clonal lineage (only
501 MAT-1), we found a single MAT-Proximal haplotype (MAT-Prox1; Fig S9B). In two
502 invasive European clonal lineages of American origin (re092 and re103), the strains with a
503 mating type different from their clone-mates due to localized introgression from other clonal
504 lineages (*31*) also had their MAT-Proximal region introgressed (Fig S9C), reinforcing the
505 view of the strong linkage between the mating-type locus and the MAT-Proximal region, and
506 of an advantage of maintaining an association between MAT-proximal haplotypes and mating
507 types within populations. The few 2019 Swiss strains that exhibited an intermediate sequence
508 in the MAT-Proximal region clustered with strains from Europe with a direct Asian origin
509 (Fig S10C). This may represent a third MAT-Proximal haplotype, associated with the second
510 invasion in Europe, directly from Asia.
511
512 **No evidence of introgression in the MAT-proximal region from other species**
513 The stronger differentiation in the MAT-Proximal region in the Swiss invasive population
514 than in other populations could also be due to an introgression event from a closely related

515 species; however, the small genetic distance and the reticulations observed in the neighbor-net
516 tree between the M6697 Swiss reference genome and the Asian genomes M8444 and XA19
517 rather indicate that the MAT-Prox1 haplotype in Switzerland originated from Asian *C.*
518 *parasitica* populations, likely a population different from those at the origin of the CL2
519 invasive cluster (Fig 7).
520
521 Gene genealogies also provided no evidence of introgression in the MAT-proximal region
522 from closely related *Cryphonectria* species. We performed a gene orthology analysis
523 including three genomes of the closely related species *C. japonica*, C. *carpinicola*, and *C.*
524 *naterciae* that were previously sequenced with short read technologies (*43*). We retrieved
525 only nine genes from the MAT proximal region that were shared between the M6697 and
526 M1400 genomes that had orthologs in one of the related species. This may be due to low
527 quality assemblies in the outgroups (*43*) and/or a region without recombination also in the
528 outgroups with gene losses and many repeats. We found no gene with a topology consistent
529 with introgression from another species into the MAT-Proximal region, i.e. with a MAT-
530 Prox1 or MAT-Prox2 haplotype that would branch with an outgroup allele rather than with
531 the alternative haplotype of *C. parasitica*. The high load of TEs and gene disruptions in the
532 two MAT-Proximal haplotypes does not fit with the introgression hypothesis either, unless
533 the introgression occurred from a species with recombination suppression in the MAT-
534 Proximal region.
535
536 **Discussion**
537 ***A large and million-year-old non-recombining region partially linked to the mating-type***
538 ***locus in* Cryphonectria parasitica**
539 We found strong evidence for the complete cessation of recombination in a region of more
540 than 1 Mb (1.2 Mb to 1.5 Mb depending on the haplotype), strongly but not completely linked
541 to the mating-type locus in *C. parasitica*. Indeed, we detected maximal levels of linkage
542 disequilibrium in otherwise recombining populations and the existence of two highly
543 differentiated haplotypes, without shared polymorphism, while even low rates of
544 recombination can homogenize alleles and prevent LD build-up (*35*). The full cessation of
545 recombination was further supported by the existence of a previously unknown, large
546 rearrangement (an inversion) in the center of the non-recombining region in invasive strains.
547 The higher TE load in this region constitutes additional evidence for full recombination
548 cessation, as the lack of recombination renders selection against TE accumulation less

549 efficient, in particular due to Muller's ratchet: two chromosomes with different TE insertions
550 cannot recombine to produce a chromosome free of TE insertions. TE takes hundreds of
551 thousands of years to accumulate in non-recombining regions (*15*), so that the high TE load in
552 the MAT-Proximal region in *C. parasitica* indicates old recombination cessation. The
553 particularly high frequency of missing genes and non-synonymous substitutions in the MAT-
554 Proximal region worldwide further indicates degeneration and ancient recombination
555 suppression. The estimates for the age of recombination suppression based on the divergence
556 between haplotypes and on TE insertions were in fact at least a few million years. However,
557 in contrast to other known evolutionary strata on sex or mating-type chromosomes (*4*, *8*), the
558 region without recombination was not completely linked to the mating-type locus. Indeed, the
559 occurrence of rare recombination events between the non-recombining MAT-Proximal region
560 and the mating-type locus are supported by previous segregation analyses (*34*) and the lack of
561 association found here between the MAT-Proximal haplotypes and the mating types at the
562 worldwide scale, as well as their incomplete association within populations.
563
564 ***Large inversion in the MAT-proximal non-recombining region***
565 We found a large inversion between the two highly differentiated MAT-proximal haplotypes
566 in invasive populations originating from North America (CL1 genetic cluster, Swiss 1990 and
567 2019 populations and US populations). The chromosomal arrangement of the M1400 MAT-
568 Prox2 haplotype is likely the ancestral state as it is the most frequent worldwide, although a
569 more complete sampling in the native range is required to obtain definitive conclusion. The
570 inversion would therefore be the M6697 MAT-Prox1 haplotype and may contribute to
571 recombination suppression. However, the non-recombining region was larger than the
572 inversion, which indicates that additional proximal mechanisms suppress recombination. The
573 inversion may even be a consequence rather than a cause of recombination cessation, as
574 previously found in other fungal mating-type chromosomes (*37*, *39*, *40*, *44*). Other proximal
575 causes of recombination cessation may for example be genetic recombination modifiers,
576 epigenetic marks or histone modifications (*45*, *46*). Alternatively, the inversion may have
577 constituted a cause for a first step of recombination suppression (i.e., for a first evolutionary
578 stratum), as actually suggested by the higher d_S and d_{XY} values in the middle of the inversion
579 compared to the rest of the non-recombining region. Evolutionary strata have been reported in
580 many sex chromosomes in plants and animals (*13*, *47–49*) and also around mating-type loci in
581 fungi (*8*, *50*). The lack of rearrangements at the edge of the MAT-proximal region may allow
582 rare recombination or gene conversion events, which may explain the presence of two distinct

583 LD blocks with lower LD between them and the lower differentiation at the edges of the
584 MAT-proximal region.
585
586
587 ***The MAT-proximal non-recombining region is million-year-old and degenerated***
588 Following recombination suppression, selection efficacy decreases. We detected footprints of
589 degeneration, with particularly high occurrence of gene disruptions and of TE insertions in the
590 MAT-proximal region, both in Europe and in the native Asian range of the fungus. (*33*) had
591 already reported an enrichment of LTR-Ty3 (*Gypsy*) element in the region near the mating-
592 type locus in *C. parasitica*. We found LTR-Ty3 elements at the inversion breakpoints and an
593 enrichment in the MAT-Prox1 haplotype. Interestingly, the same TE family (LTR-Ty3) seem
594 to have expanded in the non-recombining regions in other fungi with recombination
595 suppression on their mating-type chromosomes (*15*), suggesting they may be more prone to
596 accumulate in non-recombining regions. Accumulations of transposable elements have also
597 been detected in other mating-type or sex chromosomes and supergenes (*13–15*, *51*), although
598 some other large supergenes do not seem to exhibit TE load, likely due to recent
599 introgression, gene conversion or double crossovers (*52–55*).
600
601 The transposable element accumulation in the MAT-proximal region and the enrichment in
602 gene disruptions support the estimated age of recombination suppression at a few million
603 years ago based on mean pairwise gene sequence divergence between haplotypes. The finding
604 of the two haplotypes and of the inversion in the two invaded continents, Europe and North
605 America, as well as the occurrence of degeneration footprints in the MAT-Proximal region in
606 the native Asian range, further support the inference of an ancient recombination suppression
607 maintained by balancing selection. Indeed, even rare recombination events allow purging
608 deleterious mutations and TE insertions. This means that the association of the MAT-
609 Proximal haplotypes with mating types is very strong within all recombining populations
610 (even if not at the worldwide scale, as founding effects may generate different associations
611 between mating types and MAT-Proximal haplotypes); it may also be that recombination is
612 suppressed within the MAT-Proximal region even for homozygotes in this region. This is in
613 agreement with previous genetic maps reporting lack of recombination near the mating-type
614 locus in crosses involving Japanese, North American and Italian isolates (*22*). The high TE
615 load and degeneration level in terms of gene disruption in strains from Asian and North

American origins indicate a worldwide recombination cessation, so likely ancestral to all these native and invasive populations.

The differentiation between the MAT-Proximal haplotypes appeared greater in the CL1 cluster and Swiss populations (from the first invasion event, from North America) than in the other clusters, which may be due to a founder effect, the lack of sampling in the source population in Asia and/or a selective sweep. The MAT-Prox1 haplotype in the invasive European population from the first introduction event shows an inversion, a strong selective sweep, and the most degenerated locus. These findings point to a strong selection maintaining the two haplotypes differentiated and non-recombining, and perhaps different selection regimes or successive selection events during the various invasion waves of *C. parasitica.*

The higher differentiation between the MAT-Proximal haplotypes in the first invasion wave in Europe could also have resulted from introgression from a closely related species. The evolution of supergenes by introgression has been reported for example in ants and butterflies (*56–58*). However, we found little evidence for an introgression: i) the two haplotypes are present in the CL2 cluster of invasive European populations directly originating from Asia, ii) the MAT-Prox1 haplotype in the CL1 cluster more differentiated from the MAT-Prox2 haplotype than in the CL2 cluster is genetically close to an Asian haplotype, iii) footprints of recombination suppression are present in the native range in terms of degeneration and particular genomic structure in the MAT-Proximal region, and iv) we found no signatures of introgression from closely related species in gene genealogies.

The high differentiation into two distinct haplotypes, their occurrence in at least two genetic clusters, their balanced frequencies within populations, their degeneration in terms of TE load and disrupted genes on different continents, and the strikingly different genomic subdivision in the MAT-Proximal regions compared to the rest of the genome, altogether suggest an ancient recombination suppression maintained by balancing selection for at least a few million years. The degeneration in both haplotypes indeed indicates a lack of homozygotes at the diploid stage, in which recombination could occur and purge deleterious mutations, or suppressed recombination even in homozygotes.

The selective forces potentially maintaining the two haplotypes in the MAT-proximal region

650 The genes found missing in the non-recombining region in one or the other haplotype may
651 help maintaining the two haplotypes by selection due to deleterious mutation sheltering, i.e.
652 thanks to a heterozygote advantage. Such overdominance advantage may only be a
653 consequence of recombination suppression, and contribute to its maintenance, or may even be
654 its initial cause (*11*, *19–21*). Indeed, this situation of a non-recombining region partially
655 linked to the mating-type locus and with two haplotypes maintained by selection fits
656 predictions from a recent theoretical model showing that recombination suppression can be
657 selected for near permanently heterozygous alleles for protecting haplotypes with fewer
658 deleterious mutations than average in the population, while sheltering their few recessive
659 deleterious alleles. Importantly, the model predicts that selection can also occur for regions
660 partially linked to heterozygous alleles (*19*). Similar overdominance effect due to the
661 sheltering of deleterious mutations can lead to balancing selection in the context of local
662 adaptation (*59*). These models require the existence of a substantial diploid or dikaryotic
663 phase for sheltering recessive deleterious alleles in a heterozygous state, although the
664 mechanism also works with a substantial haploid phase (*19*). In fact, although *C. parasitica*
665 has a predominant haploid phase, strains can be found as dikaryons heterozygous for the
666 mating-type locus in some natural populations, and not only as monokaryons (*23*, *24*, *26*, *27*).
667 The importance and frequency of heterokaryons in *C. parasitica* in nature would deserve
668 further investigations. It may be sufficient if recessive deleterious alleles are sheltered in ca
669 20% of individuals for selecting recombination suppression (*19*), and/or if the genes involved
670 are preferentially expressed during the diploid or dikaryotic phases. This hypothesis would
671 explain the long-term balancing selection, and also the selective sweep by a recent extension
672 of recombination suppression, possibly corresponding to the inversion found in invasive
673 population from the first introduction. This hypothesis of balancing selection, or the
674 hypothesis of overdominance (i.e. heterozygote advantage), does not require the same
675 association between the MAT-Proximal haplotypes and mating types across all populations
676 worldwide, it is sufficient that the MAT-Proximal haplotypes are strongly associated to
677 mating types within populations. Further sampling will be required to test whether the MAT-
678 Proximal haplotypes are associated to mating types within local populations in the CL3 and
679 CL4 clusters and in the native Asian range.
680
681 It could indeed also be that genes evolve under overdominance in this region for other reasons
682 than deleterious mutations, so that a partial linkage to a permanently heterozygous locus is
683 beneficial, perhaps in relationship with the pathogenic lifestyle of the fungus. One could

684 imagine, for example, that heterozygosity could be advantageous at genes involved in
685 virulence against the tree host, and especially the novel hosts colonized in invasive ranges, or
686 in resistance against a parasitic virus known to negatively affect fitness in *C. parasitica* (*60*,
687 *61*). As an alternative to a heterozygote advantage, the selection maintaining the
688 recombination suppression and the two haplotypes in the MAT-proximal region could be
689 some kind of negative-frequency dependent selection of beneficial allelic combinations,
690 possibly linked to a trench-warfare-like arms race with the host tree or the microbial
691 community, where partial linkage to the mating-type locus would help maintenance in
692 balanced proportions and therefore would prevent allele loss (*59*, *62*). Such a role in host-
693 pathogen interactions would also be consistent with the selective sweep footprints detected in
694 the MAT-proximal region if there is recurrent positive selection for improving the efficiency
695 of the pathogen weapons within each of one of the two MAT-proximal haplotypes. Balancing
696 selection and location adjacent to a supergene was recently found for loci involved in host
697 resistance in the *Daphnia–Pasteuria* system (*12*).
698
699 Another hypothesis for explaining such extension of recombination suppression beyond the
700 mating-type locus is antagonistic selection, i.e. linkage of alleles that would improve fitness
701 of a MAT-1 gamete while being deleterious in a MAT-2 gamete, or vice-versa. However, full
702 recombination suppression with the mating-type locus would be expected under such
703 antagonistic selection, as well as the same association between the MAT-Proximal alleles and
704 the mating-type in all populations, in contrast to our findings. In addition, we found no
705 particular predicted function in the MAT-proximal region that could be related to mating
706 compatibility and there are very little functions, if any, with possible antagonistic functions
707 between mating types in fungi (*8*, *63*).
708
709 **Conclusion**
710 In conclusion, we provide strong evidence for the existence of a non-recombining region
711 partially linked to the mating-type locus and maintained by selection in the chestnut blight
712 fungus *C. parasitica*. Such supergenes are increasingly reported in a variety of organisms
713 thanks to new sequencing technologies (*2*, *10*, *56*, *57*, *64*). This study therefore contributes to
714 our understanding of how selection for recombination suppression promotes the maintenance
715 of two divergent haplotypes despite degeneration. The resulting supergene in *C. parasitica*
716 displayed contrasting differentiation levels between haplotypes and different chromosomal
717 arrangements in native and invasive populations. We found strong selection footprints in the

718 MAT-proximal region in the first introduction of the pathogen in Europe from North America
719 and a chromosomal inversion. The supergene displayed lower differentiation between
720 haplotypes in a second invasion wave directly from Asia. This supergene may thus underlie
721 important adaptive traits and thereby provide important applications for the control of a
722 devastating tree pathogen. In addition, the high-quality genome assemblies provided here,
723 from the native and introduced ranges, will more generally be useful for studies aiming at
724 understanding the evolution of this pathogen.
725
726 **Material and Methods**
727
728 **Strains and genomic data**
729 For population genomic analyses, we studied a collection of 386 monokaryotic *C. parasitica*
730 strains sampled worldwide, from the native and introduced ranges of the pathogen, and
731 sequenced previously with the short-read Illumina technology (strain information are
732 presented in Table S1; (*23*, *24*, *31*)). Data were downloaded from NCBI Bioproject numbers
733 PRJNA604575, PRJNA644891 and PRJNA706885. We focused our analyses first on
734 European invasive strains originating from North America. We studied 88 strains belonging
735 to the CL1 genetic cluster, in central and southeastern Europe (*23*). We excluded the putative
736 clonal genotypes previously identified (*23*). We also studied 71 strains sampled in the 1990
737 and 62 strains sampled in 2019 in southern Switzerland (*24*). Mating-type ratios close to ½
738 and population structure analyses of the CL1 genetic cluster and the Swiss 1990 population
739 suggest frequent recombination in these populations (*23*, *24*). The mating-type ratio was 1/3
740 *MAT-1* - 2/3 *MAT-2* in the 2019 Swiss population and population structure analyses suggested
741 regular sexual reproduction and recent population bottleneck . To study more broadly the
742 presence of the two haplotypes in the MAT-proximal region, we analyzed monokaryotic
743 strains belonging to the CL2, CL3, CL4 genetic clusters, as well as the S12 European invasive
744 lineage (*23*) and additional monokaryotic strains from the US and Europe (with an Asian or
745 North American origin (*31*)). We excluded heterokaryotic strains, i.e. strains having both
746 MAT-1 and MAT-2 alleles, as the phase was challenging to infer.
747
748 For mapping and SNP calling, we first used as reference the 43.9 Mb genome sequence
749 EP155 v2.0 of the strain EP155 (*MAT-2,* North America, CL1 cluster) available at the Joint
750 Genome Institute (http://jgi.doe.gov/)(*32*). For comparative genomics, we used the published
751 genome of the strain ESM15 (MAT-2, Japan, CL2 cluster; (*33*)) available at

752 DDBJ/ENA/GenBank on the bioproject PRJNA700518 under the accession
753 JAGDFO000000000. We additionally sequenced *de novo,* with long-read technologies, five
754 strains from the native and introduced ranges of the pathogen. We sequenced with PacBio
755 Hifi the genomes of the strains M1400 (MAT-2) and M6697 (MAT-1) sampled in Gnosca,
756 southern Switzerland (*23*). Mycelia were stored as glycerol stocks at -80 C after strain
757 isolation. To produce mycelium for DNA extraction, isolates were inoculated onto
758 cellophane-covered potato dextrose agar plates (PDA, 39 g/L; BD Becton, Dickinson and
759 Company, Franklin Lakes, USA) (*65*) and incubated for a minimum of 1 week at 24°C, at a
760 14 hr light and 10 hr darkness cycle. Mycelium and spores were harvested by scratching the
761 mycelial mass off the cellophane, transferring it into 2 mL tubes and freeze-drying it for 24 hr
762 (*23*). DNA extraction was performed with the NucleoBond High Molecular Weight DNA kit
763 from Macherey-Nagel, with the mechanical disruption of about 30 mg of lyophilized
764 mycelium with two beads of 3 mm for 5 min at 30 Hz. Sequencing was outsourced to
765 Novogene, the Netherlands. We additionally sequenced with Oxford Nanopore MinION
766 technology the strains XIM9508 (China, CL4 cluster, MAT-1), MRC10 (South Western
767 France introduced directly from Asia, CL2 cluster, MAT-2) and DUM-005 (USA, CL1
768 cluster, MAT-2). These isolates had been collected for previous studies (*30*, *66*, *67*). The
769 protocols for mycelium production, DNA extraction and sequencing for these three strains
770 were the same as in (*33*).
771
772 **Short-read data processing and SNP calling**
773 We used SNP calling datasets of the genomes from monokaryotic strains of the CL1, CL2,
774 CL3, CL4 genetic clusters, the S12 invasive lineage and the Swiss populations against the
775 EP155 reference v2.0 genome obtained in (*23*, *24*). We performed mapping and raw SNP
776 calling of short-read data using the M1400 new genome assembly as a reference as described
777 in (*23*, *24*). Briefly, we trimmed reads with Trimmomatic v0.36 (*68*) and aligned them with
778 Bowtie 2 v2.3.5.1 (*69*) and SAMtools v1.9 (*70*) to the EP155 v2.0 genome. Raw SNP calling
779 and filtration for quality were conducted with the genome analysis toolkit GATK v3.8 and
780 v4.0.2.0 (*71*). We used the filtration parameters described in (*23*): QUAL>=100,
781 MQRankSum (lower) >= -2.0, QD:_20.0, MQRankSum (upper) <=2.0, MQ:_20.0,
782 BaseQRankSum (lower) >=-2.0, ReadPosRankSum (lower) >=-2.0, ReadPosRankSum
783 (upper) <=2.0, BaseQRankSum (upper) <=2.0. We further removed SNPs overlapping with
784 transposable elements predicted *de novo* in the EP155 v2.0 genome by (*23*).
785

786 **Population genomics analyses**

787 For all population genomics analyses, we excluded SNPs with missing data in more than 10%
788 of the strains and kept only polymorphic strains with vcftools v0.1.16 (*72*). To study linkage
789 disequilibrium, we further excluded rare variants (minor allele frequency <0.1) with the
790 vcftools (*72*) option --maf 0.1. We computed LD with the --hap-r2 option of vcftools (*72*) for
791 each scaffold and each population separately. We used the R package LDheatmap v1.0-6 (*73*)
792 to plot LD r^2 values among SNP pairs. To perform principal component analyses (PCAs), we
793 first used vcftools (*72*) to convert VCF format files in Plink format. We then used the Plink
794 v1.90b5.3 (*74*) –pca command to run PCA analysis. We used the R package PopGenome
795 v2.7.5 (*75*) to compute nucleotide diversity, Tajima's D values and the F_{ST} and d_{xy} index. We
796 used the R package ggplot2 (2_3.5.0) to plot results. Pairwise Wilcoxon tests were performed
797 in R with Bonferroni or false discovery rate correction.

798

799

800 We built neighbornet trees using SplitsTree4 v 4.19.2 (*76*). VCF file were converted to nexus
801 format using PGDSpider v1.0 tool (*77*). For the study of other populations, we included only
802 one strain of each haplotype of the invasive European population but not all CL1 and Swiss
803 strains for tree readability; we nevertheless checked that the results were the same with all
804 CL1 and Swiss strains.

805

806 **Long-read based assemblies**

807 PacBio Hifi reads of strains M1400 and M6697 were assembled using canu v1.8 (*78*) program
808 with a set genome size of 44 Mb. Oxford Nanopore MinION reads and Illumina reads of the
809 genomes of XIM9508 and MRC10 strains were assembled using HybridSPAdes (*79*).
810 Assemblies were manually curated as previously described (*33*) and scaffolds were cut when
811 an evidence of a chimeric connection was detected (i.e. mis-paired short reads). For
812 DUM005, the assembly was generated by Ra (https://github.com/lbcb-sci/ra) with basic
813 parameters as the Ra assembly quality outperformed the HybridSPades assembly quality (*33*).
814 Ra uses Oxford Nanopore MinION reads and corrects the assembly with Illumina reads
815 Assemblies statistics were obtained with quast v5.1 (*80*). We assessed the completeness of
816 each assembly using the Benchmarking of Universal Single-Copy Orthologue (BUSCO) tools
817 with the Sordariomyceta ortholog set (sordariomycetes_odb10, 2020-08-05, n = 3817
818 orthologous groups searched) (*81*). Gene models were predicted with the Helixer v0.3.1
819 pipeline (*82*). We also run Helixer pipeline for ESM15 and EP155 strains for the gene

820 orthology analysis. Statistics of the obtained gene annotation was obtained with the Agat
821 v1.0.0 tool (*83*). Transposable elements of all genomes were annotated using Repeatmasker
822 v4-0-7 (*84*) and the customized library built for *C. parasitica* in (*33*) contained in the
823 "Curated_TEs_database_ESM015_EP155.fa" file (available on the "Portail Data INRAE:
824 Chromosomal rearrangements but no change of genes and transposable elements repertoires
825 in an invasive forest-pathogenic fungus" at https://doi.org/10.15454/UTIB8U). The class
826 Gipsy invader was renamed LTR-Ty3. We filtered out TE copies shorter than 100bp.
827
828 **Estimation of retrotranposon insertion time and nucleotide divergence time**
829
830 To get an estimation of the insertion date of the transposable elements present in the MAT-
831 proximal non-recombining region, we applied two complementary methods. We first used the
832 divergence between the LTR sequences in retrotransposons, as these LTR sequences at their
833 edges are identical at the time of TE insertion and then diverge with time. For this, we used a
834 *de-novo* prediction of LTR retrotransposons using LTRharvest GenomeTools 1.6.2 (*85*, *86*)
835 and LTR_Finder v1.07 (*87*). To prepare data for LTRharvest, we first created enhanced suffix
836 for the M1400 and M6697 genome assemblies using the GenomeTools Suffixerator (-dna -suf
837 -lcp -bwt -bck -mirrored parameters). We ran LTRharvest using two pipelines, designed to
838 identify LTR retrotransposons with a minimum and maximum LTR length of 100 bp and
839 7000 bp respectively and at least 85% identity between two LTR regions, with and without
840 canonical TGCA motifs, respectively: i) -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6
841 -similar 85 -vic 10 -seed 20 -motif TGCA -motifmis 1; ii) -minlentltr 100, -maxlenltr 7000, -
842 mintsd 4, -maxtsd 6, -similar 85, -vic 10, -seed 20. Similarly, we ran LTR_Finder on the
843 M1400 and M6697 genome assemblies to retrotransposons with both TGCA and non-TGCA
844 motifs and a minimum and maximum LTR length of 100 bp and 7000 bp respectively and at
845 least 85% identity between two LTR regions (-D 15000, -d 1000, -l 100, -L 7000, -p 20, -C, -
846 M 0.85). Finally, we used LTR_retriever v2.8.5 (*88*) with default parameters to filter out false
847 positive LTR candidates identified by LTRharvest and LTR_Finder and get an estimation of
848 each element insertion date.
849
850 As a second method to estimate the insertion date of the transposable elements present in the
851 MAT-proximal non-recombining region, we used the set of curated consensus sequences from
852 (*33*) to annotate the inversion sequence or its surroundings. We first used samtools faidx v1.9
853 (*70*) to extract the sequence corresponding to inversion and their 50 kb surroundings in
854 M1400 and M6697 genome assemblies. We annotated the sequences corresponding to the

855 inversions and the concatenated 50 kb surroundings in both isolates using RepeatMasker
856 version 4.1.5 and rmblast as search engine (v2.10.0) with -no_is -pa 20 -cutoff 250 -a
857 parameters. Finally, we parsed the RepeatMasker .out file using the helped script
858 parseRM_merge_interrupted.pl and omitting Simple_repeat and Low_complexity regions
859 (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs). We then built a summary of the
860 alignments using the RepeatMasker helper script buildSummary.pl and calculated sequence
861 divergence using the calcDivergenceFromAlign.pl script to finally render results with the
862 createRepeatLandscape.pl from the same helper suite.
863
864
865 **Comparative genomics analyses**
866 Genome synteny between long reads assemblies were studied using the nucmer v3.1 program
867 (*89*). Outputs were plotted with the R programs ggplot2 (*90*), genoPlotR (*91*) and RIdeogram
868 (*92*).
869
870 For MRC010, we investigate the rearrangements by mapping the longest reads (>10 kb) from
871 the ONT fastq files using minimap2 (-ax map-ont option) (*93*). Reads were sorted and
872 indexed with samtools (*70*) and Jbrowse2 (*94*) was used for displaying the pileups. We parsed
873 the reads mapping to the different contigs of the MAT-proximal region (i.e. split reads) and
874 obtained the overlaps again with minimap (-x ava-ont option) and plot with pafr to visually
875 study the pileup of split reads. A total of 56 reads join contigs MRC010_10 and MRC010_15
876 on the same orientation as presented in Fig 6A. We found no split reads with contig
877 MRC010_01 but we found telomeric repeats in the same orientation on contig MRC010_15
878 as the telomeric repeats in contig MRC010_01, supporting that mating-type chromosome of
879 MRC010 starts with contig MRC010_01 and finishes with contig MRC010_15. For contig
880 MRC010_14 reads mapping at both ends of this contig also map on contig MRC010_10,
881 indicating it should be in between MRC010_01 and MRC010_10, but its orientation was not
882 resolved.
883
884 To study gene disruption, synonymous and non-synonymous divergence (dS and dN) and
885 testing introgression in the MAT-proximal region, we build orthology relationships for genes
886 of the genome assemblies of *C. parasitica* strains M1400, M6697, XIM9508, MRC10, DUM-
887 005, ESM15 and EP155 and included three genomes of the closely related species
888 *Cryphonectria japonica* (IF-6), *Cryphonectria carpinicola* (CS3), and *Cryphonectria*

889 *naterciae* (M3664). Genome, gene annotation and species tree of these closely related species
890 were previously published by (*43*). Genome data were retrieved from NCBI bioproject
891 number PRJNA644891 and accession IDs JACWRX000000000 for IF-6,
892 JACWRQ000000000 for CS3 and JACWST000000000 for M3664. We run OrthoFinder
893 v2.3.7 (*95*) analysis on protein sequences. We used the translatorX v1.1 program (*96*) with
894 default parameters that use a codon-based approach to align orthologous gene coding
895 sequences. To compute dS and dN vaues of one-to-one orthologs between pairs of genome
896 assemblies, we use the yn00 program of PAML (*97*). Estimation of divergence time between
897 haplotypes was performed using computed gene dS values and the formula Tgenerations =
898 dS/2μ. We used previous estimates of substitution rates in fungi (*41*, *42*) and considered that
899 *C. parasitica* undergoes one generation a year (*98*). We build gene coding sequences trees
900 with the outgroup genomes in the MAT-proximal region using iqtree2 v2.2.2.6 (*99*) with 1000
901 bootraps and used the Newick Utilities (https://github.com/tjunier/newick_utils) for
902 displaying phylogenetics tree.
903
904 **References**

905 1. S. P. Otto, T. Lenormand, Resolving the paradox of sex and recombination. *Nat. Rev.*
906 *Genet.* **3**, 252–261 (2002).

907 2. T. Schwander, R. Libbrecht, L. Keller, Supergenes and complex phenotypes. *Curr. Biol.*
908 **24**, R288–R294 (2014).

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

911 4. R. Bergero, D. Charlesworth, The evolution of restricted recombination in sex
912 chromosomes. *Trends Ecol. Evol.* **24**, 94–102 (2009).

913 5. J. G. Umen, Evolution of sex and mating loci: An expanded view from Volvocine algae.
914 *Curr. Opin. Microbiol.* **14**, 634–641 (2011).

915 6. D. Charlesworth, The status of supergenes in the 21st century: recombination
916 suppression in Batesian mimicry and sex chromosomes and other complex adaptations.
917 *Evol. Appl.* **9**, 74–90 (2016).

918 7. Y. Dussert, L. Legrand, I. D. Mazet, C. Couture, M.-C. Piron, R.-F. Serre, O. Bouchez,
919 P. Mestre, S. L. Toffolatti, T. Giraud, F. Delmotte, Identification of the first oomycete
920 mating-type locus sequence in the grapevine downy mildew pathogen, *Plasmopara*
921 *viticola*. *Curr. Biol.* **30**, 3897-3907.e4 (2020).

922 8. F. E. Hartmann, M. Duhamel, F. Carpentier, M. E. Hood, M. Foulongne-Oriol, P. Silar,
923 F. Malagnac, P. Grognet, T. Giraud, Recombination suppression and evolutionary strata

924 around mating-type loci in fungi: documenting patterns and understanding evolutionary
925 and mechanistic causes. *New Phytol.* **229**, 2470–2491 (2021).

926 9. C. Küpper, M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell, S. B. McRae, T. C.
927 Morgan, N. Karlionova, P. Pinchuk, Y. I. Verkuil, A. S. Kitaysky, J. C. Wingfield, T.
928 Piersma, K. Zeng, J. Slate, M. Blaxter, D. B. Lank, T. Burke, A supergene determines
929 highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).

930 10. Z. Yan, S. H. Martin, D. Gotzek, S. V. Arsenault, P. Duchen, Q. Helleu, O. Riba-
931 Grognuz, B. G. Hunt, N. Salamin, D. Shoemaker, K. G. Ross, L. Keller, Evolution of a
932 supergene that regulates a trans-species social polymorphism. *Nat. Ecol. Evol.* **4**, 240–
933 249 (2020).

934 11. P. Jay, M. Chouteau, A. Whibley, H. Bastide, H. Parrinello, V. Llaurens, M. Joron,
935 Mutation load at a mimicry supergene sheds new light on the evolution of inversion
936 polymorphisms. *Nat. Genet.* **53**, 288–293 (2021).

937 12. M. Fredericksen, P. D. Fields, L. D. Pasquier, V. Ricci, D. Ebert, QTL study reveals
938 candidate genes underlying host resistance in a Red Queen model system. *PLOS Genet.*
939 **19**, e1010570 (2023).

940 13. D. Bachtrog, Y-chromosome evolution: emerging insights into processes of Y-
941 chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).

942 14. F. Carpentier, R. C. Rodríguez de la Vega, P. Jay, M. Duhamel, J. A. Shykoff, M. H.
943 Perlin, R. M. Wallen, M. E. Hood, T. Giraud, Tempo of degeneration across
944 independently evolved nonrecombining regions. *Mol. Biol. Evol.* **39**, msac060 (2022).

945 15. M. Duhamel, M. E. Hood, R. C. R. de la Vega, T. Giraud, Dynamics of transposable
946 element accumulation in the non-recombining regions of mating-type chromosomes in
947 anther-smut fungi. bioRxiv [Preprint] (2023).
948 https://doi.org/10.1101/2022.08.03.502670.

949 16. E. L. Berdan, A. Blanckaert, R. K. Butlin, T. Flatt, T. Slotte, B. Wielstra, Mutation
950 accumulation opposes polymorphism: supergenes and the curious case of balanced
951 lethals. *Philos. Trans. R. Soc. B Biol. Sci.* **377**, 20210199 (2022).

952 17. J. E. Ironside, No amicable divorce? Challenging the notion that sexual antagonism
953 drives sex chromosome evolution. *BioEssays* **32**, 718–726 (2010).

954 18. B. L. S. Furman, D. C. H. Metzger, I. Darolti, A. E. Wright, B. A. Sandkam, P. Almeida,
955 J. J. Shu, J. E. Mank, Sex chromosome evolution: So many exceptions to the rules.
956 *Genome Biol. Evol.*, doi: 10.1093/gbe/evaa081 (2020).

957 19. P. Jay, E. Tezenas, A. Véber, T. Giraud, Sheltering of deleterious mutations explains the
958 stepwise extension of recombination suppression on sex chromosomes and other
959 supergenes. *PLOS Biol.* **20**, e3001698 (2022).

960 20. B. Charlesworth, The fitness consequences of genetic divergence between polymorphic
961 gene arrangements. *Genetics*, iyad218 (2023).

962 21. P. Jay, D. L. Jeffries, F. E. Hartmann, A. Véber, T. Giraud, Why do sex chromosomes
963 progressively lose recombination? *Trends in Genetics*, in press (2024).

964 22. T. L. Kubisiak, M. G. Milgroom, Markers linked to vegetative incompatibility (vic)
965 genes and a region of high heterogeneity and reduced recombination near the mating
966 type locus (MAT) in *Cryphonectria parasitica*. *Fungal Genet. Biol.* **43**, 453–463 (2006).

967 23. L. Stauber, T. Badet, A. Feurtey, S. Prospero, D. Croll, Emergence and diversification of
968 a highly invasive chestnut pathogen lineage across southeastern Europe. *eLife* **10**,
969 e56279 (2021).

970 24. L. Stauber, D. Croll, S. Prospero, Temporal changes in pathogen diversity in a perennial
971 plant–pathogen–hyperparasite system. *Mol. Ecol.* **31**, 2073–2088 (2022).

972 25. I. C. McGuire, R. E. Marra, B. G. Turgeon, M. G. Milgroom, Analysis of mating-Type
973 genes in the chestnut blight fungus, *Cryphonectria parasitica*. *Fungal Genet. Biol.* **34**,
974 131–144 (2001).

975 26. I. C. McGuire, R. E. Marra, M. G. Milgroom, Mating-type heterokaryosis and selfing in
976 *Cryphonectria parasitica*. *Fungal Genet. Biol.* **41**, 521–533 (2004).

977 27. I. C. McGuire, J. E. Davis, M. L. Double, W. L. MacDonald, J. T. Rauscher, S.
978 McCawley, M. G. Milgroom, Heterokaryon formation and parasexual recombination
979 between vegetatively incompatible lineages in a population of the chestnut blight fungus,
980 *Cryphonectria parasitica*. *Mol. Ecol.* **14**, 3657–3669 (2005).

981 28. C. Dutech, O. Fabreguettes, X. Capdevielle, C. Robin, Multiple introductions of
982 divergent genetic lineages in an invasive fungal pathogen, Cryphonectria parasitica, in
983 France. *Heredity* **105**, 220–228 (2010).

984 29. S. L. Anagnostakis, Chestnut blight: The classical problem of an introduced pathogen.
985 *Mycologia* (1987).

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

989 31. A. Demené, L. Legrand, J. Gouzy, R. Debuchy, G. Saint-Jean, O. Fabreguettes, C.
990 Dutech, Whole-genome sequencing reveals recent and frequent genetic recombination
991 between clonal lineages of Cryphonectria parasitica in western Europe. *Fungal Genet.*
992 *Biol.* **130**, 122–133 (2019).

993 32. J. A. Crouch, A. Dawe, A. Aerts, K. Barry, A. C. L. Churchill, J. Grimwood, B. I.
994 Hillman, M. G. Milgroom, J. Pangilinan, M. Smith, A. Salamov, J. Schmutz, J. S.
995 Yadav, I. V. Grigoriev, D. L. Nuss, Genome sequence of the chestnut blight fungus
996 Cryphonectria parasitica EP155: A fundamental resource for an archetypical invasive
997 plant pathogen. *Phytopathology®* **110**, 1180–1188 (2020).

998 33. A. Demené, B. Laurent, S. Cros-Arteil, C. Boury, C. Dutech, Chromosomal
999 rearrangements with stable repertoires of genes and transposable elements in an invasive
1000 forest-pathogenic fungus. *Peer Community J.* **2** (2022).

1001 34. T. L. Kubisiak, M. G. Milgroom, Markers linked to vegetative incompatibility (*vic*)
1002 genes and a region of high heterogeneity and reduced recombination near the mating
1003 type locus (*MAT*) in *Cryphonectria parasitica*. *Fungal Genet. Biol. FG B* **43**, 453–463
1004 (2006).

1005 35. C. Dufresnes, A. Borzée, A. Horn, M. Stöck, M. Ostini, R. Sermier, J. Wassef, S. N.
1006 Litvinchuck, T. A. Kosch, B. Waldman, Y. Jang, A. Brelsford, N. Perrin, Sex-
1007 chromosome homomorphy in palearctic tree frogs results from both turnovers and X-Y
1008 recombination. *Mol. Biol. Evol.* **32**, 2328–2337 (2015).

1009 36. M. G. Milgroom, M. L. Smith, M. T. Drott, D. L. Nuss, Balancing selection at nonself
1010 recognition loci in the chestnut blight fungus, *Cryphonectria parasitica*, demonstrated
1011 by trans-species polymorphisms, positive selection, and even allele frequencies.
1012 *Heredity* **121**, 511–523 (2018).

1013 37. Y. Sun, J. Svedberg, M. Hiltunen, P. Corcoran, H. Johannesson, Large-scale suppression
1014 of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat.*
1015 *Commun.* **8**, 1–8 (2017).

1016 38. A. Menkis, D. J. Jacobson, T. Gustafsson, H. Johannesson, The mating-type
1017 chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model
1018 for early evolution of sex chromosomes. *PLOS Genet.* **4**, e1000030 (2008).

1019 39. P. Grognet, F. Bidard, C. Kuchly, L. C. H. Tong, E. Coppin, J. A. Benkhali, A. Couloux,
1020 P. Wincker, R. Debuchy, P. Silar, Maintaining two mating types: structure of the mating
1021 type locus and its role in heterokaryosis in *Podospora anserina*. *Genetics* **197**, 421–432
1022 (2014).

1023 40. N. Vittorelli, Hartmann, T. Giraud, Stepwise recombination suppression around the
1024 mating-type locus in the fungus *Schizothecium tetrasporum* (Ascomycota, Sordariales).
1025 (2022).

1026 41. T. Kasuga, T. J. White, J. W. Taylor, Estimation of nucleotide substitution rates in
1027 eurotiomycete fungi. *Mol. Biol. Evol.* **19**, 2318–2324 (2002).

1028 42. J. W. Taylor, M. L. Berbee, Dating divergences in the Fungal Tree of Life: review and
1029 new analyses. *Mycologia* **98**, 838–849 (2006).

1030 43. L. Stauber, S. Prospero, D. Croll, Comparative genomics analyses of lifestyle transitions
1031 at the origin of an invasive fungal pathogen in the genus *Cryphonectria*. *mSphere* **5**,
1032 10.1128/msphere.00737-20 (2020).

1033 44. S. Branco, F. Carpentier, R. C. Rodríguez de la Vega, H. Badouin, A. Snirc, S. L. Prieur,
1034 M. A. Coelho, D. M. de Vienne, F. E. Hartmann, D. Begerow, M. E. Hood, T. Giraud,
1035 Multiple convergent supergene evolution events in mating-type chromosomes. *Nat.*
1036 *Commun.* **9**, 2000 (2018).

1037 45. L. Maloisel, J. L. Rossignol, Suppression of crossing-over by DNA methylation in
1038 *Ascobolus*. *Genes Dev.* **12**, 1381–1389 (1998).

1039 46. F. Boideau, G. Richard, O. Coriton, V. Huteau, C. Belser, G. Deniot, F. Eber, C.
1040 Falentin, J. Ferreira de Carvalho, M. Gilet, M. Lodé-Taburel, L. Maillet, J. Morice, G.

Trotoux, J.-M. Aury, A.-M. Chèvre, M. Rousseau-Gueutin, Epigenomic and structural events preclude recombination in Brassica napus. *New Phytol.* **234**, 545–559 (2022).

47. C. L. Peichel, S. R. McCann, J. A. Ross, A. F. S. Naftaly, J. R. Urton, J. N. Cech, J. Grimwood, J. Schmutz, R. M. Myers, D. M. Kingsley, M. A. White, Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* **21**, 177 (2020).

48. N. Hu, Evolution of a ZW sex chromosome system in willows. *Nat. Commun.* (2023).

49. C. Moraga, C. Branco, Q. Rougemont, P. Veltsos, P. Jedlička, A. Muyle, M. Hanique, E. Tannier, X. Liu, E. Mendoza-Galindo, C. Lemaitre, P. D. Fields, C. Cruaud, K. Labadie, C. Belser, J. Briolay, S. Santoni, R. Cegan, R. Linheiro, R. C. R. de la Vega, G. Adam, A. E. Filali, V. Mossion, A. Boualem, R. Tavares, A. Chebbi, R. Cordaux, C. Fruchard, D. Prentout, A. Velt, B. Spataro, S. Delmotte, L. Weingartner, H. Toegelová, Z. Tulpová, P. Cápal, H. Šimková, H. Štorchová, M. Krüger, O. A. J. Abeyawardana, D. R. Taylor, M. S. Olson, D. B. Sloan, S. Karrenberg, L. F. Delph, D. Charlesworth, T. Giraud, A. Bendahmane, A. D. Genova, A. Madoui, R. Hobza, G. A. B. Marais, The Silene latifolia genome and its giant Y chromosome. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.09.21.558754.

50. S. Branco, H. Badouin, R. C. Rodríguez de la Vega, J. Gouzy, F. Carpentier, G. Aguileta, S. Siguenza, J.-T. Brandenburg, M. A. Coelho, M. E. Hood, T. Giraud, Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc. Natl. Acad. Sci.* **114**, 7067–7072 (2017).

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

52. O. S. Harringmeyer, H. E. Hoekstra, Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nat. Ecol. Evol.* **6**, 1965–1979 (2022).

53. K. Huang, K. L. Ostevik, C. Elphinstone, M. Todesco, N. Bercovich, G. L. Owens, L. H. Rieseberg, Mutation load in sunflower inversions is negatively correlated with inversion heterozygosity. *Mol. Biol. Evol.* **39**, msac101 (2022).

54. M. Matschiner, J. M. I. Barth, O. K. Tørresen, B. Star, H. T. Baalsrud, M. S. O. Brieuc, C. Pampoulie, I. Bradbury, K. S. Jakobsen, S. Jentoft, Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* **6**, 469–481 (2022).

55. J. Hill, E. D. Enbody, H. Bi, S. Lamichhaney, W. Lei, J. Chen, C. Wei, Y. Liu, D. Schwochow, S. Younis, F. Widemo, L. Andersson, Low mutation load in a supergene underpinning alternative male mating strategies in ruff (*Calidris pugnax*). *Mol. Biol. Evol.* **40**, msad224 (2023).

56. P. Jay, A. Whibley, L. Frézal, M. Á. Rodríguez de Cara, R. W. Nowell, J. Mallet, K. K. Dasmahapatra, M. Joron, Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr. Biol.* **28**, 1839-1845.e3 (2018).

57. Q. Helleu, C. Roux, K. G. Ross, L. Keller, Radiation and hybridization underpin the spread of the fire ant social supergene. *Proc. Natl. Acad. Sci.* **119**, e2201040119 (2022).

1082 58. E. Stolle, R. Pracana, F. López-Osorio, M. K. Priebe, G. L. Hernández, C. Castillo-
1083 Carrillo, M. C. Arias, C. I. Paris, M. Bollazzi, A. Priyam, Y. Wurm, Recurring adaptive
1084 introgression of a supergene variant that determines social organization. *Nat. Commun.*
1085 **13**, 1180 (2022).

1086 59. P. Jay, T. G. Aubier, M. Joron, The interplay of local adaptation and gene flow may lead
1087 to the formation of supergenes. *Mol. Ecol.*, e17297 (2024).

1088 60. G. H. Choi, D. L. Nuss, Hypovirulence of chestnut blight fungus conferred by an
1089 infectious viral cDNA. *Science* **257**, 800–803 (1992).

1090 61. J. Brusini, M. L. Wayne, A. Franc, C. Robin, The impact of parasitism on resource
1091 allocation in a fungal host: the case of *Cryphonectria parasitica* and its mycovirus,
1092 Cryphonectria Hypovirus 1. *Ecol. Evol.* **7**, 5967–5976 (2017).

1093 62. A. Tellier, S. Moreno-Gámez, W. Stephan, Speed of adaptation and genomic footprints
1094 of host–parasite coevolution under arms race and trench warfare dynamics. *Evolution*
1095 **68**, 2211–2224 (2014).

1096 63. A. L. Bazzicalupo, F. Carpentier, S. P. Otto, T. Giraud, Little evidence of antagonistic
1097 selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum*
1098 *lychnidis-dioicae*). *G3 Genes Genomes Genet.* **9**, 1987–1998 (2019).

1099 64. J. Gutiérrez-Valencia, P. W. Hughes, E. L. Berdan, T. Slotte, The genomic architecture
1100 and evolutionary fates of supergenes. *Genome Biol. Evol.* **13** (2021).

1101 65. P. J. Hoegger, D. Rigling, O. Holdenrieder, U. Heiniger, Genetic structure of newly
1102 established populations of *Cryphonectria parasitica*. *Mycol. Res.* **104**, 1108–1116
1103 (2000).

1104 66. M. G. Milgroom, K. Wang, Y. Zhou, S. E. Lipari, S. Kaneko, Intercontinental
1105 population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia*
1106 **88**, 179–190 (1996).

1107 67. Y. C. Liu, D. Linder-Basso, B. I. Hillman, S. Kaneko, M. G. Milgroom, Evidence for
1108 interspecies transmission of viruses in natural populations of filamentous fungi in the
1109 genus Cryphonectria. *Mol. Ecol.* **12**, 1619–1628 (2003).

1110 68. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina
1111 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

1112 69. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient
1113 alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

1114 70. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G.
1115 Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools.
1116 *Bioinformatics* **25**, 2078–2079 (2009).

1117 71. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K.
1118 Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis
1119 Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.
1120 *Genome Res.* **20**, 1297–1303 (2010).

1121 72. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E.
1122 Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes
1123 Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–
1124 2158 (2011).

1125 73. J.-H. Shin, S. Blay, B. McNeney, J. Graham, others, LDheatmap: an R function for
1126 graphical display of pairwise linkage disequilibria between single nucleotide
1127 polymorphisms. *J. Stat. Softw.* **16**, 1–10 (2006).

1128 74. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J.
1129 Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for
1130 whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*
1131 **81**, 559–575 (2007).

1132 75. B. Pfeifer, U. Wittelsbürger, S. E. Ramos-Onsins, M. J. Lercher, PopGenome: an
1133 efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**,
1134 1929–1936 (2014).

1135 76. D. H. Huson, SplitsTree: analyzing and visualizing evolutionary data. *Bioinforma. Oxf.*
1136 *Engl.* **14**, 68–73 (1998).

1137 77. H. E. L. Lischer, L. Excoffier, PGDSpider: an automated data conversion tool for
1138 connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299
1139 (2012).

1140 78. S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, Canu:
1141 scalable and accurate long-read assembly via adaptive k-mer weighting and repeat
1142 separation. *Genome Res.* **27**, 722–736 (2017).

1143 79. D. Antipov, A. Korobeynikov, J. S. McLean, P. A. Pevzner, hybridSPAdes: an algorithm
1144 for hybrid assembly of short and long reads. *Bioinformatics* **32**, 1009–1015 (2016).

1145 80. A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for
1146 genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

1147 81. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, E. M. Zdobnov, BUSCO update:
1148 Novel and streamlined workflows along with broader and deeper phylogenetic coverage
1149 for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–
1150 4654 (2021).

1151 82. F. Holst, A. Bolger, C. Günther, J. Maß, S. Triesch, F. Kindel, N. Kiel, N. Saadat, O.
1152 Ebenhöh, B. Usadel, R. Schwacke, M. Bolger, A. P. M. Weber, A. K. Denton, Helixer–
1153 de novo prediction of primary Eukaryotic gene models combining deep learning and a
1154 hidden markov model. bioRxiv [Preprint] (2023).
1155 https://doi.org/10.1101/2023.02.06.527280.

1156 83. J. Dainat, D. Hereñú, P. Pucholt, AGAT:Another Gff Analysis Toolkit to handle
1157 annotations in any GTF. *GFF Format* (2020).

1158 84. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. Available at:
1159 http://www.repeatmasker.org Accessed: November 30, 2017. (2013).

1160 85. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for
1161 de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).

1162 86. G. Gremme, S. Steinbiss, S. Kurtz, GenomeTools: A Comprehensive Software Library
1163 for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans. Comput.*
1164 *Biol. Bioinform.* **10**, 645–656 (2013).

1165 87. Z. Xu, H. Wang, LTR_FINDER: an efficient tool for the prediction of full-length LTR
1166 retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).

1167 88. S. Ou, N. Jiang, LTR_retriever: A highly accurate and sensitive program for
1168 identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422
1169 (2018).

1170 89. G. Marçais, A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, A. Zimin,
1171 MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**,
1172 e1005944 (2018).

1173 90. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer Science &
1174 Business Media, 2009).

1175 91. L. Guy, J. Roat Kultima, S. G. E. Andersson, genoPlotR: comparative gene and genome
1176 visualization in R. *Bioinformatics* **26**, 2334–2335 (2010).

1177 92. Z. Hao, D. Lv, Y. Ge, J. Shi, D. Weijers, G. Yu, J. Chen, RIdeogram: drawing SVG
1178 graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.*
1179 **6**, e251 (2020).

1180 93. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**,
1181 3094–3100 (2018).

1182 94. C. Diesh, G. J. Stevens, P. Xie, T. De Jesus Martinez, E. A. Hershberg, A. Leung, E.
1183 Guo, S. Dider, J. Zhang, C. Bridge, G. Hogue, A. Duncan, M. Morgan, T. Flores, B. N.
1184 Bimber, R. Haw, S. Cain, R. M. Buels, L. D. Stein, I. H. Holmes, JBrowse 2: a modular
1185 genome browser with views of synteny and structural variation. *Genome Biol.* **24**, 74
1186 (2023).

1187 95. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative
1188 genomics. *Genome Biol.* **20**, 238 (2019).

1189 96. F. Abascal, R. Zardoya, M. J. Telford, TranslatorX: multiple alignment of nucleotide
1190 sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7-13 (2010).

1191 97. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**,
1192 1586–1591 (2007).

1193 98. L. Guerin, G. Froidefond, X.-M. Xu, Seasonal patterns of dispersal of ascospores of
1194 Cryphonectria parasitica (chestnut blight). *Plant Pathol.* **50**, 717–724 (2001).

1195 99. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von
1196 Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic
1197 inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

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. The authors declare that they have no competing interests.

TG and FEH conceptualized the study and acquired funding. TG supervised the study. AL, AS, CD, AD, SP, DC, LS and TB contributed to data and their availability. AL and AS performed *in vitro* culture and DNA extraction. FEH, RRDLV, QR, LS and TB analyzed genomes. FEH and TG wrote the original draft. All authors edited the manuscript.

The raw data and the new assemblies produced in this study will be published pending scientific review.

Tables

Table 1: Distribution of mating types among non-recombining haplotypes of the MAT-proximal region in the European invasive *Cryphonectria parasitica* CL1 genetic cluster and the Swiss populations.

Figures

Figure 1: Linkage disequilibrium (LD) analysis along the contig carrying the mating-type locus in a *Cryphonectria parasitica* European invasive population. LD heatmaps using single nucleotide polymorphisms (SNPs; n=3815) located on the contig carrying the mating-type locus (scaffold_2 of the EP155 genome) in the CL1 genetic cluster (European invasive population introduced from North America); pairs of SNPs with high LD, i.e. r^2 >0.9, correspond to the red triangle. The mating-type locus location is shown with a green triangle and the MAT-proximal region lacking recombination is shown with a red arrow. The two high-LD blocks within the MAT-proximal region are shown with orange arrows. SNPs at the limit of the MAT-proximal region and the two high LD blocks were manually highlighted with red dotted lines and an orange line.

Figure 2: Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-proximal region lacking recombination and other genomic regions in a *Cryphonectria*

1233 ***parasitica* European invasive population. A-C.** Principal component analysis (PCA). Two
1234 principal components are presented. Percentage of variance explained by each PC is indicated
1235 into brackets. Strains are colored according to their mating type (*MAT1-1* or *MAT1-2*). **D-E.**
1236 Neighbor-net tree from a SplitsTree analysis. These analyses were performed in the CL1
1237 genetic cluster (European invasive populations introduced from North America) based on: **A**
1238 **and D.** SNPs (n=2,220) located within the MAT-proximal region along the contig carrying
1239 the mating-type locus (scaffold_2 of the EP155 genome).; **B.** SNPs (n=1,595) located in other
1240 regions along the contig carrying the mating-type locus; **C and E** SNPs (n=11,289) located on
1241 other contigs of the EP155 genome. On panels A and D, the two clusters corresponding to the
1242 MAT-Prox1 and MAT-Prox2 haplotypes are shown with red circles. The identified haplotype
1243 of each strain is indicated in Table S1. The number of strains within the MAT-Prox1
1244 haplotype is indicated on panel A by the letter n. The newly sequenced M1400 and M6697
1245 strains are highlighted with a black rectangle
1246
1247 **Figure 3: Synteny and rearrangements between the two newly sequenced genomes of**
1248 **European invasive strains introduced from North America (M1400 and M6697) in the**
1249 **chromosome carrying the MAT-proximal region lacking recombination in**
1250 ***Cryphonectria parasitica*.** The mating-type locus is located with a green diamond. The MAT-
1251 proximal region defined from LD analyses is indicated with red arrows.
1252
1253 **Figure 4: Genetic diversity and divergence between non-recombining haplotypes of the**
1254 **MAT-proximal region along the contig carrying the mating-type locus in a**
1255 ***Cryphonectria parasitica* European invasive population.** A-B. GC content (A) and
1256 transposable element densities (B) along the contig carrying the mating-type locus
1257 (tig_0000001) of the M1400 *C. parasitica* genome. C. Relative divergence (F_{ST}) between
1258 strains of the MAT-Prox1 and MAT-Prox2 haplotypes: D. Absolute divergence d_{XY} between
1259 strains of the two MAT-Proximal haplotypes. E-F. Nucleotide diversity within pools of strains
1260 for each MAT-Proximal haplotype; G-H-I. Tajima's D for all strains pooled and within pools
1261 of strains for each MAT-Proximal haplotype; The MAT-proximal region defined from LD
1262 analyses and the inversion between M1400 and M6697 genomes are indicated with red and
1263 blue arrows respectively. The mating-type locus location is shown with a green triangle. The
1264 location of the putative centromere is indicated by a yellow circle. All population statistics
1265 were computed for the 1990 Swiss population along the mating-type contig (tig_0000001) of

1266 the M1400 genome per 50-kb window overlapping over 10 kbp. Windows containing fewer
1267 than 5 SNPs were removed from the analysis.
1268
1269 **Figure 5: Annotation of transposable elements (TEs) and estimates of their insertion**
1270 **time in the MAT-proximal region lacking recombination and other genomic regions in**
1271 ***Cryphonectria parasitica*.** A. Annotation of transposable elements (TEs) and gene density
1272 along the MAT-proximal region in M1400 and M6697 genomes. The figure only shows the
1273 genomic region from 7 Mb on tig_0000001 of the M1400 genome and from 2.8 Mb on
1274 tig_0000060 of the M6697 genome. The mating-type locus is located with a green triangle.
1275 The two high-LD blocks within the MAT-proximal region are shown with orange arrows.
1276 Synteny for 10 kb segments with identity > 90 % is shown in red and inversion in blue.
1277 Transposons larger than 5 kb are shown in purple for the transposons annotated as LTR-Ty3
1278 by and in green for other transposons. Orthologous genes shared between genomes are shown
1279 in black and unique to each each genome in grey. B. Relative percentage of each TE
1280 annotation (in %) in the non-recombining MAT-proximal region and other recombining
1281 regions of M6697 and M1400 genomes. C. Pairwise genetic distance between TE copies
1282 within the inversion and around (50 kb) in both M1400 and M6697 genomes. Distribution of
1283 the Kimura substitution levels computed using the consensus sequence for the TEs. D.
1284 Estimates of the age of intact copies of LTR within and around the inversion.
1285
1286 **Figure 6: Transposable elements content and gene content variation in seven high**
1287 **quality genome assemblies from other *Cryphonectria parasitica* populations from the**
1288 **native and the introduced range. A.** TE load (percentage of base pairs occupied by
1289 transposable elements) in the seven high quality genome assemblies in the MAT proximal
1290 region and other recombining regions. **B.** Number of orthologous genes present in the seven
1291 high quality genome assemblies. Only single copy orthologous genes shared between M6697
1292 and M1400 or present only in one this genome are shown.
1293
1294 **Figure 7: Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-**
1295 **proximal region lacking recombination (A) and other genomic regions (B) in other**
1296 **resequenced *Cryphonectria parasitica* populations from the native and the introduced**
1297 **range.** Neighbor-net tree from a SplitsTree analysis based on: **A** SNPs (n=4,120) located
1298 within the MAT-proximal region along the contig carrying the mating-type locus (scaffold_2
1299 of the EP155 genome); **B** SNPs (n= 103,058) located on other contigs of the EP155 genome.

1300 Color of empty circles around strain ID indicate the genetic clusters the strains belong to
1301 (CL1, CL2, CL3, CL4; n= 33 strains). Dotted circles indicate strains of the native range and
1302 plain circles indicate strains of the introduced range. Red and blue dots near strain ID indicate
1303 the mating type (*MAT-1* or *MAT-2*, respectively).
1304
1305

1306 Table 1

1307

	Population	CL1 genetic cluster		Swiss 1990 population		Swiss 2019 population	
	PCA cluster	PCA cluster 1 (PC>0)	PCA cluster 2 (PC<0)	PCA cluster 1 (PC>0)	PCA cluster 2 (PC<0)	PCA cluster 1 (PC>0)	PCA cluster 2 (PC<0)
	Haplotype (based on PCA cluster)	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
% of strains in each PCA cluster	mat-1	83	21	90	33	63	28
	mat-2	17	79	10	67	38	72

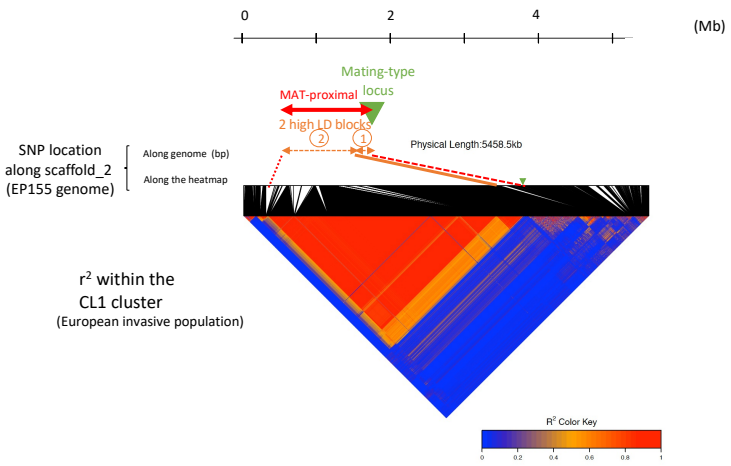
1308

1309

1310

1311 Fig 1

1312



1313

1314

Fig 2

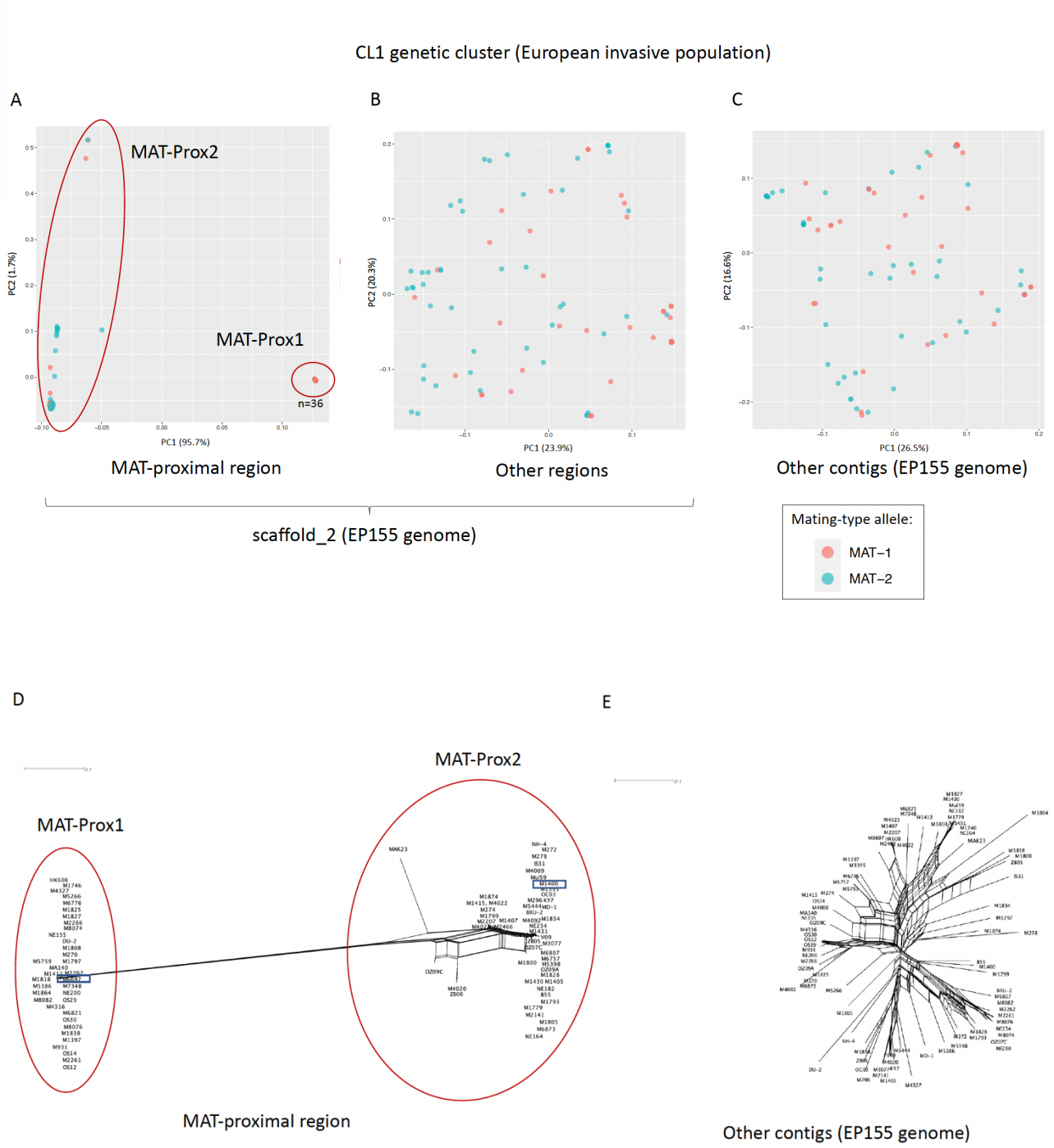
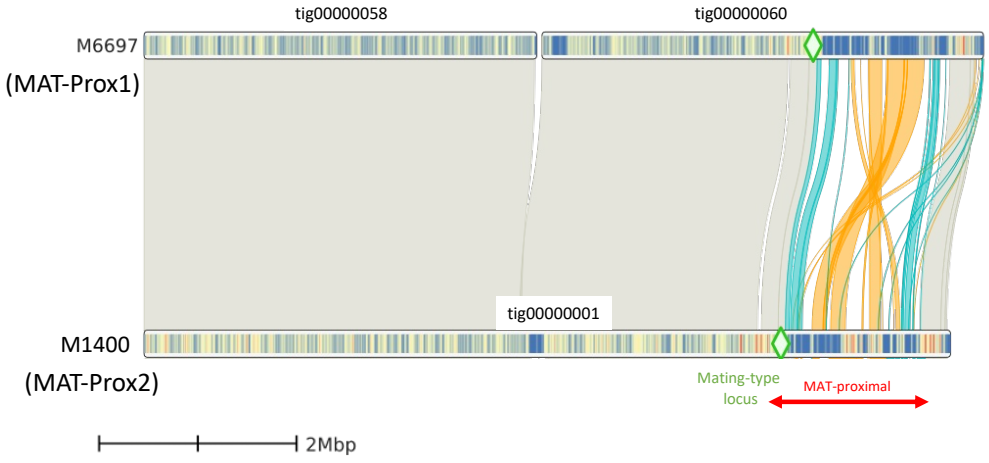
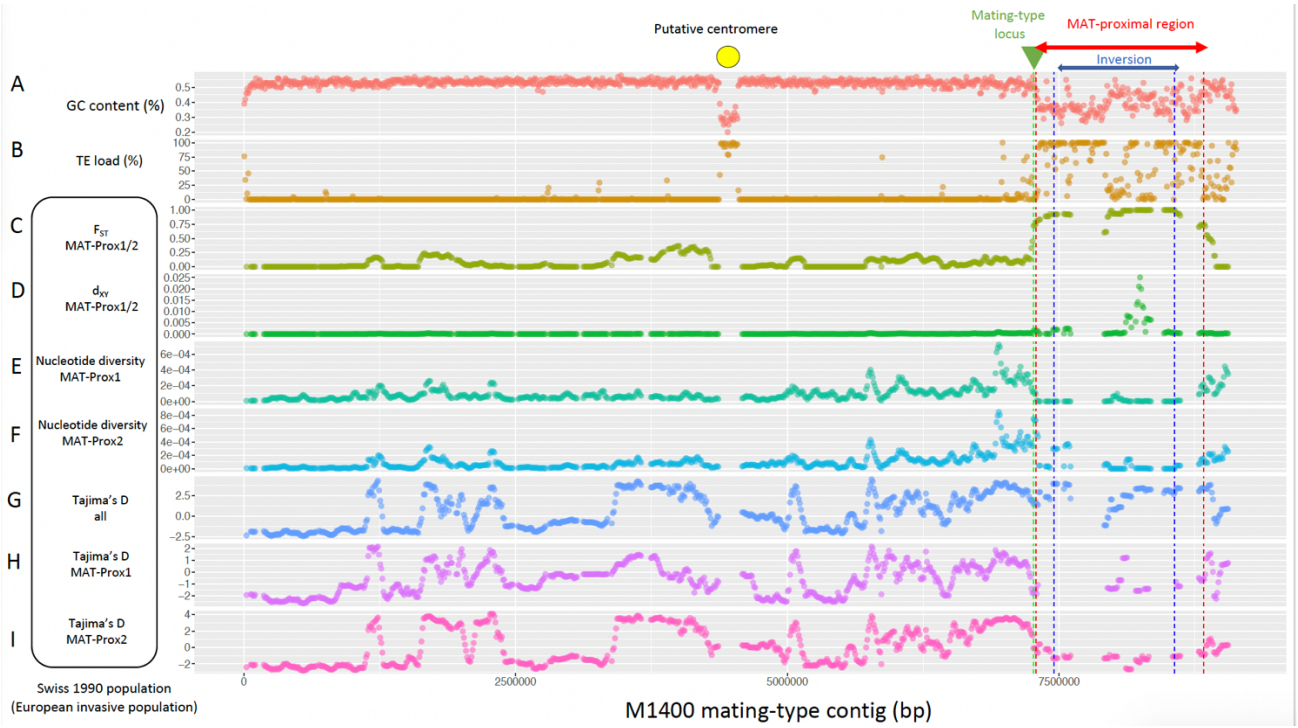

CL1 genetic cluster (European invasive population)
A
B
C
MAT-Prox2
MAT-Prox1
n=36
PC2 (1.7%)
PC1 (95.7%)
PC2 (20.3%)
PC1 (23.9%)
PC2 (16.6%)
PC1 (26.5%)
MAT-proximal region
Other regions
Other contigs (EP155 genome)
scaffold_2 (EP155 genome)
Mating-type allele:
MAT−1
MAT−2
D
E
MAT-Prox2
MAT-Prox1
MAT-proximal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

Fig 3


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

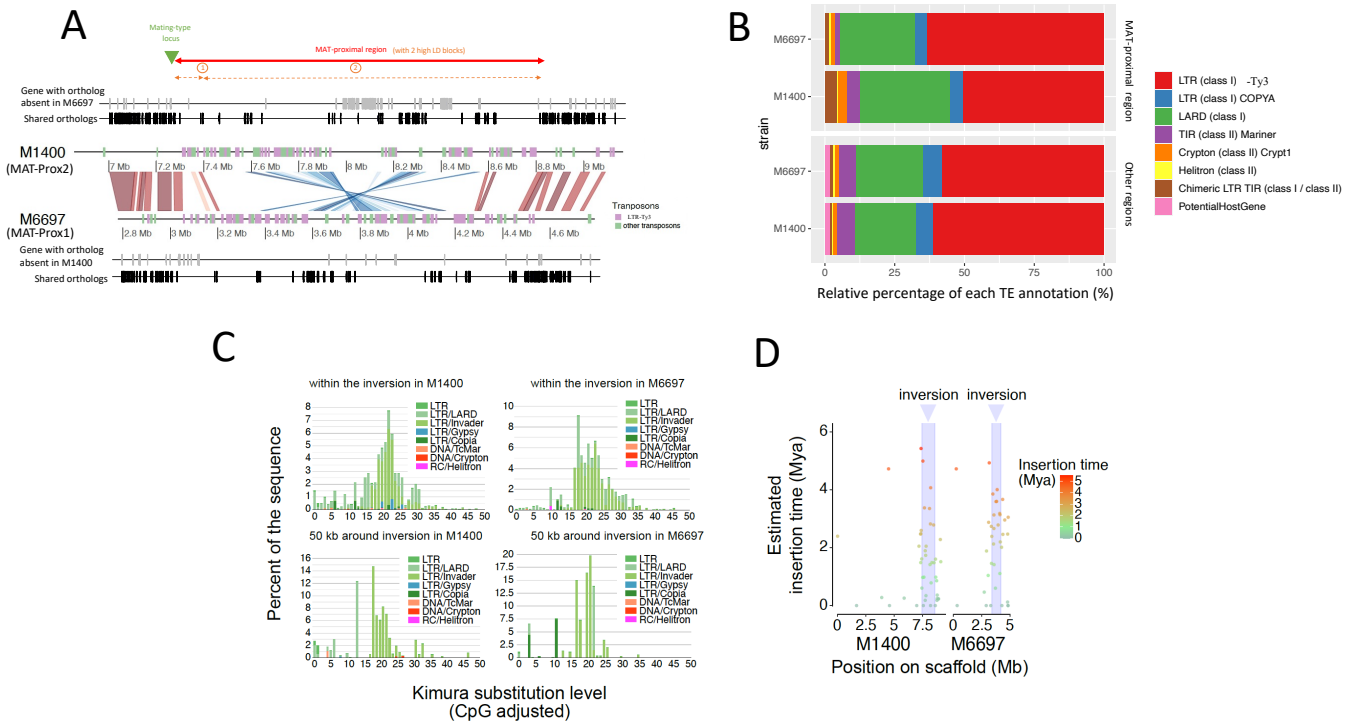
1325 Fig 4



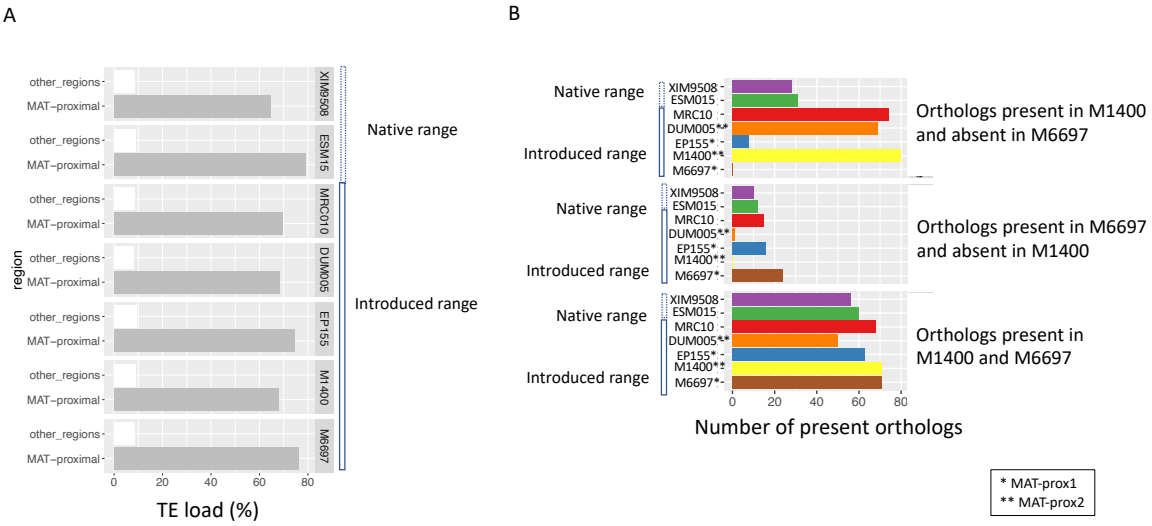
1326

1327

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
LTR-Ty3
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 (%)
LTR (class I) -Ty3
LTR (class I) COPYA
LARD (class I)
TIR (class II) Mariner
Crypton (class II) Crypt1
Helitron (class II)
Chimeric LTR TIR (class I / class II)
PotentialHostGene
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
Kimura substitution level (CpG adjusted)
LTR
LTR/LARD
LTR/Invader
LTR/Gypsy
LTR/Copia
DNA/TcMar
DNA/Crypton
RC/Helitron
D
inversion
inversion
Estimated insertion time (Mya)
Insertion time (Mya)
M1400
M6697
Position on scaffold (Mb)

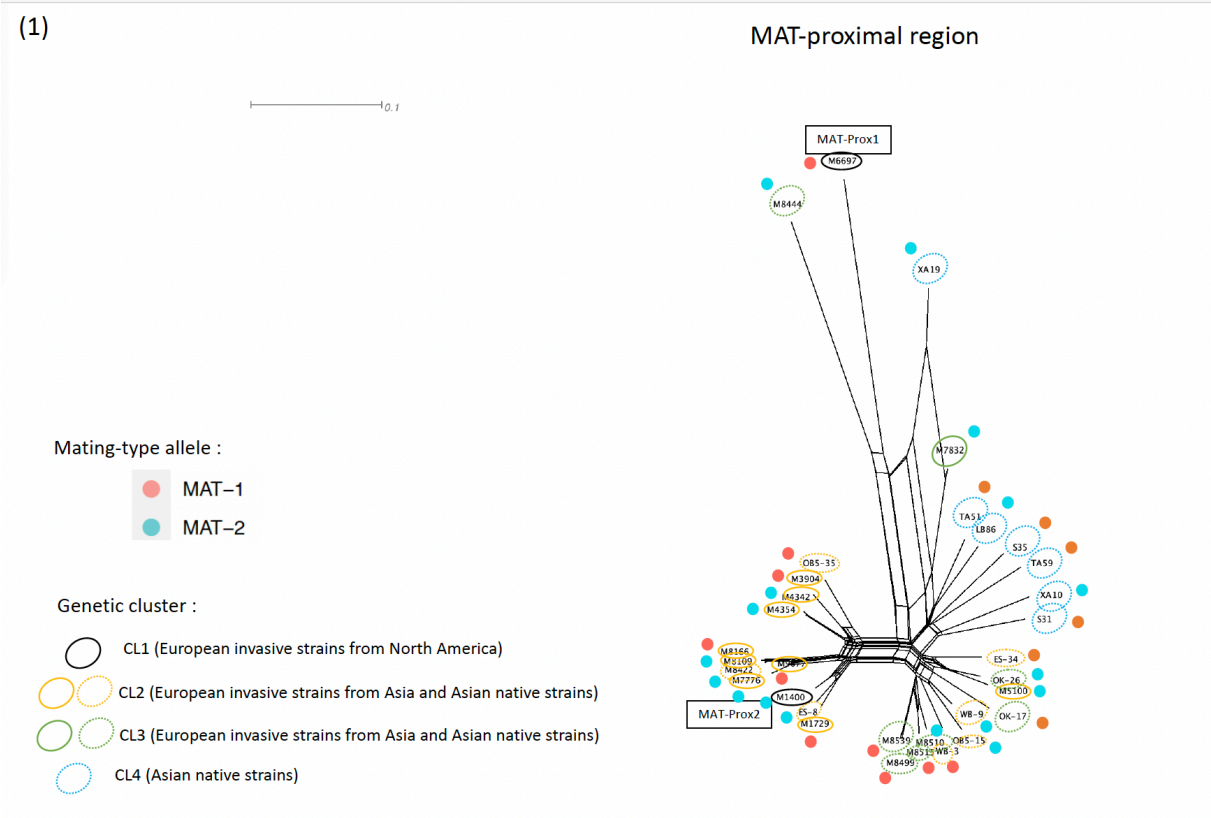
1332 Fig 6

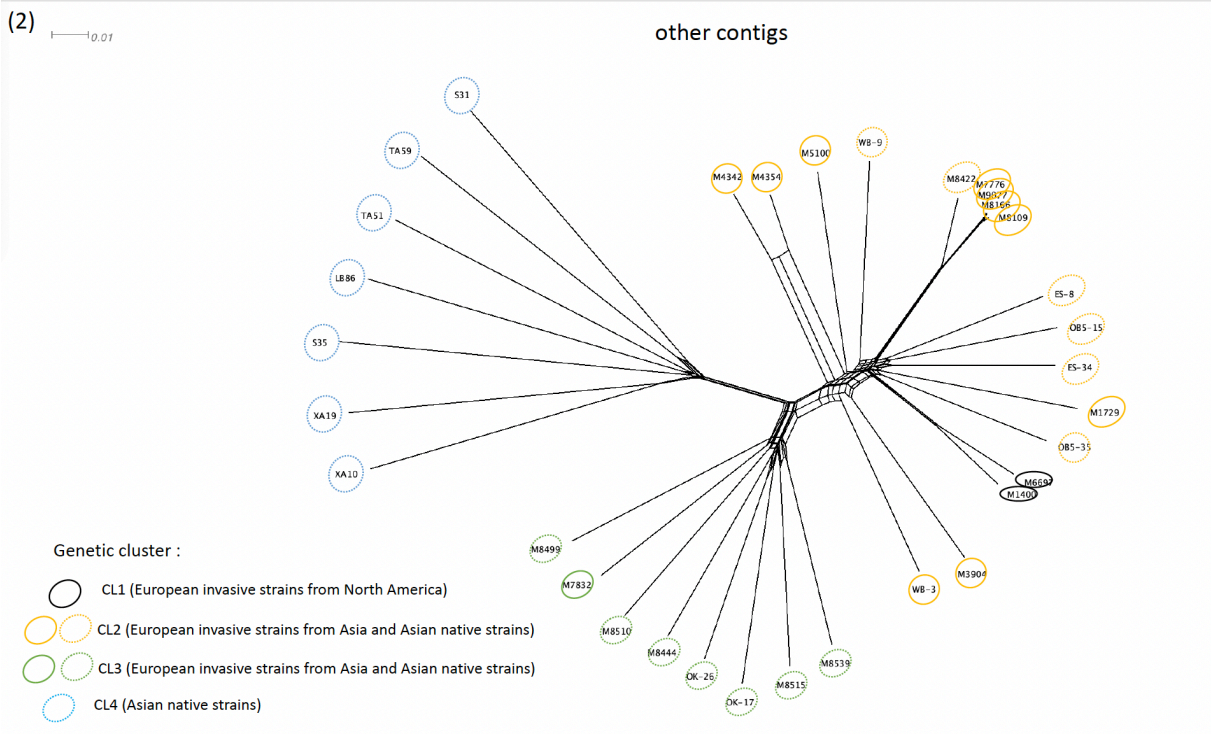


1333

1334

Fig 7


(1)
MAT-proximal region
0.1
MAT-Prox1
MAT-Prox2
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)
other contigs
0.01
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)