An old inversion polymorphism involving giant mobile elements in an invasive fungal pathogen

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

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

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

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

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

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

*Corresponding author: Fanny E. Hartmann

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

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

Abstract

Recombination suppression can evolve in sex or mating-type chromosomes, or in autosomal supergenes, with different haplotypes being maintained by balancing selection. In the invasive chestnut blight fungus *Cryphonectria parasitica*, a genomic region was suggested to lack recombination and to be partially linked to the mating-type (MAT) locus based on segregation analyses. Using hundreds of available *C. parasitica* genomes and generating new high-quality genome assemblies, we show that a ca. 1.2 Mb genomic region proximal to the mating-type locus lacks recombination, with the segregation of two highly differentiated haplotypes in balanced proportions in invasive populations. High-quality genome assemblies further revealed an inversion in one of the haplotypes in the invaded range. The two haplotypes were estimated to have diverged 1.5 million years ago, and each harboured specific genes, some of which likely belonging to *Starship* elements, that are large mobile elements, mobilized by tyrosine recombinases, able to move accessory genes, and involved in adaptation in multiple fungi. The MAT-proximal region carried genes upregulated under virus infection or vegetative incompatibility reaction. In the native range, the MAT-proximal region also appeared to have a different evolutionary history than the rest of the genome. In all continents, the MAT-Proximal region was enriched in non-synonymous substitutions, in gene presence/absence polymorphism, in tyrosine recombinases and in transposable elements. This study thus sheds light on a case of a large non-recombining region partially linked to a mating compatibility locus, with likely balancing selection maintaining differentiated haplotypes, possibly involved in adaptation in a devastating tree pathogen.

Introduction

Recombination increases the efficiency of selection, the purging of deleterious alleles and the generation of potentially beneficial allelic combinations (Otto and Lenormand 2002). However, recombination can also break up beneficial allelic combinations and, therefore, can be selected against, generating supergenes, i.e., large regions without recombination encompassing multiple genes (Schwander et al. 2014). This is often the case in genomic regions controlling mating compatibility, such as regions determining sex, mating type or self-incompatibility in plants, animals, fungi, algae and oomycetes (Charlesworth et al. 2005; Bergero and Charlesworth 2009; Umen 2011; Charlesworth 2016; Dussert et al. 2020; Hartmann et al. 2021). Such recombination suppression keeps alleles linked across different genes that prevent self-compatibility or intermediate sexual phenotypes, and can sometimes extend away from sex-determining genes or mating-type loci (Bergero and Charlesworth 2009; Furman et al. 2020; Hartmann et al. 2021). Even beyond sex and mating-type chromosomes, there is an increasing number of supergene reports in autosomes (Schwander et al. 2014) with highly differentiated haplotypes maintained by balancing selection. Striking examples include supergenes controlling social structure in ants, wing color patterns in butterflies, reproductive morphs in birds or host-parasite interactions (Küpper et al. 2016; Yan et al. 2020; Jay et al. 2021; Fredericksen et al. 2023).

In the long term, however, because of less efficient selection, genomic regions without recombination show signs of degeneration, such as transposable element accumulation, rearrangements, non-synonymous substitutions, decreased gene expression, reduced frequency of optimal codons and gene losses (Bachtrog 2013; Carpentier et al. 2022; Duhamel et al. 2023). The non-recombining regions can nevertheless be maintained by balancing selection, with different allelic combinations having contrasting advantages in different situations or experiencing negative frequency dependent selection (Jay et al. 2021; Berdan et al. 2022). The evolutionary mechanisms leading to recombination suppression and allowing their persistence remain however poorly known (Jay, D. Jeffries, Hartmann, et al. 2024). Reporting more diverse constellations of recombination suppression in a variety of organisms with contrasting life-history traits is important for understanding the general patterns of recombination suppression, their evolution and maintenance (Ironside 2010; Charlesworth 2016; Furman et al. 2020; Hartmann et al. 2021).

83

84 In the invasive chestnut blight fungal pathogen *Cryphonectria parasitica*, mating
85 compatibility is determined by a mating-type locus that displays two alleles, called *MAT-1*
86 and *MAT-2,* that are both permanently heterozygous in the diploid and dikaryotic stages
87 (McGuire et al. 2001). A large region without recombination was suggested to occur near the
88 mating-type locus, although not completely linked to it (*i.e.,* at 3.9 cM of the mating-type
89 locus), based on segregation of RAPD and RFLP markers in three crosses involving five
90 different parents, from the USA, Japan or Italy (Kubisiak and Milgroom 2006a). Recent
91 genome-wide association studies in a worldwide collection and in local *C. parasitica*
92 populations in southern Switzerland found significant association of SNPs with the mating-
93 type locus across a large region (>1 Mb), further suggesting the existence of reduced
94 recombination in this region (Stauber et al. 2021; Stauber et al. 2022). This genomic region
95 was enriched in SNPs, transposable elements and copy-number variants, such as deletions,
96 which are consistent with sequence degeneration and a lack of recombination (Stauber et al.
97 2021), and which prevented its assembly so far.
98

99 The evolutionary history of the species is well documented as it is an invasive and highly
100 damaging pathogen, having almost caused the extinction of American chestnut (*Castanea*
101 *dentata*) in North America (Anagnostakis 1987). From its center of origin in Asia and its
102 original hosts, the Chinese chestnut (*Ca. mollissima*) and the Japanese chestnut (*Ca. crenata*),
103 *C. parasitica* has first invaded North America, killing millions of American chestnuts (*Ca.*
104 *dentata*). In Europe, at least two distinct introduction events occurred, on the European
105 chestnut (*Ca. sativa*): one introduction from North America to Italy, and one directly from
106 Asia, probably to the Pyrenees Mountains. European strains will be therefore hereafter
107 referred to as invasive European strains introduced from North America or invasive European
108 strains introduced directly from Asia. The chestnut blight symptoms have been less severe in
109 Europe than in North America, due to lower susceptibility of *Ca. sativa* and a virus infecting
110 *C. parasitica* and causing hypovirulence (Dutech et al. 2010; Dutech et al. 2012). The genetic
111 determinants of the adaptation of the pathogen to its new environments and hosts remain
112 largely unknown (Lovat and Donnelly 2019).
113

114 Lineages with an apparently clonal structure have been identified in invasive populations
115 while sexually reproducing populations occur both in the native and invaded ranges (Dutech
116 et al. 2012; Stauber et al. 2021). The lineages with a predominant clonal structure likely still

undergo rare sex events, as shown by the presence of the two mating-type alleles in most of them (Demené et al. 2019). Furthermore, although this ascomycete fungus is mostly found as haploid mycelia, some isolates have been reported to be heterokaryotic at the mating-type locus in several invasive European and North American populations, i.e., with cells carrying different nuclei, of the opposite mating types *MAT-1* and *MAT-2* (McGuire et al. 2004; McGuire et al. 2005; Dutech et al. 2010; Stauber et al. 2021; Stauber et al. 2022).

Here, we therefore studied the occurrence of recombination suppression in the genomic region proximal to the mating-type locus in *C. parasitica,* using the previously published genomes and further generated six new high-quality long-read based genome assemblies of strains from the native and invaded range of the pathogen. More specifically, we tested whether recombination was completely suppressed in the genomic region proximal to the mating-type locus in *C. parasitica*. We first analyze sexually reproducing invasive European populations, for which extensive genomic datasets are available in local populations. We assessed, in these populations, the level of differentiation between haplotypes, as well as the frequency of the two non-recombining haplotypes and their association with mating types. We investigated whether other regions of the genome showed reduced recombination rates and whether genomic footprints of degeneration were present in the non-recombining region, such as genomic rearrangements, non-synonymous substitutions and transposable element accumulation. We also tested whether this non-recombining region has been gradually expanding, estimated the age of recombination suppression and investigated whether the predicted gene functions in the MAT-proximal region could help understanding the evolutionary cause for recombination suppression, polymorphism maintenance and partial linkage to the mating-type locus. We looked in particular for *Starships*, that are giant mobile elements recently discovered in ascomycete fungi and able to move accessory genes as cargo within and between genomes (Gluck-Thaler et al. 2022; Urquhart et al. 2024). These cargo genes can be involved in adaptation (Gluck-Thaler et al. 2022; Urquhart et al. 2024), as shown in fungal pathogens of coffee (Peck et al. 2023) and of wheat (Bucknell et al. 2024; Tralamazza et al. 2024), as well as in molds used for making cheeses (Cheeseman et al. 2014; Ropars et al. 2015) and those maturing dry-cured meat (Lo et al. 2023). *Starships* are characterized by their captain, a tyrosine recombinase with a DUF3435 domain, being the first gene at the 5' edge of the elements and allowing their excision and insertion. *Starships* can contain "cargo" genes, highly variable in nature and number among *Starship* elements (Urquhart et al. 2024). *Starship* diversity is partitioned into 11 major families, targeting

151 specific genomic niches for their insertions, e.g., AT-rich regions (Gluck-Thaler et Vogan,
152 2024). We then extended the analyses to other populations with less extensive genomic
153 resources, i.e. other invasive populations and native Asian populations, investigating the
154 presence of recombination suppression footprints and the occurrence of the two differentiated
155 haplotypes worldwide, the presence of *Starships*, as well as the possibility of introgression,
156 using available genomes of closely related species, pathogenic and non-pathogenic (Stauber et
157 al. 2021). Finally, we used available expression data to investigate whether the MAT-
158 Proximal region carried genes upregulated under infection by the virus responsible for
159 hypovirulence or during vegetative incompatibility reactions, a phenomenon avoiding hyphal
160 fusions between individuals, considered to protect against virus transmission.
161
162 **Results**
163 ***Footprints of recombination suppression in a large region (> 1 Mb) proximal to the mating-***
164 ***type locus in sexually reproducing invasive populations***
165 We first performed population genomic analyses on available genome sequences within local
166 populations sampled in southern Switzerland over two temporal frames (early 1990 and
167 2019), and within the more broadly geographically distributed CL1 cluster in central and
168 southeastern Europe, both having a recombining genetic structure and being invasive
169 European populations introduced from North America (Table S1 (Stauber et al. 2021)). We
170 studied only monokaryotic strains, i.e. having either *MAT-1* or *MAT-2* in their genomes, for
171 phasing haplotypes and performed stringent SNPs filtering by masking repeats on the EP155
172 reference genome (Crouch et al. 2020) and removing missing data and rare variants to ensure
173 robustness of our analyses. We identified 8,900 SNPs segregating among 71 strains of the
174 1990 Swiss population (35 MAT-1; 36 MAT-2; Table S1), 9,646 SNPs segregating among 62
175 strains of the Swiss 2019 population (20 MAT-1; 42 MAT-2; Table S1) and 15,104 SNPs
176 segregating among 88 strains of the CL1 cluster (41 MAT-1; 47 MAT-2; Table S1), which
177 represent an average density of about 20 SNPs/100 kb, in agreement with previous studies in
178 this invasive fungus (Stauber et al. 2021; Stauber et al. 2022). We investigated the genome-
179 wide linkage disequilibrium (LD) landscape and looked for large blocks (>1 Mb) of high LD
180 (r2>0.9) among SNPs within contigs, to look for signatures of reduced recombination along
181 the genome. LD indeed decays by half across a few 100 kb on average in *C. parasitica* in
182 recombining regions of the genome (Demené et al. 2019). Therefore, high LD beyond >1 Mb
183 represents strong evidence of recombination cessation and the SNP density is sufficient to
184 study such LD variation.

185

186 We found a large block of high LD on the chromosome called scaffold_2 of the EP155
187 genome in the three populations, corresponding to the mating-type chromosome (see Fig 1 for
188 the CL1 population, the pattern in the other populations being similar). The high-LD block
189 was located near the mating-type locus but did not encompass the mating-type locus itself and
190 will be called hereafter the "MAT-proximal region". In the 1990 and 2019 Swiss populations,
191 SNPs between 0.493 Mb and 1.720 Mb in the mating-type chromosome were all in high LD
192 (with r^2>0.9 between SNPs distant by up to 1 Mb; in this region, mean r^2 of all SNP pairs in
193 the Swiss 1990 population = 0.89; mean r^2 of all SNP pairs in the Swiss 2019 population =
194 0.83; Fig S1). In the CL1 cluster, the large high-LD block was also present, but was split into
195 two blocks, a large one and a smaller, peripheral one (orange arrows in Fig 1). We indeed
196 detected values of r^2>0.9 between SNPs distant by up to 1 Mb, between 0.532 Mb and 1.535
197 Mb (mean r^2 of all SNP pairs = 0.94). Between 1.535 and 1.717 Mb, SNPs were also in high
198 LD (mean r^2 = 0.84). The LD between the two high-LD blocks was lower (r^2= 0.59) than
199 within blocks but was still higher than elsewhere along the genome (orange rectangle at the
200 right border of the red triangle; Fig 1). A similar pattern also existed in the Swiss populations
201 (orange rectangle at the right border of the red triangle; Fig S1), although less marked. This is
202 likely due to rare recombination events at one specific locus near the edge of the large fully
203 non-recombining region, lowering the LD level in populations between the two parts of the
204 otherwise fully non-recombining region. The mating-type locus was located at 1.737 Mb in
205 the EP155 genome, distant by 17 kb and 20 kb from the LD block in the Swiss populations
206 and CL1 cluster, respectively. SNP density was on average 48 SNPs/100 kb in the high-LD
207 block region, i.e. higher than in the rest of the chromosome (Fig S2C). When sampling SNPs
208 distant of at least 50 kb in this region to have the same SNP density all over the
209 chromosome, we also detected the high-LD block, which indicates that the SNP density did
210 not generated biases in the LD pattern (Fig S2D).

211

212 We found no other regions of the genomes with r^2>0.9 between SNPs across such a large
213 genomic distance in any population. On scaffold_6 in the CL1 cluster, two smaller SNP
214 blocks had a r^2>0.9 between each other despite being distant (Fig S2C; between SNPs blocks
215 at 3-20 kb and 2.151-2.278 kb), which is likely due to the major intra-scaffold translocation
216 described in this region between the EP155 and ESM15 strains (Demené et al. 2022). The
217 maximum size of the blocks with r^2>0.9 on other scaffolds ranged from 135 to 540 kb in the
218 CL1 cluster, as previously described (Demené et al. 2019). The average distance between

219 SNPs in high LD (r^2>0.9) across the genome was higher on the scaffold_2 than on other
220 scaffolds (in the CL1 cluster, pairwise Wilcoxon test p-value < 2e-16 with Bonferroni
221 correction). Furthermore, the proportion of SNPs in high LD (r^2>0.9) across the genome was
222 also much higher on the scaffold_2 than on other scaffolds due to the high LD in the MAT-
223 proximal region (Fig S2D, Table S2). Therefore, the large (1Mb) and localized block with
224 maximal LD values proximal to the mating-type locus both in the Swiss populations and the
225 CL1 genetic cluster stands out as exceptional in the genome and indicates full recombination
226 cessation, which supports previous inferences from progeny segregation analyses (Kubisiak
227 and Milgroom 2006b). Indeed, even low rates of recombination homogenize allelic
228 frequencies and prevents LD building (Dufresnes et al. 2015). 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 Consistent with a lack of recombination, the MAT-proximal region formed two genetically
233 highly differentiated haplotypes in these invasive Swiss and CL1 populations (Fig 2A; Fig S3
234 A(1)-C(1)), as shown by the two clusters on the principal component analysis (PCA) using
235 only SNPs located in the MAT-proximal region, while no structure was detected in the rest of
236 the genome (Fig 2B-C; Fig S3A(2)-C(2)). The neighbor-net networks further supported the
237 existence of two differentiated haplotypes in the MAT-Proximal region, contrasting with an
238 otherwise recombining structure genome-wide (Fig 2D-E; Fig S3B-D). In the Swiss 2019
239 population, a few reticulations between haplotypes were found and three strains (LU3, Nov10,
240 Nov4) appeared to have an intermediate sequence in the MAT-Proximal region between the
241 two haplotypes (Fig S3D). This intermediate haplotype was also found in a few other invasive
242 strains introduced directly from Asia (see below).
243
244 The two genetic clusters in the MAT-proximal region appeared associated with mating types
245 in the CL1 and Swiss 1990 recombining populations (Figs 2A and D; Fig S3, Table 1).
246 Indeed, the distribution of mating types between the two PCA clusters strongly deviated from
247 expectations under random association (Table 1; chi-squared test: χ^2 = 33.3; p-value = 5.876e-
248 08 in CL1; χ^2 = 18.76; p-value = 8.441e-05 in Swiss 1990). We thus hereafter call the two
249 differentiated haplotypes MAT-Prox1 and MAT-Prox2, referring to the mating-type
250 association. However, the association between mating types and MAT-Proximal haplotypes
251 was not complete, with for example only 83% of MAT1-1 strains carrying the MAT-Prox1
252 haplotype in CL1. The two MAT-Proximal haplotypes were not significantly associated with

253 mating types in the Swiss 2019 recombining populations (Table 1; chi-squared test: χ^2 =
254 3.8427; p-value = 0.1464). The EP155 genome used as reference for SNP calling above
255 carried the MAT-Prox1 haplotype but the MAT1-2 allele.
256
257 ***Inversion in the non-recombining MAT-proximal region***
258 To investigate whether rearrangements between haplotypes were present in the MAT-
259 proximal region, we sequenced the genome, using PacBio HiFi, of two strains with alternative
260 MAT-proximal haplotypes: M1400 (MAT-2; MAT-Prox2) and M6697 (MAT-1; MAT-
261 Prox1), originating from the Swiss population in Gnosca and belonging to the CL1 cluster, i.e.
262 European invasive populations introduced from North America (Stauber et al. 2021; Stauber
263 et al. 2022). We built high-quality genome assemblies: statistics of the assemblies were in the
264 same range as those of the genome assemblies of the ESM15 and EP155 strains previously
265 sequenced (Table S3; Demené et al 2022; Crouch et al 2020). The mating-type chromosome
266 could be assembled as a single contig for the first time, in the M1400 genome, likely
267 corresponding to a full chromosome, and was assembled into two contigs in the M6697
268 genome. We therefore used the M1400 genome as a reference for subsequent analyses. To
269 identify the location of the MAT-proximal region in the M1400 genome, we computed LD by
270 mapping the reads of the 1990 Swiss population to the new M1400 genome. We found the
271 large block (> 1 Mb) of high LD (r^2>0.9) between 7,285,137 and 8,828,934 bp (red arrow;
272 Fig. S4), showing that calling SNPs on either a MAT-Prox1 or MAT-Prox2 haplotype yielded
273 similar LD patterns. The MAT-proximal region was also located 20 kb away from the mating-
274 type locus (located at 7.265 Mb) and was 1.540 Mb long using the M1400 genome as
275 reference. Consistent with the results using the EP155 genome as reference, we found that the
276 high-LD region was divided into two higher-LD blocks near the edge at 7.392 Mb (orange
277 arrows in Fig. S4).
278
279 The two new high-quality genome assemblies from the invasive Swiss population (introduced
280 from North America) revealed an inversion between the two haplotypes in the MAT-proximal
281 region (Fig 3). The two newly sequenced PacBio genomes were indeed collinear except for
282 the mating-type chromosome (Fig S5A-B-C), where we found a large region (>1 Mb) that
283 seem inverted between the M1400 and M6697 genomes (Fig 3; Fig S5A-B-C). Breakpoints of
284 the inversion were located at ca. 7.455 and 8.563 Mb of the tig00000001 contig in the M1400
285 genome and at ca. 3.380 and 4.185 Mb of the tig00000060 contig in the M16697 genome. The
286 region affected by the inversion was ca. 300 kb smaller in the M6697 genome than in the

287 M1400 genome, while the whole MAT-proximal region was ca. 100 kb smaller. The split into
288 the two high-LD blocks in natural populations reported above was outside of the inversion,
289 i.e. 70 kb from the inversion breakpoint (orange arrows in Fig S4). LD in natural populations
290 was higher within the inversion (median value of r^2 = 1) than in other regions of the MAT-
291 proximal region (median value of r^2 = 0.64; Wilcoxon rank sum test with continuity
292 correction; W = 4.8714e+10; p-value < 2e-16).
293
294 We looked for centromeres, as the non-recombining regions near the mating-type locus in
295 other ascomycetes, when they occur, either capture the centromere (Menkis et al. 2008; Sun et
296 al. 2017), or are associated to the occurrence of a single crossing-over between the centromere
297 and the mating-type locus (Grognet et al. 2014; Hartmann et al. 2021; Vittorelli et al. 2022).
298 The centromere in *C. parasitica* may be at 4.380-4.536 Mb on the mating-type chromosome
299 in the M1400 genome, as we detected here a peak in TE density and a drop in GC content
300 (Fig 4A-B). A dotplot of repeats in this region also presented a pattern typical of centromeres
301 (Fig S5D). The MAT-proximal region and the inversion thus did not include the putative
302 centromere and was instead located about 2.9 Mb away of the putative centromere.
303
304 ***Higher genetic differentiation and lower genetic diversity within the MAT-proximal region***
305 ***than in recombining regions***
306 When analyzing polymorphism within the invasive 1990 Swiss population using the M1400
307 genome as reference, we found much higher genetic differentiation between strains of the two
308 MAT-proximal haplotypes in this region than elsewhere along the genome (Fig 4C; Fig S6A;
309 Table S4), as expected for a non-recombining region in LD with the mating-type locus. For
310 example, the F_{ST} values between the two non-recombining haplotypes within the MAT-
311 proximal region in the 1990 Swiss population were nearly maximal (median F_{ST}=0.93 per 50
312 kb window), while they were near zero in the rest of the genome (median F_{ST}=0.02; Wilcoxon
313 test p-value < 2e-16 with Bonferroni correction; Fig 4C; Fig S6A). Such F_{ST} values close to 1
314 indicate a lack of shared polymorphism and therefore support the inference of recombination
315 suppression in the MAT-proximal region.
316
317 Within each haplotype, the genetic diversity at the MAT-proximal region was lower than in
318 the rest of the genome (Wilcoxon test p-value < 2e-16 for the MAT-Prox1 haplotype; p-value
319 = 2.3e-07 for the MAT-Prox2 haplotype; Fig 4E-F; Fig S6C; Table S4), as expected for a
320 region without recombination associated with the mating-type locus, as its effective

321 population size is half as in recombining regions. The diversity was especially low in the pool
322 of isolates with the MAT-Prox1 haplotype (see all MAT-Prox1 sequences clustering on a
323 single point on the PCA in Fig 2A and in the neighbor-net network in Fig 2D, in contrast to
324 the more scattered MAT-Prox2 sequences). Such a very low diversity may be due to a recent
325 selective sweep or to particularly strong bottleneck in this haplotype during the invasion.
326
327 We detected signals for long-term balancing selection in the MAT-proximal haplotypes.
328 Computation of the Tajima's D statistics within all strains suggested signatures of balancing
329 selection in the MAT-proximal region as expected in a region without recombination
330 associated with the mating-type locus maintained at a frequency close to 0.5. Tajima's D
331 values in the MAT-proximal region pooling all sequences (median D=3.0 per 50 kb window)
332 were higher than in the rest of the genome (median D=-0.14; Wilcoxon test p-value < 2e-16
333 with Bonferroni correction; Fig 4G; Fig S6D). In contrast, Tajima's D was significantly lower
334 than the rest of the genome within each pool of haplotype, including in the MAT-Prox2
335 haplotype, suggestive of positive selection (Wilcoxon test p-value = 0.0037 for the MAT-
336 Prox1 haplotype; p-value = 0.0022 for the MAT-Prox2 haplotype with Bonferroni correction;
337 Fig 4H-I; Fig S6D; Table S4). We found no difference in population diversity statistics
338 between the recombining part of the mating-type contig and the other contigs (Table S4A).
339
340 The study of synonymous divergence (d_S) between the shared orthologs in the M1400 and
341 M6697 haplotypes suggests that their differentiation is at least 1.5 million years old and that
342 there is no pattern of evolutionary strata within the MAT-proximal region, i.e. segments with
343 different levels of differentiation between haplotypes that would indicate stepwise expansion
344 of recombination suppression away from the mating-type locus. Per-gene synonymous
345 divergence (d_S) is typically used for detecting gradual expansion of recombination cessation
346 and for estimating its age, as it is considered a good proxy for the time since recombination
347 suppression. Indeed, when recombination is suppressed, mutations accumulate independently
348 in the two non-recombining haplotypes. We plotted the per-gene synonymous divergence (d_S)
349 between M1400 and M6697 along the M1400 genome, as its mating-type contig is likely
350 being assembled as a full chromosome (Fig S7A&E; Table S4). Consistent with
351 recombination suppression, we found significantly higher d_S values in the MAT-proximal
352 region (computed for 71 genes) than in the other regions of the mating-type contig (pairwise
353 Wilcoxon test p-value <2e-16 with Bonferroni correction) and other contigs (pairwise
354 Wilcoxon test p-value <2e-16 with Bonferroni correction). We found no significant

355 differences between the other regions of the mating-type contig and other contigs (pairwise
356 Wilcoxon test p-value = 0.23 with Bonferroni correction). The d_S pattern displayed no
357 indication of gradual expansion of recombination cessation within the MAT-proximal region,
358 as there was no stair-like pattern. Using synonymous substitution rate estimates across
359 multiple ascomycetes (from 0.9 x 10^{-9} to 16.7 x 10^{-9} substitutions site per year (Kasuga et al.
360 2002; Taylor and Berbee 2006) and the mean d_S value (0.0495) across the genes shared
361 between the MAT-proximal haplotypes, we estimated the age of their divergence to be at least
362 1.5 Million years (Table S4).
363
364 ***Enrichment in transposable elements in the non-recombining region and inversion***
365 ***breakpoints***
366 When studying the M1400 and M6697 high-quality assemblies, we detected an enrichment in
367 transposable elements (TEs) in the MAT-proximal region in the two haplotypes compared to
368 the rest of the genome (Fig 4B). The percentage of bp occupied by TEs (TE load) was higher
369 than 50% in the MAT-proximal region, while it was only 9% on average in other regions. TE
370 load in the MAT proximal region was higher in the M6697 haplotype (76%, MAT-1; MAT-
371 Prox1) than the M1400 haplotype (68% MAT-2; MAT-Prox2). Class I retrotransposons with
372 Gypsy (LTR-Ty3) and LARD elements were the most abundant TEs in all genomic regions
373 (autosomes, MAT-Proximal region and the rest of the mating-type contig), representing >70%
374 of TE annotations (Fig 5A-B), as shown previously (Demené et al. 2022). The TE load
375 however varied significantly among TE families and genomic regions (ANOVA; Table S5).
376 In M6697 for example, TIR elements appeared less abundant in the non-recombining than
377 recombining regions while LTR-Ty3 elements were more frequent in the MAT-proximal
378 region (Fig 5B). In M1400, LARD elements were more frequent in the MAT-proximal region
379 than in recombining regions (Fig 5B).
380
381 The presence of transposable elements at the two inversion breakpoints in the two genomes
382 (Fig 5A) suggests that the inversion may have occurred via non-homologous recombination
383 mediated by these elements. By dating retrotransposon insertions using between-copy
384 divergence and within-copy LTR divergence in both M1400 and M6697 genomes (Fig 5C and
385 D), we found that TEs within the inversion and around the inversion breakpoints were older
386 than TEs farther from the breakpoints. This is expected in non-recombining regions as
387 selection is less efficient to purge TE insertions in such regions so that they remain there
388 longer (Duhamel et al. 2023). The date estimates of LTR element insertions (Fig 5D) further

389 indicated that the TE accumulation in the MAT-proximal region was as old as a few million
390 years, confirming the age estimated based on sequence divergence between MAT-proximal
391 haplotypes.
392
393 ***Gene function in the MAT-proximal region and association to known phenotypes***
394 We found no gene function known to be involved in mating or mating compatibility, nor in
395 pathogenicity, in the MAT-proximal region (Table S7-Table S8). We nevertheless identified
396 genes potentially coding for proteins with kinase domain or proteins with homeodomains
397 (Table S6), that are often involved in the regulation of developmental pathways.
398 By re-analysing data from a previous study in the light the two haplotypes (Stauber et al.
399 2021; Stauber et al. 2022), we found no association of the MAT-Proximal haplotypes to
400 previously described vegetative compatibility groups or to the sensitivity to virus infection,
401 but more specific studies need to be performed to specifically test hypotheses.
402
403 ***The MAT-proximal region harbours* Starship *elements***
404 In addition to a high density of transposable elements, gene annotations of the MAT-proximal
405 region in the M1400 and M6697 genome assemblies revealed multiple genes encoding
406 tyrosine recombinases (genes with DUF3435 domains, that can be *Starship* captains) and gene
407 content variation (Tables S6-S7), suggesting the presence of *Starships*. By running the Gene
408 Finder Module of the starfish pipeline designed to identify *Starships* (Gluck-Thaler and
409 Vogan 2024), we in fact identified the presence of putative *Starship* captains in the M1400
410 haplotype (MAT-Prox2) and three putative *Starship* captains in the M6697 haplotype (MAT-
411 Prox1). We detected eight and ten other putative *Starship* captains elsewhere in the M1400
412 and M6697 genomes, respectively, but the MAT-proximal region appeared enriched in these
413 elements (Fisher's exact test p-value =0.0004). A phylogenetic analysis with previously
414 described *Starship* captains (Gluck-Thaler and Vogan 2024) suggested that the captains in the
415 MAT-Proximal region belonged to the Phoenix *Starship* family, that preferentially target AT-
416 rich sites (Fig S7 ; (Gluck-Thaler and Vogan 2024)). As the Starfish pipeline performs poorly
417 in TE-rich regions (Gluck-Thaler and Vogan 2024), we investigated manually the delimitation
418 of putative starships. Three putative captains shared orthology relationships between the two
419 haplotypes, one being just at the border of the MAT-proximal region near the mating-type
420 locus (Figure 3). The high similarity between captains shared by the two haplotypes (>90%
421 identity), their similar positions and reverse strand orientations suggested that these captains
422 were present before the differentiation between haplotypes and before the inversion (Figure

423 5). The genes near these three shared captains are mostly common to the two haplotypes. The
424 fact that these elements were present in all high-quality assemblies at the same locus
425 prevented delimiting Starship boundaries and these genes may represent solo Starship
426 captains or pseudogenes as found in high frequency in Pezizomycotina genomes (Gluck-
427 Thaler and Vogan 2024). The MAT-Prox2 haplotype carried an additional putative captain,
428 followed by 18 genes lacking in the MAT-Prox1 haplotype (Figure 5A), suggesting the
429 presence of a *Starship* element, together with its cargo genes, only in the MAT-Prox2
430 haplotype. Functional annotation of the putative cargo genes included two genes with
431 DUF3723 domain and one gene with ferredoxin reductase-type domain, that are often found
432 in *Starships* (Table S8 (Gluck-Thaler and Vogan 2024). This putative *Starship* was about 44
433 kb in the M1400 strain.
434
435 In addition to the genes only present in the putative Starship specific to the MAT-Prox2
436 haplotype, the two haplotypes carried additional specific genes, especially the MAT-Prox2
437 haplotype. Based on automatic gene annotation and reconstruction of orthologous
438 relationships, we found that, out of 175 groups of single-copy genes in the MAT-proximal
439 region, 80 (46%) were only present in M1400 (MAT-Prox2) and 24 (14%) were only present
440 in M6697 (MAT-Prox1). This level or presence/absence gene polymorphism was higher than
441 in other genomic regions, in which only 3% of genes were only present in one of the two
442 genomes genome (Fisher's exact test p-value < 2.2e-16). However, not all genes specific to
443 one haplotype were located close to a putative captain (Figure 3), so that part of the observed
444 gene presence/absence polymorphism between the two MAT-Proximal haplotypes may also
445 be due to gene losses due to degeneration or pseudogenisation after *Starship* insertions, to
446 recombination suppression or to gene gains by other means than *Starships.*
447
448 Among the genes specific to M6697 (MAT-Prox1), we found a gene encoding a protein of the
449 Sirtuin family (IPR003000). A Sirtuin homolog was found in a *Starship* associated with local
450 thermal climate adaptation in the wheat fungal pathogen *Zymoseptoria tritici* (Tralamazza et
451 al. 2024).
452
453 ***Gene expression in the MAT-proximal region***
454 We analysed the expression data available for the EP155 strain (Chun et al. 2020), carrying
455 the MAT-Prox1 haplotype, to test whether some of the genes in the MAT-Proximal region
456 were upregulated under certain conditions. Out of 127 genes functionally annotated in the

457 MAT-proximal region of the EP155 genome, 50 genes were supported by RNAseq data
458 acquired *in vitro* for this strain (Chun et al. 2020), suggesting that the region contains
459 functionally active genes (Table S6). Among these 50 genes, nine were upregulated during
460 barrage allorecognition of the EP155 strain with a compatible strain (Belov et al. 2021) and
461 five were upregulated during infection of the EP155 strain with the hypovirus CHV1 (Chun et
462 al. 2020) (Table S6), suggesting that the MAT-Proximal region may have a role in fungus-
463 mycovirus interactions and vegetative incompatibility. Unfortunately, no expression data are
464 available for a strain with the MAT-Prox2 haplotype, that harbours the additional *Starship*
465 and the highest number of specific genes and no expression data is available during plant
466 infection.
467
468 ***Degeneration in the non-recombining region***
469 The MAT-Proximal region was enriched in TEs and depleted in genes (Fig 4 and Fig 5A),
470 which supports the inference of an old full recombination suppression. We also found overall
471 higher non-synonymous substitution rate compared to the baseline substitution rates (d_N/d_S
472 values) between the M1400 and M6697 genomes in the MAT-proximal region (median per
473 gene $d_N/d_S = 0.80$) than in other contigs ($d_N/d_S = 0.43$; pairwise Wilcoxon test p-value = 1.7e-
474 07 with Bonferroni correction; Fig S8A; Table S4). This is consistent with relaxed purifying
475 selection due to recombination suppression in the MAT-proximal region. In addition, 13
476 genes displayed $d_N/d_S > 1$ between M1400 and M6697, suggesting that they may evolve under
477 positive selection (Table S9).
478
479 ***Recombination suppression and* Starships *in the MAT-Proximal region in other***
480 ***populations assessed from high-quality genome assemblies***
481 We analyzed five additional high-quality genome assemblies for strains from the Asian native
482 range (CL2 and CL4 clusters) and the invasion range (including two North American strains
483 belonging to the CL1 cluster and one European strain introduced directly from Asia and
484 belonging to the CL2 cluster), three of which were generated for the present study (Table S3).
485 Most of the high-quality genomes displayed the same MAT-Proximal chromosomal
486 arrangement as M1400 (MAT-Prox2, Swiss population and belonging to the CL1 cluster),
487 suggesting that it is the ancestral state (Fig S9A-D), although the incomplete assemblies may
488 prevent detecting other rearrangements. The inversion detected in M6697 (MAT-Prox1, Swiss
489 population and belonging to CL1 cluster) was also present in the EP155 strain (MAT-Prox1,
490 North America and belonging to CL1 cluster, Crouch et al, 2020; Fig S8D), indicating that

491 this inversion was already segregating in early invasive populations in North America. In
492 other genomic regions outside of the MAT-Proximal region, all analyzed high-quality
493 genomes were colinear to the ESM15 reference genome, except the highly rearranged EP155
494 genome, as previously reported (Demené et al. 2022).
495
496 The MAT-Proximal region in all the high-quality genomes, from Asian native strains and
497 invasive strains, displayed signs of degeneration, with a high TE load (Fig6; Fig S9),
498 confirming the ancient age of recombination suppression and its occurrence in the native
499 range. Study of d_S and d_N/d_S between the MRC10 strain, belonging to the European clonal
500 lineages of direct Asian origin (CL2 cluster), the M6697, M1400 and the two Asian native
501 strains further confirmed recombination suppression in the native range and in invasive strains
502 from different origins (North America or directly from Asia). The d_S and d_N/d_S values were
503 indeed significantly higher in the MAT-proximal region than other genomic regions (Fig S7
504 B-E). Mean d_S values in the MAT-proximal region were in the same range as d_S values
505 between the M6697 and M1400 genomes, suggesting a similar date of recombination
506 suppression and, therefore, that recombination suppression was ancestral. The MAT-proximal
507 region appeared smaller in the two Asian strains XIM9508 and ESM15 (<1 Mb) than in
508 invasive strains (Table S3), but this may be due to the incomplete genome assemblies, and/or
509 to additional *Starship* presence/absence polymorphism.
510
511 We detected the three captains shared between the M6697 and M1400 haplotypes in the
512 MAT-proximal region of all high-quality genome assemblies, indicating that their insertions
513 in the MAT-proximal region were old, being already present in the native range. We found
514 the putative *Starship* specific to M1400 (MAT-Prox2), i.e., the captain and the 18 cargo
515 genes, in the MAT-proximal region of DUM-005 (MAT-Prox2) and in the haplotypes of
516 MRC10 and the Japanese strain ESM15 (both genetically close to MAT-Prox2; Figure
517 S10D1), while it was absent in the MAT-Proximal region of the EP155 strain (MAT-Prox1).
518 In the XIM9508 Chinese strain (genetically close to MAT-Prox1; Figure S10D1), the captain
519 and four genes of the *Starship* specific to M1400 were absent in the MAT-proximal region.
520 However, 14 of its putative cargo genes were present and a very similar captain was found
521 elsewhere in the genome, suggesting that this *Starship* may have moved away from the MAT-
522 Proximal region after its insertion there. In the seven high-quality assemblies, the putative
523 *Starship* presence/absence polymorphism detected between the 1400 and M6697 genomes
524 thus seem associated to the MAT-Proximal haplotypes in the invaded range and also in the

525 native Asian range (Figure S10D), although this would need to be confirmed with additional
526 high-quality assemblies.
527
528 ***Presence of the two MAT-Proximal haplotypes in other populations, but less differentiated***
529 ***in the native range***
530 By analyzing available additional Illumina genomes of invasive strains from the first invasion
531 wave via North America (strains from the US and European clonal lineages), we recovered
532 the two MAT-proximal haplotypes in balanced frequencies. We indeed observed high
533 differentiation between two clusters in the MAT-Proximal region, contrasting with a lack of
534 genetic subdivision in other genomic regions or a completely different genetic subdivision,
535 corresponding to previously described population genetic structures (Figs S10A and S10C).
536 This reinforces the view of balancing selection on the MAT-Proximal haplotypes.
537
538 In the native range, the MAT-Proximal region also seemed to have undergone a very different
539 evolutionary history than the rest of the genome (Figure 7). In particular, the neighbor-net
540 network with multiple strains sequenced previously from the CL2 and CL3 clusters (from the
541 invaded range and the native range, in South Korean and Japan) and from the CL4 cluster
542 (from the native range, in China) indicated that sequences from these three clusters were
543 intermingled in the MAT-proximal region (Fig 7A), while the structure in the rest of the
544 genome instead corresponded to the different population clusters, i.e., CL1, CL2, CL3 and
545 CL4 (Fig 7B). However, the two haplotypes were much less differentiated in the native range
546 and in invasive populations originating directly from Asia than in invasive populations from
547 the first invasion wave via North America. For example, branch lengths on the neighbor-net
548 indicate a lower differentiation between haplotypes within CL2 (Fig 7A) than within the
549 European CL1 or Swiss invasive populations (Fig 2D; Fig S3) and some reticulations were
550 present between MAT-proximal haplotypes.
551
552 We found no association of the MAT-Proximal haplotypes with mating types at the world
553 scale (Figs 7 and S10). In the S12 European invasive clonal lineage (only MAT-1), we found
554 a single MAT-Proximal haplotype (MAT-Prox1; Fig S10B), but we detected the two MAT-
555 Proximal haplotypes in all other populations. Even in two invasive European lineages of
556 American origin with a predominantly clonal structure (re092 and re103), the strains with a
557 mating type different from their clone-mates, due to localized introgression from other clonal
558 lineages (Demené et al. 2019), also had their MAT-Proximal region introgressed (Fig S10C);

559 this finding reinforces the view of the strong linkage between the mating-type locus and the
560 MAT-Proximal region, of an advantage of maintaining the two MAT-Proximal haplotypes in
561 populations, and perhaps of maintaining an association between MAT-proximal haplotypes
562 and mating types within populations. Three 2019 Swiss strains exhibited a similar
563 intermediate sequence in the MAT-Proximal region as the other strains from Europe with a
564 direct Asian origin described above (Fig S10C). This intermediate haplotype thus seemed
565 associated with the second invasion in Europe, directly from Asia, and carried the same
566 captains as the MAT-Prox2 haplotype from the M1400 strain. The location as an intermediate
567 in the neighbor-net network may be due to recombination between the haplotypes, or may just
568 reflect the fact that the two haplotypes are less differentiated in the native Asian area.
569
570 ***No evidence of introgression in the MAT-proximal region from* closely related**
571 ***Cryphonectria species***
572 The stronger differentiation in the MAT-Proximal region in the Swiss and CL1 invasive
573 populations than in other populations could also be due to an introgression event from a
574 closely related species; however, the small genetic distance and the reticulations observed in
575 the neighbor-net network between the M6697 Swiss reference genome and the Asian
576 genomes M8444 and XA19 rather indicate that the MAT-Prox1 haplotype in Switzerland
577 originated from Asian *C. parasitica* populations, likely a population different from those at
578 the origin of the CL2 invasive cluster (Fig 7).
579
580 Gene genealogies also provided no evidence of introgression in the MAT-proximal region
581 from closely related *Cryphonectria* species. We performed a gene orthology analysis
582 including three genomes of the closely related species *C. japonica*, C. *carpinicola*, and *C.*
583 *naterciae* that were previously sequenced with short read technologies (Stauber et al. 2020).
584 We retrieved only nine genes from the MAT proximal region that were shared between the
585 M6697 and M1400 genomes that had orthologs in one of the related species. This may be due
586 to low quality assemblies in the outgroups (Stauber et al. 2020) and/or to many gene gains in
587 this region in *C. parasitica* because of *Starships*. We found no gene with a topology
588 consistent with introgression from another species into the MAT-Proximal region, i.e. with a
589 MAT-Prox1 or MAT-Prox2 haplotype that would branch with an outgroup allele rather than
590 with the alternative haplotype of *C. parasitica*. The high load of TEs and gene disruptions in
591 the two MAT-Proximal haplotypes does not fit with the introgression hypothesis either,
592 unless the introgression occurred from a species with recombination suppression in the MAT-

593 Proximal region. The Starships may have moved from a distant species by horizontal gene
594 transfer together with their cargo genes, but we found no evidence for this by blasting the
595 captains and cargo genes in public databases.
596
597
598 **Discussion**
599 ***A large non-recombining region near the mating-type locus, with two highly differentiated***
600 ***haplotypes maintained polymorphic in the native and introduced ranges***
601 We found strong evidence for the complete cessation of recombination in a region of more
602 than 1 Mb (1.2 Mb to 1.5 Mb depending on the haplotype), proximal to the mating-type locus
603 in *C. parasitica*. Indeed, we detected maximal levels of linkage disequilibrium in otherwise
604 recombining populations and the existence of two differentiated haplotypes, without shared
605 polymorphism, while even low rates of recombination can homogenize alleles and prevent LD
606 build-up (Dufresnes et al. 2015). The full cessation of recombination was further supported by
607 the existence of a previously unknown, large inversion in the center of the non-recombining
608 region in invasive strains. This is in agreement with previous genetic maps reporting lack of
609 recombination near the mating-type locus in crosses involving Japanese, North American and
610 Italian isolates (Kubisiak and Milgroom 2006a). The higher TE load in this region constitutes
611 additional evidence for full recombination cessation, as the lack of recombination renders
612 selection against TE accumulation less efficient, in particular due to Muller's ratchet: two
613 chromosomes with different TE insertions cannot recombine to produce a chromosome free of
614 TE insertions. TE takes hundreds of thousands of years to accumulate in non-recombining
615 regions (Duhamel et al. 2023), so that the high TE load in the MAT-Proximal region in *C.*
616 *parasitica* indicates old recombination cessation. The particularly high frequency of non-
617 synonymous substitutions in the MAT-Proximal region worldwide further indicates
618 degeneration and ancient recombination suppression. The existence of two differentiated
619 haplotypes appeared less clear in the native range in Asia and in the second-wave invasive
620 populations originating directly from Asia than in invasive populations from the first
621 introduction via North America, but the MAT-Proximal region nevertheless appeared to
622 display a different evolutionary history than the rest of the genome and a high TE load.
623
624 The estimates for the differentiation between the two MAT-Proximal haplotypes based on
625 synonymous divergence between shared single-copy orthologs and based on TE insertion
626 dates were at least 1.5 million years. The suppression of recombination in *C. parasitica* may

627 be younger than this estimate if one of the two haplotypes was introgressed from a distant
628 species. The evolution of non-recombining regions by introgression has been reported for
629 example in ants and butterflies (Jay et al. 2018; Helleu et al. 2022; Stolle et al. 2022).
630 However, we found little evidence for an introgression: i) the two haplotypes were detected in
631 native populations and in the CL2 cluster of invasive European populations directly
632 originating from Asia, even if less differentiated than in the invasive populations from the first
633 introduction, ii) the MAT-Prox1 haplotype in the CL1 cluster that is the most differentiated
634 from the MAT-Prox2 haplotype is genetically close to an Asian haplotype, iii) footprints of
635 recombination suppression are present in the native range in terms of degeneration and
636 particular genomic structure in the MAT-Proximal region, and iv) we found no signatures of
637 introgression from closely related species in gene genealogies or by blast in public databases.
638 In any case, the findings of recombination suppression footprints and of the presence of two
639 differentiated haplotypes in the two invaded continents, Europe and North America, as well as
640 in the native range of *C. parasitica*, with intermingled sequences from different Asian
641 populations in neighbor-net networks, indicate ancient recombination suppression and long-
642 term polymorphism maintenance in the MAT-Proximal region. These findings point to a
643 strong balancing selection maintaining two differentiated haplotypes in *C. parasitica.*
644
645 In contrast to other known non-recombining regions on sex or mating-type chromosomes
646 (Bergero and Charlesworth 2009; Hartmann et al. 2021), the region without recombination
647 was not completely linked to the mating-type locus. Indeed, the occurrence of rare
648 recombination events between the non-recombining MAT-Proximal region and the mating-
649 type locus are supported by previous segregation analyses (Kubisiak and Milgroom 2006b)
650 and the lack of association found here between the MAT-Proximal haplotypes and the mating
651 types at the worldwide scale, as well as their incomplete association within populations.
652
653 ***Proximal cause of recombination suppression***
654 The large inversion in the MAT-Proximal region may contribute to recombination
655 suppression. There may also be additional nested inversions and/or translocations in the
656 MAT-Proximal region as there are breaks of synteny between the inverted fragments with
657 many transposable element insertions. The chromosomal arrangement of the M1400 MAT-
658 Prox2 haplotype is likely the ancestral state as it is the most frequent worldwide, although a
659 more complete sampling in the native range is required to obtain definitive conclusion. The
660 non-recombining region was defined based on LD pattern and was larger than the inversion.

661 This may indicate that additional proximal mechanisms suppress recombination. The
662 inversion may even be a consequence rather than a cause of recombination cessation, as
663 previously found in other fungal mating-type chromosomes (Grognet et al. 2014; Sun et al.
664 2017; Branco et al. 2018; Vittorelli et al. 2022), especially as the inversion was not detected
665 so far in the native range despite the presence of footprints of recombination suppression.
666 Other proximal causes of recombination cessation may for example be genetic recombination
667 modifiers, epigenetic marks, histone modifications or lack of recruitment of proteins
668 responsible for double-strand breaks (Maloisel and Rossignol 1998; Boideau et al. 2022;
669 Legrand et al. 2024). The lack of rearrangements at the edge of the MAT-proximal region
670 may allow rare recombination or gene conversion events, which may explain the presence of
671 two distinct LD blocks with lower LD between them and the lower differentiation at the edges
672 of the MAT-proximal region. Alternatively, the LD around the inversion may be formed by
673 less frequent recombination as recombination is often modified around inversion breakpoints
674 (Pegueroles et al. 2010; Stevison et al. 2011) and the actual size of the non-recombining
675 region may be smaller than suggested from the LD pattern.
676
677 ***The MAT-proximal haplotypes carry putative* Starship *elements***
678 We detected signatures of *Starship* elements in the non-recombining MAT-proximal region,
679 with likely three "solo" captains shared between the two haplotypes, but also an additional
680 captain associated to 18 specific genes, strongly suggesting the presence of a cargo-
681 mobilizing *Starship*, only present in the MAT-Prox2 haplotype. The difference in gene
682 content between the two MAT-proximal haplotypes thus likely at least partly result from
683 specific cargo genes inserted with *Starships,* although there could also be gene losses because
684 of the less efficient selection due to recombination suppression, and/or additional gene gains.
685 One of the captains shared between haplotypes was located just nearby the mating-type locus
686 and at the edge of the MAT-proximal region. This altogether strongly suggest a role of
687 *Starship* elements in the formation of the MAT-proximal haplotypes and in their long-term
688 maintenance, although experiments are required to confirm this hypothesis and elucidate their
689 role. High-quality assemblies suggested that the presence/absence of the polymorphic
690 *Starship* insertion was associated to the MAT-Proximal haplotypes, although this also needs
691 to be checked in a larger set of strains. The presence of the inversion in the native range also
692 remains to be investigated. Additional sampling in the native range and high-quality genome
693 assemblies are required to obtain a more comprehensive view of the history of the MAT-
694 Proximal region and of the putative *Starship* insertions.

695 *Starships* may have been introgressed from another species, as frequent with these elements
696 (Ropars et al. 2015; Peck et al. 2023), but searches in databases returned no significant blast
697 results. Such horizontal gene transfers should not have, however, inflated the date estimates of
698 divergence of recombination suppression in the MAT-Proximal region, as these were
699 computed based on shared single-copy orthologs between haplotypes and the shared putative
700 *Starship* elements seem to correspond to homologous insertions.
701
702 ***The selective forces potentially maintaining the two haplotypes in the MAT-proximal***
703 ***region***
704 Our findings indicate that selection maintains the two haplotypes polymorphic in the MAT-
705 Proximal region. There can be several hypotheses regarding the nature of this balancing
706 selection. The degeneration of the non-recombining region may help maintaining the two
707 haplotypes by selection due to the sheltering of combinations of deleterious mutations in
708 repulsion, i.e. thanks to a heterozygote advantage called pseudo-overdominance (Abu-Awad
709 and Waller 2023). Such pseudo-overdominance advantage may only be a consequence of
710 recombination suppression, and contribute to its maintenance, or may even be its initial cause
711 (Branco et al. 2017; Jay et al. 2021; Charlesworth 2023; Jay, D.L. Jeffries, Hartmann, et al.
712 2024). These hypotheses require the existence of a substantial diploid or dikaryotic phase for
713 sheltering recessive deleterious alleles in a heterozygous state. In fact, although *C. parasitica*
714 has a predominant haploid phase, strains can be found as dikaryons heterozygous for the
715 mating-type locus in some natural populations, and not only as monokaryons (McGuire et al.
716 2004; McGuire et al. 2005; Stauber et al. 2021; Stauber et al. 2022). The importance and
717 frequency of heterokaryons in *C. parasitica* in nature would deserve further investigations. It
718 may be sufficient if recessive deleterious alleles are sheltered in a substantial percentage of
719 individuals for selecting recombination suppression, and/or if the genes involved are
720 preferentially expressed during the diploid or dikaryotic phases. This hypothesis of balancing
721 selection, and in particular the hypothesis of overdominance (i.e. heterozygote advantage),
722 does not require the same association between the MAT-Proximal haplotypes and mating
723 types across all populations worldwide, it is sufficient that the MAT-Proximal haplotypes are
724 strongly associated to mating types within populations. Further sampling will be required to
725 test whether the MAT-Proximal haplotypes are associated to mating types within local
726 populations in the CL3 and CL4 clusters and in the native Asian range. It will also be
727 interesting in future studies to compare the fitness between strains carrying the two MAT-
728 Proximal haplotypes and dikaryotic strains being homozygous for the MAT-Proximal

729 haplotype. Such experiments in Sordariales fungi reported a heterozygote advantage (Guyot et
730 al. 2024). Results from previous studies suggest that strains carrying either the same or
731 different MAT-Proximal haplotypes can be crossed in *C. parasitica* (Stauber et al. 2021).
732
733 It could indeed also be that genes evolve under overdominance in this region for other reasons
734 than deleterious mutations, so that a partial linkage to a permanently heterozygous locus is
735 beneficial, perhaps in relationship with the pathogenic lifestyle of the fungus. One could
736 imagine, for example, that heterozygosity could be advantageous at genes involved in
737 virulence against the tree host, and especially the novel hosts colonized in invasive ranges, or
738 in resistance against a parasitic virus known to negatively affect fitness in *C. parasitica* (Choi
739 and Nuss 1992; Brusini et al. 2017). As a matter of fact, multiple genes present in the MAT-
740 Proximal region are up-regulated under infection by the hypovirus CHV1 or during the
741 vegetative incompatibility reaction considered to play a role in the prevention of virus
742 transmission (Choi et al. 2012; Rigling and Prospero 2018). The specific genes unique to one
743 or the other MAT-proximal haplotypes, and possibly the cargo genes inserted with the
744 *Starships*, may contribute to such a heterozygous advantage.
745
746 As an alternative to a heterozygote advantage, the selection maintaining the recombination
747 suppression and the two haplotypes in the MAT-proximal region could be some kind of
748 negative-frequency dependent selection of beneficial allelic combinations, possibly linked to a
749 trench-warfare-like arms race with the host tree, the virus or the microbial community, or to
750 self-incompatibility, where partial linkage to the mating-type locus would help maintenance
751 in balanced proportions and therefore would prevent allele loss (Tellier et al. 2014; Jay,
752 Aubier, et al. 2024). The MAT-proximal region did not include any of the genes previously
753 identified as controlling vegetative incompatibility in *C. parasitica* but not all self-
754 incompatibility genes have been identified yet. The MAT-proximal region actually carried
755 genes upregulated under virus infection or vegetative incompatibility reaction. Such a role in
756 host-pathogen interactions or vegetative incompatibility would also be consistent with the
757 selective sweep footprints detected in the MAT-proximal region in a previous study (Stauber
758 et al. 2021), if there is recurrent positive selection for improving the efficiency of the
759 pathogen weapons within each of one of the two MAT-proximal haplotypes or new, rare self-
760 recognition alleles. Such negative-frequency dependent selection, could also explain the
761 particular population structure in the MAT-proximal region in the native region, with a mix of
762 sequences from the different genetic clusters, if alleles introgressed between populations are

763 favored by a positive selection of rare alleles, or with the long-term maintenance of self-
764 incompatibility alleles, and therefore of ancestral polymorphism. Much higher differentiation
765 between MAT-Proximal haplotypes in the invasive than the native populations may be due to
766 a selective sweep of a rare and differentiated MAT-Prox1 allele present in the native range or
767 having evolved rapidly, by particular demographic effect during the invasion (Moinet et al.
768 2022), both of which are consistent with the very low genetic diversity in the MAT-Prox1
769 haplotype in the introduce range. Balancing selection and location adjacent to a non-
770 recombining region has been reported for loci involved in host resistance in the *Daphnia–*
771 *Pasteuria* system (Fredericksen et al. 2023). The MAT-Proximal region may in this case even
772 include genes involved different traits under balancing selection, and would then constitute a
773 supergene (Schwander et al. 2014). The MAT-Proximal region also contained, only in the
774 MAT-Prox1 haplotype, a gene of the Sirtuin family, a homolog of which had previously been
775 found in a *Starship* associated with local thermal climate adaptation in the wheat fungal
776 pathogen *Zymoseptoria tritici* _(Tralamazza et al. 2024). Another hypothesis to explain the
777 balancing selection in the MAT-Proximal region may thus be a heterogeneous selection, with
778 different genes in the two haplotypes providing contrasted advantages in different conditions
779 or different phases of the life cycle. The proximity of the mating-type locus, permanently
780 heterozygous in dikaryotic and diploid stages, may help maintaining balanced frequencies of
781 the two MAT-proximal haplotypes. Here too, the genes inserted by the *Starships* may play a
782 role in such balancing selection, especially the genes that are present in a single of the two
783 MAT-Proximal haplotypes. The other transposable elements insertion polymorphism
784 observed from the high quality assemblies may also be adaptive (Casacuberta and González
785 2013; Orteu et al. 2024).
786
787 Another hypothesis for explaining such extension of recombination suppression beyond the
788 mating-type locus is antagonistic selection, i.e. linkage of alleles that would improve fitness
789 of a MAT-1 gamete while being deleterious in a MAT-2 gamete, or vice-versa. However, full
790 recombination suppression with the mating-type locus would be expected under such
791 antagonistic selection, as well as the same association between the MAT-Proximal alleles and
792 the mating-type in all populations, in contrast to our findings. In addition, we found no
793 particular predicted function in the MAT-proximal region that could be related to mating
794 compatibility and there are very little functions, if any, with possible antagonistic functions
795 between mating types in fungi (Bazzicalupo et al. 2019; Hartmann et al. 2021).
796

797 **Conclusion**

798 In conclusion, we provide strong evidence for the existence of a non-recombining region
799 partially linked to the mating-type locus in the chestnut blight fungus *C. parasitica*, with two
800 highly differentiated haplotypes, each carrying specific genes, maintained polymorphic by
801 selection. We found footprints of balancing selection in the MAT-proximal region in the first
802 introduction of the pathogen in Europe from North America and a chromosomal inversion.
803 The non-recombining region also displayed footprints of particular evolution in Asia and in
804 the second invasion wave directly from Asia, although the levels of differentiation between
805 haplotypes was lower than in the populations from the first introduction wave. This non-
806 recombining region may underlie important adaptive traits and thereby provide important
807 applications for the control of a devastating tree pathogen. This is supported by the finding of
808 putative *Starships* elements in the MAT-proximal region, i.e., giant mobile elements recently
809 discovered in ascomycete fungi, containing multiple cargo genes (Gluck-Thaler et al. 2022;
810 Urquhart et al. 2024), that can be involved in adaptation. In addition, the high-quality genome
811 assemblies provided here, from the native and invaded ranges, will more generally be useful
812 for studies aiming at understanding the evolution of this invasive and damaging pathogen.

813

814 **Material and Methods**

815

816 **Strains and genomic data**

817 For population genomic analyses, we studied a collection of 386 monokaryotic *C. parasitica*
818 strains sampled worldwide, from the native and invaded ranges of the pathogen, and
819 sequenced previously with the short-read Illumina technology (strain information are
820 presented in Table S1; (Demené et al. 2019; Stauber et al. 2021; Stauber et al. 2022)). Data
821 were downloaded from NCBI Bioproject numbers PRJNA604575, PRJNA644891 and
822 PRJNA706885. We focused our analyses first on European invasive strains originating from
823 North America. We studied 88 strains belonging to the CL1 genetic cluster, in central and
824 southeastern Europe (Stauber et al. 2021). We excluded the putative clonal genotypes
825 previously identified (Stauber et al. 2021). We also studied 71 strains sampled in the 1990 and
826 62 strains sampled in 2019 in southern Switzerland (Stauber et al. 2022). Mating-type ratios
827 close to ½ and population structure analyses of the CL1 genetic cluster and the Swiss 1990
828 population suggest frequent recombination in these populations (Stauber et al. 2021; Stauber
829 et al. 2022). The mating-type ratio was 33% *MAT-1* in the 2019 Swiss population and
830 population structure analyses suggested regular sexual reproduction and recent population

831 bottleneck. To study the presence of the two haplotypes in the MAT-proximal region more
832 broadly, we analyzed monokaryotic strains belonging to the CL2, CL3, CL4 genetic clusters,
833 as well as the S12 European invasive lineage (Stauber et al. 2021) and additional
834 monokaryotic strains from the US and Europe (with an Asian or North American origin
835 (Demené et al. 2019)). We excluded heterokaryotic strains, i.e. strains having both MAT-1
836 and MAT-2 alleles, as the phase was challenging to infer.
837
838 For mapping and SNP calling, we first used as reference the 43.9 Mb genome sequence
839 EP155 v2.0 of the strain EP155 (*MAT-2,* North America, CL1 cluster) available at the Joint
840 Genome Institute (http://jgi.doe.gov/)(Crouch et al. 2020). For comparative genomics, we
841 used the published genome of the strain ESM15 (MAT-2, Japan, CL2 cluster; (Demené et al.
842 2022)) available at DDBJ/ENA/GenBank on the bioproject PRJNA700518 under the
843 accession JAGDFO000000000. We additionally sequenced *de novo,* with long-read
844 technologies, five strains from the native and invaded ranges of the pathogen. We sequenced
845 with PacBio Hifi the genomes of the strains M1400 (MAT-2) and M6697 (MAT-1) sampled
846 in Gnosca, southern Switzerland (Stauber et al. 2021). Mycelia were stored as glycerol stocks
847 at -80 C after strain isolation. To produce mycelium for DNA extraction, isolates were
848 inoculated onto cellophane-covered potato dextrose agar plates (PDA, 39 g/L; BD Becton,
849 Dickinson and Company, Franklin Lakes, USA) (Hoegger et al. 2000) and incubated for a
850 minimum of 1 week at 24°C, at a 14 hr light and 10 hr darkness cycle. Mycelium and spores
851 were harvested by scratching the mycelial mass off the cellophane, transferring it into 2 mL
852 tubes and freeze-drying it for 24 hr (Stauber et al. 2021). DNA extraction was performed with
853 the NucleoBond High Molecular Weight DNA kit from Macherey-Nagel, with the mechanical
854 disruption of about 30 mg of lyophilized mycelium with two beads of 3 mm for 5 min at 30
855 Hz. Sequencing was outsourced to Novogene, the Netherlands. We additionally sequenced
856 with Oxford Nanopore MinION technology the strains XIM9508 (China, CL4 cluster, MAT-
857 1), MRC10 (South Western France introduced directly from Asia, CL2 cluster, MAT-2) and
858 DUM-005 (USA, CL1 cluster, MAT-2). These isolates had been collected for previous
859 studies (Milgroom et al. 1996; Liu et al. 2003; Dutech et al. 2012). The protocols for
860 mycelium production, DNA extraction and sequencing for these three strains were the same
861 as in (Demené et al. 2022).
862
863 **Short-read data processing and SNP calling**

We used SNP calling datasets of the genomes from monokaryotic strains of the CL1, CL2, CL3, CL4 genetic clusters, the S12 invasive lineage and the Swiss populations against the EP155 reference v2.0 genome obtained in (Stauber et al. 2021; Stauber et al. 2022). We performed mapping and raw SNP calling of short-read data using the M1400 new genome assembly as a reference as described in (Stauber et al. 2021; Stauber et al. 2022). Briefly, we trimmed reads with Trimmomatic v0.36 (Bolger et al. 2014) and aligned them with Bowtie 2 v2.3.5.1 (Langmead et al. 2009) and SAMtools v1.9 (Li et al. 2009) to the EP155 v2.0 genome. Raw SNP calling and filtration for quality were conducted with the genome analysis toolkit GATK v3.8 and v4.0.2.0 (McKenna et al. 2010). We used the filtration parameters described in (Stauber et al. 2021): QUAL>=100, MQRankSum (lower) >= -2.0, QD:_20.0, MQRankSum (upper) <=2.0, MQ:_20.0, BaseQRankSum (lower) >=-2.0, ReadPosRankSum (lower) >=-2.0, ReadPosRankSum (upper) <=2.0, BaseQRankSum (upper) <=2.0. We further removed SNPs overlapping with transposable elements predicted *de novo* in the EP155 v2.0 genome by (Stauber et al. 2021).

Population genomics analyses

For all population genomics analyses, we excluded SNPs with missing data in more than 10% of the strains and kept only polymorphic strains with vcftools v0.1.16 (Danecek et al. 2011). To study linkage disequilibrium, we further excluded rare variants (minor allele frequency <0.1) with the vcftools (Danecek et al. 2011) option --maf 0.1. We computed LD with the --hap-r2 option of vcftools (Danecek et al. 2011) for each scaffold and each population separately. We used the --thin 50000 option of vcftools (Danecek et al. 2011) to sample SNPs distant of at least 50 kb. We used the R package LDheatmap v1.0-6 (Shin et al. 2006) to plot LD r^2 values among SNP pairs. To perform principal component analyses (PCAs), we first used vcftools (Danecek et al. 2011) to convert VCF format files in Plink format. We then used the Plink v1.90b5.3 (Purcell et al. 2007) –pca command to run PCA analysis. We used the R package PopGenome v2.7.5 (Pfeifer et al. 2014) to compute nucleotide diversity, Tajima's D values and the F_{ST} index in 50 kb window overlapping over 10 kbp. Windows containing fewer than 5 SNPs were removed from the analysis. We used the R package ggplot2 (2_3.5.0) to plot results. Pairwise Wilcoxon tests were performed in R with Bonferroni or false discovery rate correction.

We built neighbornet networks using SplitsTree4 v 4.19.2 (Huson 1998). VCF file were converted to nexus format using PGDSpider v1.0 tool (Lischer and Excoffier 2012). For the

898 study of other populations, we included only one strain of each haplotype of the invasive
899 European population but not all CL1 and Swiss strains for network readability; we
900 nevertheless checked that the results were the same with all CL1 and Swiss strains.
901
902 **Long-read based assemblies**
903 PacBio Hifi reads of strains M1400 and M6697 were both assembled using canu v1.8 (Koren
904 et al. 2017) program with a set genome size of 44 Mb. Multiple assembly pipelines were used
905 for the other strains. Oxford Nanopore MinION reads and Illumina reads of the genomes of
906 both XIM9508 and MRC10 strains were assembled using HybridSPAdes (Antipov et al.
907 2016). Assemblies were manually curated and scaffolds were cut when an evidence of a
908 chimeric connection was detected (i.e. mis-paired short reads) as previously described
909 (Demené et al. 2022). For DUM005, the assembly was generated by Ra with basic parameters
910 that uses Oxford Nanopore MinION reads and corrects the assembly with Illumina reads
911 (https://github.com/lbcb-sci/ra) as it outperformed the HybridSPAdes assembly. As the
912 HybridSpades assembly of the MRC10 strain suggested absence of collinearity with the
913 MAT-proximal region M1400 and M6697, we further checked the assembly of the MAT-
914 proximal region of MRC010 by generating a meta-assembly of this strain. We used the
915 assembler tool canu v2.2 with an estimated genome size of 42 Mb. We also used Flye v2.9.3-
916 b1797 (Kolmogorov et al. 2019) with an estimated genome size of 42 Mb, --nano-raw for
917 reads with an error rate below 20 and a coverage for initial disjointig assembly of 50. Then we
918 used ragtag (Alonge et al. 2019) to patch the canu assembly with the Flye assembly as query
919 in a first loop. In a second loop, the canu assembly was patched with the first loop ragtag
920 assembly. Finally, we polished this second output assembly of ragtag with short reads and the
921 consensus part of medaka v1.11.3 with no change in the parameters
922 (https://github.com/nanoporetech/medaka). Assemblies statistics were obtained with quast
923 v5.1 (Gurevich et al. 2013). We assessed the completeness of each assembly using the
924 Benchmarking of Universal Single-Copy Orthologue (BUSCO) tools with the Sordariomyceta
925 ortholog set (sordariomycetes_odb10, 2020-08-05, n = 3817 orthologous groups searched)
926 (Manni et al. 2021). Gene models were predicted with the Helixer v0.3.1 pipeline (Holst et al.
927 2023). We also run Helixer pipeline for ESM15 and EP155 strains for the gene orthology
928 analysis. Statistics of the obtained gene annotation was obtained with the Agat v1.0.0 tool
929 (Dainat et al. 2020). Transposable elements of all genomes were annotated using
930 Repeatmasker v4-0-7 (Smit et al. 2013) and the customized library built for *C. parasitica* in
931 (Demené et al. 2022) contained in the "Curated_TEs_database_ESM015_EP155.fa" file

932 (available on the “Portail Data INRAE: Chromosomal rearrangements but no change of genes
933 and transposable elements repertoires in an invasive forest-pathogenic fungus" at
934 https://doi.org/10.15454/UTIB8U). The class Gipsy invader was renamed LTR-Ty3. We
935 filtered out TE copies shorter than 100bp. We used the HybridSPAdes preliminary assembly
936 of MRC10 for gene model annotations. To study gene functions of the EP155 genome, we
937 used the gene annotation available at http://jgi.doe.gov/ (Crouch et al. 2020). To study support
938 from RNAseq data, we used RNAseq data from the EP155 strain cultivated *in vitro* available
939 at Genebank under Project ID number PRJNA588887 and accessions numbers SRR10428542,
940 SRR10428543, SRR10428544 (Chun et al. 2020). Raw reads were mapped using STAR
941 v2.7.10a (Dobin et al. 2013) with the settings. --alignIntronMax 1000 --limitBAMsortRAM
942 1629219965--quantMode GeneCounts. We used the program featureCounts (Liao et al. 2013)
943 v2.0.6 with the options -p --countReadPairs -M - -B -O --largestOverlap. We considered a
944 gene to be supported for read count >10. To predict gene functions of the protein predicted by
945 Helixer, we used the funannotate pipeline with default options. To look for *Starships*, we ran
946 the Gene Finder Module of Starfish pipeline with default options (Gluck-Thaler and Vogan
947 2024). To predict gene functions of the protein predicted by Helixer, we used the funannotate
948 pipeline with default options. To identify putative *Starships*, we ran the Starfish pipeline with
949 default options (Gluck-Thaler and Vogan 2024). We studied the phylogenetic relationships of
950 the putative Starships captains by aligning them to the YRsuperfamRefs.faa from the starfish
951 database (Gluck-Thaler and Vogan 2024). Protein sequences were aligned using Clustal
952 Omega (Sievers and Higgins 2021)version 1.2.4 allowing for five iterations (--iterations 5).
953 Gaps in the resulting alignment were trimmed using trimAl v1.4.rev15 and the -gappyout
954 option. The phylogenetic relationship among proteins was inferred from the trimmed
955 alignment using FastTree (Price et al. 2009) under the Whelan-And-Goldman 2001 model
956 after 1,000 bootstraps (-boot 1000 and -wag options). Plots were performed in R v 4.1.2 using
957 ggtree v3.9.1.
958
959 **Estimation of retrotranposon insertion time and nucleotide divergence time**
960
961 To get an estimation of the insertion date of the transposable elements present in the MAT-
962 proximal non-recombining region, we applied two complementary methods. We first used the
963 divergence between the LTR sequences in retrotransposons, as these LTR sequences at their
964 edges are identical at the time of TE insertion and then diverge with time. For this, we used a
965 *de-novo* prediction of LTR retrotransposons using LTRharvest GenomeTools 1.6.2

966 (Ellinghaus et al. 2008; Gremme et al. 2013) and LTR_Finder v1.07 (Xu and Wang 2007). To
967 prepare data for LTRharvest, we first created enhanced suffix for the M1400 and M6697
968 genome assemblies using the GenomeTools Suffixerator (-dna -suf -lcp -bwt -bck -mirrored
969 parameters). We ran LTRharvest using two pipelines, designed to identify LTR
970 retrotransposons with a minimum and maximum LTR length of 100 bp and 7000 bp
971 respectively and at least 85% identity between two LTR regions, with and without canonical
972 TGCA motifs, respectively: i) -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -similar 85
973 -vic 10 -seed 20 -motif TGCA -motifmis 1; ii) -minlentltr 100, -maxlenltr 7000, -mintsd 4, -
974 maxtsd 6, -similar 85, -vic 10, -seed 20. Similarly, we ran LTR_Finder on the M1400 and
975 M6697 genome assemblies to retrotransposons with both TGCA and non-TGCA motifs and a
976 minimum and maximum LTR length of 100 bp and 7000 bp respectively and at least 85%
977 identity between two LTR regions (-D 15000, -d 1000, -l 100, -L 7000, -p 20, -C, -M 0.85).
978 Finally, we used LTR_retriever v2.8.5 (Ou and Jiang 2018) with default parameters to filter
979 out false positive LTR candidates identified by LTRharvest and LTR_Finder and get an
980 estimation of each element insertion date.
981
982 As a second method to estimate the insertion date of the transposable elements present in the
983 MAT-proximal non-recombining region, we used the set of curated consensus sequences from
984 (Demené et al. 2022) to annotate the inversion sequence or its surroundings. We first used
985 samtools faidx v1.9 (Li et al. 2009) to extract the sequence corresponding to inversion and
986 their 50 kb surroundings in M1400 and M6697 genome assemblies. We annotated the
987 sequences corresponding to the inversions and the concatenated 50 kb surroundings in both
988 isolates using RepeatMasker version 4.1.5 and rmblast as search engine (v2.10.0) with -no_is
989 -pa 20 -cutoff 250 -a parameters. Finally, we parsed the RepeatMasker .out file using the
990 helped script parseRM_merge_interrupted.pl and omitting Simple_repeat and
991 Low_complexity regions (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs). We
992 then built a summary of the alignments using the RepeatMasker helper script
993 buildSummary.pl and calculated sequence divergence using the calcDivergenceFromAlign.pl
994 script to finally render results with the createRepeatLandscape.pl from the same helper suite.
995
996
997 **Comparative genomics analyses**
998 Genome synteny between long reads assemblies were studied using the nucmer v3.1 program
999 (Marçais et al. 2018). Outputs were plotted with the R programs ggplot2 (Wickham 2009),

genoPlotR (Guy et al. 2010) and RIdeogram (Hao et al. 2020). The dotplot in the putative centromere region of the M1400 mating-type contig was performed using the online megablast alignment tool available at https://blast.ncbi.nlm.nih.gov/(last accessed 13th May 2024).

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

References

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

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

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

1036 Anagnostakis SL. 1987. Chestnut blight: The classical problem of an introduced pathogen.
1037 *Mycologia* [Internet]. Available from:
1038 https://www.tandfonline.com/doi/abs/10.1080/00275514.1987.12025367

1039 Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for
1040 hybrid assembly of short and long reads. *Bioinformatics* 32:1009–1015.

1041 Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-
1042 chromosome degeneration. *Nat. Rev. Genet.* 14:113–124.

1043 Bazzicalupo AL, Carpentier F, Otto SP, Giraud T. 2019. Little evidence of antagonistic
1044 selection in the evolutionary strata of fungal mating-type chromosomes
1045 (*Microbotryum lychnidis-dioicae*). *G3 Genes Genomes Genet.* 9:1987–1998.

1046 Belov AA, Witte TE, Overy DP, Smith ML. 2021. Transcriptome analysis implicates
1047 secondary metabolite production, redox reactions, and programmed cell death during
1048 allorecognition in Cryphonectria parasitica. *G3 GenesGenomesGenetics* 11:jkaa021.

1049 Berdan EL, Blanckaert A, Butlin RK, Flatt T, Slotte T, Wielstra B. 2022. Mutation
1050 accumulation opposes polymorphism: supergenes and the curious case of balanced
1051 lethals. *Philos. Trans. R. Soc. B Biol. Sci.* 377:20210199.

1052 Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex
1053 chromosomes. *Trends Ecol. Evol.* 24:94–102.

1054 Boideau F, Richard G, Coriton O, Huteau V, Belser C, Deniot G, Eber F, Falentin C, Ferreira
1055 de Carvalho J, Gilet M, et al. 2022. Epigenomic and structural events preclude
1056 recombination in Brassica napus. *New Phytol.* 234:545–559.

1057 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
1058 sequence data. *Bioinformatics* 30:2114–2120.

1059 Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza
1060 S, Brandenburg J-T, Coelho MA, Hood ME, et al. 2017. Evolutionary strata on young
1061 mating-type chromosomes despite the lack of sexual antagonism. *Proc. Natl. Acad.*
1062 *Sci.* 114:7067–7072.

1063 Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, Prieur SL, Coelho
1064 MA, Vienne DM de, Hartmann FE, Begerow D, et al. 2018. Multiple convergent
1065 supergene evolution events in mating-type chromosomes. *Nat. Commun.* 9:2000.

1066 Brusini J, Wayne ML, Franc A, Robin C. 2017. The impact of parasitism on resource
1067 allocation in a fungal host: the case of *Cryphonectria parasitica* and its mycovirus,
1068 Cryphonectria Hypovirus 1. *Ecol. Evol.* 7:5967–5976.

1069 Bucknell A, Wilson HM, Santos KCG do, Simpfendorfer S, Milgate A, Germain H, Solomon
1070 PS, Bentham A, McDonald MC. 2024. Sanctuary: A Starship transposon facilitating
1071 the movement of the virulence factor ToxA in fungal wheat pathogens.
1072 :2024.03.04.583430. Available from:
1073 https://www.biorxiv.org/content/10.1101/2024.03.04.583430v1

1074 Carpentier F, Rodríguez de la Vega RC, Jay P, Duhamel M, Shykoff JA, Perlin MH, Wallen
1075 RM, Hood ME, Giraud T. 2022. Tempo of degeneration across independently evolved
1076 nonrecombining regions. *Mol. Biol. Evol.* 39:msac060.

1077 Casacuberta E, González J. 2013. The impact of transposable elements in environmental
1078 adaptation. *Mol. Ecol.* 22:1503–1517.

1079 Charlesworth B. 2023. The fitness consequences of genetic divergence between polymorphic
1080 gene arrangements. *Genetics*:iyad218.

1081 Charlesworth D. 2016. The status of supergenes in the 21st century: recombination
1082 suppression in Batesian mimicry and sex chromosomes and other complex
1083 adaptations. *Evol. Appl.* 9:74–90.

1084 Charlesworth D, Vekemans X, Castric V, Glémin S. 2005. Plant self-incompatibility systems:
1085 a molecular evolutionary perspective. *New Phytol.* 168:61–69.

1086 Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M,
1087 Conseiller E, Debuchy R, et al. 2014. Multiple recent horizontal transfers of a large
1088 genomic region in cheese making fungi. *Nat. Commun.* 5:2876.

1089 Choi GH, Dawe AL, Churbanov A, Smith ML, Milgroom MG, Nuss DL. 2012. Molecular
1090 characterization of vegetative incompatibility genes that restrict hypovirus
1091 transmission in the chestnut blight fungus *Cryphonectria parasitica*. *Genetics*
1092 190:113–127.

1093 Choi GH, Nuss DL. 1992. Hypovirulence of chestnut blight fungus conferred by an infectious
1094 viral cDNA. *Science* 257:800–803.

1095 Chun J, Ko Y-H, Kim D-H. 2020. Transcriptome Analysis of Cryphonectria parasitica
1096 Infected With Cryphonectria hypovirus 1 (CHV1) Reveals Distinct Genes Related to
1097 Fungal Metabolites, Virulence, Antiviral RNA-Silencing, and Their Regulation. *Front.*
1098 *Microbiol.* 11:1711.

1099 Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, Hillman BI, Milgroom
1100 MG, Pangilinan J, Smith M, et al. 2020. Genome sequence of the chestnut blight
1101 fungus Cryphonectria parasitica EP155: A fundamental resource for an archetypical
1102 invasive plant pathogen. *Phytopathology®* 110:1180–1188.

1103 Dainat J, Hereñú D, Pucholt P. 2020. AGAT:Another Gff Analysis Toolkit to handle
1104 annotations in any GTF. *GFF Format*.

1105 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter
1106 G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools.
1107 *Bioinformatics* 27:2156–2158.

1108 Demené A, Laurent B, Cros-Arteil S, Boury C, Dutech C. 2022. Chromosomal
1109 rearrangements with stable repertoires of genes and transposable elements in an
1110 invasive forest-pathogenic fungus. *Peer Community J.* [Internet] 2. Available from:
1111 https://peercommunityjournal.org/articles/10.24072/pcjournal.127/

1112 Demené A, Legrand L, Gouzy J, Debuchy R, Saint-Jean G, Fabreguettes O, Dutech C. 2019.
1113 Whole-genome sequencing reveals recent and frequent genetic recombination between
1114 clonal lineages of Cryphonectria parasitica in western Europe. *Fungal Genet. Biol.*
1115 130:122–133.

1116 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
1117 Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*
1118 29:15–21.

1119 Dufresnes C, Borzée A, Horn A, Stöck M, Ostini M, Sermier R, Wassef J, Litvinchuck SN,
1120 Kosch TA, Waldman B, et al. 2015. Sex-chromosome homomorphy in palearctic tree
1121 frogs results from both turnovers and X-Y recombination. *Mol. Biol. Evol.* 32:2328–
1122 2337.

1123 Duhamel M, Hood ME, Vega RCR de la, Giraud T. 2023. Dynamics of transposable element
1124 accumulation in the non-recombining regions of mating-type chromosomes in anther-
1125 smut fungi. :2022.08.03.502670. Available from:
1126 https://www.biorxiv.org/content/10.1101/2022.08.03.502670v2

1127 Dussert Y, Legrand L, Mazet ID, Couture C, Piron M-C, Serre R-F, Bouchez O, Mestre P,
1128 Toffolatti SL, Giraud T, et al. 2020. Identification of the first oomycete mating-type
1129 locus sequence in the grapevine downy mildew pathogen, *Plasmopara viticola*. *Curr.*
1130 *Biol.* 30:3897-3907.e4.

1131 Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V. 2012. The chestnut blight
1132 fungus world tour: successive introduction events from diverse origins in an invasive
1133 plant fungal pathogen. *Mol. Ecol.* 21:3931–3946.

1134 Dutech C, Fabreguettes O, Capdevielle X, Robin C. 2010. Multiple introductions of divergent
1135 genetic lineages in an invasive fungal pathogen, Cryphonectria parasitica, in France.
1136 *Heredity* 105:220–228.

1137 Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for
1138 de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.

1139 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
1140 genomics. *Genome Biol.* 20:238.

1141 Fredericksen M, Fields PD, Pasquier LD, Ricci V, Ebert D. 2023. QTL study reveals
1142 candidate genes underlying host resistance in a Red Queen model system. *PLOS*
1143 *Genet.* 19:e1010570.

1144 Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank
1145 JE. 2020. Sex chromosome evolution: So many exceptions to the rules. *Genome Biol.*
1146 *Evol.*

1147 Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack
1148 TL, Wood CW, Slot JC, Lopez-Nicora HD, et al. 2022. Giant Starship Elements
1149 Mobilize Accessory Genes in Fungal Genomes. *Mol. Biol. Evol.* 39:msac109.

1150 Gluck-Thaler E, Vogan AA. 2024. Systematic identification of cargo-mobilizing genetic
1151 elements reveals new dimensions of eukaryotic diversity. *Nucleic Acids Res.* 52:5496–
1152 5513.

1153 Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A Comprehensive Software Library
1154 for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans.*
1155 *Comput. Biol. Bioinform.* 10:645–656.

1156 Grognet P, Bidard F, Kuchly C, Tong LCH, Coppin E, Benkhali JA, Couloux A, Wincker P,
1157 Debuchy R, Silar P. 2014. Maintaining two mating types: structure of the mating type
1158 locus and its role in heterokaryosis in *Podospora anserina*. *Genetics* 197:421–432.

1159 Guerin L, Froidefond G, Xu X-M. 2001. Seasonal patterns of dispersal of ascospores of
1160 Cryphonectria parasitica (chestnut blight). *Plant Pathol.* 50:717–724.

1161 Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for
1162 genome assemblies. *Bioinformatics* 29:1072–1075.

1163 Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome
1164 visualization in R. *Bioinformatics* 26:2334–2335.

1165 Guyot L, Chahine E, Lalanne C, Hartmann FE, Giraud T. 2024. Sheltered load in fungal
1166 mating-type chromosomes revealed by fitness experiments. :2024.09.10.612177.
1167 Available from: https://www.biorxiv.org/content/10.1101/2024.09.10.612177v1

1168 Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. RIdeogram: drawing SVG
1169 graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.*
1170 6:e251.

1171 Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, Malagnac F,
1172 Grognet P, Giraud T. 2021. Recombination suppression and evolutionary strata around
1173 mating-type loci in fungi: documenting patterns and understanding evolutionary and
1174 mechanistic causes. *New Phytol.* 229:2470–2491.

1175 Helleu Q, Roux C, Ross KG, Keller L. 2022. Radiation and hybridization underpin the spread
1176 of the fire ant social supergene. *Proc. Natl. Acad. Sci.* 119:e2201040119.

1177 Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. 2000. Genetic structure of newly
1178 established populations of *Cryphonectria parasitica*. *Mycol. Res.* 104:1108–1116.

1179 Holst F, Bolger A, Günther C, Maß J, Triesch S, Kindel F, Kiel N, Saadat N, Ebenhöh O,
1180 Usadel B, et al. 2023. Helixer–de novo prediction of primary Eukaryotic gene models
1181 combining deep learning and a hidden markov model. :2023.02.06.527280. Available
1182 from: https://www.biorxiv.org/content/10.1101/2023.02.06.527280v2

1183 Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinforma. Oxf.*
1184 *Engl.* 14:68–73.

1185 Ironside JE. 2010. No amicable divorce? Challenging the notion that sexual antagonism
1186 drives sex chromosome evolution. *BioEssays* 32:718–726.

1187 Jay P, Aubier TG, Joron M. 2024. The interplay of local adaptation and gene flow may lead to
1188 the formation of supergenes. *Mol. Ecol.*:e17297.

1189 Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M. 2021.
1190 Mutation load at a mimicry supergene sheds new light on the evolution of inversion
1191 polymorphisms. *Nat. Genet.* 53:288–293.

1192 Jay P, Jeffries D, Hartmann FE, Véber A, Giraud T. 2024. Why do sex chromosomes
1193 progressively lose recombination? *Trends Genet.* [Internet]. Available from:
1194 https://www.sciencedirect.com/science/article/pii/S0168952524000672

1195 Jay P, Jeffries DL, Hartmann FE, Véber A, Giraud T. 2024. Why do sex chromosomes
1196 progressively lose recombination? *Trends in Genetics*:in press.

1197 Jay P, Tezenas E, Véber A, Giraud T. 2022. Sheltering of deleterious mutations explains the
1198 stepwise extension of recombination suppression on sex chromosomes and other
1199 supergenes. *PLOS Biol.* 20:e3001698.

1200 Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J, Dasmahapatra
1201 KK, Joron M. 2018. Supergene evolution triggered by the introgression of a
1202 chromosomal inversion. *Curr. Biol.* 28:1839-1845.e3.

1203 Kasuga T, White TJ, Taylor JW. 2002. Estimation of nucleotide substitution rates in
1204 eurotiomycete fungi. *Mol. Biol. Evol.* 19:2318–2324.

1205 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using
1206 repeat graphs. *Nat. Biotechnol.* 37:540–546.

1207 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable
1208 and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
1209 *Genome Res.* 27:722–736.

1210 Kubisiak TL, Milgroom MG. 2006a. Markers linked to vegetative incompatibility (vic) genes
1211 and a region of high heterogeneity and reduced recombination near the mating type
1212 locus (MAT) in *Cryphonectria parasitica*. *Fungal Genet. Biol.* 43:453–463.

1213 Kubisiak TL, Milgroom MG. 2006b. Markers linked to vegetative incompatibility (*vic*) genes
1214 and a region of high heterogeneity and reduced recombination near the mating type
1215 locus (*MAT*) in *Cryphonectria parasitica*. *Fungal Genet. Biol. FG B* 43:453–463.

1216 Küpper C, Stocks M, Risse JE, Remedios N dos, Farrell LL, McRae SB, Morgan TC,
1217 Karlionova N, Pinchuk P, Verkuil YI, et al. 2016. A supergene determines highly
1218 divergent male reproductive morphs in the ruff. *Nat. Genet.* 48:79–83.

1219 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
1220 alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

1221 Legrand S, Saifudeen A, Bordelet H, Vernerey J, Guille A, Bignaud A, Thierry A, Acquaviva
1222 L, Gaudin M, Sanchez A, et al. 2024. Absence of chromosome axis protein
1223 recruitment prevents meiotic recombination chromosome-wide in the budding yeast
1224 Lachancea kluyveri. *Proc. Natl. Acad. Sci.* 121:e2312820121.

1225 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
1226 R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
1227 25:2078–2079.

1228 Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read
1229 mapping by seed-and-vote. *Nucleic Acids Res.* 41:e108–e108.

1230 Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for
1231 connecting population genetics and genomics programs. *Bioinformatics* 28:298–299.

1232 Liu YC, Linder-Basso D, Hillman BI, Kaneko S, Milgroom MG. 2003. Evidence for
1233 interspecies transmission of viruses in natural populations of filamentous fungi in the
1234 genus Cryphonectria. *Mol. Ecol.* 12:1619–1628.

1235 Lo Y-C, Bruxaux J, Rodríguez de la Vega RC, O'Donnell S, Snirc A, Coton M, Le Piver M,
1236 Le Prieur S, Roueyre D, Dupont J, et al. 2023. Domestication in dry-cured meat
1237 Penicillium fungi: Convergent specific phenotypes and horizontal gene transfers
1238 without strong genetic subdivision. *Evol. Appl.* 16:1637–1660.

1239 Lovat C-A, Donnelly DJ. 2019. Mechanisms and metabolomics of the host–pathogen
1240 interactions between Chestnut (Castanea species) and Chestnut blight (Cryphonectria
1241 parasitica). *For. Pathol.* 49:e12562.

1242 Maloisel L, Rossignol JL. 1998. Suppression of crossing-over by DNA methylation in
1243 *Ascobolus*. *Genes Dev.* 12:1381–1389.

1244 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: Novel
1245 and streamlined workflows along with broader and deeper phylogenetic coverage for
1246 scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38:4647–4654.

1247 Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4:
1248 A fast and versatile genome alignment system. *PLOS Comput. Biol.* 14:e1005944.

1249 McGuire IC, Davis JE, Double ML, MacDonald WL, Rauscher JT, McCawley S, Milgroom
1250 MG. 2005. Heterokaryon formation and parasexual recombination between
1251 vegetatively incompatible lineages in a population of the chestnut blight fungus,
1252 *Cryphonectria parasitica*. *Mol. Ecol.* 14:3657–3669.

1253 McGuire IC, Marra RE, Milgroom MG. 2004. Mating-type heterokaryosis and selfing in
1254 *Cryphonectria parasitica*. *Fungal Genet. Biol.* 41:521–533.

1255 McGuire IC, Marra RE, Turgeon BG, Milgroom MG. 2001. Analysis of mating-Type genes
1256 in the chestnut blight fungus, *Cryphonectria parasitica*. *Fungal Genet. Biol.* 34:131–
1257 144.

1258 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
1259 Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a
1260 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
1261 *Res.* 20:1297–1303.

1262 Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. 2008. The mating-type chromosome
1263 in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early
1264 evolution of sex chromosomes. *PLOS Genet.* 4:e1000030.

1265 Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneko S. 1996. Intercontinental population
1266 structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* 88:179–
1267 190.

1268 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A,
1269 Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic
1270 inference in the genomic era. *Mol. Biol. Evol.* 37:1530–1534.

1271 Moinet A, Schlichta F, Peischl S, Excoffier L. 2022. Strong neutral sweeps occurring during a
1272 population contraction. *Genetics* 220:iyac021.

1273 Orteu A, Kucka M, Gordon IJ, Ng'iru I, van der Heijden ESM, Talavera G, Warren IA,
1274 Collins S, ffrench-Constant RH, Martins DJ, et al. 2024. Transposable element
1275 insertions are associated with Batesian mimicry in the pantropical butterfly
1276 *Hypolimnas misippus*. *Mol. Biol. Evol.* 41:msae041.

1277 Otto SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. *Nat. Rev.*
1278 *Genet.* 3:252–261.

1279 Ou S, Jiang N. 2018. LTR_retriever: A highly accurate and sensitive program for
1280 identification of long terminal repeat retrotransposons. *Plant Physiol.* 176:1410–1422.

1281 Peck LD, Llewellyn T, Bennetot B, O'Donnell S, Nowell RW, Ryan MJ, Flood J, Vega RCR
1282 de la, Ropars J, Giraud T, et al. 2023. Horizontal transfers between fungal Fusarium
1283 species contributed to successive outbreaks of coffee wilt disease.
1284 :2023.12.22.572981. Available from:
1285 https://www.biorxiv.org/content/10.1101/2023.12.22.572981v1

1286 Pegueroles C, Ordóñez V, Mestres F, Pascual M. 2010. Recombination and selection in the
1287 maintenance of the adaptive value of inversions. *J. Evol. Biol.* 23:2709–2717.

1288 Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient
1289 Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31:1929–
1290 1936.

1291 Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing Large Minimum Evolution Trees
1292 with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* 26:1641–1650.

1293 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P,
1294 de Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome
1295 association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.

1296 Rigling D, Prospero S. 2018. Cryphonectria parasitica, the causal agent of chestnut blight:
1297 invasion history, population biology and disease control. *Mol. Plant Pathol.* 19:7–20.

1298 Ropars J, Rodríguez de la Vega RC, López-Villavicencio M, Gouzy J, Sallet E, Dumas É,
1299 Lacoste S, Debuchy R, Dupont J, Branca A, et al. 2015. Adaptive horizontal gene
1300 transfers between multiple cheese-associated fungi. *Curr. Biol.* 25:2562–2569.

1301 Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr. Biol.*
1302 24:R288–R294.

1303 Shin J-H, Blay S, McNeney B, Graham J, others. 2006. LDheatmap: an R function for
1304 graphical display of pairwise linkage disequilibria between single nucleotide
1305 polymorphisms. *J. Stat. Softw.* 16:1–10.

1306 Sievers F, Higgins DG. 2021. The Clustal Omega Multiple Alignment Package. In: Katoh K,
1307 editor. Multiple Sequence Alignment: Methods and Protocols. New York, NY:
1308 Springer US. p. 3–16. Available from: https://doi.org/10.1007/978-1-0716-1036-7_1

1309 Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available at:
1310 http://www.repeatmasker.org Accessed: November 30, 2017.

1311 Stauber L, Badet T, Feurtey A, Prospero S, Croll D. 2021. Emergence and diversification of a
1312 highly invasive chestnut pathogen lineage across southeastern Europe.Castric V,
1313 Weigel D, Laine A-L, Castric V, editors. *eLife* 10:e56279.

1314 Stauber L, Croll D, Prospero S. 2022. Temporal changes in pathogen diversity in a perennial
1315 plant–pathogen–hyperparasite system. *Mol. Ecol.* 31:2073–2088.

1316 Stauber L, Prospero S, Croll D. 2020. Comparative genomics analyses of lifestyle transitions
1317 at the origin of an invasive fungal pathogen in the genus *Cryphonectria*. *mSphere*
1318 5:10.1128/msphere.00737-20.

1319 Stevison LS, Hoehn KB, Noor MAF. 2011. Effects of Inversions on Within- and Between-
1320 Species Recombination and Divergence. *Genome Biol. Evol.* 3:830–841.

1321 Stolle E, Pracana R, López-Osorio F, Priebe MK, Hernández GL, Castillo-Carrillo C, Arias
1322 MC, Paris CI, Bollazzi M, Priyam A, et al. 2022. Recurring adaptive introgression of a
1323 supergene variant that determines social organization. *Nat. Commun.* 13:1180.

1324 Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. 2017. Large-scale suppression
1325 of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat.*
1326 *Commun.* 8:1–8.

1327 Taylor JW, Berbee ML. 2006. Dating divergences in the Fungal Tree of Life: review and new
1328 analyses. *Mycologia* 98:838–849.

1329 Tellier A, Moreno-Gámez S, Stephan W. 2014. Speed of adaptation and genomic footprints of
1330 host–parasite coevolution under arms race and trench warfare dynamics. *Evolution*
1331 68:2211–2224.

1332 Tralamazza SM, Gluck-Thaler E, Feurtey A, Croll D. 2024. Copy number variation
1333 introduced by a massive mobile element facilitates global thermal adaptation in a
1334 fungal wheat pathogen. *Nat. Commun.* 15:5728.

1335 Umen JG. 2011. Evolution of sex and mating loci: An expanded view from Volvocine algae.
1336 *Curr. Opin. Microbiol.* 14:634–641.

1337 Urquhart A, Vogan AA, Gluck-Thaler E. 2024. Starships: a new frontier for fungal biology.
1338 *Trends Genet.* [Internet] 0. Available from:
1339 https://www.cell.com/trends/genetics/abstract/S0168-9525(24)00183-5

1340 Vittorelli N, Hartmann, Giraud T. 2022. Stepwise recombination suppression around the
1341 mating-type locus in the fungus *Schizothecium tetrasporum* (Ascomycota,
1342 Sordariales).

1343 Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer Science & Business
1344 Media

1345 Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR
1346 retrotransposons. *Nucleic Acids Res.* 35:W265–W268.

1347 Yan Z, Martin SH, Gotzek D, Arsenault SV, Duchen P, Helleu Q, Riba-Grognuz O, Hunt BG,
1348 Salamin N, Shoemaker D, et al. 2020. Evolution of a supergene that regulates a trans-
1349 species social polymorphism. *Nat. Ecol. Evol.* 4:240–249.

1350 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*
1351 24:1586–1591.

1352

1353

1354

1355 **Acknowledgments**

1356 This work was supported by the European Research Council (ERC) EvolSexChrom (832352) grant
1357 and a Fondation Louis D grant from the Institut de France to TG, and by the ANR PIA grant # ANR-
1358 20-IDEES-0002 to F.E.H. We acknowledge the GenOuest bioinformatics core facility
1359 (https://www.genouest.org) for providing the computing infrastructure, (GPU) for genome
1360 annotation". We thank Jeanne Ropars, Aaron Vogan and Marie Foulongne for discussions and
1361 suggestions on the analyses. The authors declare that they have no competing interests.

1362

1363 TG and FEH conceptualized the study and acquired funding. TG supervised the study. AL, AS, CD,
1364 AD, SP, DC, LS and TB contributed to data and their availability. AL and AS performed *in vitro*
1365 culture and DNA extraction. FEH, RRDLV, QR, JPV, LS and TB analyzed genomes. FEH and TG
1366 wrote the original draft. All authors edited the manuscript.

1367

1368 The raw data and the new assemblies produced in this study will be published pending scientific
1369 review.

1370

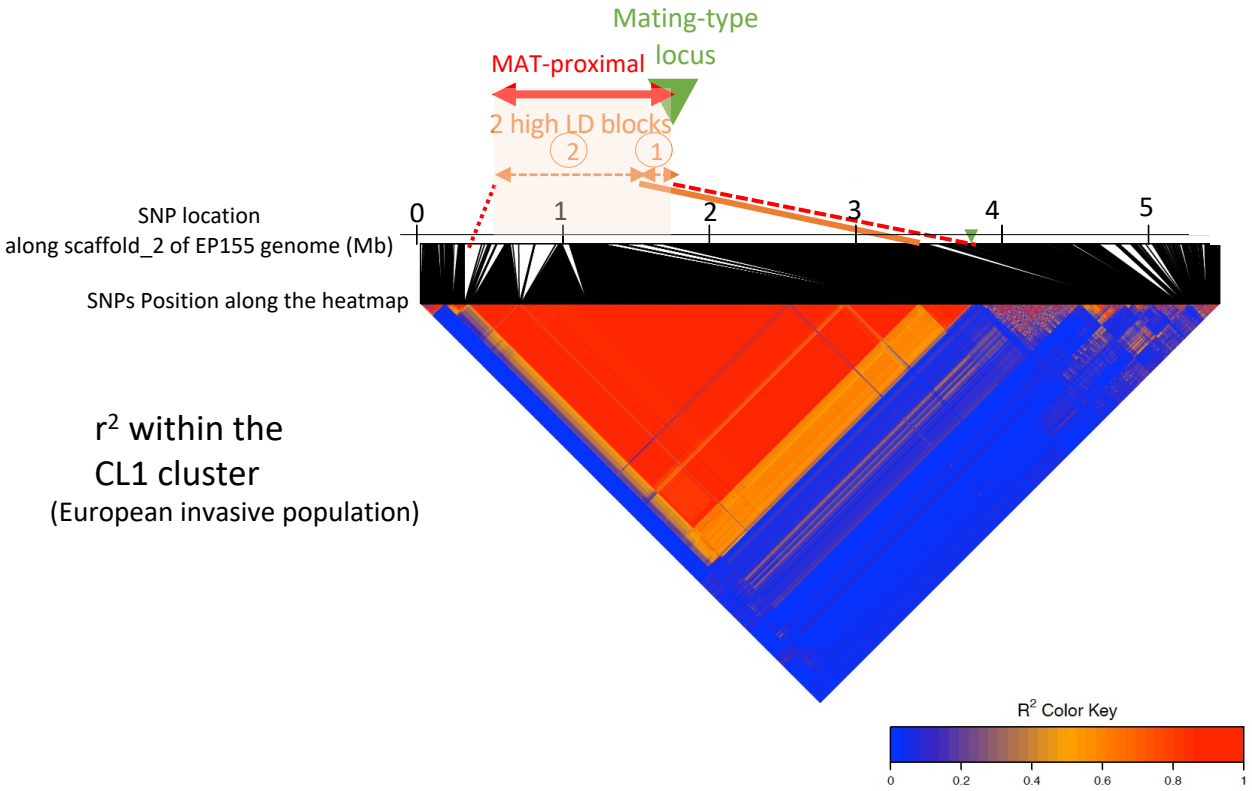
1371 **Tables**

1372 **Table 1: Distribution of mating types (*MAT-1* and *MAT-2*) among non-recombining**
1373 **haplotypes of the MAT-proximal region in the European invasive *Cryphonectria***
1374 ***parasitica* CL1 genetic cluster and the Swiss populations.** MAT-Prox1 and MAT-Prox2
1375 haplotypes were defined based on the clusters of the principal component analysis.
1376
1377 **Figures**
1378
1379 **Figure 1: Linkage disequilibrium (LD) analysis along the contig carrying the mating-**
1380 **type locus in a *Cryphonectria parasitica* European invasive population.** LD heatmaps
1381 using single nucleotide polymorphisms (SNPs; n=3815) located on the contig carrying the
1382 mating-type locus (scaffold_2 of the EP155 genome) in the CL1 genetic cluster (European
1383 invasive population introduced from North America); pairs of SNPs with high LD, i.e. r^2
1384 >0.9, correspond to the red triangle. The mating-type locus location is shown with a green
1385 triangle and the MAT-proximal region lacking recombination is shown with a red arrow. The
1386 two high-LD blocks within the MAT-proximal region are shown with orange arrows. SNPs at
1387 the limit of the MAT-proximal region and the two high LD blocks were manually highlighted
1388 with red dotted lines and an orange line.
1389
1390 **Figure 2: Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-**
1391 **proximal region lacking recombination and other genomic regions in a *Cryphonectria***
1392 ***parasitica* European invasive population. A-C.** Principal component analysis (PCA). Two
1393 principal components are presented. Percentage of variance explained by each PC is indicated
1394 into brackets. Strains are colored according to their mating type (*MAT-1* or *MAT-2*). **D-E.**
1395 Neighbor-net network from a SplitsTree analysis. These analyses were performed in the CL1
1396 genetic cluster (European invasive populations introduced from North America) based on: **A**
1397 **and D.** SNPs (n=2,220) located within the MAT-proximal region along the contig carrying
1398 the mating-type locus (scaffold_2 of the EP155 genome).; **B.** SNPs (n=1,595) located in other
1399 regions along the contig carrying the mating-type locus; **C and E** SNPs (n=11,289) located on
1400 other contigs of the EP155 genome. On panels A and D, the two clusters corresponding to the
1401 MAT-Prox1 and MAT-Prox2 haplotypes are shown with red circles. The identified haplotype
1402 of each strain is indicated in Table S1. The number of strains within the MAT-Prox1
1403 haplotype is indicated on panel A by the letter n. The newly sequenced M1400 and M6697
1404 strains are highlighted with a black rectangle
1405

1406 **Figure 3: Synteny and rearrangements between the two newly sequenced genomes of**
1407 **European invasive strains introduced from North America (M1400 and M6697) in the**
1408 **chromosome carrying the MAT-proximal region lacking recombination in**
1409 ***Cryphonectria parasitica*.** Blue links show colinear regions and orange links show inverted
1410 regions in the MAT-proximal region. Grey links show other regions. The mating-type locus is
1411 located with a green diamond. The MAT-proximal region defined from LD analyses is
1412 indicated with red arrows. Gene density tracks is shown with a color gradient (blue with low
1413 density, orange with high density).
1414
1415 **Figure 4: Genetic diversity and divergence between non-recombining haplotypes of the**
1416 **MAT-proximal region along the contig carrying the mating-type locus in a**
1417 ***Cryphonectria parasitica* European invasive population.** A-B. GC content (A) and
1418 transposable element densities (B) along the contig carrying the mating-type locus
1419 (tig_0000001) of the M1400 *C. parasitica* genome. C. Relative divergence (F_{ST}) between
1420 strains of the MAT-Prox1 and MAT-Prox2 haplotypes: D-E. Nucleotide diversity within
1421 pools of strains for each MAT-Proximal haplotype; F-G-H Tajima's D for all strains pooled
1422 and within pools of strains for each MAT-Proximal haplotype; The MAT-proximal region
1423 defined from LD analyses and the inversion between M1400 and M6697 genomes are
1424 indicated with red and blue arrows respectively. The mating-type locus location is shown with
1425 a green triangle. The location of the putative centromere is indicated by a yellow circle. All
1426 population statistics were computed for the 1990 Swiss population along the mating-type
1427 contig (tig_0000001) of the M1400 genome per 50-kb window overlapping over 10 kbp.
1428 Windows containing fewer than 5 SNPs were removed from the analysis.
1429
1430 **Figure 5: Annotation of transposable elements (TEs) and estimates of their insertion**
1431 **time in the MAT-proximal region lacking recombination and other genomic regions in**
1432 ***Cryphonectria parasitica*.** A. Annotation of transposable elements (TEs) and gene density
1433 along the MAT-proximal region in M1400 and M6697 genomes. On the left panel (1) the
1434 figure shows the genomic region from 7 Mb on tig_0000001 of the M1400 genome and from
1435 2.8 Mb on tig_0000060 of the M6697 genome. The mating-type locus is located with a green
1436 triangle. The two high-LD blocks within the MAT-proximal region are shown with orange
1437 arrows. Synteny for 10 kb segments with identity > 90 % is shown in red and inversion in
1438 blue. Transposons larger than 5 kb are shown in purple for the transposons annotated as LTR-
1439 Ty3 by and in green for other transposons. Orthologous genes shared between genomes are

1440 shown in black and unique to each each genome in grey. Genes with DU3435 domains
1441 (putative starship captains) are shown in red. On the right panel (2) the figure shows a zoom
1442 on the starship present in M1400 and absent in M6697. Transposons larger than 150 pb are
1443 shown in orange; orthologous genes shared between genomes are shown in black and unique
1444 to one genome in grey and genes with DU3435 domains are shown in red. Synteny for
1445 segments with identity > 90 % is shown in red and inversion in blue. B. Relative percentage
1446 of each TE annotation (in %) in the non-recombining MAT-proximal region and other
1447 recombining regions of M6697 and M1400 genomes. C. Pairwise genetic distance between
1448 TE copies within the inversion and around (50 kb) in both M1400 and M6697 genomes.
1449 Distribution of the Kimura substitution levels computed using the consensus sequence for the
1450 TEs. D. Estimates of the age of intact copies of LTR within and around the inversion.
1451
1452 **Figure 6: Transposable elements content and Starship content in high quality genome**
1453 **assemblies from other *Cryphonectria parasitica* populations from the native and the**
1454 **introduced range. A.** TE load (percentage of base pairs occupied by transposable elements)
1455 in the seven high quality genome assemblies in the MAT proximal region and other
1456 recombining regions. **B.** Annotation of transposable elements (TEs) and gene density along
1457 the MAT-proximal region in M1400 and the two Asian genomes ESM15 and XIM9508.
1458 Orthologous genes shared between M1400 and M6697 genomes are shown in black and
1459 unique to M1400 in grey as in Figure 5. Genes of ESM15 and XIM9508 are shown in
1460 darkblue? Genes with DU3435 domains (putative starship captains) are shown in red.
1461 Transposons larger than 150 pb are shown in orange. Synteny for segments with identity > 90
1462 % is shown in red and inversion in blue.
1463
1464 **Figure 7: Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-**
1465 **proximal region lacking recombination (A) and other genomic regions (B) in other**
1466 **resequenced *Cryphonectria parasitica* populations from the native and the introduced**
1467 **range.** Neighbor-net network from a SplitsTree analysis based on: **A** SNPs (n=4,120) located
1468 within the MAT-proximal region along the contig carrying the mating-type locus (scaffold_2
1469 of the EP155 genome); **B** SNPs (n= 103,058) located on other contigs of the EP155 genome.
1470 Color of empty circles around strain ID indicate the genetic clusters the strains belong to
1471 (CL1, CL2, CL3, CL4; n= 33 strains**)**. Dotted circles indicate strains of the native range and
1472 plain circles indicate strains of the introduced range. Red and blue dots near strain ID indicate
1473 the mating type (*MAT-1* or *MAT-2*, respectively).

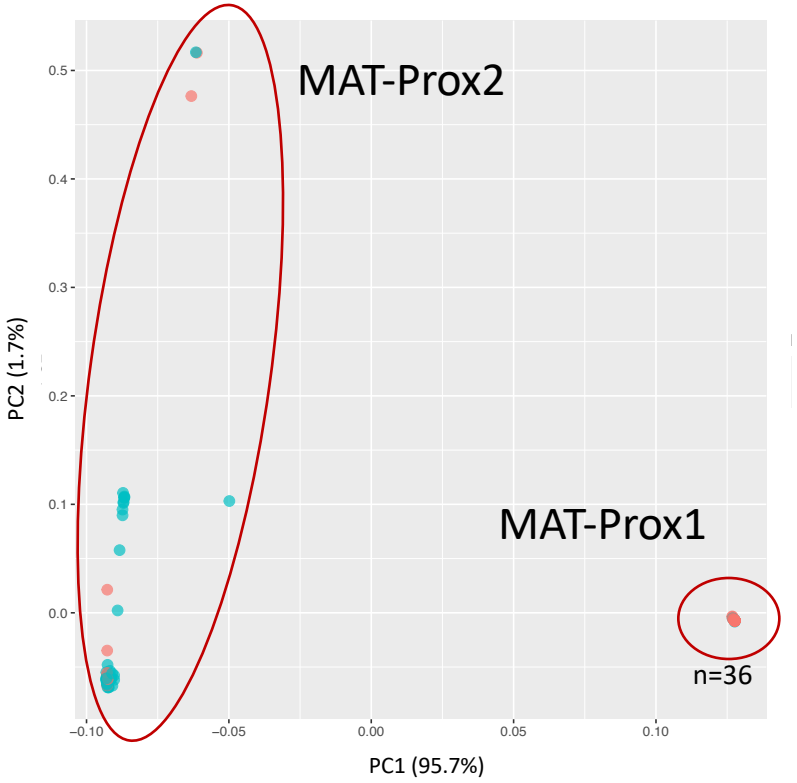
	Population	CL1 genetic cluster		Swiss 1990 population		Swiss 2019 population	
	Haplotype	MAT-Prox1	MAT-Prox2	MAT-Prox1	MAT-Prox2	MAT-Prox1	MAT-Prox2
Number of strains in each PCA cluster	total	36	52	20	51	8	54
	MAT-1	30	11	18	17	5	15
	MAT-2	6	41	2	34	3	39
% of strains in each PCA cluster	*MAT-1*	83	21	90	33	63	28
	MAT-2	17	79	10	67	38	72



r^2 within the
CL1 cluster
(European invasive population)

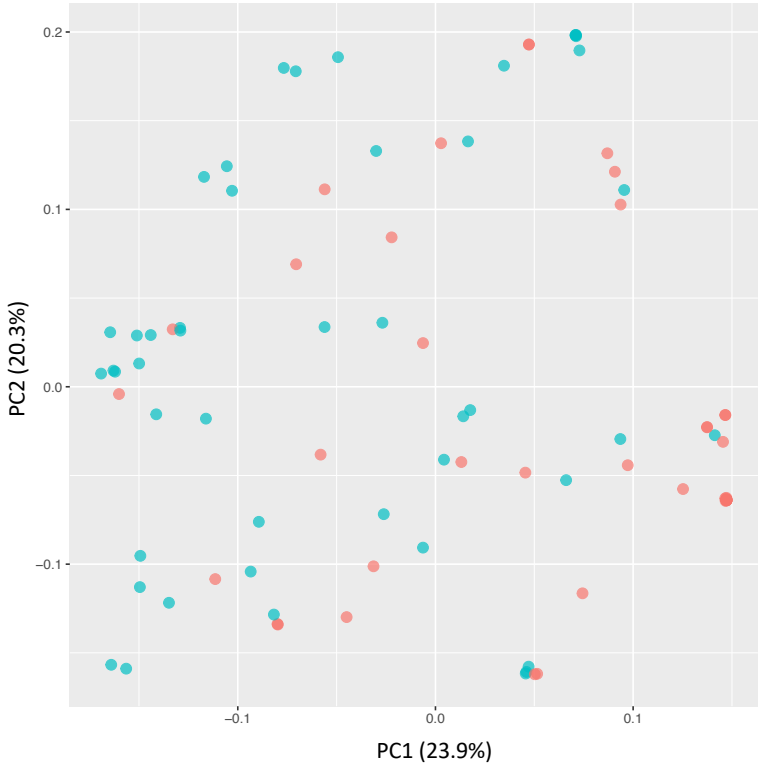
CL1 genetic cluster (European invasive population)

A



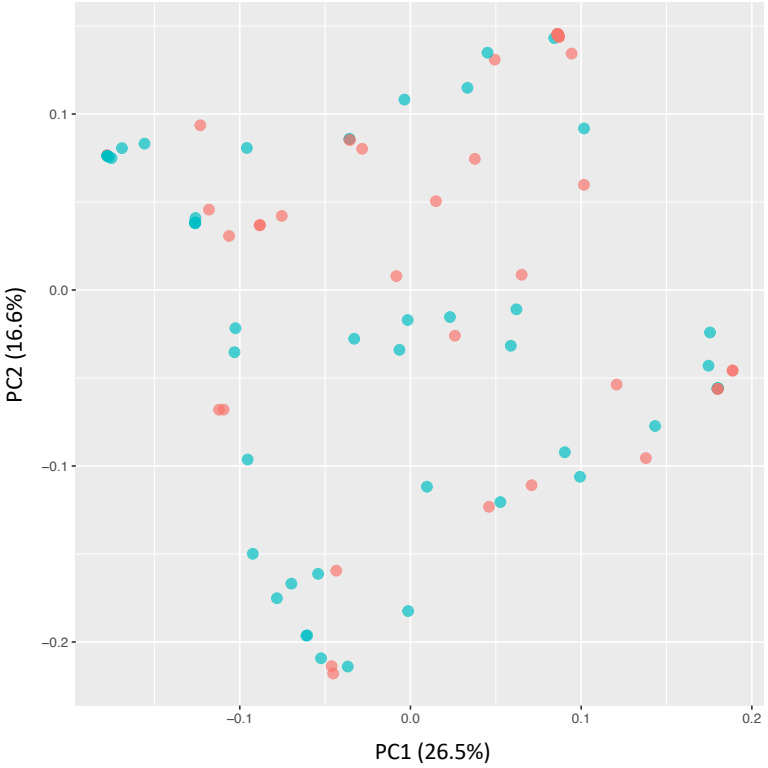
MAT-proximal region

B



Other regions

C



Other contigs (EP155 genome)

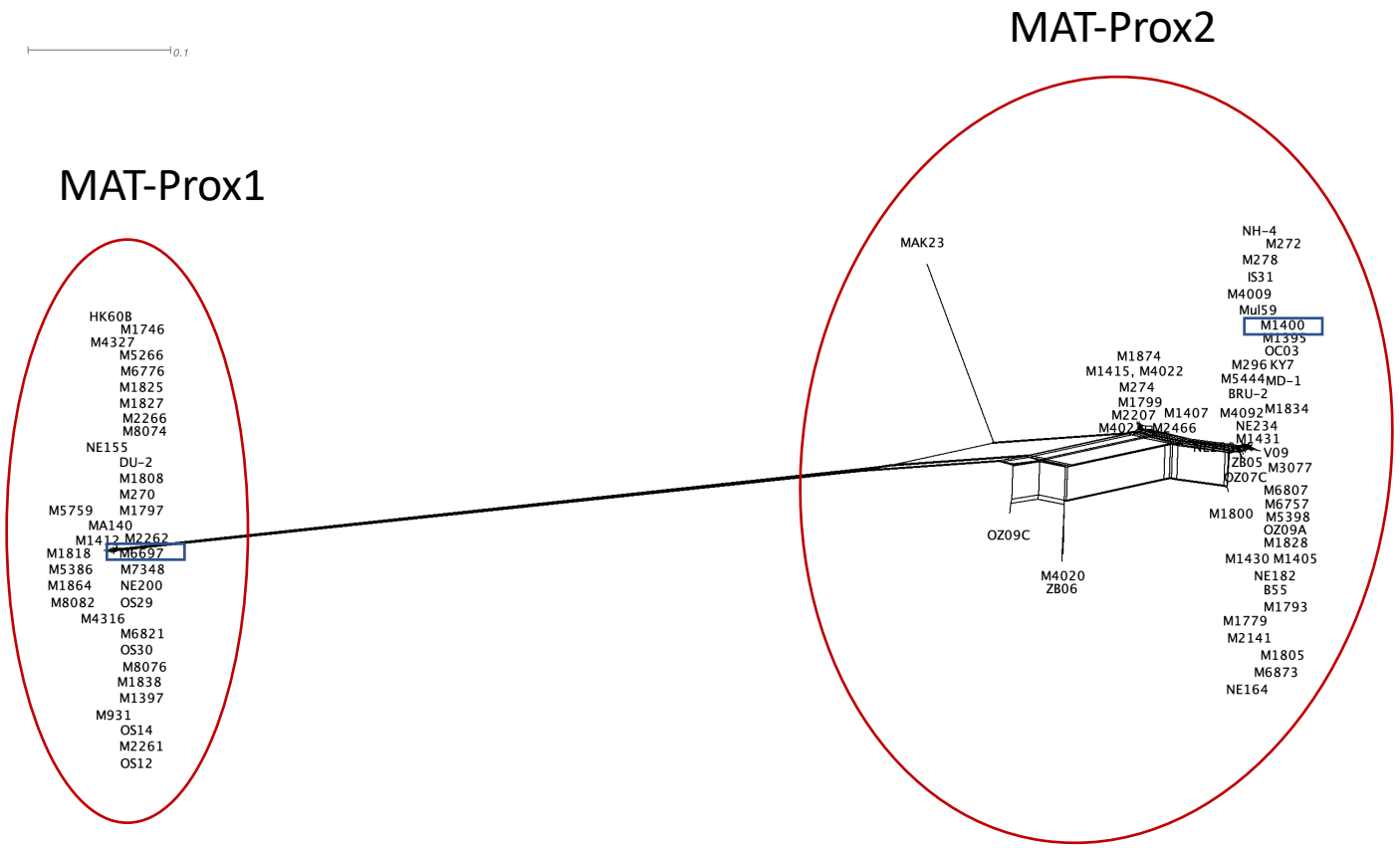
scaffold_2 (EP155 genome)

Mating-type allele:

MAT−1

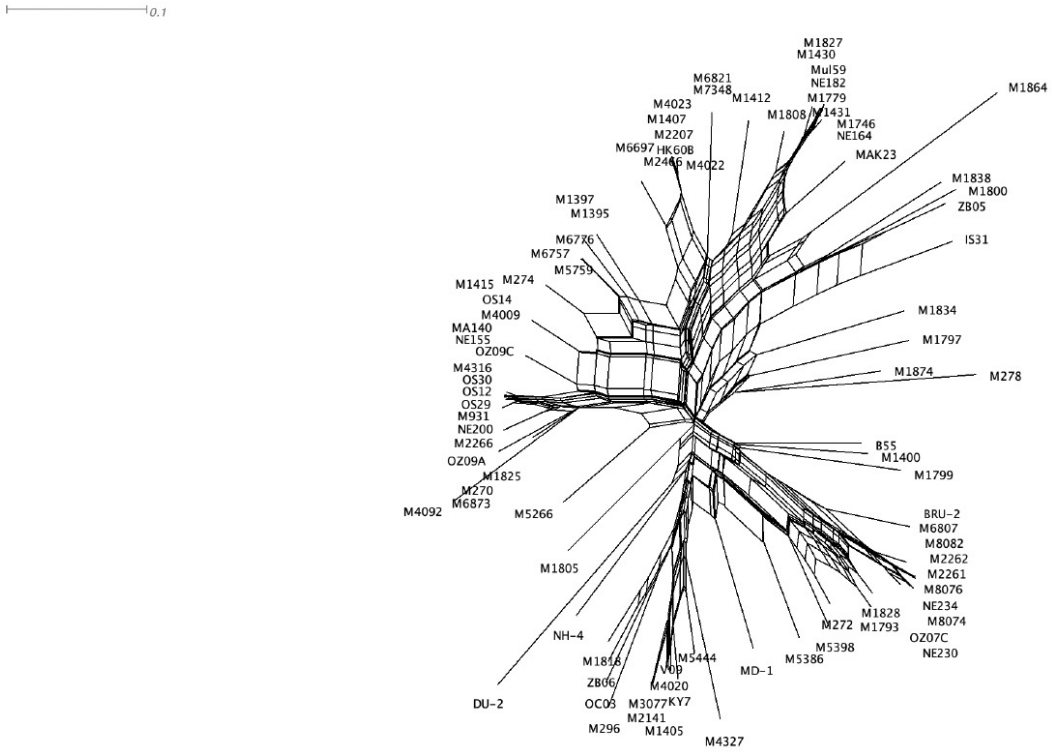
MAT−2

D

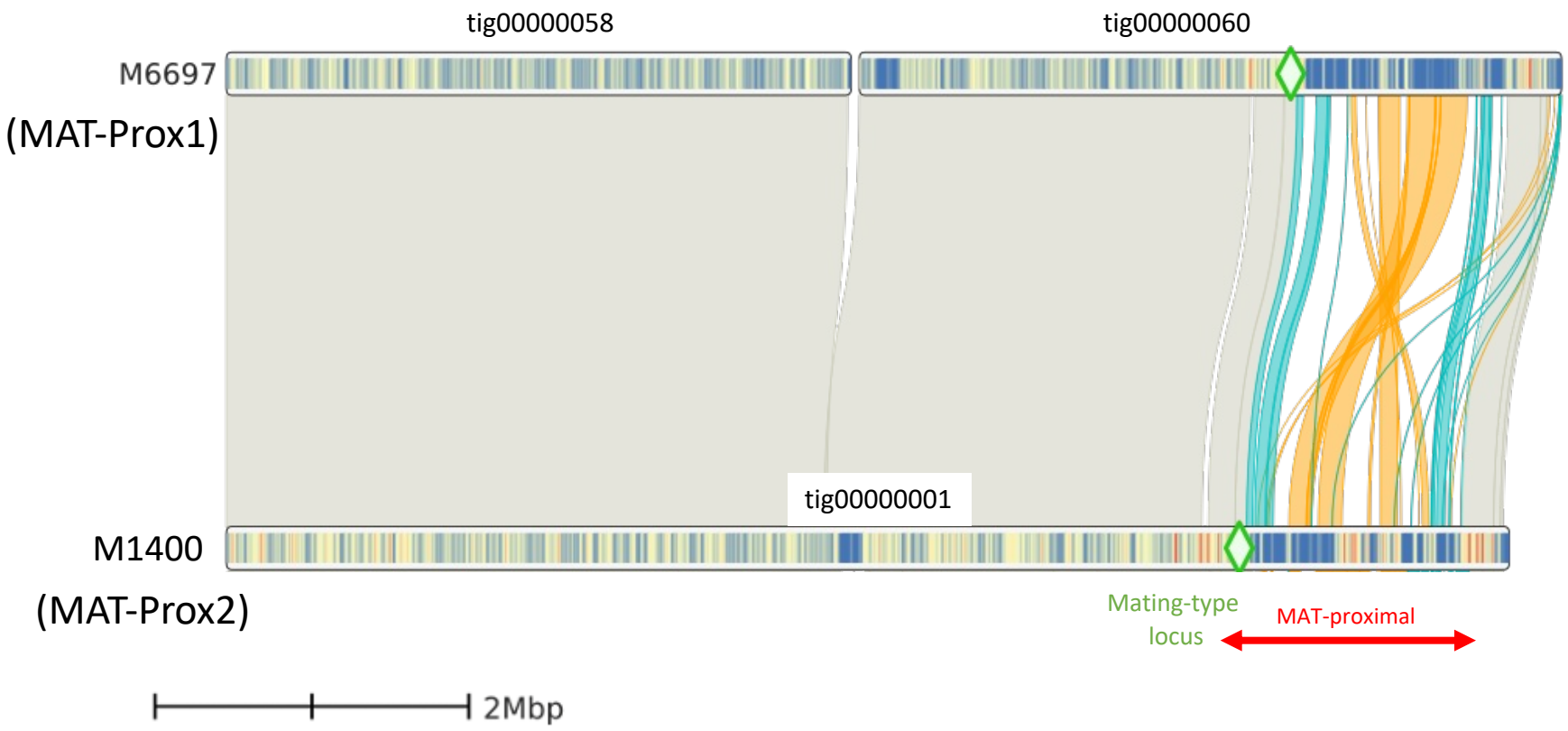


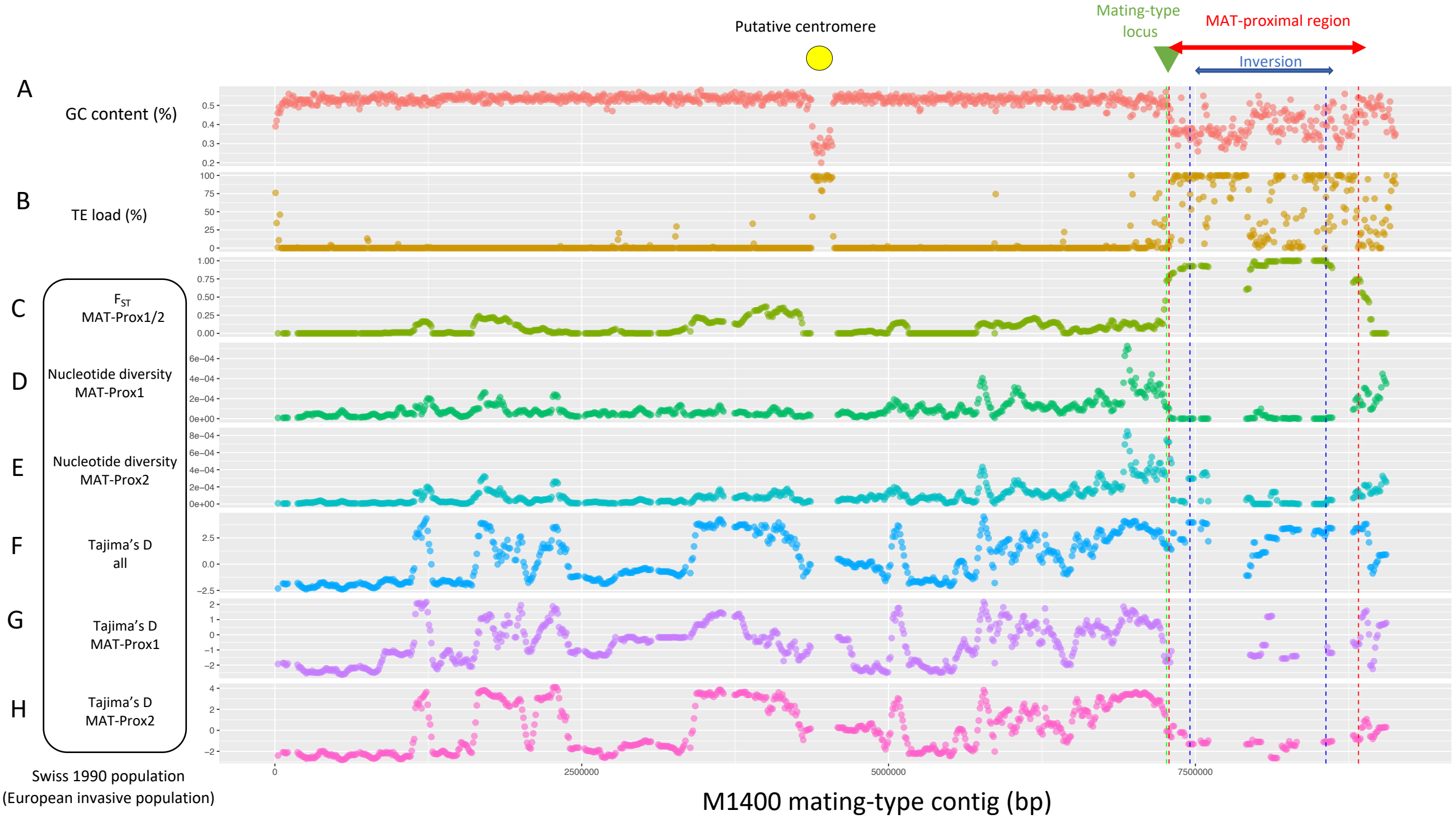
MAT-proximal region

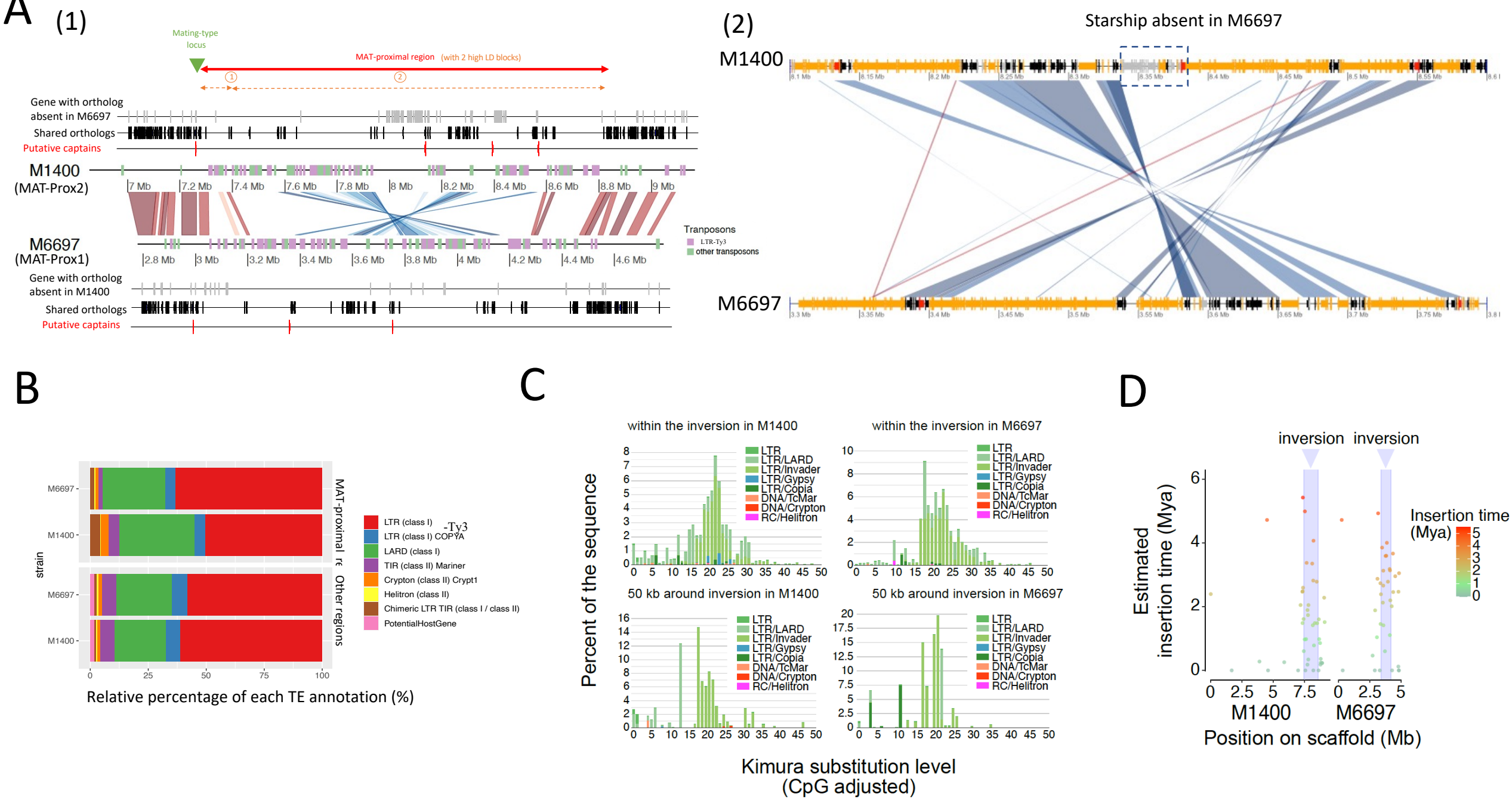
E



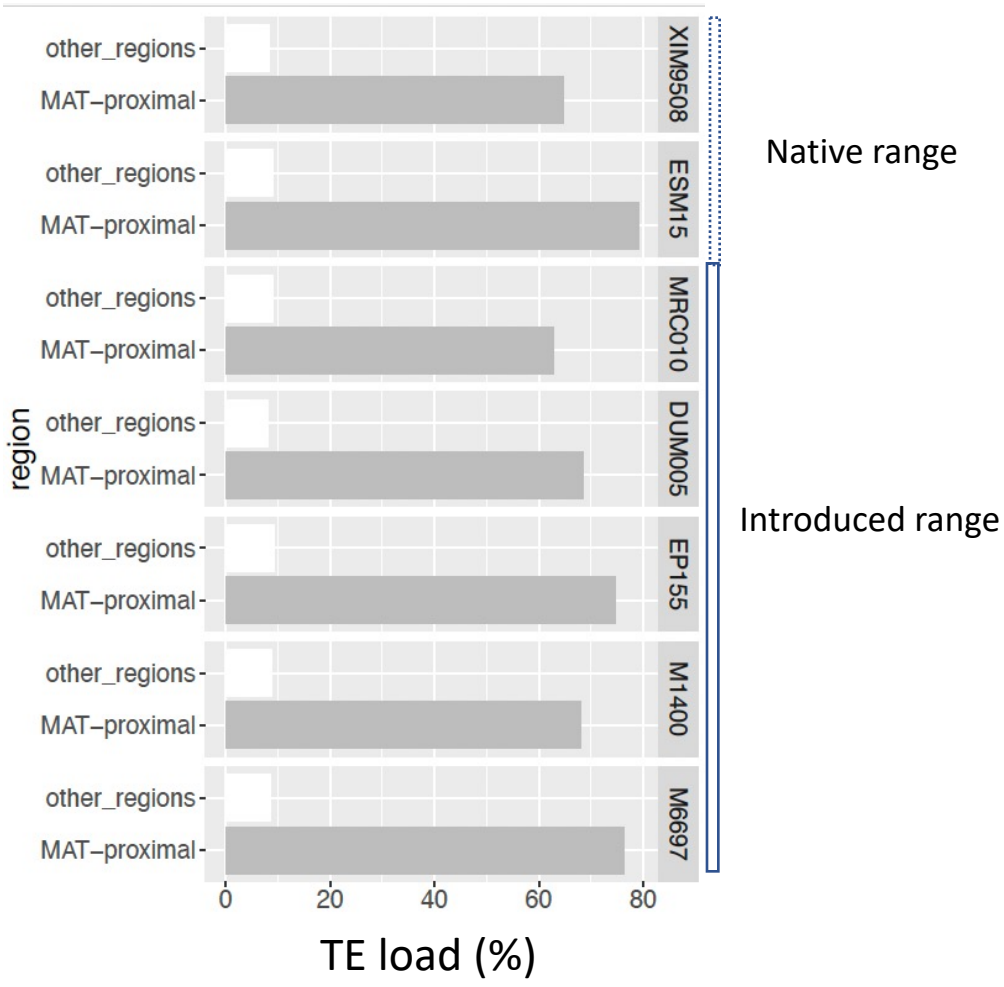
Other contigs (EP155 genome)


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

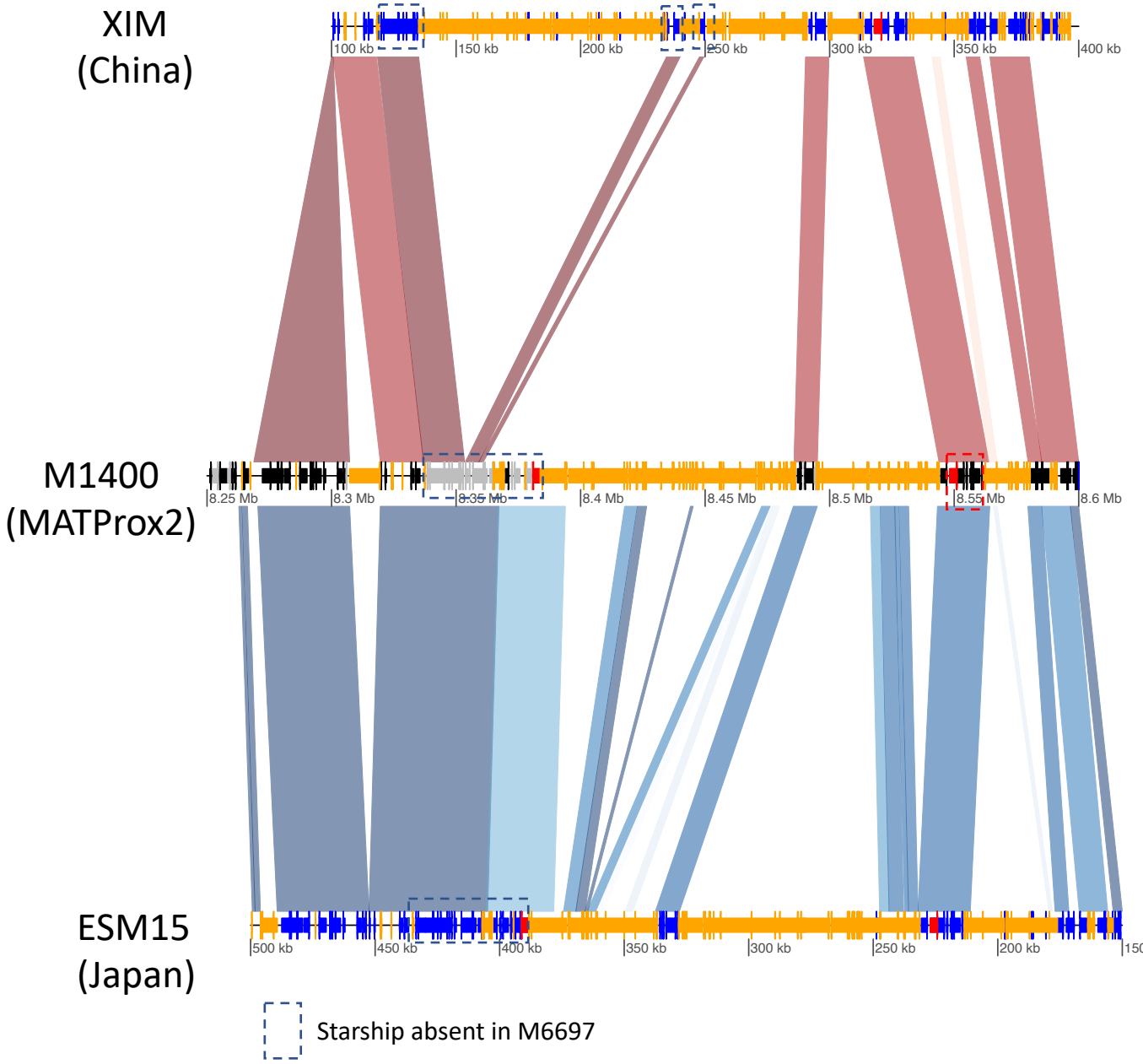

Putative centromere
Mating-type locus
MAT-proximal region
Inversion
A
GC content (%)
B
TE load (%)
C
F_{ST} MAT-Prox1/2
D
Nucleotide diversity MAT-Prox1
E
Nucleotide diversity MAT-Prox2
F
Tajima's D all
G
Tajima's D MAT-Prox1
H
Tajima's D MAT-Prox2
0
2500000
5000000
7500000
Swiss 1990 population
(European invasive population)
M1400 mating-type contig (bp)


A
(1)
Mating-type locus
MAT-proximal region (with 2 high LD blocks)
Gene with ortholog absent in M6697
Shared orthologs
Putative captains
M1400 (MAT-Prox2)
M6697 (MAT-Prox1)
Tranposons
LTR-Ty3
other transposons
Gene with ortholog absent in M1400
(2)
Starship absent in M6697
M1400
M6697
B
strain
MAT-proximal re
Other regions
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
Relative percentage of each TE annotation (%)
C
within the inversion in M1400
within the inversion in M6697
50 kb around inversion in M1400
50 kb around inversion in M6697
LTR
LTR/LARD
LTR/Invader
LTR/Gypsy
LTR/Copia
DNA/TcMar
DNA/Crypton
RC/Helitron
Percent of the sequence
Kimura substitution level (CpG adjusted)
D
inversion
inversion
Estimated insertion time (Mya)
Insertion time (Mya)
M1400
M6697
Position on scaffold (Mb)

A

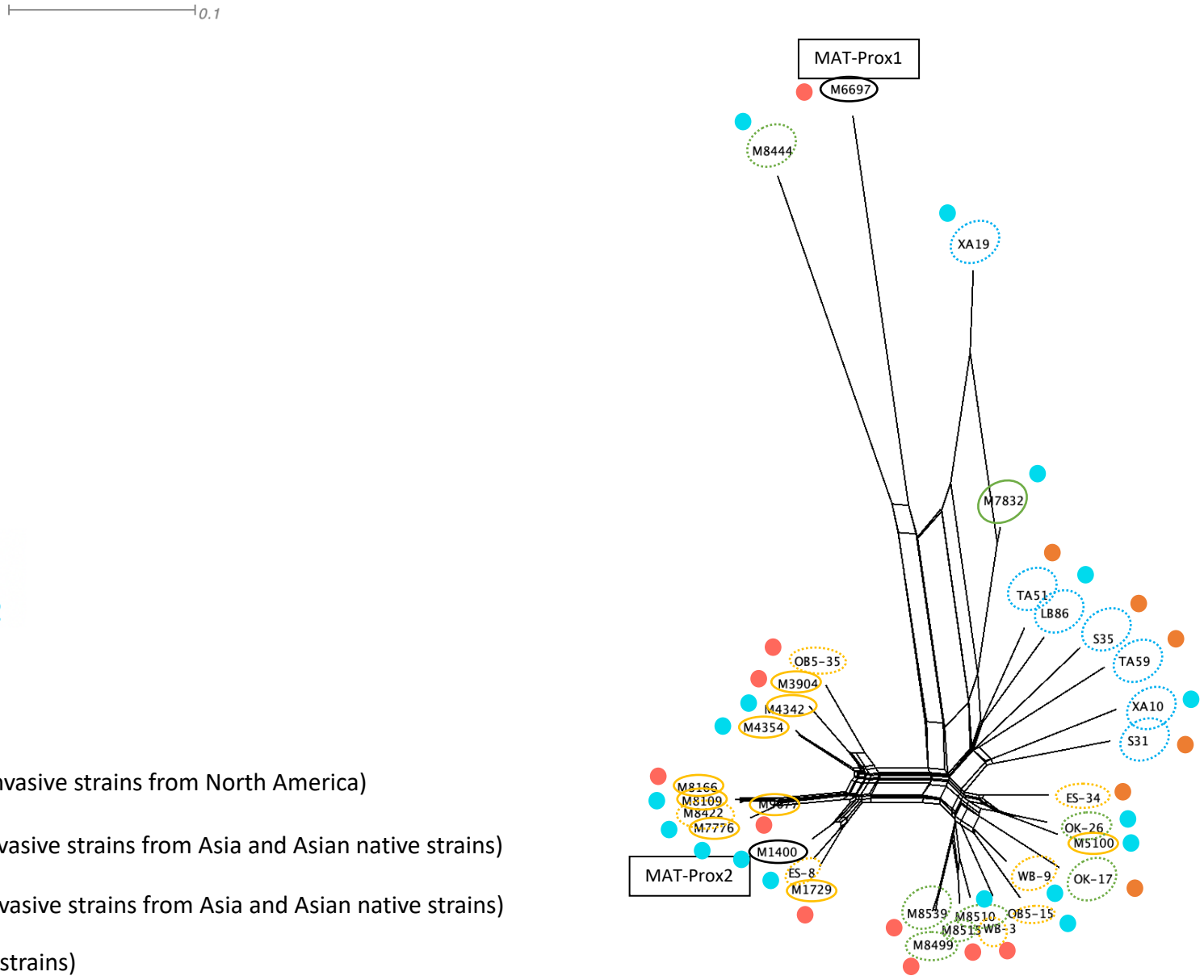

other_regions
MAT–proximal
XIM9508
ESM15
MRC010
DUM005
EP155
M1400
M6697
region
0
20
40
60
80
TE load (%)
Native range
Introduced range

B


XIM
(China)
M1400
(MATProx2)
ESM15
(Japan)
100 kb
150 kb
200 kb
250 kb
300 kb
350 kb
400 kb
8.25 Mb
8.3 Mb
8.35 Mb
8.4 Mb
8.45 Mb
8.5 Mb
8.55 Mb
8.6 Mb
500 kb
450 kb
400 kb
350 kb
300 kb
250 kb
200 kb
Starship absent in M6697

(1)

MAT-proximal region



Mating-type allele :

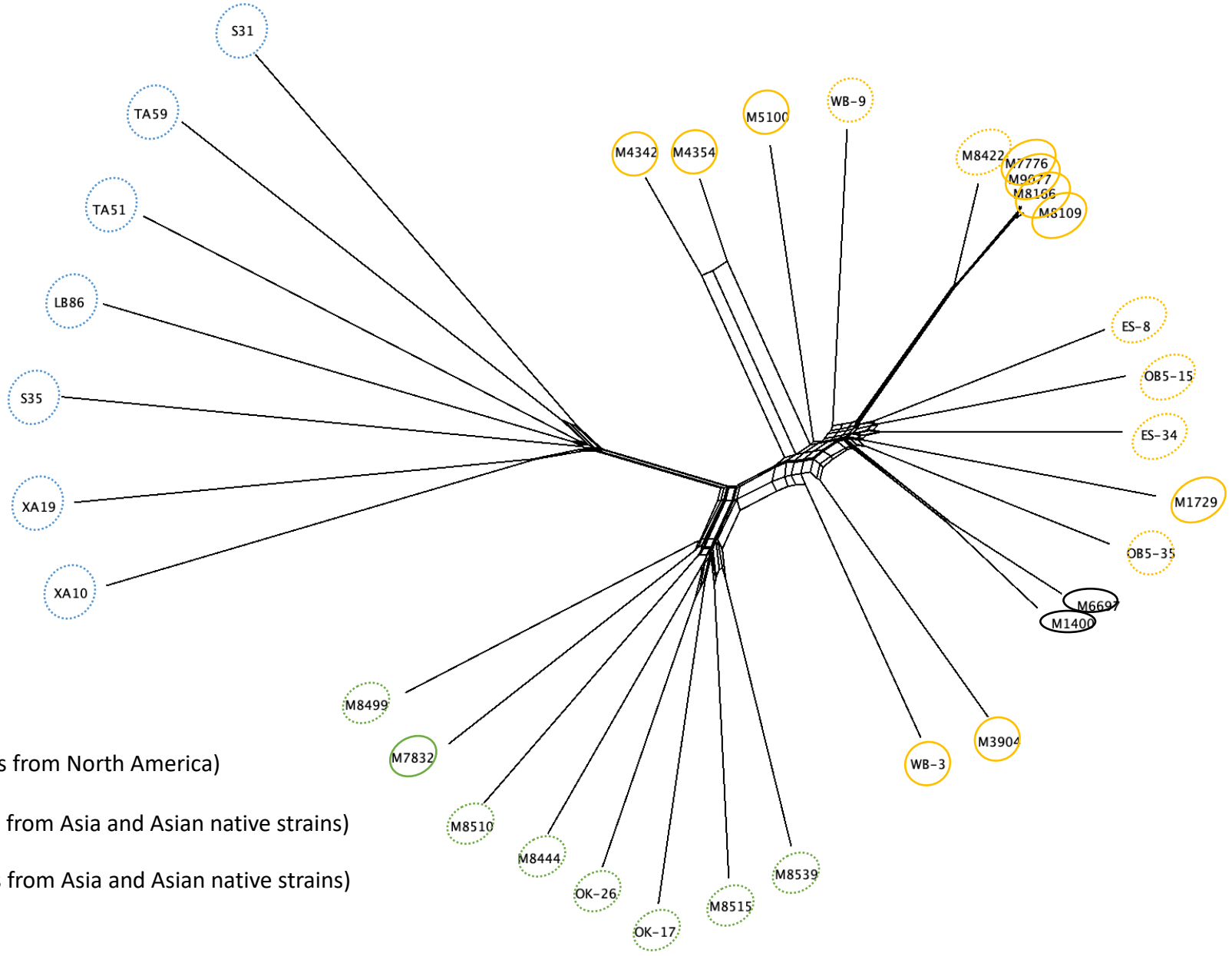
- MAT−1
- MAT−2

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



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)