Molecular Biology and Evolution, 2025, **42**, 1–21
https://doi.org/10.1093/molbev/msaf026
Advance access publication 5 February 2025
Discoveries

MBE OXFORD

An Inversion Polymorphism Under Balancing Selection, Involving Giant Mobile Elements, in an Invasive Fungal Pathogen

Fanny E. Hartmann [1,*] **Ricardo C. Rodríguez 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** [4] **Daniel Croll** [3] **Simone Prospero** [5] **Cyril Dutech,**[2] **Tatiana Giraud** [1]

[1]Ecologie Systematique et Evolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette F-91190, France
[2]Biodiversité Gènes & Communautés, INRAE, Univ. Bordeaux, Cestas F-33610, France
[3]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[4]Institute for Infectious Diseases, University of Bern, Bern, Switzerland
[5]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

***Corresponding author:** E-mail: fanny.hartmann@universite-paris-saclay.fr.
Associate editor: Melissa Wilson

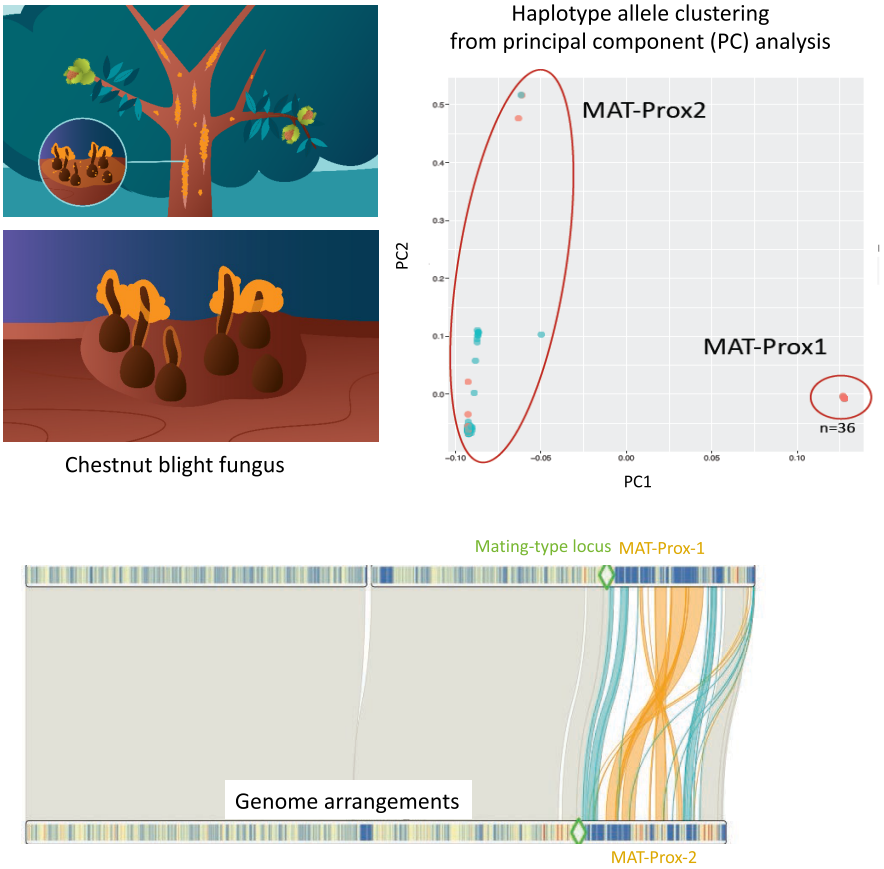
Abstract

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 physically 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 *Starships*. These are large transposable 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 nonsynonymous substitutions, in gene presence/absence polymorphism, in tyrosine recombinases and in transposable elements. This study thus sheds light on a case of a large nonrecombining region partially linked to a mating compatibility locus, likely maintained by balancing selection on differentiated haplotypes, possibly involved in adaptation in a devastating tree pathogen.

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

Received: May 24, 2024. **Revised:** January 23, 2025. **Accepted:** January 23, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Graphical Abstract



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

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, 2016; Bergero and Charlesworth 2009; Umen 2011; 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, nonsynonymous substitutions, decreased gene expression, reduced frequency of optimal codons and gene losses (Bachtrog 2013; Carpentier et al. 2022; Duhamel et al. 2023). The nonrecombining 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 et al. 2024b). 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).

In the invasive chestnut blight fungal pathogen *Cryphonectria parasitica*, mating compatibility is determined by a mating-type locus that displays two alleles, called *MAT-1* and *MAT-2*, that are both permanently heterozygous in the diploid and dikaryotic stages (McGuire et al. 2001). A large region without recombination was suggested to occur near the mating-type locus, although not completely linked to it (i.e. at 3.9 cM of the mating-

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

type locus), based on segregation of RAPD and RFLP markers in three crosses involving five different parents, from the USA, Japan or Italy (Kubisiak and Milgroom 2006). Recent genome-wide association studies in *C. parasitica* found significant association of SNPs with the mating-type locus across a large region (>1 Mb), further suggesting the existence of reduced recombination in this region (Stauber et al. 2021, 2022). This genomic region was enriched in SNPs, transposable elements and copy-number variants, such as deletions. Such enrichment is consistent with sequence degeneration following a lack of recombination (Stauber et al. 2021), and has prevented assembly of the region near the mating-type locus so far.

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

Lineages with an apparently clonal structure have been identified in invasive populations while sexually reproducing populations occur both in the native and invaded ranges (Dutech 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 types 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, 2005; Dutech et al. 2010; Stauber et al. 2021, 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. We further generated six new high-quality long-read based genome assemblies of strains from the native and invaded range of the pathogen. 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 in the genomic region proximal to the mating-type locus (MAT-Proximal), as well as the frequency of the two nonrecombining haplotypes and their association with mating types in populations. We investigated whether other regions of the genome showed reduced recombination rates and whether genomic footprints of degeneration were present in the nonrecombining region, such as genomic rearrangements, nonsynonymous substitutions and transposable element accumulation. We also tested whether this nonrecombining region has been gradually expanding, estimated the age of recombination suppression and investigated the predicted gene functions in the MAT-proximal region. We looked in particular for *Starships*, which 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. 2024) 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 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 specific genomic niches for their insertions, e.g. AT-rich regions (Gluck-Thaler et Vogan 2024). We then extended the analyses to other populations with less extensive genomic resources, i.e. other invasive populations and native Asian populations, investigating the presence of recombination suppression footprints and the occurrence of the two differentiated haplotypes worldwide, as well as the presence of *Starships*. We also investigated the possibility of introgression, using available genomes of closely related species, pathogenic and nonpathogenic (Stauber et al. 2021). Finally, we used available expression data to investigate whether the MAT-Proximal region carried genes upregulated under infection by the virus responsible for hypovirulence or during vegetative incompatibility reactions, a phenomenon avoiding hyphal fusions between individuals, considered to protect against virus transmission.

Results

Footprints of Recombination Suppression in a Large Region (> 1 Mb) Proximal to the Mating-type Locus in Sexually Reproducing Invasive Populations

We first performed population genomic analyses on available genome sequences within local populations sampled in southern Switzerland over two periods of time (early 1990 and 2019), and within the more broadly geographically distributed CL1 cluster in central and southeastern Europe. All these populations have a recombining genetic structure and are invasive European populations introduced from North America (supplementary table S1, Supplementary Material online [Stauber et al. 2021]). We studied only monokaryotic strains, i.e. having either *MAT-1* or *MAT-2* in their genomes, for phasing haplotypes and performed stringent SNPs filtering by masking repeats on the EP155 reference genome (Crouch et al. 2020) and removing missing data and rare variants to ensure robustness of our analyses. We identified 8,900 SNPs segregating among 71 strains of the 1990 Swiss population (35 *MAT-1*; 36 MAT-2; supplementary table S1, Supplementary Material online), 9,646 SNPs segregating among 62 strains of the Swiss 2019 population (20 *MAT-1*; 42 MAT-2;

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

supplementary table S1, Supplementary Material online) and 15,104 SNPs segregating among 88 strains of the CL1 cluster (41 *MAT-1;* 47 *MAT-2*; supplementary table S1, Supplementary Material online). This represents an average density of about 20 SNPs/100 kb, in agreement with previous studies in this invasive fungus (Stauber et al. 2021, 2022). We investigated the genome-wide linkage disequilibrium (LD) landscape and looked for large blocks (>1 Mb) of high LD ($r^2 > 0.9$) among SNPs within contigs, to look for signatures of reduced recombination along the genome. LD indeed decays by half across a few 100 kb on average in *C. parasitica* in recombining regions of the genome (Demené et al. 2019). Therefore, high LD beyond >1 Mb represents strong evidence of recombination cessation and the SNP density is sufficient to study such LD variation.

We found a large block of high LD on the chromosome called scaffold_2 of the EP155 genome in the three populations, corresponding to the mating-type chromosome (see Fig. 1 for the CL1 population, the pattern in the other populations being similar as presented in supplementary fig. S1, Supplementary Material online). The high-LD block was located near the mating-type locus but did not encompass the mating-type locus itself and will be called hereafter the "MAT-proximal region". In the 1990 and 2019 Swiss populations, SNPs between 0.493 Mb and 1.720 Mb in the mating-type chromosome were all in high LD (with $r^2 > 0.9$ between SNPs distant by up to 1 Mb; in the MAT-Proximal region: mean r^2 of all SNP pairs in the Swiss 1990 population = 0.89; mean r^2 of all SNP pairs in the Swiss 2019 population = 0.83; supplementary fig. S1, Supplementary Material online). In the CL1 cluster, the large high-LD block was also present, but was split into two blocks, a large one and a smaller, peripheral one (orange arrows in Fig. 1). We indeed detected values of $r^2 > 0.9$ between SNPs distant by up to 1 Mb, between 0.532 Mb and 1.535 Mb (mean r^2 of all SNP pairs = 0.94). Between 1.535 and 1.717 Mb, SNPs were also in high LD (mean $r^2 = 0.84$). The LD between the two high-LD blocks was lower ($r^2 = 0.59$) than within blocks but was still higher than elsewhere along the genome (orange rectangle at the right border of the red triangle; Fig. 1). A similar pattern also existed in the Swiss populations (orange rectangle at the right border of the red triangle; supplementary fig. S1, Supplementary Material online), although less marked. This is likely due to rare recombination events at one specific locus near the edge of the large fully nonrecombining region, lowering the LD level in populations between the two parts of the otherwise fully nonrecombining region. The mating-type locus was located at 1.737 Mb in the EP155 genome, distant by 17 and 20 kb from the LD block in the Swiss populations and CL1 cluster, respectively. SNP density was on average 48 SNPs/100 kb in the high-LD block region, i.e. higher than in the rest of the chromosome or the other chromosomes (supplementary fig. S2a, Supplementary Material online). When sampling SNPs distant of at least 50 kb in this region to have the same SNP density all over the chromosome, we also detected the high-LD block, which indicates that the SNP density did not generated biases in the LD pattern (supplementary fig. S2b, Supplementary Material online).

We found no other regions of the genomes with $r^2 > 0.9$ between SNPs across such a large genomic distance in any population. On scaffold_6 in the CL1 cluster, two smaller SNP blocks had a $r^2 > 0.9$ between each other despite being distant

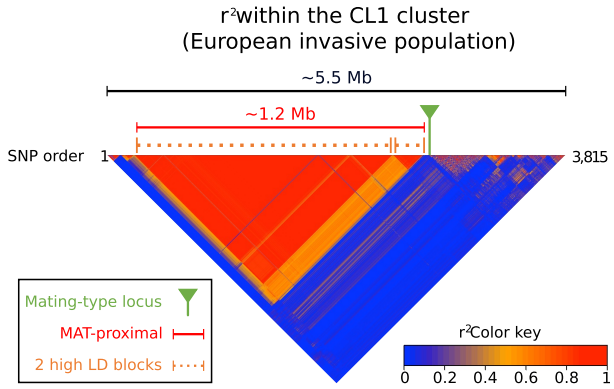


Fig. 1. Linkage disequilibrium (LD) analysis along the contig carrying the mating-type locus in a *Cryphonectria parasitica* European invasive population. LD heatmap using single nucleotide polymorphisms (SNPs; $n = 3{,}815$) located on the contig carrying the mating-type locus (scaffold_2 of the EP155 genome) in the CL1 genetic cluster (European invasive population introduced from North America); pairs of SNPs with high LD, i.e. $r^2 > 0.9$, correspond to the red triangle. The mating-type locus location is shown with a pin and the MAT-proximal region lacking recombination is shown with a plain bar. The two high-LD blocks within the MAT-proximal region are shown with dotted bars. The heatmap is plotted along SNP positions (1 to 3,815 SNPs). The size of the genomic region corresponding to the MAT-proximal region with high LD is indicated at the top in Megabase (Mb), as well as the total length of the contig. The MAT-Proximal region is enriched in SNPs, and therefore appears larger in the heatmap relative to the contig than it is based on physical size.

(supplementary fig. S2c, Supplementary Material online; between SNPs blocks at 3 to 20 kb and 2.151 to 2.278 Mb), which is likely due to the major intra-scaffold translocation described in this region between the EP155 and ESM15 strains (Demené et al. 2022). The maximum size of the blocks with $r^2 > 0.9$ on other scaffolds ranged from 135 to 540 kb in the CL1 cluster, as previously described (Demené et al. 2019). The average distance between SNPs in high LD ($r^2 > 0.9$) across the genome was higher on the scaffold_2 than on other scaffolds (in the CL1 cluster, pairwise Wilcoxon test *P*-value < 2e-16 with Bonferroni correction). Furthermore, the proportion of SNPs in high LD ($r^2 > 0.9$) across the genome was also much higher on the scaffold_2 than on other scaffolds due to the high LD in the MAT-proximal region (supplementary fig. S2d, Supplementary Material online, supplementary table S2, Supplementary Material online). Therefore, the large (1Mb) and localized block with maximal LD values proximal to the mating-type locus in the Swiss populations and the CL1 genetic cluster stands out as exceptional in the genome and indicates full recombination cessation, which supports previous inferences from progeny segregation analyses (Kubisiak and Milgroom 2006). Indeed, even low rates of recombination homogenize allelic frequencies and prevent LD building (Dufresnes et al. 2015). The MAT-proximal nonrecombining region was actually larger (1.23 Mb), but with likely rare recombination events at one precise locus near its edge, lowering LD between its two fully nonrecombining parts.

Consistent with a lack of recombination, the MAT-proximal region formed two genetically highly differentiated haplotypes in these invasive Swiss and CL1 populations (Fig. 2a; supplementary fig. S3 a(S1)-c(S1), Supplementary Material online), as shown by the two clusters on the principal component analysis (PCA) using only SNPs located in the MAT-proximal region, while no structure was detected in

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

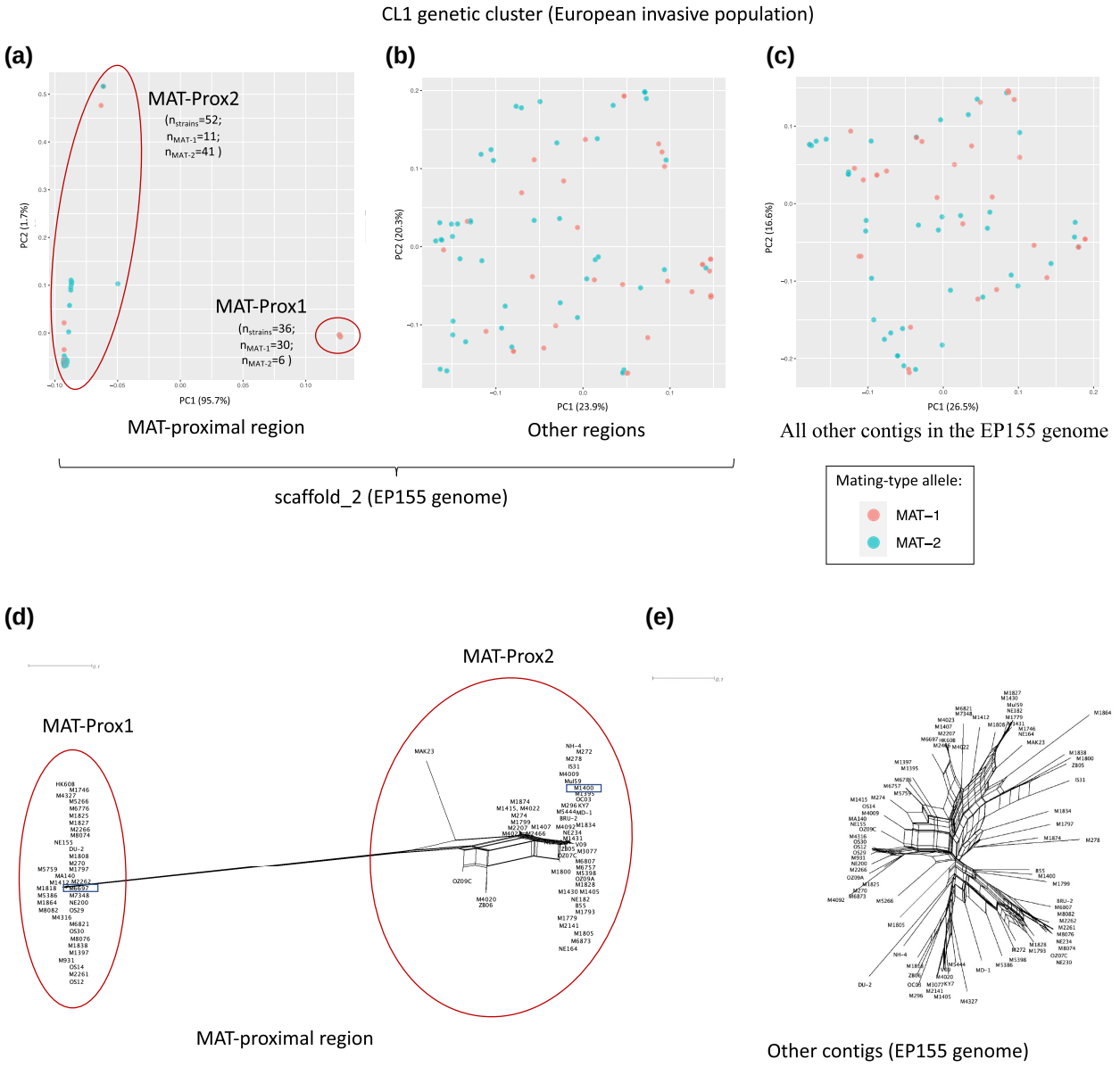


Fig. 2. Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-proximal region lacking recombination and other genomic regions in a *Cryphonectria parasitica* European invasive population. Panels a, b and c) Principal component analysis (PCA). Two principal components are presented. Percentage of variance explained by each PC is indicated into brackets. Strains are colored according to their mating-type (*MAT-1* or *MAT-2*). Panels d and e) Neighbor-net network from a SplitsTree analysis. These analyses were performed in the CL1 genetic cluster (European invasive populations introduced from North America) based on: a and d) SNPs (n = 2,220) located within the MAT-proximal region along the contig carrying the mating-type locus (scaffold_2 of the EP155 genome); b) SNPs (n = 1,595) located in other regions along the contig carrying the mating-type locus; c and e) SNPs (n = 11,289) located on other contigs of the EP155 genome. On panels a) and d), the two clusters corresponding to the MAT-Prox1 and MAT-Prox2 haplotypes are shown with empty circles. The identified haplotype of each strain is indicated in supplementary table S1, Supplementary Material online. The number of strains within the MAT-Prox1 haplotype is indicated on panel a) by the letter n. The newly sequenced M1400 and M6697 strains are highlighted with a empty rectangle.

the rest of the genome (Fig. 2b-c; supplementary fig. S3a(S2)-c(S2), Supplementary Material online). The neighbor-net networks further supported the existence of two differentiated haplotypes in the MAT-Proximal region, contrasting with an otherwise recombining structure genome-wide (Fig. 2d-e; supplementary fig. S3b-d, Supplementary Material online). In the Swiss 2019 population, a few reticulations between haplotypes were found and three strains (LU3, Nov10, Nov4, all three being *MAT-1*) appeared to have an intermediate sequence in the MAT-Proximal region between the two haplotypes (supplementary fig. S3d, Supplementary Material online). This intermediate haplotype was also found in a few other invasive strains introduced directly from Asia, of either *MAT-2* or *MAT-1* mating types (see below).

The two genetic clusters in the MAT-proximal region appeared statistically associated with mating types in the CL1 and Swiss 1990 recombining populations (Figs. 2a and d; supplementary fig. S3, Supplementary Material online, Table 1). Indeed, the distribution of mating types between the two PCA clusters strongly deviated from expectations under random association (Table 1; chi-squared test: χ^2 = 33.3; P-value = 5.876e-08 in CL1; χ^2 = 18.76; P-value = 8.441e-05 in Swiss 1990). We thus hereafter call the two differentiated haplotypes MAT-Prox1 and MAT-Prox2,

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

Table 1 Distribution of mating types (*MAT-1* and *MAT-2*) among nonrecombining haplotypes of the MAT-proximal region in the European invasive *Cryphonectria parasitica* CL1 genetic cluster and the Swiss populations

	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

MAT-Prox1 and MAT-Prox2 haplotypes were defined based on the clusters of the principal component analysis.

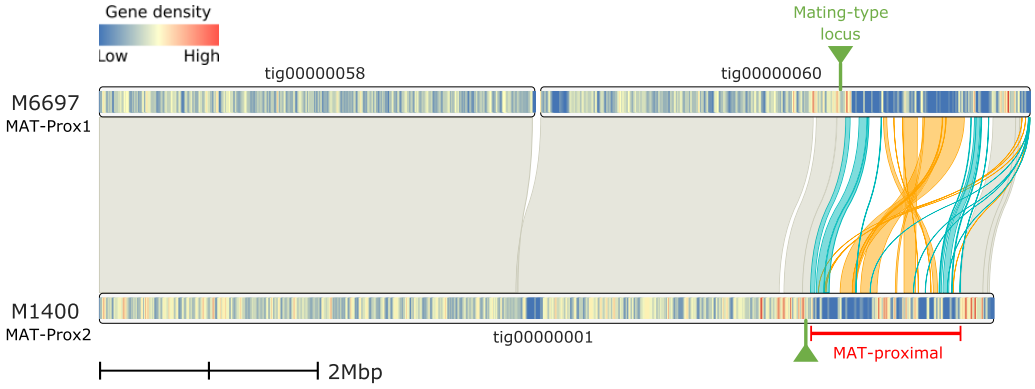


Fig. 3. Synteny and rearrangements between the two newly sequenced genomes of European invasive strains introduced from North America (M1400 and M6697) in the chromosome carrying the MAT-proximal region lacking recombination in *Cryphonectria parasitica*. Blue links show collinear regions and orange links show inverted regions in the MAT-proximal region. Grey links show other regions. The mating-type locus is indicated by a pin. The MAT-proximal region defined from LD analyses is indicated by a bar. The gene density track is shown with a color gradient (blue with low density, orange with high density).

referring to the mating-type association. However, the association between mating types and MAT-Proximal haplotypes was not complete, with for example only 83% of *MAT-1* strains carrying the MAT-Prox1 haplotype in CL1. The two MAT-Proximal haplotypes were actually not significantly associated with mating types in the Swiss 2019 recombining populations (Table 1; chi-squared test: $\chi^2 = 3.8427$; *P*-value = 0.1464). The EP155 genome used as reference for SNP calling above carried the MAT-Prox1 haplotype but the *MAT-2* allele.

An Inversion in the nonrecombining MAT-proximal Region

To investigate whether rearrangements between haplotypes were present in the MAT-proximal region, we sequenced the genome, using PacBio HiFi, of two strains with alternative MAT-proximal haplotypes: M1400 (*MAT-2*; MAT-Prox2) and M6697 (*MAT-1*; MAT-Prox1), originating from the Swiss population in Gnosca and belonging to the CL1 cluster, i.e. European invasive populations introduced from North America (Stauber et al. 2021, 2022). We built high-quality genome assemblies: statistics of the assemblies were in the same range as those of the genome assemblies of the ESM15 and EP155 strains previously sequenced (supplementary table S3, Supplementary Material online; Demené et al 2022; Crouch et al 2020). The mating-type chromosome could be assembled as a single contig for the first time, in the M1400 genome, likely corresponding to a full chromosome as we identified telomeric repeats in both ends. The mating-type chromosome was assembled into two contigs in the M6697 genome with telomeric repeats in one single end of each contig. We therefore used the M1400 genome as a reference for subsequent analyses. To identify the location of the MAT-proximal region in the M1400 genome, we computed LD by mapping the reads of the 1990 Swiss population to the new M1400 genome. We again found the same large block (> 1 Mb) of high LD ($r^2 > 0.9$) between 7,285,137 and 8,828,934 bp (red arrow; supplementary fig. S4, Supplementary Material online), showing that calling SNPs on either a MAT-Prox1 or MAT-Prox2 haplotype yielded similar LD patterns. The MAT-proximal region was also located 20 kb away from the mating-type locus (located at 7.265 Mb) and was 1.540 Mb long using the M1400 genome as reference. Consistent with the results using the EP155 genome as reference, we found that the high-LD region was divided into two higher-LD blocks near the edge at 7.392 Mb (orange arrows in supplementary fig. S4, Supplementary Material online).

The two new high-quality genome assemblies from the invasive Swiss population (introduced from North America) revealed an inversion between the two haplotypes in the MAT-proximal region (Fig. 3). The two newly sequenced genomes (PacBio's HiFi) were indeed collinear except for the mating-type chromosome (supplementary fig. S5a-b-c, Supplementary Material online), where we found a large region (>1 Mb) that was inverted between the M1400 and M6697 genomes (Fig. 3; supplementary fig. S5a-b-c, Supplementary

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025



Fig. 4. Genetic diversity and divergence between nonrecombining haplotypes of the MAT-proximal region along the contig carrying the mating-type locus in a *Cryphonectria parasitica* European invasive population. A-B. GC content a) and transposable element densities b) along the contig carrying the mating-type locus (tig_0000001) of the M1400 *C. parasitica* genome. c) Relative divergence (F_{ST}) between strains of the MAT-Prox1 and MAT-Prox2 haplotypes: d-e) Nucleotide diversity within pools of strains for each MAT-Proximal haplotype; F-G-H Tajima's D for all strains pooled and within pools of strains for each MAT-Proximal haplotype; The MAT-proximal region defined from LD analyses and the inversion between M1400 and M6697 genomes are indicated by bars and vertical dotted lines, respectively. The mating-type locus location is shown with a pin and a vertical dotted line. The location of the putative centromere is indicated by a circle. All population statistics were computed for the 1990 Swiss population along the mating-type contig (tig_0000001) of the M1400 genome per 50-kb window overlapping over 10 kbp. Windows containing fewer than 5 SNPs were removed from the analysis.

Material online). Breakpoints of the inversion were located at ca. 7.455 and 8.563 Mb of the tig00000001 contig in the M1400 genome and at ca. 3.380 and 4.185 Mb of the tig00000060 contig in the M16697 genome. The region affected by the inversion was ca. 300 kb smaller in the M6697 genome than in the M1400 genome, while the whole MAT-proximal region was ca. 100 kb smaller. The split into the two high-LD blocks in natural populations reported above was outside of the inversion, i.e. 70 kb from the inversion breakpoint (orange arrows in supplementary fig. S4, Supplementary Material online). LD in natural populations was higher within the inversion (median value of $r^2 = 1$) than in other regions of the MAT-proximal region (median value of $r^2 = 0.64$; Wilcoxon rank sum test with continuity correction; W = 4.8714e + 10; *P*-value < 2e-16).

We looked for centromeres, as the nonrecombining regions near the mating-type locus in other ascomycetes, when they occur, either capture the centromere (Menkis et al. 2008; Sun et al. 2017), or are associated with the occurrence of a single crossing-over between the centromere and the mating-type locus (Grognet et al. 2014; Hartmann et al. 2021; Vittorelli et al. 2023). The centromere in *C. parasitica* may be at 4.380 to 4.536 Mb on the mating-type chromosome in the M1400 genome, as we detected here a peak in TE density and a drop in GC content (Fig. 4a-b). A dotplot of repeats in this region also presented a pattern typical of centromeres (supplementary fig. S5d, Supplementary Material online). The MAT-proximal region and the inversion thus did not include the putative centromere and was instead located about 2.9 Mb away from the putative centromere.

Higher Genetic Differentiation and Lower Genetic Diversity Within the MAT-proximal Region Than in Recombining Regions

When analyzing polymorphism within the invasive 1990 Swiss population using the M1400 genome as reference, we found much higher genetic differentiation between strains of the two MAT-proximal haplotypes in this region than elsewhere along the genome (Fig. 4c; supplementary fig. S6a, Supplementary Material online; supplementary table S4, Supplementary Material online), as expected for a nonrecombining region in LD with the mating-type locus. For example, the F_{ST} values between the two nonrecombining haplotypes within the MAT-proximal region in the 1990 Swiss population were nearly maximal (median $F_{ST} = 0.93$ per 50 kb window), while they were near zero in the rest of the genome (median $F_{ST} = 0.02$; Wilcoxon test *P*-value < 2e-16 with Bonferroni correction; Fig. 4c; supplementary fig. S6a, Supplementary Material online). Such F_{ST} values close to 1 indicate a lack of shared polymorphism and therefore support the inference of recombination suppression in the MAT-proximal region.

Within each haplotype, the genetic diversity at the MAT-proximal region was lower than in the rest of the genome (Wilcoxon test *P*-value < 2e-16 for the MAT-Prox1 haplotype; *P*-value = 2.3e-07 for the MAT-Prox2 haplotype; Fig. 4e-f; supplementary fig. S6c, Supplementary Material online; supplementary table S4, Supplementary Material online). Such low diversity is expected for a region without recombination associated with the mating-type locus, as its effective population size is half as in recombining regions. The diversity was especially low in the pool of isolates with the MAT-Prox1 haplotype.

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

For example, all MAT-Prox1 sequences clustered on a single point on the PCA in Fig. 2a and in the neighbor-net network in Fig. 2d, in contrast to the more scattered MAT-Prox2 sequences. Such a very low diversity may be due to a recent selective sweep or to a particularly strong bottleneck in this haplotype during the invasion.

We detected signals of long-term balancing selection in the MAT-proximal haplotypes. Computation of the Tajima's D statistics within all strains suggested signatures of balancing selection in the MAT-proximal region. Tajima's D values in the MAT-proximal region pooling all sequences (median D = 3.0 per 50 kb window) were higher than in the rest of the genome (median D = -0.14; Wilcoxon test *P*-value < 2e-16 with Bonferroni correction; Fig. 4g; supplementary fig. S6d, Supplementary Material online). In contrast, Tajima's D was significantly lower than the rest of the genome within each pool of haplotype, including in the MAT-Prox2 haplotype, suggesting positive selection (Wilcoxon test *P*-value = 0.0037 for the MAT-Prox1 haplotype; *P*-value = 0.0022 for the MAT-Prox2 haplotype with Bonferroni correction; Fig. 4h-i; supplementary fig. S6d, Supplementary Material online; supplementary table S4, Supplementary Material online). We found no difference in population diversity statistics between the recombining part of the mating-type contig and the other contigs (supplementary table S4a, Supplementary Material online).

The study of synonymous divergence (d_S) between the shared orthologs in the M1400 and M6697 haplotypes suggests that their differentiation is at least 1.5 million years old and that there is no pattern of evolutionary strata within the MAT-proximal region. Indeed, we found no segments with different levels of differentiation between haplotypes that would indicate stepwise expansion of recombination suppression away from the mating-type locus. Per-gene synonymous divergence (d_S) is typically used for detecting gradual expansion of recombination cessation and for estimating its age, as it is considered a good proxy for the time since recombination suppression. Indeed, when recombination is suppressed, mutations accumulate independently in the two nonrecombining haplotypes. We plotted the per-gene synonymous divergence (d_S) between M1400 and M6697 along the M1400 genome, as its mating-type contig is likely being assembled as a full chromosome (supplementary fig. S7a&e, Supplementary Material online; supplementary table S4, Supplementary Material online). Consistent with recombination suppression, we found significantly higher d_S values in the MAT-proximal region (computed for 71 genes) than in the other regions of the mating-type contig (pairwise Wilcoxon test *P*-value <2e-16 with Bonferroni correction) and other contigs (pairwise Wilcoxon test *P*-value <2e-16 with Bonferroni correction). We found no significant differences between the other regions of the mating-type contig and other contigs (pairwise Wilcoxon test *P*-value = 0.23 with Bonferroni correction). The d_S pattern displayed no indication of gradual expansion of recombination cessation within the MAT-proximal region, as there was no stair-like pattern. Using synonymous substitution rate estimates across multiple ascomycetes (from 0.9×10^{-9} to 16.7×10^{-9} substitutions site per year (Kasuga et al. 2002; Taylor and Berbee 2006) and the mean d_S value (0.0495) across the genes shared between the MAT-proximal haplotypes, we estimated the age of their divergence to be at least 1.5 Million years (supplementary table S4, Supplementary Material online).

Enrichment in Transposable Elements in the Nonrecombining Region and Inversion Breakpoints

When studying the M1400 and M6697 high-quality assemblies, we detected an enrichment in transposable elements (TEs) in the MAT-proximal region in the two haplotypes compared to the rest of the genome (Fig. 4b). The percentage of bp occupied by TEs (TE load) was higher than 50% in the MAT-proximal region, while it was only 9% on average in other regions. The TE load in the MAT-proximal region was higher in the M6697 haplotype (76%, *MAT-1*; MAT-Prox1) than the M1400 haplotype (68% *MAT-2*; MAT-Prox2). Class I retrotransposons with Gypsy (LTR-Ty3) and LARD elements were the most abundant TEs in all genomic regions (autosomes, MAT-Proximal region and the rest of the mating-type contig), representing >70% of TE annotations (Fig. 5a-b), as shown previously (Demené et al. 2022). The TE load however varied significantly among TE families and genomic regions (ANOVA; supplementary table S5, Supplementary Material online). In M6697 for example, TIR elements appeared less abundant in the nonrecombining than recombining regions while LTR-Ty3 elements were more frequent in the MAT-proximal region (*Cestas*). In M1400, LARD elements were more frequent in the MAT-proximal region than in recombining regions (Fig. 5b).

The presence of TEs at the two inversion breakpoints in the two genomes (Fig. 5a) suggests that the inversion may have occurred via nonhomologous recombination mediated by these elements. By dating retrotransposon insertions using between-copy divergence and within-copy LTR divergence in both M1400 and M6697 genomes (Fig. 5c and d), we found that TEs within the inversion and around the inversion breakpoints were older than TEs farther from the breakpoints. This is expected in nonrecombining regions as selection is less efficient to purge TE insertions in such regions so that they remain there longer (Duhamel et al. 2023). The date estimates of LTR element insertions (Fig. 5d) further indicated that the TE accumulation in the MAT-proximal region was as old as a few million years, confirming the age estimated based on sequence divergence between MAT-proximal haplotypes.

The MAT-proximal Region Harbours *Starship* Elements

In addition to a high density of TEs, gene annotations of the MAT-proximal region in the M1400 and M6697 genome assemblies revealed multiple genes encoding tyrosine recombinases (genes with DUF3435 domains, that can be *Starship* captains), and gene content variation (supplementary tables S6 to S7, Supplementary Material online; Fig. 5a), suggesting the presence of *Starships*. By running the Gene Finder Module of the starfish pipeline designed to identify *Starships* (Gluck-Thaler and Vogan 2024), we in fact identified the presence of four putative *Starship* captains in the M1400 haplotype (MAT-Prox2) and three putative *Starship* captains in the M6697 haplotype (MAT-Prox1) (supplementary tables S6 to S7, Supplementary Material online; Fig. 5a[1]). We detected eight and ten other putative *Starship* captains elsewhere in the M1400 and M6697 genomes, respectively, but the MAT-proximal region appeared enriched in these elements (Fisher's exact test *P*-value =0.0004). A phylogenetic analysis with previously described *Starship* captains (Gluck-Thaler and Vogan 2024) suggested that the captains in the MAT-Proximal region belonged to the Phoenix *Starship* family, that preferentially target AT-rich sites (supplementary fig. S7, Supplementary Material online ; (Gluck-Thaler and Vogan

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

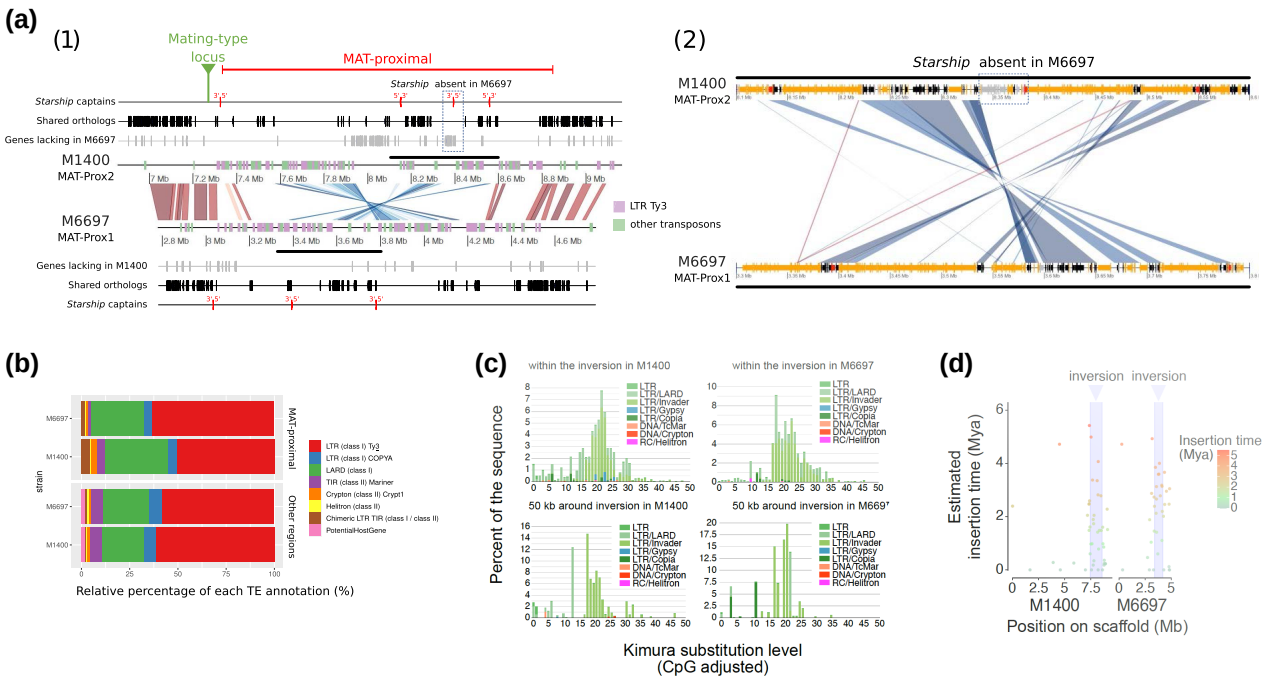


Fig. 5. Annotation of Transposable Elements (TEs) and estimates of their insertion time in the MAT-proximal region lacking recombination and other genomic regions in *Cryphonectria parasitica*. a) Annotation of TEs and gene density along the MAT-proximal region in M1400 and M6697 genomes. On the left panel (1) the figure shows the genomic region from 7 Mb on tig_0000001 of the M1400 genome and from 2.8 Mb on tig_0000060 of the M6697 genome. The mating-type locus is indicated by a pin. The MAT-proximal region is indicated by a bar. Genes with DUF3435 domains (putative *Starship* captains) are shown in red. Orthologous genes shared between genomes are shown in black and those unique to each genome in grey. Synteny for 10 kb segments with identity > 90% is shown in red and inversion in blue. Transposons larger than 5 kb are shown in purple for the transposons annotated as LTR-Ty3 and in green for other transposons. The region with the putative *Starship* absent in M6697 is highlighted by a dotted rectangle and the region zoomed on the panel (2) is highlighted by thick horizontal bars. On the right panel (2) the figure shows a zoom on the *Starship* present in M1400 and absent in M6697 (i.e. the genomic region from 8.1 Mb to 8.6 Mb on tig_0000001 of the M1400 genome and from 3.3 to 3.8 Mb on tig_0000060 of the M6697 genome). In this panel, transposons larger than 150 pb are shown in orange; orthologous genes shared between genomes are shown in black, the genes unique to one genome are in grey. The synteny for segments with identity > 90% is shown in red and inversions in blue. b) Relative percentage of each TE annotation (in %) in the nonrecombining MAT-proximal region and other recombining regions of M6697 and M1400 genomes. The LTR-Ty3 category on panels A and B includes LTR/Invader and LTR/Gypsy categories of panel C. c) Pairwise genetic distance between TE copies within the inversion and around (50 kb) in both M1400 and M6697 genomes. Distribution of the Kimura substitution levels computed using the consensus sequence for the TEs. d) Estimates of the age of intact copies of LTR within and around the inversion.

2024)). As the starfish pipeline performs poorly in TE-rich regions (Gluck-Thaler and Vogan 2024), we investigated manually the delimitation of putative *Starships*. Three putative captains shared orthology relationships between the two haplotypes, one being just at the border of the MAT-proximal region near the mating-type locus (Fig. 3). The high similarity between captains shared by the two haplotypes (>90% identity), their similar positions and reverse strand orientations suggested that these captains were present before the differentiation between haplotypes and before the inversion. The genes near these three shared captains are mostly common to the two haplotypes (supplementary tables S6 to S7, Supplementary Material online; Fig. 5a(1)). The fact that these elements were present in all high-quality assemblies at the same locus prevented delimiting *Starship* boundaries and these genes may represent solo *Starship* captains or pseudogenes as found in Pezizomycotina genomes (Gluck-Thaler and Vogan 2024). The M1400 haplotype (MAT-Prox2) carried an additional putative captain, followed in 3′ by 19 genes lacking in the M6697 haplotype (MAT-Prox1) (Fig. 5a(2)), suggesting the presence of a *Starship* element, together with its cargo genes, only in the MAT-Prox2 haplotype. Functional annotation of the putative cargo genes included two genes with DUF3723 domain and one gene with ferredoxin reductase-type domain, that are often found in *Starships* (supplementary table S7, Supplementary Material online (Gluck-Thaler and Vogan 2024)). This putative *Starship* was about 44 kb long in the M1400 strain.

In addition to the genes only present in the putative *Starship* specific to the MAT-Prox2 haplotype, the two haplotypes carried additional specific genes, especially again in the MAT-Prox2 haplotype. Based on automatic gene annotation and reconstruction of orthologous relationships, we found that, out of 175 single-copy orthologous groups in the MAT-proximal region, 80 (46%) were only present in M1400 (MAT-Prox2) and 24 (14%) were only present in M6697 (MAT-Prox1). This level or presence/absence gene polymorphism was higher than in other genomic regions, in which only 3% of genes were only present in one of the two genomes (Fisher's exact test *P*-value < 2.2e-16). However, not all genes specific to one haplotype were located close to the 3′ end of a putative captain (Fig. 3), so that part of the observed gene presence/absence polymorphism between the two MAT-Proximal haplotypes may also be due to gene losses due to degeneration or pseudogenisation after *Starship* insertions, to recombination suppression or to gene gains by other means than *Starships*. Among the genes specific to M6697 (MAT-Prox1), we found a cluster of 14 genes including a gene encoding a protein of the Sirtuin family (IPR003000). A Sirtuin homolog was found in a *Starship* associated with local thermal climate adaptation

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

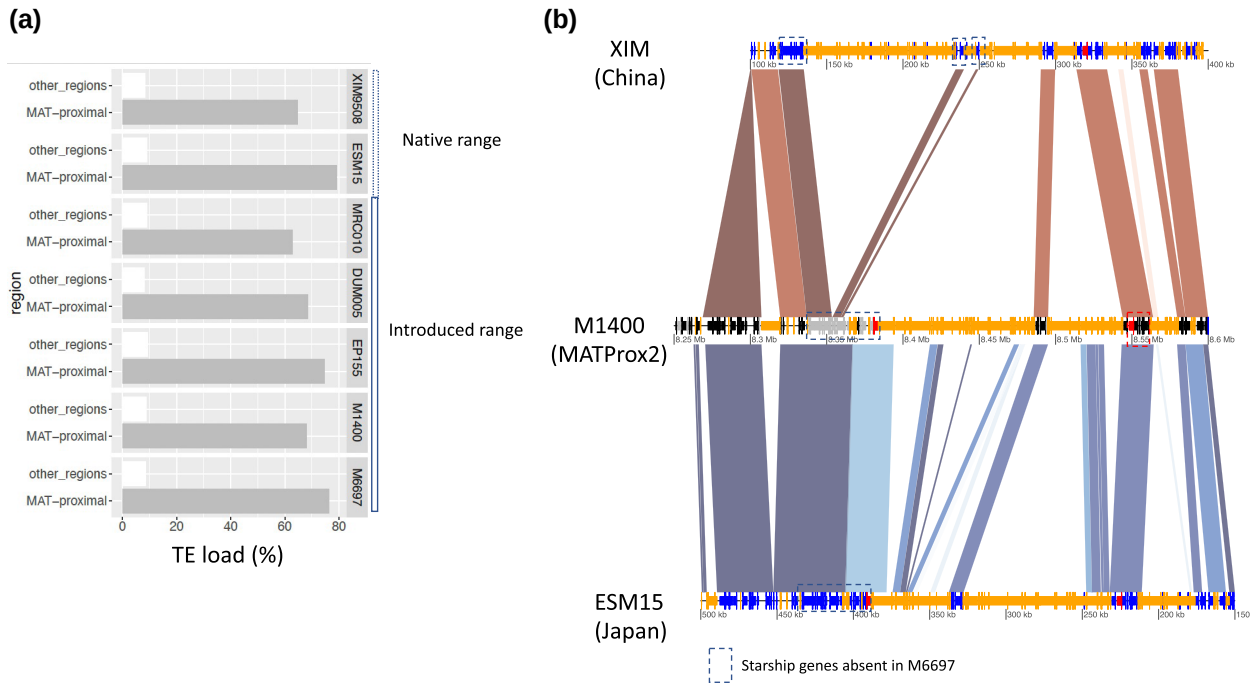


Fig. 6. Transposable element (TE) content and ***Starship*** content in high-quality genome assemblies from other *Cryphonectria parasitica* populations from the native and the introduced range. a) TE load (percentage of base pairs occupied by TEs) in the seven high-quality genome assemblies in the MAT-proximal region and other recombining regions. b) Annotation of TEs and gene density along the MAT-proximal region in M1400 (MAT-Prox2) and the two Asian strains ESM15 (Japan) and XIM9508 (China). For all three strains, genes with a DUF3435 domain (indicating putative *Starship* captains) are shown in red. Transposons larger than 150 pb are in orange. The synteny for segments with identity > 90% is shown in red and inversions in blue. Genes of M1400 are shown in black and grey colors as in Fig. 5, i.e. orthologous genes shared between the M1400 and M6697 genomes are shown in black and those unique to M1400 (i.e. absent in M6697) in grey. The genes of ESM15 and XIM9508 are shown in dark blue.

in the wheat fungal pathogen *Zymoseptoria tritici* (Tralamazza et al. 2024).

We found no gene function known to be involved in mating or mating compatibility, nor in pathogenicity, in the MAT-proximal region. We nevertheless identified genes potentially coding for proteins with kinase domain or proteins with homeodomains (supplementary table S7 to S8, Supplementary Material online), that are often involved in the regulation of developmental pathways.

Gene Expression in the MAT-proximal Region

We analyzed the expression data available, in the EP155 strain (Chun et al. 2020) carrying the MAT-Prox1 haplotype, to test whether some of the genes in the MAT-Proximal region were upregulated under certain conditions. Out of 127 genes functionally annotated in the MAT-proximal region of the EP155 genome, 50 genes were supported by RNAseq data acquired *in vitro* for this strain (Chun et al. 2020), suggesting that the region contains functionally active genes (supplementary table S8, Supplementary Material online). Among these 50 genes, nine were upregulated during barrage allorecognition of the EP155 strain with a compatible strain (Belov et al. 2021) and five were upregulated during infection of the EP155 strain with the hypovirus CHV1 (Chun et al. 2020) (supplementary table S8, Supplementary Material online), suggesting that the MAT-Proximal region may have a role in fungus-mycovirus interactions and vegetative incompatibility. Unfortunately, no expression data are available for a strain with the MAT-Prox2 haplotype, that harbors the additional *Starship* and the highest number of specific genes and no expression data is available during plant infection.

Degeneration in the Nonrecombining Region

The MAT-Proximal region was enriched in TEs and depleted in genes (Fig. 4 and Fig. 5a), which supports the inference of an old full recombination suppression. We also found overall higher nonsynonymous substitution over synonymous substitution rates (d_N/d_S values) between the M1400 and M6697 genomes in the MAT-proximal region (median per gene $d_N/d_S = 0.80$) than in other contigs ($d_N/d_S = 0.43$; pairwise Wilcoxon test *P*-value = 1.7e-07 with Bonferroni correction; supplementary fig. S8a, Supplementary Material online; supplementary table S4, Supplementary Material online). This is consistent with relaxed purifying selection due to recombination suppression in the MAT-proximal region. In addition, 13 genes displayed $d_N/d_S > 1$ between M1400 and M6697, suggesting that they may evolve under positive selection (supplementary table S9, Supplementary Material online).

Recombination Suppression and *Starships* in the MAT-Proximal Region in Other High-quality Genome Assemblies

We analyzed five additional high-quality genome assemblies for strains from the Asian native range (CL2 and CL4 clusters) and the invasion range (including two North American strains belonging to the CL1 cluster and one European strain introduced directly from Asia and belonging to the CL2 cluster), three of which were generated for the present study (supplementary table S3, Supplementary Material online). Most of the high-quality genomes displayed the same MAT-Proximal chromosomal arrangement as M1400 (MAT-Prox2, Swiss population and belonging to the CL1 cluster), suggesting that it is the ancestral state (supplementary fig. S9a-d, Supplementary Material online); the incomplete assemblies may however prevent detecting other

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

rearrangements. The inversion detected in M6697 (MAT-Prox1, Swiss population and belonging to CL1 cluster) was also present in the EP155 strain (MAT-Prox1, North America and belonging to CL1 cluster, Crouch et al 2020; supplementary fig. S8d, Supplementary Material online), indicating that this inversion was already segregating in early invasive populations in North America. In other genomic regions outside of the MAT-Proximal region, all analyzed high-quality genomes were collinear to the ESM15 reference genome, except the highly rearranged EP155 genome, as previously reported (Demené et al. 2022).

The MAT-Proximal region in all the high-quality genomes, from Asian native strains and invasive strains, displayed signs of degeneration, with a high TE load (Fig. 6; supplementary fig. S9, Supplementary Material online), confirming the ancient age of recombination suppression and its occurrence in the native range. Study of d_S and d_N/d_S between the MRC10 strain, belonging to the European clonal lineages of direct Asian origin (CL2 cluster), M6697, M1400 and the two Asian native strains further confirmed recombination suppression in the native range and in invasive strains from different origins (North America or directly from Asia). The d_S and d_N/d_S values were indeed significantly higher in the MAT-proximal region than other genomic regions (supplementary fig. S7 b-e, Supplementary Material online). Mean d_S values in the MAT-proximal region were in the same range as d_S values between the M6697 and M1400 genomes, suggesting a similar date of recombination suppression and, therefore, that recombination suppression was ancestral. The MAT-proximal region appeared smaller in the two Asian strains XIM9508 and ESM15 (<1 Mb) than in invasive strains (supplementary table S3, Supplementary Material online); this smaller size may however be due to the incomplete genome assemblies and/or to gene content variation and additional *Starship* presence/absence polymorphism.

The three captains shared between the M6697 and M1400 haplotypes were present in the MAT-proximal region of all high-quality genome assemblies, indicating that their insertions were old, being already present in the native range. The putative *Starship* specific to M1400 (MAT-Prox2), i.e. the captain and the 19 cargo genes, was also present in the MAT-proximal region of DUM-005 (MAT-Prox2) and of MRC10 and the Japanese strain ESM15 (both genetically close to MAT-Prox2; supplementary fig. S10D1, Supplementary Material online). In contrast, this specific *Starship* was absent in the MAT-Proximal region of the EP155 strain (MAT-Prox1). In the XIM9508 Chinese strain (genetically close to MAT-Prox1; supplementary fig. S10D1, Supplementary Material online), the captain and the four adjacent genes of the *Starship* specific to M1400 were absent in the MAT-proximal region. However, 14 of its putative cargo genes were present in the MAT-proximal region. In the seven high-quality assemblies, the putative *Starship* presence/absence polymorphism detected between the 1400 and M6697 genomes thus seem associated with the MAT-Proximal haplotypes in the invaded range and also in the native Asian range (supplementary fig. S10d, Supplementary Material online), although this would need to be confirmed with additional high-quality assemblies.

Presence of the Two MAT-Proximal Haplotypes in Other Populations, but Less Differentiated in the Native Range

The analysis of available additional Illumina genomes of invasive strains from the first invasion wave via North America (i.e. strains from the US and European clonal lineages) reinforced the view of the existence of balancing selection on the MAT-Proximal haplotypes. We indeed observed high differentiation between two clusters in the MAT-Proximal region in these populations, contrasting with patterns in other genomic regions (supplementary figs. S10a and S10c, Supplementary Material online), and the two MAT-proximal haplotypes were present in balanced frequencies. Only in the S12 European invasive clonal lineage (only *MAT-1*), we found a single MAT-Proximal haplotype (MAT-Prox1; supplementary fig. S10b, Supplementary Material online).

In the native populations and the invasive populations coming directly from Asia, the MAT-Proximal region also seemed to have undergone a very different evolutionary history than the rest of the genome (Fig. 7). In particular, the neighbor-net network indicated that sequences from the different genetic clusters previously described were intermingled in the MAT-proximal region (Fig. 7a), while the structure in the rest of the genome corresponded to the different clusters (Fig. 7b), i.e. CL1, CL2, CL3 and CL4. However, the two haplotypes were much less differentiated in the native range and in invasive populations originating directly from Asia than in invasive populations from the first invasion wave via North America. For example, branch lengths on the neighbor-net indicated a lower differentiation between haplotypes within CL2 (Fig. 7a) than within the European CL1 or within Swiss invasive populations (Fig. 2d; supplementary fig. S3, Supplementary Material online) and reticulations were present between MAT-proximal haplotypes. The European strains with a direct Asian origin exhibited a similar intermediate sequence in the MAT-Proximal region (supplementary fig. S10c, Supplementary Material online) as the three 2019 Swiss strains described above. This intermediate haplotype thus seemed associated with the second invasion in Europe, directly from Asia, and carried the same captains as the MAT-Prox2 haplotype from the M1400 strain. The intermediate location in the neighbor-net network may therefore just reflect the fact that the two haplotypes are less differentiated in the native Asian area.

We found no association of the MAT-Proximal haplotypes with mating types at the world scale (Figs 7 and supplementary fig. S10, Supplementary Material online). This may be due to small sampling sizes at the population level in North America and in Asia. Interestingly, in two invasive European lineages of American origin with a predominantly clonal structure (re092 and re103), the strains with a mating-type different from their clone-mates, due to localized introgression from other clonal lineages (Demené et al. 2019), also had their MAT-Proximal region introgressed (supplementary fig. S10c, Supplementary Material online).

No Evidence of Introgression in the MAT-proximal Region From Closely Related *Cryphonectria* Species

The stronger differentiation in the MAT-Proximal region in the Swiss and CL1 invasive populations than in other populations could be due to an introgression event from a closely related species. However, the small genetic distance and the reticulations observed in the neighbor-net network between the M6697 Swiss reference genome and the Asian genomes M8444 and XA19 rather indicate that the MAT-Prox1 haplotype in Switzerland originated from Asian *C. parasitica* populations (Fig. 7).

Gene genealogies also provided no evidence of introgression in the MAT-proximal region from closely related *Cryphonectria*

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

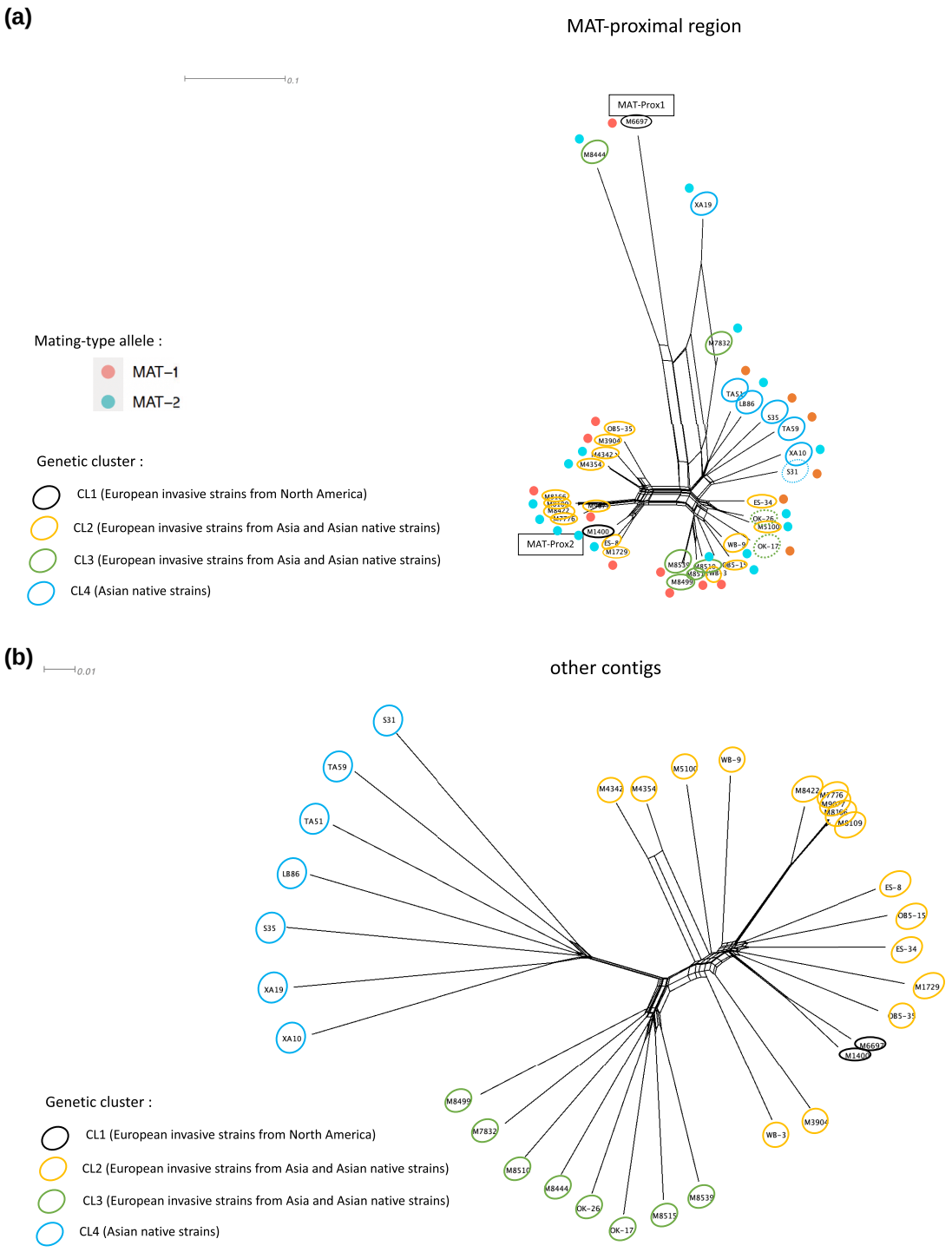


Fig. 7. Genetic structure using single nucleotide polymorphisms (SNPs) in the MAT-proximal region lacking recombination (a) and other genomic regions (b) in other resequenced *Cryphonectria parasitica* populations from the native and the introduced range. Neighbor-net network from a SplitsTree analysis based on: (a) SNPs ($n = 4{,}120$) located within the MAT-proximal region along the contig carrying the mating-type locus (scaffold_2 of the EP155 genome); (b) SNPs ($n = 103{,}058$) located on other contigs of the EP155 genome. Color of empty circles around strain ID indicate the genetic clusters the strains belong to (CL1, CL2, CL3, CL4; $n = 33$ strains). Dotted circles indicate strains of the native range and plain circles indicate strains of the introduced range. Red and blue dots near strain ID indicate the mating-type (*MAT-1* or *MAT-2*, respectively).

species. We performed a gene orthology analysis including three genomes of the closely related species *C. japonica*, *C. carpinicola* and *C. naterciae* that were previously sequenced with short-read technologies (Stauber et al. 2020). We retrieved only nine genes from the MAT-proximal region that were shared between the M6697 and M1400 genomes that had orthologs in one of the related species. This may be due to low quality assemblies in the outgroups (Stauber et al. 2020) and/or to many gene gains in this region in *C. parasitica* because of *Starships*. We found no gene with a topology consistent with introgression from another species into the MAT-Proximal region, i.e. with a MAT-Prox1 or MAT-Prox2 haplotype that would branch with an outgroup allele rather than with the alternative haplotype of *C. parasitica*. The high load of TEs and gene disruptions in the two MAT-Proximal haplotypes does not fit with the introgression hypothesis either, unless the introgression occurred from a species

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

with recombination suppression in the MAT-Proximal region. The *Starships* may have moved from a distant species by horizontal gene transfer together with their cargo genes, but we found no evidence for this by blasting the captains and cargo genes in public databases.

Discussion

A Large Nonrecombining Region Near the Mating-type Locus, With Two Highly Differentiated Haplotypes Maintained Polymorphic in the Native and Introduced Ranges

We found strong evidence for the complete cessation of recombination, in *C. parasitica*, in a region of more than 1 Mb (1.2 Mb to 1.5 Mb depending on the haplotype), proximal to the mating-type locus. Indeed, we detected maximal levels of linkage disequilibrium in otherwise recombining populations. We further revealed the existence of two differentiated haplotypes, without shared polymorphism, while even low rates of recombination can homogenize alleles and prevent LD build-up (Dufresnes et al. 2015). The full cessation of recombination was further supported by the existence of a previously unknown, large inversion in the center of the nonrecombining region in invasive strains. This is in agreement with previous genetic maps reporting lack of recombination near the mating-type locus in crosses involving Japanese, North American and Italian isolates (Kubisiak and Milgroom 2006). The higher TE load in this region constitutes additional evidence for full recombination cessation, as the lack of recombination renders selection against TE accumulation less efficient. TE takes hundreds of thousands of years to accumulate in nonrecombining regions (Duhamel et al. 2023), so that the high TE load in the MAT-Proximal region in *C. parasitica* indicates old recombination cessation. The particularly high frequency of nonsynonymous substitutions in the MAT-Proximal region worldwide further indicates degeneration and ancient recombination suppression. The existence of two differentiated haplotypes appeared less clear in the native range in Asia and in the invasive populations originating directly from Asia than in invasive populations from the first introduction via North America. The MAT-Proximal region nevertheless appeared to display an evolutionary history very different from the rest of the genome and a high TE load in the native range.

The estimates for the differentiation between the two MAT-Proximal haplotypes based on synonymous divergence between shared single-copy orthologs and based on TE insertion dates were at least 1.5 million years. The suppression of recombination in *C. parasitica* may be younger than this estimate if one of the two haplotypes was introgressed from a distant species. The evolution of nonrecombining regions by introgression has been reported for example in ants and butterflies (Jay et al. 2018; Helleu et al. 2022; Stolle et al. 2022). However, we found little evidence for an introgression: i) footprints of recombination suppression are present in the native range in terms of degeneration and particular genomic structure in the MAT-Proximal region, and ii) we found no signatures of introgression from closely related species in gene genealogies or by blast in public databases. In any case, the findings of recombination suppression footprints in the two invaded continents and in the native range of *C. parasitica* indicate ancient recombination suppression and long-term polymorphism maintenance in the MAT-Proximal region. These findings point to a strong balancing selection maintaining two differentiated haplotypes in *C. parasitica*.

In contrast to other known nonrecombining regions on sex or mating-type chromosomes (Bergero and Charlesworth 2009; Hartmann et al. 2021), the region without recombination was not completely linked to the mating-type locus. Indeed, the occurrence of rare recombination events between the nonrecombining MAT-Proximal region and the mating-type locus are supported by previous segregation analyses (Kubisiak and Milgroom 2006) and the lack of association found here between the MAT-Proximal haplotypes and the mating types at the worldwide scale, as well as their incomplete association within populations.

Proximal Cause of Recombination Suppression

The large inversion in the MAT-Proximal region may contribute to recombination suppression. The chromosomal arrangement of the M1400 MAT-Prox2 haplotype is likely the ancestral state as it is the most frequent worldwide, although a more complete sampling in the native range is required to obtain definitive conclusion. The nonrecombining region was defined based on the LD pattern and was larger than the inversion. This may indicate that additional proximal mechanisms suppress recombination. The inversion may even be a consequence rather than a cause of recombination cessation, as previously found in other fungal mating-type chromosomes (Grognet et al. 2014; Sun et al. 2017; Branco et al. 2018; Vittorelli et al. 2023), especially as the inversion was not detected so far in the native range despite the presence of footprints of recombination suppression. Other proximal causes of recombination cessation may for example be genetic recombination modifiers, epigenetic marks, histone modifications or lack of recruitment of proteins responsible for double-strand breaks (Maloisel and Rossignol 1998; Boideau et al. 2022; Legrand et al. 2024). The lack of rearrangements at the edge of the MAT-proximal region may allow rare recombination or gene conversion events, which may explain the presence of two distinct LD blocks with lower LD between them and the lower differentiation at the edges of the MAT-proximal region. Alternatively, the LD around the inversion may be formed by less frequent recombination as recombination is often modified around inversion breakpoints (Pegueroles et al. 2010; Stevison et al. 2011).

The MAT-proximal Haplotypes Carry Putative *Starship* Elements

We detected signatures of *Starship* elements in the nonrecombining MAT-proximal region, with likely three "solo" captains shared between the two haplotypes. We also identified an additional captain associated with 18 specific genes, strongly suggesting the presence of a cargo-mobilizing *Starship*, only present in the MAT-Prox2 haplotype. The difference in gene content between the two MAT-proximal haplotypes thus likely results, at least partly, from specific cargo genes inserted with *Starships*. There could also be gene losses because of the less efficient selection due to recombination suppression, and/or additional gene gains. One of the captains shared between haplotypes was located just nearby the mating-type locus and at the edge of the MAT-proximal region. This altogether strongly suggests a role of *Starship* elements in the formation of the MAT-proximal haplotypes and in their long-term maintenance. Experiments are however required to confirm this hypothesis and elucidate their role. High-quality assemblies suggested that the presence/absence of the polymorphic *Starship* insertion was associated with the MAT-Proximal haplotypes. Additional sampling in

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

the native range and high-quality genome assemblies are however required to obtain a more comprehensive view of the history of the MAT-Proximal region and of the putative *Starship* insertions.

Starships may have been introgressed from another species, as frequent with these elements (Ropars et al. 2015; Peck et al. 2024), but searches in databases returned no significant blast results. Such horizontal gene transfers should not have, however, inflated the date estimates of divergence of recombination suppression in the MAT-Proximal region, as these were computed based on shared single-copy orthologs between haplotypes and the shared putative *Starship* elements seem to correspond to homologous insertions.

The Selective Forces Potentially Maintaining the Two Haplotypes in the MAT-proximal Region

Our findings indicate that selection maintains the two haplotypes polymorphic in the MAT-Proximal region. There can be several hypotheses regarding the nature of this balancing selection. The degeneration of the nonrecombining region may help maintaining the two haplotypes by selection due to the sheltering of combinations of deleterious mutations in repulsion, i.e. thanks to a heterozygote advantage called pseudo-overdominance (Abu-Awad and Waller 2023). Such pseudo-overdominance advantage may only be a consequence of recombination suppression, and contribute to its maintenance, or may even be its initial cause (Branco et al. 2017; Jay et al. 2021, 2024b; Charlesworth 2024). These hypotheses require the existence of a substantial diploid or dikaryotic phase for sheltering recessive deleterious alleles in a heterozygous state. In fact, although *C. parasitica* has a predominant haploid phase, strains can be found as dikaryons heterozygous for the mating-type locus in some natural populations, and not only as monokaryons (McGuire et al. 2004, 2005; Stauber et al. 2021, 2022). The importance and frequency of heterokaryons in *C. parasitica* in nature would deserve further investigations. It may be sufficient if recessive deleterious alleles are sheltered in a substantial percentage of individuals for selecting recombination suppression, and/or if the genes involved are preferentially expressed during the diploid or dikaryotic phases. For a heterozygous advantage, it is sufficient that the MAT-Proximal haplotypes are associated with mating types within populations, but not necessarily the same mating types across populations. Further sampling will be required to test the association between the MAT-Proximal haplotypes and the mating types within local populations in the CL3 and CL4 clusters and in the native Asian range. It will also be interesting in future studies to compare the fitness between strains carrying the two MAT-Proximal haplotypes and dikaryotic strains being homozygous for the MAT-Proximal haplotype. Such experiments in Sordariales fungi reported a heterozygote advantage compared to one mating type (Guyot et al. 2024). Results from previous studies suggest that strains carrying either the same or different MAT-Proximal haplotypes can be crossed in *C. parasitica* (Stauber et al. 2021).

It could indeed also be that genes evolve under overdominance in this region for other reasons than deleterious mutations, perhaps in relationship with the pathogenic lifestyle of the fungus. One could imagine, for example, that heterozygosity could be advantageous at genes involved in virulence against the host tree, and especially the novel hosts colonized in invasive ranges, or in resistance against a parasitic virus known to negatively affect fitness in *C. parasitica* (Choi and Nuss 1992; Brusini et al. 2017). As a matter of fact, multiple genes present in the MAT-Proximal region are upregulated under infection by the hypovirus CHV1 or during the vegetative incompatibility reaction considered to play a role in the prevention of virus transmission (Choi et al. 2012; Rigling and Prospero 2018). The specific genes unique to one or the other MAT-proximal haplotypes, and possibly the cargo genes inserted with the *Starships*, may contribute to such a heterozygous advantage.

As an alternative to a heterozygote advantage, the selection maintaining the recombination suppression and the two haplotypes in the MAT-proximal region could be some kind of negative-frequency dependent selection of beneficial allelic combinations, possibly linked to a trench-warfare-like arms race with the host tree, the virus or the microbial community, or to self-incompatibility (Tellier et al. 2014; Jay et al. 2024a). Here too, the genes upregulated under virus infection or vegetative incompatibility reaction in the MAT-Proximal region may be relevant. The MAT-proximal region did not include any of the genes previously identified as controlling vegetative incompatibility in *C. parasitica* but not all self-incompatibility genes have been identified yet (Bragança et al. 2007; Robin et al. 2009; Short et al. 2015). Such a role in host–pathogen interactions or vegetative incompatibility would also be consistent with the selective sweep footprints detected in the MAT-proximal region in a previous study (Stauber et al. 2021). Balancing selection and location adjacent to a nonrecombining region has been reported for loci involved in host resistance in the *Daphnia–Pasteuria* system (Fredericksen et al. 2023). Such negative-frequency dependent selection could also explain the particular population structure in the MAT-proximal region in the native region, with a mix of sequences from the different genetic clusters. Indeed, introgression between populations could be favored by a positive selection of rare alleles, or there could be a long-term maintenance of alleles, and therefore of ancestral polymorphism. The much higher differentiation between MAT-Proximal haplotypes in the invasive populations than the native populations may be due to a selective sweep, in this context, of a rare and differentiated MAT-Prox1 allele present in the native range or having evolved rapidly. Particular demographic effects during the invasion, such as bottlenecks, may also have increased the differentiation between haplotypes (Moinet et al. 2022). These two phenomena are consistent with the very low genetic diversity in the MAT-Prox1 haplotype in the introduced range.

The MAT-Proximal region also contained, only in the MAT-Prox1 haplotype, a gene of the Sirtuin family, a homolog of which had previously been found in a *Starship* associated with local thermal climate adaptation in the wheat fungal pathogen *Zymoseptoria tritici* (Tralamazza et al. 2024). Another hypothesis to explain the balancing selection in the MAT-Proximal region may thus be a heterogeneous selection, with different genes in the two haplotypes providing contrasted advantages in different conditions or different phases of the life cycle. The proximity of the mating-type locus, permanently heterozygous in dikaryotic and diploid stages, may help maintain balanced frequencies of the two MAT-proximal haplotypes. The other TE insertion polymorphism observed from the high-quality assemblies may also be adaptive (Casacuberta and González 2013; Orteu et al. 2024).

Another hypothesis for explaining such extension of recombination suppression beyond the mating-type locus is antagonistic selection, i.e. linkage of alleles that would improve fitness of a *MAT-1* gamete while being deleterious in a *MAT-2* gamete, or vice-versa. However, full recombination suppression with the mating-type locus would be expected under such antagonistic selection, as well as the same association between the

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

MAT-Proximal alleles and the mating-type in all populations, in contrast to our findings. In addition, we found no particular predicted function in the MAT-proximal region that could be related to mating compatibility and there are very little functions, if any, with possible antagonistic functions between mating types in fungi (Bazzicalupo et al. 2019; Hartmann et al. 2021).

Conclusion

In conclusion, we provide strong evidence for the existence of a nonrecombining region partially physically linked to the mating-type locus in the chestnut blight fungus *C. parasitica*, with two highly differentiated haplotypes, each carrying specific genes, maintained polymorphic by selection. We found footprints of balancing selection in the MAT-proximal region in the first introduction of the pathogen in Europe from North America and a chromosomal inversion. The nonrecombining region also displayed footprints of particular evolution in Asia and in the second invasion wave directly from Asia, although the levels of differentiation between haplotypes was lower than in the populations from the first introduction wave. This nonrecombining region may underlie important adaptive traits and thereby provide important applications for the control of a devastating tree pathogen. This is supported by the finding of putative *Starship* elements in the MAT-proximal region, that can be involved in adaptation. Examples include fungal pathogens of coffee (Peck et al. 2024) and of wheat (Bucknell and McDonald 2023), and in molds maturing cheeses (Cheeseman et al. 2014; Ropars et al. 2015) or dry-cured meat (Lo et al. 2023). In addition, the high-quality genome assemblies provided here, from the native and invaded ranges, will more generally be useful for studies aiming at understanding the evolution of this invasive and damaging pathogen.

Materials and Methods

Strains and Genomic Data

For population genomic analyses, we studied a collection of 386 monokaryotic *C. parasitica* strains sampled worldwide, from the native and invaded ranges of the pathogen, and sequenced previously with the short-read Illumina technology (strain information are presented in supplementary table S1, Supplementary Material online; (Demené et al. 2019; Stauber et al. 2021; Stauber et al. 2022)). Data were downloaded from GenBank Bioproject numbers PRJNA604575, PRJNA644891 and PRJNA706885. We focused our analyses first on European invasive strains originating from North America. We studied 88 strains belonging to the CL1 genetic cluster, in central and south-eastern Europe (Stauber et al. 2021). We excluded the putative clonal genotypes previously identified (Stauber et al. 2021). We also studied 71 strains sampled in the 1990 and 62 strains sampled in 2019 in southern Switzerland (Stauber et al. 2022). Mating-type ratios close to ½ and population structure analyses of the CL1 genetic cluster and the Swiss 1990 population suggest frequent recombination in these populations (Stauber et al. 2021, 2022). The mating-type ratio was 33% *MAT-1* - 66 % *MAT-2* in the 2019 Swiss population and population structure analyses suggested regular sexual reproduction and a recent population bottleneck. To study the presence of the two haplotypes in the MAT-proximal region more broadly, we analyzed monokaryotic strains belonging to the CL2, CL3, CL4 genetic clusters, as well as the S12 European invasive lineage (Stauber et al. 2021) and additional monokaryotic strains from the US and Europe (with an Asian or North American origin (Demené et al. 2019)). We excluded heterokaryotic strains, i.e. strains having both *MAT-1* and *MAT-2* alleles, as the phase was challenging to infer.

For mapping and SNP calling, we first used as reference the 43.9 Mb genome sequence EP155 v2.0 of the strain EP155 (*MAT-2*, North America, CL1 cluster) available at the Joint Genome Institute (http://jgi.doe.gov/)(Crouch et al. 2020). For comparative genomics, we used the published genome of the strain ESM15 (*MAT-2*, Japan, CL2 cluster; (Demené et al. 2022)) available at GenBank Bioproject PRJNA700518 under the accession JAGDFO000000000. We additionally sequenced de novo, with long-read technologies, five strains from the native and invaded ranges of the pathogen. We sequenced with PacBio Hifi the genomes of the strains M1400 (*MAT-2*) and M6697 (*MAT-1*) sampled in Gnosca, southern Switzerland (Stauber et al. 2021). Mycelia were stored as glycerol stocks at −80 C after strain isolation. To produce mycelium for DNA extraction, isolates were inoculated onto cellophane-covered potato dextrose agar plates (PDA, 39 g/L; BD Becton, Dickinson and Company, Franklin Lakes, USA) (Hoegger et al. 2000) and incubated for a minimum of 1 wk at 24 °C, at a 14 hr light and 10 hr darkness cycle. Mycelium and spores were harvested by scratching the mycelial mass off the cellophane, transferring it into 2 mL tubes and freeze-drying it for 24 hr (Stauber et al. 2021). DNA extraction was performed with the NucleoBond High Molecular Weight DNA kit from Macherey-Nagel, with the mechanical disruption of about 30 mg of lyophilized mycelium with two beads of 3 mm for 5 min at 30 Hz. Sequencing was outsourced to Novogene, the Netherlands. We additionally sequenced with Oxford Nanopore MinION technology the strains XIM9508 (China, CL4 cluster, *MAT-1*), MRC10 (South Western France introduced directly from Asia, CL2 cluster, *MAT-2*) and DUM-005 (USA, CL1 cluster, *MAT-2*). These isolates had been collected for previous studies (Milgroom et al. 1996; Liu et al. 2003; Dutech et al. 2012). The protocols for mycelium production, DNA extraction and sequencing for these three strains were the same as in (Demené et al. 2022).

Short-read Data Processing and SNP Calling

We performed mapping and SNP calling of short-read data using the EP155 reference v2.0 genome and then the M1400 new genome assembly as a reference. Mapping and SNP calling was performed as described in (Stauber et al. 2021, 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 TEs.

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

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

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 computed SNP density in 100kb nonoverlapping windows using the --SNPdensity of vcftools (Danecek et al. 2011). 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.

We built neighbor-net 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 study of other populations, we included only one strain of each haplotype of the invasive European population but not all CL1 and Swiss strains for network readability; we nevertheless checked that the results were the same with all CL1 and Swiss strains.

Long-read based Assemblies

PacBio Hifi reads of strains M1400 and M6697 were both assembled using canu v1.8 (Koren et al. 2017) program with a set genome size of 44 Mb. Multiple assembly pipelines were used for the other strains. Oxford Nanopore MinION reads and Illumina reads of the genomes of both XIM9508 and MRC10 strains were assembled using HybridSPAdes (Antipov et al. 2016). Assemblies were manually curated and scaffolds were cut when evidence of a chimeric connection was detected (i.e. mis-paired short reads) as previously described (Demené et al. 2022). For DUM005, the assembly was generated by Ra with basic parameters that uses Oxford Nanopore MinION reads and corrects the assembly with Illumina reads (https://github.com/lbcb-sci/ra) as it outperformed the HybridSPAdes assembly. As the HybridSpades assembly of the MRC10 strain suggested absence of collinearity with the MAT-proximal region M1400 and M6697, we further checked the assembly of the MAT-proximal region of MRC010 by generating a meta-assembly of this strain. We used the assembler tool canu v2.2 with an estimated genome size of 42 Mb. We also used Flye v2.9.3-b1797 (Kolmogorov et al. 2019) with an estimated genome size of 42 Mb, --nano-raw for reads with an error rate below 20 and a coverage for initial disjointing assembly of 50. Then we used ragtag (Alonge et al. 2019) to patch the canu assembly with the Flye assembly as query in a first loop. In a second loop, the canu assembly was patched with the first loop ragtag assembly. Finally, we polished this second output assembly of ragtag with short reads and the consensus part of medaka v1.11.3 with no change in the parameters (https://github.com/nanoporetech/medaka). Assemblies statistics were obtained with quast v5.1 (Gurevich et al. 2013). We assessed the completeness of each assembly using the Benchmarking of Universal Single-Copy Orthologue (BUSCO) tools with the Sordariomyceta ortholog set (sordariomycetes_odb10, 2020 to 2008–05, $n = 3817$ orthologous groups searched) (Manni et al. 2021). To identify telomeres, we search for telomeric repeats "TTAGGG" and "CCCTAA". Gene models were predicted with the Helixer v0.3.1 pipeline (Holst et al. 2023). We also run Helixer pipeline for ESM15 and EP155 strains for the gene orthology analysis. Statistics of the obtained gene annotation was obtained with the Agat v1.0.0 tool (Dainat et al. 2020). TEs of all genomes were annotated using Repeatmasker v4–0–7 (Smit et al. 2013) and the customized library built for *C. parasitica* in (Demené et al. 2022) contained in the "Curated_TEs_database_ESM015_EP155.fa" file (available on the "Portail Data INRAE: Chromosomal rearrangements but no change of genes and TEs repertoires in an invasive forest-pathogenic fungus" at https://doi.org/10.15454/UTIB8U). The class Gypsy invader was renamed LTR-Ty3. We filtered out TE copies shorter than 100bp. We used the HybridSPAdes preliminary assembly of MRC10 for gene model annotations. To study gene functions of the EP155 genome, we used the gene annotation available at http://jgi.doe.gov/ (Crouch et al. 2020). To study support from RNAseq data, we used RNAseq data from the EP155 strain cultivated in vitro available at Genebank under Project ID number PRJNA588887 and accessions numbers SRR10428542, SRR10428543, SRR10428544 (Chun et al. 2020). Raw reads were mapped using STAR v2.7.10a (Dobin et al. 2013) with the settings. --alignIntronMax 1000 --limitBAMsortRAM 1629219965--quantMode GeneCounts. We used the program featureCounts (Liao et al. 2013) v2.0.6 with the options *-P* --countReadPairs -M - -B -O --largestOverlap. We considered a gene to be supported for read count >10. To predict gene functions of the protein predicted by Helixer, we used InterProScan version 5.56–89.0 (Jones et al. 2014) and the funannotate pipeline version 1.8.15 with default options (Palmer and Stajich 2020). To look for *Starship* captains, we ran the Gene Finder Module of starfish pipeline version 1.0.0 with default options (Gluck-Thaler and Vogan 2024). We studied the phylogenetic relationships of the putative *Starships* captains by aligning them to 1222 representative *Starship* captains of the YRsuperfamRefs.faa file from the starfish database (Gluck-Thaler and Vogan 2024). Protein sequences were aligned using Clustal Omega (Sievers and Higgins 2021) version 1.2.4 allowing for five iterations (–iterations 5). Gaps in the resulting alignment were trimmed using trimAl v1.4.rev15 and the -gappyout option. The phylogenetic relationship among proteins was inferred from the trimmed alignment using FastTree (Price et al. 2009) under the Whelan-And-Goldman 2001 model after 1,000 bootstraps (-boot 1000 and -wag options). Plots were performed in R v 4.1.2 using ggtree v3.9.1.

Estimation of Retrotranposon Insertion Time and Nucleotide Divergence Time

To get an estimation of the insertion date of the TEs present in the MAT-proximal nonrecombining region, we applied two complementary methods. We first used the divergence between the LTR sequences in retrotransposons, as these LTR sequences at their edges are identical at the time of TE insertion and then diverge with time. For this, we used a *de novo* prediction of LTR retrotransposons using LTRharvest GenomeTools 1.6.2 (Ellinghaus et al. 2008; Gremme et al. 2013) and LTR_Finder v1.07 (Xu and Wang 2007). To prepare data for LTRharvest, we first created enhanced suffix for the M1400 and M6697 genome assemblies using the GenomeTools Suffixerator (-dna -suf -lcp -bwt -bck -mirrored parameters). We ran LTRharvest using two pipelines, designed to identify LTR retrotransposons with a minimum and maximum LTR length of 100 bp and 7000 bp

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

respectively and at least 85% identity between two LTR regions, with and without canonical TGCA motifs, respectively: i) -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -similar 85 -vic 10 -seed 20 -motif TGCA -motifmis 1; ii) -minlentltr 100, -maxlenltr 7000, -mintsd 4, -maxtsd 6, -similar 85, -vic 10, -seed 20. Similarly, we ran LTR_Finder on the M1400 and M6697 genome assemblies to retrotransposons with both TGCA and non-TGCA motifs and a minimum and maximum LTR length of 100 bp and 7000 bp respectively and at least 85% identity between two LTR regions (-D 15000, -d 1000, -l 100, -L 7000, -*P* 20, -C, -M 0.85). Finally, we used LTR_retriever v2.8.5 (Ou and Jiang 2018) with default parameters to filter out false positive LTR candidates identified by LTRharvest and LTR_Finder and get an estimation of each element insertion date.

As a second method to estimate the insertion date of the TEs present in the MAT-proximal nonrecombining region, we used the set of curated consensus sequences from (Demené et al. 2022) to annotate the inversion sequence or its surroundings. We first used samtools faidx v1.9 (Li et al. 2009) to extract the sequence corresponding to inversion and their 50 kb surroundings in M1400 and M6697 genome assemblies. We annotated the sequences corresponding to the inversions and the concatenated 50 kb surroundings in both isolates using RepeatMasker version 4.1.5 and rmblast as search engine (v2.10.0) with -no_is -pa 20 -cutoff 250 -a parameters. Finally, we parsed the RepeatMasker .out file using the script parseRM_merge_interrupted.pl (https://github.com/4ureliek/Parsing-RepeatMasker-Outputs) and omitted Simple_repeat and Low_complexity regions. We then built a summary of the alignments using the RepeatMasker helper script buildSummary.pl and calculated sequence divergence using the calcDivergenceFromAlign.pl script to finally render results with the createRepeatLandscape.pl from the same helper suite.

Comparative Genomics Analyses

Genome synteny between long reads assemblies were studied using nucmer v3.1 (Marçais et al. 2018) and minimap2 -cx asm10 (Li 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 alignment tool available at https://blast.ncbi.nlm.nih.gov/Blast.cgi(last accessed 13th May 2024).

To study gene disruption, synonymous and nonsynonymous 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. For the seven *C. parasitica* assemblies, we used the annotation obtained from the Helixer v0.3.1 pipeline (Holst et al. 2023) for this analysis. 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 values 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 $T\text{generations} = d_S/2\mu$. 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.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

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

Author Contributions

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

Data Availability

The genomic data newly generated in this article are available in the GenBank database under BioProject number PRJNA1198726. Raw reads accession numbers and biosample ID of each strains are described in supplementary table S3, Supplementary Material online.

References

Abascal F, Zardoya R, Telford MJ. Translatorx: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*. 2010:38(suppl_2):W7–13. https://doi.org/10.1093/nar/gkq291.

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

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol*. 2019:20(1):224. https://doi.org/10.1186/s13059-019-1829-6.

Anagnostakis SL. Chestnut blight: the classical problem of an introduced pathogen. *Mycologia*. 1987:79:23–37. https://doi.org/10.1080/00275514.1987.12025367.

Antipov D, Korobeynikov A, McLean JS, Pevzner PA. hybridSPAdes: an algorithm for hybrid assembly of short and long reads.

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

Bioinformatics. 2016:32(7):1009–1015. https://doi.org/10.1093/bioinformatics/btv688.

Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet*. 2013:14(2):113–124. https://doi.org/10.1038/nrg3366.

Bazzicalupo AL, Carpentier F, Otto SP, Giraud T. Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-dioicae*). *G3 Genes Genomes Genet*. 2019:9(6):1987–1998. https://doi.org/10.1534/g3.119.400242.

Belov AA, Witte TE, Overy DP, Smith ML. Transcriptome analysis implicates secondary metabolite production, redox reactions, and programmed cell death during allorecognition in *Cryphonectria parasitica*. *G3 GenesGenomesGenetics*. 2021:11:jkaa021. https://doi.org/10.1093/g3journal/jkaa021.

Berdan EL, Blanckaert A, Butlin RK, Flatt T, Slotte T, Wielstra B. Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals. *Philos Trans R Soc B Biol Sci*. 2022:377(1856):20210199. https://doi.org/10.1098/rstb.2021.0199.

Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol*. 2009:24(2):94–102. https://doi.org/10.1016/j.tree.2008.09.010.

Boideau F, Richard G, Coriton O, Huteau V, Belser C, Deniot G, Eber F, Falentin C, Ferreira de Carvalho J, Gilet M, *et al*. Epigenomic and structural events preclude recombination in *Brassica napus*. *New Phytol*. 2022:234(2):545–559. https://doi.org/10.1111/nph.18004.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 2014:30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bragança H, Simões S, Onofre N, Tenreiro R, Rigling D. *Cryphonectria parasitica* in Portugal: diversity of vegetative compatibility types, mating types, and occurrence of hypovirulence. *For Pathol*. 2007:37(6):391–402. https://doi.org/10.1111/j.1439-0329.2007.00513.x.

Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza S, Brandenburg J-T, Coelho MA, Hood ME, *et al*. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc Natl Acad Sci U S A*. 2017:114(27):7067–7072. https://doi.org/10.1073/pnas.1701658114.

Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, Prieur SL, Coelho MA, de Vienne DM, Hartmann FE, Begerow D, *et al*. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat Commun*. 2018:9(1):2000. https://doi.org/10.1038/s41467-018-04380-9.

Brusini J, Wayne ML, Franc A, Robin C. The impact of parasitism on resource allocation in a fungal host: the case of *Cryphonectria parasitica* and its mycovirus, *Cryphonectria hypovirus 1*. *Ecol Evol*. 2017:7(15):5967–5976. https://doi.org/10.1002/ece3.3143.

Bucknell A, Wilson HM, do Santos KCG, Simpfendorfer S, Milgate A, Germain H, Solomon PS, Bentham A, McDonald MC. *Sanctuary*: A *Starship* transposon facilitating the movement of the virulence factor ToxA in fungal wheat pathogens. *bioRxiv 583430*. 2024. https://doi.org/10.1101/2024.03.04.583430.

Bucknell AH, McDonald MC. That's no moon, it's a starship: giant transposons driving fungal horizontal gene transfer. *Mol Microbiol*. 2023:120(4):555–563. https://doi.org/10.1111/mmi.15118.

Carpentier F, Rodríguez de la Vega RC, Jay P, Duhamel M, Shykoff JA, Perlin MH, Wallen RM, Hood ME, Giraud T. Tempo of degeneration across independently evolved nonrecombining regions. *Mol Biol Evol*. 2022:39(4):msac060. https://doi.org/10.1093/molbev/msac060.

Casacuberta E, González J. The impact of transposable elements in environmental adaptation. *Mol Ecol*. 2013:22(6):1503–1517. https://doi.org/10.1111/mec.12170.

Charlesworth B. The fitness consequences of genetic divergence between polymorphic gene arrangements. *Genetics*. 2024:226(3):iyad218. https://doi.org/10.1093/genetics/iyad218.

Charlesworth D. The status of supergenes in the 21st century: recombination suppression in batesian mimicry and sex chromosomes and other complex adaptations. *Evol Appl*. 2016:9(1):74–90. https://doi.org/10.1111/eva.12291.

Charlesworth D, Vekemans X, Castric V, Glémin S. Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytol*. 2005:168(1):61–69. https://doi.org/10.1111/j.1469-8137.2005.01443.x.

Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M, Conseiller E, Debuchy R, *et al*. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nat Commun*. 2014:5(1):2876. https://doi.org/10.1038/ncomms3876.

Choi GH, Dawe AL, Churbanov A, Smith ML, Milgroom MG, Nuss DL. Molecular characterization of vegetative incompatibility genes that restrict hypovirus transmission in the chestnut blight fungus *Cryphonectria parasitica*. *Genetics*. 2012:190(1):113–127. https://doi.org/10.1534/genetics.111.133983.

Choi GH, Nuss DL. Hypovirulence of chestnut blight fungus conferred by an infectious viral cDNA. *Science*. 1992:257(5071):800–803. https://doi.org/10.1126/science.1496400.

Chun J, Ko Y-H, Kim D-H. Transcriptome analysis of *Cryphonectria parasitica* infected with *Cryphonectria hypovirus 1* (CHV1) reveals distinct genes related to fungal metabolites, virulence, antiviral RNA-silencing, and their regulation. *Front Microbiol*. 2020:11:1711. https://doi.org/10.3389/fmicb.2020.01711.

Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, Hillman BI, Milgroom MG, Pangilinan J, Smith M, *et al*. Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: a fundamental resource for an archetypical invasive plant pathogen. *Phytopathology®*. 2020:110(6):1180–1188. https://doi.org/10.1094/PHYTO-12-19-0478-A.

Dainat J, Hereñú D, Pucholt P. *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format*. (Version v0.7.0). Zenodo; 2020. https://www.doi.org/10.5281/zenodo.3552717.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al*. The variant call format and VCFtools. *Bioinformatics*. 2011:27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

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

Demené A, Legrand L, Gouzy J, Debuchy R, Saint-Jean G, Fabreguettes O, Dutech C. Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of *Cryphonectria parasitica* in Western Europe. *Fungal Genet Biol*. 2019:130:122–133. https://doi.org/10.1016/j.fgb.2019.06.002.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics*. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dufresnes C, Borzée A, Horn A, Stöck M, Ostini M, Sermier R, Wassef J, Litvinchuck SN, Kosch TA, Waldman B, *et al*. Sex-chromosome homomorphy in palearctic tree frogs results from both turnovers and X-Y recombination. *Mol Biol Evol*. 2015:32(9):2328–2337. https://doi.org/10.1093/molbev/msv113.

Duhamel M, Hood ME, de la Vega RCR, Giraud T. Dynamics of transposable element accumulation in the non-recombining regions of mating-type chromosomes in anther-smut fungi. *Nature Communications*. 2023:14(5692). https://doi.org/10.1038/s41467-023-41413-4.

Dussert Y, Legrand L, Mazet ID, Couture C, Piron M-C, Serre R-F, Bouchez O, Mestre P, Toffolatti SL, Giraud T, *et al*. Identification of the first oomycete mating-type locus sequence in the grapevine downy mildew pathogen, *Plasmopara viticola*. *Curr Biol*. 2020:30(20):3897–3907.e4. https://doi.org/10.1016/j.cub.2020.07.057.

Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V. The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Mol*

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

Ecol. 2012:21(16):3931–3946. https://doi.org/10.1111/j.1365-294X.2012.05575.x.

Dutech C, Fabreguettes O, Capdevielle X, Robin C. Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity (Edinb)*. 2010:105(2):220–228. https://doi.org/10.1038/hdy.2009.164.

Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*. 2008:9(1):18. https://doi.org/10.1186/1471-2105-9-18.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y.

Fredericksen M, Fields PD, Pasquier LD, Ricci V, Ebert D. QTL study reveals candidate genes underlying host resistance in a red queen model system. *PLOS Genet*. 2023:19(2):e1010570. https://doi.org/10.1371/journal.pgen.1010570.

Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE. Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol*. 2020:12:750–763. https://doi.org/10.1093/gbe/evaa081.

Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack TL, Wood CW, Slot JC, Lopez-Nicora HD, *et al*. Giant starship elements mobilize accessory genes in fungal genomes. *Mol Biol Evol*. 2022:39(5):msac109. https://doi.org/10.1093/molbev/msac109.

Gluck-Thaler E, Vogan AA. Systematic identification of cargo-mobilizing genetic elements reveals new dimensions of eukaryotic diversity. *Nucleic Acids Res*. 2024:52(10):5496–5513. https://doi.org/10.1093/nar/gkae327.

Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinform*. 2013:10(3):645–656. https://doi.org/10.1109/TCBB.2013.68.

Grognet P, Bidard F, Kuchly C, Tong LCH, Coppin E, Benkhali JA, Couloux A, Wincker P, Debuchy R, Silar P. Maintaining two mating types: structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. *Genetics*. 2014:197(1):421–432. https://doi.org/10.1534/genetics.113.159988.

Guerin L, Froidefond G, Xu X-M. Seasonal patterns of dispersal of ascospores of *Cryphonectria parasitica* (chestnut blight). *Plant Pathol*. 2001:50(6):717–724. https://doi.org/10.1046/j.1365-3059.2001.00600.x.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013:29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Guy L, Roat Kultima J, Andersson SGE. genoPlotr: comparative gene and genome visualization in R. *Bioinformatics*. 2010:26(18):2334–2335. https://doi.org/10.1093/bioinformatics/btq413.

Guyot L, Chahine E, Lalanne C, Hartmann FE, Giraud T. Sheltered load in fungal mating-type chromosomes revealed by fitness experiments. *bioRxiv 612177*. https://doi.org/10.1101/2024.09.10.612177, 2024, preprint: not peer reviewed.

Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci*. 2020:6:e251. https://doi.org/10.7717/peerj-cs.251.

Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, Malagnac F, Grognet P, Giraud T. Recombination suppression and evolutionary strata around mating-type loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes. *New Phytol*. 2021:229(5):2470–2491. https://doi.org/10.1111/nph.17039.

Helleu Q, Roux C, Ross KG, Keller L. Radiation and hybridization underpin the spread of the fire ant social supergene. *Proc Natl Acad Sci U S A*. 2022:119(34):e2201040119. https://doi.org/10.1073/pnas.2201040119.

Hoegger PJ, Rigling D, Holdenrieder O, Heiniger U. Genetic structure of newly established populations of *Cryphonectria parasitica*. *Mycol Res*. 2000:104(9):1108–1116. https://doi.org/10.1017/S0953756299002397.

Holst F, Bolger A, Günther C, Maß J, Triesch S, Kindel F, Kiel N, Saadat N, Ebenhöh O, Usadel B, *et al*. Helixer–de novo prediction of primary Eukaryotic gene models combining deep learning and a hidden markov model. *bioRxiv 527280*. https://doi.org/10.1101/2023.02.06.527280, 6 February 2023, preprint: not peer reviewed.

Huson DH. SplitsTree: analyzing and visualizing evolutionary data. *Bioinforma Oxf Engl*. 1998:14(1):68–73. https://doi.org/10.1093/bioinformatics/14.1.68.

Ironside JE. No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays*. 2010:32(8):718–726. https://doi.org/10.1002/bies.200900124.

Jay P, Aubier TG, Joron M. The interplay of local adaptation and gene flow may lead to the formation of supergenes. *Mol Ecol*. 2024a:33:e17297. https://doi.org/10.1111/mec.17297.

Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M. Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nat Genet*. 2021:53(3):288–293. https://doi.org/10.1038/s41588-020-00771-1.

Jay P, Jeffries D, Hartmann FE, Véber A, Giraud T. Why do sex chromosomes progressively lose recombination? *Trends Genet*. 2024b:40:564–579. https://doi.org/10.1016/j.tig.2024.03.005.

Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J, Dasmahapatra KK, Joron M. Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr Biol*. 2018:28(11):1839–1845.e3. https://doi.org/10.1016/j.cub.2018.04.072.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014:30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Kasuga T, White TJ, Taylor JW. Estimation of nucleotide substitution rates in eurotiomycete fungi. *Mol Biol Evol*. 2002:19(12):2318–2324. https://doi.org/10.1093/oxfordjournals.molbev.a004056.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019:37(5):540–546. https://doi.org/10.1038/s41587-019-0072-8.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017:27(5):722–736. https://doi.org/10.1101/gr.215087.116.

Kubisiak TL, Milgroom MG. Markers linked to vegetative incompatibility (vic) genes and a region of high heterogeneity and reduced recombination near the mating type locus (MAT) in *Cryphonectria parasitica*. *Fungal Genet Biol*. 2006:43(6):453–463. https://doi.org/10.1016/j.fgb.2006.02.002.

Küpper C, Stocks M, Risse JE, dos Remedios N, Farrell LL, McRae SB, Morgan TC, Karlionova N, Pinchuk P, Verkuil YI, *et al*. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet*. 2016:48(1):79–83. https://doi.org/10.1038/ng.3443.

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009:10(3):R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Legrand S, Saifudeen A, Bordelet H, Vernerey J, Guille A, Bignaud A, Thierry A, Acquaviva L, Gaudin M, Sanchez A, *et al*. Absence of chromosome axis protein recruitment prevents meiotic recombination chromosome-wide in the budding yeast lachancea kluyveri. *Proc Natl Acad Sci U S A*. 2024:121(12):e2312820121. https://doi.org/10.1073/pnas.2312820121.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018:34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009:25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Liao Y, Smyth GK, Shi W. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013:41(10):e108–e108. https://doi.org/10.1093/nar/gkt214.

Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs.

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

Bioinformatics. 2012:28(2):298–299. https://doi.org/10.1093/bioinformatics/btr642.

Liu YC, Linder-Basso D, Hillman BI, Kaneko S, Milgroom MG. Evidence for interspecies transmission of viruses in natural populations of filamentous fungi in the genus *Cryphonectria*. *Mol Ecol*. 2003:12(6):1619–1628. https://doi.org/10.1046/j.1365-294X.2003.01847.x.

Lo Y-C, Bruxaux J, Rodríguez de la Vega RC, O'Donnell S, Snirc A, Coton M, Le Piver M, Le Prieur S, Roueyre D, Dupont J, *et al*. Domestication in dry-cured meat *Penicillium* fungi: convergent specific phenotypes and horizontal gene transfers without strong genetic subdivision. *Evol Appl*. 2023:16(9):1637–1660. https://doi.org/10.1111/eva.13591.

Lovat C-A, Donnelly DJ. Mechanisms and metabolomics of the host–pathogen interactions between chestnut (*Castanea* species) and chestnut blight (*Cryphonectria parasitica*). *For Pathol*. 2019:49(6):e12562. https://doi.org/10.1111/efp.12562.

Maloisel L, Rossignol JL. Suppression of crossing-over by DNA methylation in *Ascobolus*. *Genes Dev*. 1998:12(9):1381–1389. https://doi.org/10.1101/gad.12.9.1381.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:38(10):4647–4654. https://doi.org/10.1093/molbev/msab199.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 2018:14(1):e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

McGuire IC, Davis JE, Double ML, MacDonald WL, Rauscher JT, McCawley S, Milgroom MG. Heterokaryon formation and parasexual recombination between vegetatively incompatible lineages in a population of the chestnut blight fungus, *Cryphonectria parasitica*. *Mol Ecol*. 2005:14(12):3657–3669. https://doi.org/10.1111/j.1365-294X.2005.02693.x.

McGuire IC, Marra RE, Milgroom MG. Mating-type heterokaryosis and selfing in *Cryphonectria parasitica*. *Fungal Genet Biol*. 2004:41(5):521–533. https://doi.org/10.1016/j.fgb.2003.12.007.

McGuire IC, Marra RE, Turgeon BG, Milgroom MG. Analysis of mating-type genes in the chestnut blight fungus, *Cryphonectria parasitica*. *Fungal Genet Biol*. 2001:34(2):131–144. https://doi.org/10.1006/fgbi.2001.1295.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al*. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010:20(9):1297–1303. https://doi.org/10.1101/gr.107524.110.

Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLOS Genet*. 2008:4(3):e1000030. https://doi.org/10.1371/journal.pgen.1000030.

Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneko S. Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia*. 1996:88(2):179–190. https://doi.org/10.1080/00275514.1996.12026642.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Moinet A, Schlichta F, Peischl S, Excoffier L. Strong neutral sweeps occurring during a population contraction. *Genetics*. 2022:220(4):iyac021. https://doi.org/10.1093/genetics/iyac021.

Orteu A, Kucka M, Gordon IJ, Ng'iru I, van der Heijden ESM, Talavera G, Warren IA, Collins S, ffrench-Constant RH, Martins DJ, *et al*. Transposable element insertions are associated with batesian mimicry in the pantropical butterfly *Hypolimnas misippus*. *Mol Biol Evol*. 2024:41(3):msae041. https://doi.org/10.1093/molbev/msae041.

Otto SP, Lenormand T. Resolving the paradox of sex and recombination. *Nat Rev Genet*. 2002:3(4):252–261. https://doi.org/10.1038/nrg761.

Ou S, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol*. 2018:176(2):1410–1422. https://doi.org/10.1104/pp.17.01310.

Palmer JM, Stajich J. 2020. Funannotate v1.8.1: Eukaryotic genome annotation. Available from: https://zenodo.org/record/4054262

Peck LD, Llewellyn T, Bennetot B, O'Donnell S, Nowell RW, Ryan MJ, Flood J, de la Vega RCR, Ropars J, Giraud T, *et al*. Horizontal transfers between fungal *Fusarium* species contributed to successive outbreaks of coffee wilt disease. *PLOS Biology*. 2024:22(12):e3002480. https://doi.org/10.1371/journal.pbio.3002480.

Pegueroles C, Ordóñez V, Mestres F, Pascual M. Recombination and selection in the maintenance of the adaptive value of inversions. *J Evol Biol*. 2010:23(12):2709–2717. https://doi.org/10.1111/j.1420-9101.2010.02136.x.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 2014:31(7):1929–1936. https://doi.org/10.1093/molbev/msu136.

Price MN, Dehal PS, Arkin AP. FastTree: computing large Minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*. 2009:26(7):1641–1650. https://doi.org/10.1093/molbev/msp077.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007:81(3):559–575. https://doi.org/10.1086/519795.

Rigling D, Prospero S. *Cryphonectria parasitica*, the causal agent of chestnut blight: invasion history, population biology and disease control. *Mol Plant Pathol*. 2018:19(1):7–20. https://doi.org/10.1111/mpp.12542.

Robin C, Capdevielle X, Martin M, Traver C, Colinas C. *Cryphonectria parasitica* vegetative compatibility type analysis of populations in south-western France and northern Spain. *Plant Pathol*. 2009:58(3):527–535. https://doi.org/10.1111/j.1365-3059.2008.01986.x.

Ropars J, Rodríguez de la Vega RC, López-Villavicencio M, Gouzy J, Sallet E, Dumas É, Lacoste S, Debuchy R, Dupont J, Branca A, *et al*. Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Curr Biol*. 2015:25(19):2562–2569. https://doi.org/10.1016/j.cub.2015.08.025.

Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. *Curr Biol*. 2014:24(7):R288–R294. https://doi.org/10.1016/j.cub.2014.01.056.

Shin J-H, Blay S, McNeney B, Graham J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw*. 2006:16(Code Snippet 3):1–10. https://doi.org/10.18637/jss.v016.c03.

Short DPG, Double M, Nuss DL, Stauder CM, MacDonald W, Kasson MT. Multilocus PCR assays elucidate vegetative incompatibility gene profiles of *Cryphonectria parasitica* in the United States. *Appl Environ Microbiol*. 2015:81(17):5736–5742. https://doi.org/10.1128/AEM.00926-15.

Sievers F, Higgins DG. The clustal Omega multiple alignment package. In: Katoh K, editors. *Multiple sequence alignment: methods and protocols*. New York, NY: Springer US; 2021. p. 3–16.

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

Stauber L, Badet T, Feurtey A, Prospero S, Croll D. Emergence and diversification of a highly invasive chestnut pathogen lineage across Southeastern Europe. *eLife*. 2021:10:e56279. https://doi.org/10.7554/eLife.56279.

Stauber L, Croll D, Prospero S. Temporal changes in pathogen diversity in a perennial plant–pathogen–hyperparasite system. *Mol Ecol*. 2022:31(7):2073–2088. https://doi.org/10.1111/mec.16386.

Stauber L, Prospero S, Croll D. Comparative genomics analyses of lifestyle transitions at the origin of an invasive fungal pathogen in the

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025

genus *Cryphonectria*. *mSphere*. 2020:5(5):10.1128/msphere.00737–20. https://doi.org/10.1128/mSphere.00737-20.

Stevison LS, Hoehn KB, Noor MAF. Effects of inversions on Within- and Between-Species recombination and divergence. *Genome Biol Evol*. 2011:3:830–841. https://doi.org/10.1093/gbe/evr081.

Stolle E, Pracana R, López-Osorio F, Priebe MK, Hernández GL, Castillo-Carrillo C, Arias MC, Paris CI, Bollazzi M, Priyam A, *et al*. Recurring adaptive introgression of a supergene variant that determines social organization. *Nat Commun*. 2022:13(1):1180. https://doi.org/10.1038/s41467-022-28806-7.

Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat Commun*. 2017:8(1):1–8. https://doi.org/10.1038/s41467-017-01317-6.

Taylor JW, Berbee ML. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia*. 2006:98(6):838–849. https://doi.org/10.1080/15572536.2006.11832614.

Tellier A, Moreno-Gámez S, Stephan W. Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. *Evolution*. 2014:68(8):2211–2224. https://doi.org/10.1111/evo.12427.

Tralamazza SM, Gluck-Thaler E, Feurtey A, Croll D. Copy number variation introduced by a massive mobile element facilitates global thermal adaptation in a fungal wheat pathogen. *Nat Commun*. 2024:15(1):5728. https://doi.org/10.1038/s41467-024-49913-7.

Umen JG. Evolution of sex and mating loci: an expanded view from volvocine algae. *Curr Opin Microbiol*. 2011:14(6):634–641. https://doi.org/10.1016/j.mib.2011.10.005.

Urquhart A, Vogan AA, Gluck-Thaler E. Starships: a new frontier for fungal biology. *Trends Genet*. 2024(12):1060–1073. https://www.cell.com/trends/genetics/abstract/S0168-9525(24)00183-5.

Vittorelli N, Snirc A, Levert E, Gautier V, Lalanne C, De Filippo E, de la Vega RC, Gladieux P, Guillou S, Zhang Y, *et al*. Stepwise recombination suppression around the mating-type locus in an ascomycete fungus with self-fertile spores. *PLOS Genetics*. 2023:19(2):e1010347. https://doi.org/10.1371/journal.pgen.1010347.

Wickham H. *Ggplot2: elegant graphics for data analysis*. Berlin, Germany: Springer Science & Business Media; 2009.

Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res*. 2007:35(Web Server):W265–W268. https://doi.org/10.1093/nar/gkm286.

Yan Z, Martin SH, Gotzek D, Arsenault SV, Duchen P, Helleu Q, Riba-Grognuz O, Hunt BG, Salamin N, Shoemaker D, *et al*. Evolution of a supergene that regulates a trans-species social polymorphism. *Nat Ecol Evol*. 2020:4(2):240–249. https://doi.org/10.1038/s41559-019-1081-1.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Downloaded from https://academic.oup.com/mbe/article/42/2/msaf026/8002013 by Universite Paris Sud user on 04 March 2025