HAL
open science

Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of Cryphonectria parasitica in western Europe

Arthur Demené, Ludovic Legrand, Jerome Gouzy, Robert Debuchy, Gilles Saint-Jean, Olivier Fabreguettes, Cyril Dutech

▶ To cite this version:

Arthur Demené, Ludovic Legrand, Jerome Gouzy, Robert Debuchy, Gilles Saint-Jean, et al.. Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of Cryphonectria parasitica in western Europe. Fungal genetics and biology: FG & B, 2019, 130, pp.122-133. 10.1016/j.fgb.2019.06.002 . hal-02170240

HAL Id: hal-02170240

https://hal.science/hal-02170240v1

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S1087184518302925
Manuscript_442411b6cbaaac4cd9acb445e6924af6

Whole-genome sequencing reveals recent and frequent genetic recombination between clonal lineages of *Cryphonectria parasitica* in western Europe

Running title: Recombinations between *C. parasitica* lineages

Arthur Demené[1], Ludovic Legrand[2], Jérôme Gouzy[2], Robert Debuchy[3], Gilles Saint-Jean[1], Olivier Fabreguettes[1], Cyril Dutech[1]

[1] BIOGECO, INRA, Université de Bordeaux, 69 route d'Arcachon, Cestas F-33610, France.

[2] LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan F-31326, France.

[3] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France

Correspondent author: D. Arthur, E-mail: arthur.demene@u-bordeaux.fr

Postal address: INRA - UMR 1202 BIOGECO - 69 route d'Arcachon - 33610 Cestas – France

Declarations of interest: none

© 2019 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

Abstract

Changes in the mode of reproduction are frequently observed in invasive fungal populations. The ascomycete *Cryphonectria parasitica,* which causes Chestnut Blight, was introduced to Europe from North America and Asia in the 20th century. Previous genotyping studies based on ten microsatellite markers have identified several clonal lineages which have spread throughout western Europe, suggesting that asexuality was the main reproductive mode of this species during colonization, although occasional sexual reproduction is not excluded. Based on the whole-genome sequences alignment of 46 *C. parasitica* isolates from France, North America and Asia, genealogy and population structure analyses mostly confirmed these lineages as clonal. However, one of these clonal lineages showed a signal of strong recombination, suggesting different strategies of reproduction in western Europe. Signatures of several recent recombination events within all the French clonal lineages studied here were also identified, indicating that gene flow is regular between these lineages. In addition, haplotype identification of seven French clonal lineages revealed that emergences of new clonal lineages during colonization were the result of hybridization between the main expanding clonal lineages and minor haplotypes non-sequenced in the present study. This whole-genome sequencing study underlines the importance of recombination events in the invasive success of these clonal populations, and suggests that sexual reproduction may be more frequent within and between the western European clonal lineages of *C. parasitica* than previously assumed using few genetic markers.

Key-Words:

Bayesian inferences, clonal evolution, intra-haploid mating, recombination rates, whole

Introduction

Worldwide, the expansion of a few clonal lineages (i.e., identical or closely related multi-locus genotypes) has often been described regarding populations of invasive pathogenic fungi (e.g., Steimel *et al.,* 2004; Raboin *et al.,* 2007; Goss *et al.,* 2014). Such population structure is usually associated with mainly asexual reproduction in the introduced area; unlike sexual reproduction, which is more frequently reported in the native area of the species (Gladieux *et al.,* 2015). Asexual reproduction is assumed to provide at least two benefits during colonization. First, it allows the rapid multiplication and dispersal of genotypes without any mating partner (Sax and Brown, 2000; Barrett *et al.,* 2008). Second, it protects the population from intensive recombinations with genotypes non-adapted to the sink environment (i.e., migration load; Travis *et al.,* 2005), and preserves the best allelic combinations adapted to some environments. However, the lack of genetic recombination is also known to lead to the accumulation of deleterious mutations (Muller, 1964), as well as a lower adaptability to a new environment through diminished genetic variance (Burt, 2000). Actually, species with a mixed reproductive mode during their life-cycle may efficiently combine the advantages of the two reproductive modes. A recent theoretical study has shown that the most invasive species are those with an asexuality rate close to 0.95 (Bazin *et al.,* 2014). Beyond sexual reproduction, the benefits of genetic recombination may also be a result of intra-genomic re-arrangement (Feschotte, 2008; Hua-Van *et al.,* 2011; Kaessmann *et al.,* 2009; Thon *et al.,* 2006) or non-homologous mitotic recombination (i.e., parasexuality) described for several fungal species (Chuma *et al.,* 2011; Huang, 2014, McGuire *et al*., 2004). All these genetic mechanisms may have a dramatic effect on the evolutionary trajectories of the species and should be more systematically investigated in introduced populations to understand their

64 invasive success (Stuckenbrock and Dutheil, 2018). However, the detection of genetic
65 recombinations can be challenging when they are rare and sparse in the genome, as expected
66 in mainly asexual species. Although a few genetic markers, such as microsatellite loci, are
67 efficient when describing the overall clonal structure of populations (Bruford and Wayne,
68 1993, Steimel *et al.,* 2004; Raboin *et al.,* 2007), their limited number and scattered
69 distribution in the genome make them unsuitable for the accurate detection of recombination
70 signals. The analysis of single nucleotide polymorphism markers (SNPs) on the whole-
71 genome provide a more efficient method to estimate the importance of genetic recombination
72 along the genome, and to identify the main evolutionary mechanisms in invasive pathogenic
73 fungi (Milgroom *et al.,* 2014, Gladieux *et al.,* 2018).
74 The Chestnut blight fungus, *Cryphonectria parasitica*, is a textbook example of the variation
75 of reproductive modes among native and introduced areas (Milgroom *et al.,* 2008; Dutech *et*
76 *al.,* 2010, 2012). Native to eastern Asia, *C. parasitica* was probably introduced to North
77 America at the end of the 19th century most likely from Japan (Milgroom *et al.*, 1996), and it
78 almost caused the extinction of the American chestnut (*Castanea dentata*). In Europe, the
79 pathogen was first reported in 1938 in Italy (Biraghi, 1946), from where it expanded
80 throughout most of southern Europe (Rigling & Prospero, 2017), with a recent colonization
81 further north (Robin *et al*., 2017). Studies based on ten microsatellite loci found multiple
82 introduction events of the pathogen in Europe: North America was the source for
83 introductions into Italy and Switzerland (Dutech *et al.*, 2012, Prospero & Rigling 2012), while
84 Asia was the source for introduction in south-western France (Dutech *et al.*, 2012). These
85 south-western populations could derivate from strains mentioned in 1949 in the northern coast
86 of Spain (Darpoux, 1949; Robin *et al.*, 2009). In contrast to North America, the European

87 outbreak has been less destructive. This is most likely because European chestnuts (*Castanea*
88 *sativa*) are less susceptible to *C. parasitica* than American ones, and due to the presence of
89 *Cryphonectria* hypovirus 1 (CHV-1) that decreases the fungus fitness (Nuss 1992; Milgroom
90 and Cortesi 2004).
91 Contrary to Asian and North American areas (Milgroom *et al.*, 1995; Liu *et al.,* 1996), a
92 strong clonal structure was observed in most European populations (Milgroom *et al.,* 2008;
93 Dutech *et al.,* 2010; Prospero and Rigling, 2012). Since two haploid strains of *C. parasitica*
94 need the different idiomorphic alleles at the mating-type locus (*MAT1-1/MAT1-2*) to
95 reproduce sexually (Mcguire *et al.,* 2001), this clonal structure may be explained by the
96 detection of only one mating type in some eastern European populations (Milgroom *et al.,*
97 2008). In contrast, the frequent report of both mating types in the western populations
98 challenges this hypothesis (Bragança *et al.,* 2007; Robin *et al.,* 2009; Dutech *et al.,* 2010).
99 Using ten microsatellite loci, three genetic clusters were identified in western Europe, each
100 including at least one multilocus genotype highly repeated in numerous sampled populations
101 and defined as clonal lineages (Dutech *et al.,* 2008, 2010). Two of these clonal lineages were
102 located in south-eastern France (RE019, RE092; Fig. 1), and are probably related to the
103 migration of the Italian populations introduced from North America (cluster C1; Dutech *et al.,*
104 2010).The two other clusters located in south-western France are associated with the clonal
105 lineages RE043 and RE053 for C2 and RE028 for C3 (Dutech *et al.*, 2010; Fig. 1), and were
106 likely introduced from Asia. In addition, two other clonal lineages were clustered in C1, and
107 related to North American genetic pool (RE079, RE103; Fig. 1). However, these lineages
108 were sampled only in the south-central France, the only geographical area in southern France
109 where Asian and North American genetic pools are in sympatry. Associated with this

110 secondary contact, RE079 and RE103 might be the result of admixture between these two
111 heterogeneous genetic pools. More recently, in northern France, six of these seven clonal
112 lineages have been reported (RE028 has not been reported) with six additional emerging
113 clonal lineages (H13, H53, H11, H39, H28 and H58; Robin *et al.,* 2017; Fig. 1).
114 If these previous results suggested that these clonal structures are due to asexual reproduction,
115 other observations question the occurrence of sexual reproduction in French populations.
116 First, the presence of sexual structures has been reported in all sampled locations (Robin *et al*,
117 2009), as the two mating-types, which are commonly found in equal proportions in the
118 sampled chestnut stands and often identified within clonal lineages (Dutech *et al.,* 2010).
119 Second, several rare multilocus genotypes, which sometimes differ at a few microsatellite loci
120 from the clonal lineages, have been observed in 32% of the 994 French isolates genotyped
121 previously (Dutech *et al.*, 2010, Robin *et al.*, 2017). Without ruling out somatic mutations,
122 these rare genotypes could originate from crossings between the clonal lineages (Dutech *et*
123 *al.,* 2008, 2010). In the same way, these crosses could explain the emergence of new clonal
124 lineages during expansion of *C. parasitica* to the north (RE079 and RE103 in south-central
125 France, the six new clonal lineages in the northern France). These observations suggest that
126 genetic recombination possibly associated with sexual reproduction may be more frequent
127 than estimated from the genetic structure obtained from the microsatellite analysis.
128 In order to better describe the genetic relatedness between nine of the most frequent French
129 clonal lineages, especially between the earliest southern and the latest northern ones, and try
130 to identify the main mode of evolution in this western European *C. parasitica* population, we
131 conducted a genotyping by sequencing analysis. We hypothesized that genetic recombination
132 has a more important role in the evolution of western European populations than assumed by

133 the clonal structure described in the previous population genetic studies. Using the whole
134 genome sequences of 46 *C. parasitica* isolates (32 isolates from France among the seven
135 southern and two northern clonal lineages and some close genotypes, two isolates sampled at
136 the beginning of the European colonization from the Pyrenees and Italy, ten additional North
137 American and two Asian isolates), we addressed the following questions: 1) On the whole
138 genome, what are the nucleotide diversity and the genetic relationships between isolates
139 belonging to the most frequent multilocus genotypes which were initially defined by
140 analyzing ten microsatellite loci (Dutech *et al.*, 2010; Robin *et al.*, 2017); can these repeated
141 multilocus genotypes be considered as "clonal lineages"? 2) Are there evidences of recent
142 genetic recombinations within the seven clonal lineages studied here, and what are the length,
143 frequency and genetic origin of the detected genetic recombination events? 3) What is the
144 genetic origin of the new clonal lineages in south-central (RE079 and RE103) and northern
145 France (H13 and H53) and how are the Asian and North American genetic pools involved in
146 these emergences? 4) Can we estimate the size of founding population introduced from North
147 America and the timing of emergence of the clonal lineages?

148 ## Materials and Methods

149 **Sequenced isolates**

150 We chose several isolates belonging to and close to the French clonal lineages already
151 genotyped (using ten microsatellite loci) and analyzed in Dutech *et al.* (2010) and Robin *et al.*
152 (2017). Forty-nine isolates were sequenced, including 36 sampled in northern and southern
153 France, one in Italy, and 12 from the two main origins of European introductions (ten from
154 North America and two from Asia; Fig. 1; Table S1 for details). Three French isolates with
155 more than 30% of missing data (SNPs) were finally removed from the analysis. Among the

156 French isolates, 24 have the same multilocus genotype as one of the seven clonal lineages in
157 southern France, five others are different from one to two microsatellite alleles from these
158 clonal lineages (Fig. 1; Table S1 for details). We also sequenced three isolates from two new
159 emerging clonal lineages in northern France (H13 and H53; Robin *et al.,* 2017), and two
160 historical isolates sampled in Italy in 1968 (VG1896, J. Grente unpublished) and in the French
161 Pyrenees in 1975 (VG2106, J. Grente unpublished).

162 **DNA extraction, genome sequencing and assembly**

163 For each isolate, a monospore isolation was performed (Text S1 for details). DNA was
164 extracted from mycelium grown on PDA overlaid with cellophane following Hoegger *et al.*
165 (2000). Eighteen isolates were sequenced using the PGM/Proton Ion Torrent sequencer
166 (Thermo Fisher scientific Inc.) at the platform Genome-transcriptome of Bordeaux (Inra-
167 Université Bordeaux, Bordeaux, France). The 32 other isolates, including one isolate
168 (ABI005) previously sequenced with the PGM technology, were sequenced using Illumina
169 HiSeq2000 technology in paired-end at the Genome and transcriptome GenoToul facilities
170 (INRA, Toulouse, France).

171 Reads obtained from both sequencing technologies were first mapped on the reference
172 genome EP155 (available on http://genomeportal.jgi.doe.gov/Crypa2/Crypa2._home.html)
173 using glint software (Faraud and Courcelle, unpublished; lipm-
174 bioinfo.toulouse.inra.fr/download/glint/). Several regions of the EP155 reference presented no
175 mapped read from these 49 sequenced isolates, questioning the validity of this reference for
176 SNP calling (J. Gouzy, unpublished results). A new reference genome was produced. High-
177 molecular-weight genomic DNA of the French YVO003 monospore (i.e., H53) was extracted
178 following Cheeseman *et al.* (2014) and sequenced using the Pacific Bioscience (PacBio, San

179 Francisco, California, USA) long run sequencing technology (Institute for Genomic Medicine,
180 UCSD, CA, USA). Genome was assembled using PbcR wgs-8.2 with the genome size
181 specified to 45Mb, and polished using pbalign (available on
182 https://github.com/PacificBiosciences/pbalign), Quiver and SMRT analysis v2.3.0
183 (https://github.com/PacificBiosciences/DevNet/wiki/SMRT-View) with default parameters.
184 Contigs were ordered using Mauve v2.4.0 (Darling *et al.,* 2004), guided by the EP155
185 reference genome.

186 **Annotation of protein-coding genes and transposable elements**

187 Gene models were predicted with a fully automated and parallelized pipeline, egn-ep
188 (http://eugene.toulouse.inra.fr/Downloads/egnep-Linux-x86_64.1.4.tar.gz, release 1.2, Text
189 S2 for details). For each of the 46 isolates, a draft genome was constructed using CLCBio
190 with default parameters in order to blast the two mating types alleles sequences on it (*MAT1-1*
191 and *MAT1-2,* first described by McGuire *et al.,* In 2001). Transposable elements (TEs)
192 detection was performed, using the REPET software (Jamilloux *et al.,* 2017). REPET uses
193 two pipelines: first, TEdenovo (Flutre *et al.,* 2011), which build a TE consensus library, and
194 second, TEannot (Quesneville *et al.,* 2005) which annotates TEs in the genome using the
195 classification system proposed by Wicker *et al.* (2007).

196 **Mapping and SNP calling**

197 After mapping the 46 sequenced genomes on the new masked reference genome using glint
198 software, SNP calling was performed using varscan v2.3.7 (Koboldt *et al.,* 2009, 2012) with
199 parameters set as follows: minimum coverage 15 reads, minimum length of the high scoring
200 pair (hsp) ≥90, number of mismatches ≤5, no gap allowed, and only best-scoring hits taken
201 into account. Alternative variants were kept if they were present on a minimum of ten reads

202 and at least once on each of the two strands. Sites with heterozygous SNPs (i.e., alternative
203 variant frequency estimated between 0.25 and 0.75 for one isolate; *C. parasitica* monospore
204 isolates sequenced here are haploid), insertion-deletion sites and non bi-allelic SNPs were
205 removed from the dataset. Polymorphic sites with more than 10% of missing data and
206 scaffolds sized to less than 100kb were removed from this analysis.

207 **Levels of polymorphism, genealogical relationships, clustering and linkage**
208 **disequilibrium between isolates of *C. parasitica in France***

209 Based on the filtered SNP dataset, VCFTOOLS (Danecek *et al.,* 2011) were used to estimate
210 nucleotide diversity (π) and SNP density, on a non-overlaping sliding window counting
211 10,000 nucleotides along the genome. Nucleotide diversity was plotted using R-cran package
212 'qqplot' (Turner 2014). Pairwise linkage disequilibrium (LD) between SNPs was estimated
213 using VCFTOOLS LD statistics after keeping one single isolate of each clonal lineage with
214 the best coverage. Tajima's D was tested on SNPs alignments for different sub-populations
215 and seven clonal lineages using DnaSP v5 (Librado & Rozas, 2009).

216 We estimated the genealogical relationships between sequenced isolates using the neighbor-
217 net algorithm implemented in SplitsTree4 (Huson and Bryant 2006) with the uncorrected P
218 distance (i.e., in our case the proportion of sites at which two sequences differ compared to all
219 the polymorphic sites of the filtered dataset). This phylogenetic network method is used to
220 identify recombination events among haplotypes, incomplete lineages sorting, or homoplasic
221 mutations (Bryant and Moulton 2003).

222 Population structure was estimated using BAPS Version 6 (Corander *et al.,* 2008), a Bayesian
223 method for identifying the most likely number of sub-populations (k number of clusters) in a
224 given population by comparing the allele frequencies among all the polymorphic sites among

225 sub-populations and the whole population. To increase the resolution and to get the
226 substructure, which is hard to detect because of strongly divergent lineages, we used BAPS
227 through a hierarchical clustering process (Cheng *et al.,* 2013). The largest cluster obtained
228 from the first analysis of the whole dataset was used as input for the second analysis. We
229 obtained the best K value describing these two datasets. From this, we used the fixed k option
230 in BAPS with default parameters to perform four independent runs (K = 2,3,4,5) on the whole
231 dataset and five runs (K = 2,3,4,5,6) on the largest cluster.
232 Genealogies used for the recombination detection and BEAST analysis were estimated using
233 RAxML (Stamatakis 2014) with the New rapid hill-climbing algorithm [-f d]; the Generalized
234 Time-Reversible (GTR) substitution model with gamma site heterogeneity model; and other
235 parameters to the default values. Using jmodeltest2, the GTR substitution model was chosen
236 as the most suitable among others for our data (Darriba *et al.,* 2012).
237 **Detection of recombination between and within clonal lineages**
238 We used the Pairwise Homoplasy Index (PHI) test implemented in SplitTree4 to detect the
239 presence of recombination within six clonal lineages in which we had at least two isolates
240 sequenced. Using three different methods, we identified recombination events within five
241 clonal lineages (RE092 excluded because isolates are too genetically divergent; see results).
242 First, we used a home-made method. This fast and well adapted approach to detect
243 recombinations within clonal lineages uses the combined strategies of distance methods and
244 phylogenetic methods (Posada and Crandall, 2001). We measured the nucleotide diversity
245 within each French clonal lineage as described above. As the recombination between two
246 divergent sequences tends to significantly increase nucleotide diversity in the recipient strain
247 compared to the lineage genetic background, we associated the peaks of genetic diversity

248 within the clonal lineages with a putative recombination event. Considering the low average
249 variation within clonal lineages (see Results), a recombining region was assumed when at
250 least two successive 10kb windows with more than one SNP were identified. Regions
251 separated by less than 100kb were grouped as one unique recombining region. Recipient
252 strain and origin of each recombining region were determined using a genealogical tree
253 inferred with RAxML. In addition to this home-made method, we used two other methods to
254 confirm the detected recombinations: fastGEAR (Mostowy *et al.,* 2017), using a Bayesian
255 clustering method to detect dissimilar regions within each clonal lineage, and
256 ClonalFrameML (Didelot and Wilson 2015), using a phylogenetic approach to detect genomic
257 fragments introducing novel polymorphism within each lineage. We used the relative rate of
258 recombination to mutation (R/θ), the average length of DNA imports (δ), and the mean
259 divergence between donor and recipient genotypes (ν) estimated on the four south-eastern
260 clonal lineages by ClonalFrameML, to calculate the relative effect of recombination to
261 mutation (r/m*).*

262 ***Identification of the different haplotypes among the French clonal lineages***

263 We designed a pipeline combining BCFtools, VCFtools and R software (R Development Core
264 Team, 2008) to identify and compare the haplotypic sequences along the genome of the seven
265 southern French clonal lineages and the northern emergent genotype H13. For each of these
266 lineages, we chose one reference isolate for which no recent recombination had been detected
267 with the three methods described above. For the RE092 lineage, the most diverse lineage of
268 this study (see results), we chose three reference isolates (VG_1896, STC36 and YVO006).
269 Using VCFtools and per 10kb window along the genome alignment, we extracted a set of
270 variants between each pair of ten reference isolates and all singeltons were removed. The

271 historical isolate VG_1896, the oldest of this study (sampled in 1968), was used as the
272 reference haplotype. For each 10kb window, a new haplotype was defined when more than
273 one SNP differed from a previously identified haplotype.

274 ***Molecular parameter estimation of the south-eastern clonal lineages***

275 Assuming that genomic regions with identical haplotypes shared between the four south-
276 western and south-central clonal lineages (i.e., 22 isolates) have been inherited from a recent
277 common ancestor, we estimated the divergence times and evolutionary rates of these genomic
278 regions using a tips dated approach with BEAST 1.7 (Drummond and Rambaut 2007;
279 Drummond *et al.,* 2012). From these regions, we removed the remaining recombination
280 regions detected with ClonalFrameML to obtain a core alignment of 11Mb. For efficient
281 posterior estimation of parameters, we chose to minimize the possibility of recombination by
282 subdividing the 11Mb alignment according to the putative chromosomes obtained from the
283 two reference genomes EP155 and YVO003. On the basis of the Mauve alignment (Fig. S1),
284 we chose the four largest putative chromosomes lacking rearrangements between these two
285 reference strains: S1 in EP155 (i.e., MS1-1, MS1-2 and MS1-3 in YVO003), S3 (i.e., MS3-1
286 and MS3-2 in YVO003), S4 (i.e., MS4-1, MS4-2, MS4-3 and MS4-4 in YVO003) and S8
287 (i.e., MS8-1 and MS8-2 in YVO003). We constructed four genealogical trees using RAxML
288 and rooted the bestTree output with the option -f I. Rooted trees were used as the starting tree
289 in BEAST for each alignment. GTR (Generalized Time-Reversible; Lanave *et al.,* 1984)
290 substitution model and an empirical base frequencies model, as well as a strict molecular
291 clock model (Zuckerkandl and Pauling 1965) were set up. As a *C. parasitica* colony can
292 survive on its host for several years, tip dates for the 22 isolates were specified between the
293 sampling date and twenty years ago. As no data are available in Europe on the demography of

294 *C. parasitica*, the coalescent model was set to follow either an exponential population growth
295 or a constant population size. The Markov Chain Monte Carlo (MCMC) length was
296 100,000,000 iterations, thinned every 10,000 to retain 10,000 final sampled trees. BEAST log
297 files were checked using Tracer (Drummond *et al.,* 2012) to control the posterior distributions
298 and the Effective Sample Size (ESS) was used to check the independence of parameters
299 estimation through the chain. The highest posterior density (HPD) 95% given for the
300 parameters is the shortest interval that contains 95% of the posterior probability. From these
301 runs, we chose to discard the first 1,000 sampled trees as burn-in to generate a maximum
302 clade credibility tree with median node heights in treeannotator v. 1.7.5 (Drummon *et al.,*
303 2012).

304 Results

305 *Assembly of the new reference genome, gene and transposable element content*

306 The PacBio sequencing of the strain YVO003 yielded 494,384 reads, N50=12,757 bp
307 (L50=132,812 reads) with an average length of 9,064bp. The assembly produced 35 scaffolds
308 for a total length of 39.3Mb (N50=2.7Mb; L50=6). This new reference genome was shorter
309 than the EP155 reference genome v2 for 4.6Mb with several large genomic regions missing,
310 such as on scaffold 2 (1.5Mb) or scaffold 6 (0.5Mb) and on putative scaffold rearrangements
311 (Fig. S1). The number of predicted genes in this new *C. parasitica* reference genome is
312 12,146 (EP155 genome annotation v2: 11,609) comprising 52.4% of the assembly length with
313 an average gene density of one gene per 3.2kb (EP155: one gene per 3.8kb). 276 complete
314 plus 11 fragmented gene models out of a total of 290 (95.2% and 3.8% respectively) were
315 detected. Using the two *MAT* sequences, we located the *MAT* locus on the RC05 scaffold
316 between 52,138bp and 51,074bp. Except for the clonal lineage RE079 for which all

317 sequenced isolates were *MAT1-2* (including YVO003), all lineages included the two *MAT*
318 alleles (Table S1 for details).

319 The REPET package predicted 968 copies of transposable elements (TEs) from 28 TE
320 families, covering 2.3% of the genome with an average density of one TE per 44.8 kb. The
321 predicted TE copies were classified in 14 DNA transposons, 12 RNA transposons and 2
322 undefined families (Table S2 for details), the 28 families containing from two to 210 copies
323 with an average copies number of 31.4 ($ci_{95\%}$ = 16.2 – 46.6). The 11 putatively active families
324 represented 271 of the total number of copies (878) and 0.7% of the total genome size. The 17
325 other families showed incomplete consensus DNA sequence and may be the signature of
326 ancient TE burst.

327 ***Mapping and SNP calling***

328 All the 46 genome sequences from both sequencing technologies were mapped on the
329 new PacBio reference genome with a coverage of 41.0 to 187.8X (mean = 100.5; Table S1 for
330 details). We identified 118,182 SNPs using the Varscan method. Only three SNPs were
331 detected between the Illumina and the Ion Torrent sequences of the ABI005 isolate after
332 filtering, without considering missing data. This comparison suggests that the two sequencing
333 technologies produced a similar set of SNPs after filtering. Finally, we kept a dataset of 46
334 isolates with 38,592 SNPs (i.e., average 1SNPs per kb), and focused on the 26 scaffolds
335 longer than 100kb. Most of the discarded SNPs were present within TEs and incompletely
336 covered regions, while only 44.0% of SNPs were discarded in the genes (some genes are in
337 fact TEs) compared to 67.3% on the whole genome.

338 ***Levels of polymorphism, genealogical relationships, clustering and linkage disequilibrium***
339 ***between isolates of C. parasitica in France***

340 Nearly half of this polymorphism was detected among the Japanese and Chinese isolates
341 (18,263 SNPs, π=3.02E^{-4}; Table 1) while the ten American isolates showed 13,387 SNPs
342 (π=1.3E^{-4}) with only 2,913 SNPs shared between the two geographical areas. Among the 34
343 European isolates, 26,090 SNPs were identified (π=1.8E^{-4}) mainly observed among the ten
344 isolates introduced from Asia (18,201 SNPs, π=1.75E^{-4}). Only 12,807 SNPs (π=1.07E^{-4}) were
345 observed among the 24 isolates introduced from North America and 4,918 SNPs were shared
346 between the two introductions.

347 The neighbor-net network and the BAPS clustering generated from the 46 *C. parasitica*
348 sequences of SNPs both showed two main genetic clusters containing respectively south-
349 western and some northern French isolates related to Asian ones and south-eastern and other
350 northern French isolates related to North American ones (Fig. 2.a and 2.c). The reticulations
351 of the genetic network showed a greater proportion of shared markers within the North
352 American cluster than within the Asian one, suggesting a higher level of mating or a more
353 recent divergence within the North American populations analyzed. The BAPS analysis
354 indicated that the two south-eastern clonal lineages (RE019 and RE092) were genetically
355 divergent relative to a second French cluster consisting of the two south-central lineages
356 (RE103 and RE079), and to a third one associated with the northern clonal lineage H13 (Fig.
357 2.b).

358 Inside these clusters, French isolates were closely grouped following their clonal lineage
359 relatedness, previously defined using the ten microsatellite locus analysis (Dutech *et al.,*
360 2010). The genetic variability within five clonal lineages (RE019, RE043, RE053, RE079 and
361 RE103) was low with a P-distance comprising between 0.001 ($ci_{95\%}$ = 0-0.002; RE079) and
362 0.014 ($ci_{95\%}$ = 0.014-0.015; RE103) relative to the average P-distance of 0.237 ($ci_{95\%}$ = 0.224-

363 0.250) estimated between two clonal lineages. Reticulations were sometimes detected within
364 these lineages, but only due to few isolates (Fig. 2.c). The south-eastern RE092 clonal lineage
365 was more variable than expected on the basis of microsatellite analysis, with a mean
366 uncorrected P-distance of 0.064 ($ci_{95\%}$ = 0.056-0.073) and more reticulations than other
367 French lineages. It can therefore not be strictly considered as a clonal lineage. However, all
368 the RE092 isolates and close genotypes (RE093 and H68) remained clustered in the same
369 clade in the network and in the same cluster in the BAPS best partitioning (Second analysis,
370 k=6; Fig. 2.b). One of the least divergent pairs of isolates was surprisingly the historical
371 isolate sampled in the Pyrenees in 1975 (VG2106) and a RE043 isolate (BAR002) sampled in
372 2006 in northern France (both introduced from Asia) with only 25 SNPs detected. Tajima's D
373 estimates on the hole genome were non significant (P> 0.05) for nearly all the clonal lineages
374 and sub-populations considered (Table 1), except for RE019 with D=-1.766 (P<0.001)
375 indicating an excess of rare alleles.

376 A high linkage disequilibrium (LD) was estimated between pairs of SNPs separated by
377 1kb in the NA (North American isolates) and FNA (French isolates from North American
378 genetic pool) subsets (r^2=0.84 $ci_{95\%}$=9E^{-3} and 0.91 $ci_{95\%}$=5E^{-3} respectively; Fig. 3). At this
379 range of genomic distances, the estimate was only 0.54 ($ci_{95\%}$=9e^{-3}) for the Global subset. The
380 distances at which the LD was half decayed - a useful indicator of the importance of
381 recombination in a population (Nieuwenhuis and James 2016) - were approximately ~3kb,
382 ~28kb and ~400kb for the three datasets, Global, NA and FNA.

383 ***Detection of recombination between and within clonal lineages***

384 The PHI test did not show significant evidence for recombination in the six French clonal
385 lineages, except for RE092 (*p-value* = 0.0). However, the analysis of polymorphism within

386 the five other lineages showed numerous peaks of diversity in the genome (Fig. S2), allowing
387 us to identify several recombination events, confirmed by the two Bayesian methods
388 FastGEAR and ClonalFrameML (Table 2). The diversity method detected 44 putative
389 recombining regions observed on 21 of the 26 analyzed scaffolds, from which 31 were greater
390 than or equal to 20kb. The average size of these 46 regions was 158kb ($ci_{95\%}$ = 96-220kb) and
391 the largest region was 1,2Mb within the RE103 clonal lineage. Two regions on scaffolds
392 MS8-2 (10kb) and C25 (30kb) each showed signals of recombination within two different
393 clonal lineages: RE053 and RE103, and RE103 and RE079 respectively. One region on the
394 RC05 scaffold showed a signal of recombination within four clonal lineages, which were
395 estimated to be between 40 and 90 kb and to all encompass the mating-type locus. The
396 FastGEAR method detected fewer recombination events (33), but confirmed 30 of the 31
397 largest regions detected with the first method; meanwhile, ClonalFrameML detected 42
398 recombinations (27 of the 31 large recombining regions detected with the first method)
399 including seven smaller than 500bp and 28 large regions greater than 20kb (Table 2).
400 Among the 31 recombination events detected with the diversity method, we identified at
401 least one - and often two - isolates within each clonal lineage having one or more signals of
402 recombination (Fig. 4). Most of these signals of recombination (17/31) involved an exchange
403 between lineages introduced from the same area of origin. Ten recombinations involved
404 isolates from different introductions, mostly (8/10) detected in one isolate from the central
405 RE103 lineage, which had received fragments from the south-western RE053 lineage. Four
406 recombining fragments did not originate from a sequenced French isolate, but were related to
407 the same genetic pool (i.e., Asian or NA) as the recipient strain. Overall, regarding the four
408 clonal lineages of the North American pool (excluding H13), we calculated the relative effect

409 of recombination to mutation as r/m = 9.5. This means that recombination brought 9.5 times
410 more substitutions than mutation within these lineages.

411 ***Identification of the different haplotypes among the French clonal lineages***

412 Using our method of haplotype identification to estimate the proportion of 10kb
413 sequences shared between the clonal lineages (Fig. S3), we estimated that the RE019 lineage
414 and the closest RE092 isolate were genetically close with 79.2% of their genome sharing the
415 same haplotype. The RE079 and RE103 lineages were highly similar with 87.2% of identical
416 haplotypes, but 21.6% of their genome was found to be different from the other sequenced
417 lineages. The H13 lineage included 17.9% of sequences that differed from the other lineages.
418 However, these unknown haplotypes were phylogenetically related to the North American
419 population (results not shown). RE019, RE092, RE079, RE103 and H13 were strongly
420 associated with the North American introduction.

421 The genomic comparison of the five clonal lineages associated with the North American
422 introduction showed that 99.5% of their genome may be reconstructed with only three
423 different haplotypes (Fig. S3). Among these five lineages, 41.5% of the genomes was not
424 variable. However, half of the identical regions (i.e., 20% of the whole genome) was variable
425 in the North American isolates analyzed, revealing a lower level of genetic variability in the
426 French population. In contrast, the three lineages related to the south-western introduction
427 RE043, RE053 and RE028 were more diverse, with only 29.5% of their genome being
428 identical and 43.5% having two haplotypes.

429 ***Molecular parameter estimation of the south-eastern clonal lineages***

430 BEAST was used to reconstruct the history of the four French south-eastern and south-
431 central clonal lineages from when they were introduced to Europe from North America. We

432 chose four DNA fragments with no ancient signal of recombination detected between lineages
433 on scaffolds MS1 (2.4Mb, 29 SNPs), MS3 (1.4Mb, 155 SNPs), MS4 (1.Mb, 158 SNPs) and
434 MS8 (0.7Mb, 182 SNPs). We did not include the H13 lineages in this analysis, because they
435 diverged too much from the others analyzed in the different regions. The RAxML trees of the
436 four fragments showed differing topologies (Fig. S4), and the molecular parameter
437 estimations carried out on MS3, MS4 and MS8 did not converge. The assessment of
438 alignments in these regions revealed several possible explanations for this non-convergence.
439 First, the polymorphism is highly variable between lineages for the same scaffold, with no
440 variation within some lineages and up to 37SNPs. Second, the most recently sampled isolates
441 have nearly identical sequences to the ancestral state, whereas the isolates that were sampled
442 the earliest showed several variations, thus challenging the estimation of the mutation rate. On
443 the MS1 scaffold for which convergence was obtained, the posterior parameters which
444 assumed a constant population size, or an exponential growth rate, gave values consistent with
445 the population history of *C. parasitica* in western Europe. Under the hypothesis of constant
446 population size (the growth rate estimated in the exponential model was close to zero; 0.05,
447 95% highest posterior density (HPD) = -0.02 – 0.13), the effective population size of the
448 south-eastern introduced populations was estimated to be 44 individuals (95% HPD = 11 –
449 95), and the mutation rate per site and per year was $4.7E^{-8}$ (95% HPD = $1.5E^{-8}$ – $8.6E^{-8}$). The
450 Date of the most recent common ancestor (MRCA) of this south-eastern population was 1931
451 (95% HPD: 1874 – 1966), which is concordant with the first European report of *C. parasitica*
452 in 1938 in Italy (Biraghi 1946). The RE092 isolates were not clustered and it was not possible
453 to infer a date of emergence for this lineage (Fig. 5). The emergence of the RE019 lineage
454 was independent of the other lineages and dated to 1983 (95% HPD: 1959 – 2000). It was not

455 possible to conclude which of the RE079 and RE103 lineages emerged first, due to the
456 overlapping 95% HPD of the dates of emergence. However, it seems that they might have
457 diverged as from around 1992 (95% HPD: 1980 – 2000) and separately from the RE092
458 lineage.

459 **Discussion**

460 Genealogical and clustering analyses based on the filtered dataset of 38,592 SNPs
461 confirmed that isolates belonging to six multilocus genotypes (MLGs) highly repeated within
462 French populations are highly similar on the whole genome (H13 and RE028 excluded of this
463 analysis due to the sequencing of only one isolate). The clonal structure may be even stronger
464 than described by the ten microsatellite loci (Dutech *et al.*, 2010; Robin *et al.*, 2017), since
465 several MLGs differing by one or two microsatellite alleles were finally considered to belong
466 to the same clonal lineage (for example the clonal lineage H53 identical to RE079). These
467 observed differences in microsatellite alleles are likely to have resulted from recent mutations
468 of the repeated microsatellite motif. Within these clonal lineages - excluding the recombinant
469 regions identified (see below) - we detected a small number of variations (between 17 and 120
470 different SNPs within each lineage). For example, two isolates of the south-western RE043
471 clonal lineage, which was sampled between 1975 and 2012 (VG2106 and BAR002), differed
472 by only 25 SNPs on the whole genome. The genome-wide linkage disequilibrium decay
473 which was estimated for four lineages from North American genetic pool has a very similar
474 profile to some clonal species. Nieuwenhuis and James (2016) described the most sexual
475 species as having a LD decay lower than 1kb and highly clonal ones with half-decay points at
476 greater than 100kb. Global and NA subsets have a LD decay similar to yeast species which
477 suggest occasional sexual reproduction. By contrast, slow decay of LD of the FNA subset

478 (French lineages and the historical isolate from North American genetic pool) suggested a
479 mainly asexual or highly homogamic reproduction within each clonal lineage of this area. In
480 contrast, for one lineage (RE092), we estimated high polymorphism (5,386 SNPs), a large
481 number of reticulations in the neighbor-net network, and a PHI test indicating frequent
482 genetic recombinations, challenging the assumption of clonal evolution. However, all RE092
483 isolates and some other MLGs (RE093 and H68) have been assigned to the same cluster by
484 the BAPS structuring algorithm, suggesting that recombination events occur preferentially
485 within this lineage. This regional genetic structure in south-eastern France is similar to the
486 results obtained in *Magnaporthe oryzae* by Gladieux *et al.,* (2018), who reported a central
487 sexual lineage and several asexual lineages that have spread throughout the world. This could
488 suggests different colonization strategies for these different lineages in invasive areas.
489 Although clonal evolution was confirmed at the genome scale for most of the French clonal
490 lineages, the signature of recent recombination events between these lineages have been
491 clearly identified using three different methods. Using ClonalFrameML, we estimated that the
492 recombinations identified in the four south-eastern lineages introduced 9.5 times more
493 substitutions than did mutation; thus indicating that even if they are limited in size, these
494 events have important evolutionary consequences at the genome scale. In addition, since the
495 detection of recombination events is based on the difference in genetic distance or
496 phylogenetic branching between adjacent sequences, it is almost impossible to detect a
497 recombination event between two sequences nearly identical. Thus, the estimates presented
498 here are likely to be low, and the effect of recombination may be even greater than identified
499 in this study. Most of these exchanges (87%) occurred between the seven clonal lineages
500 studied here, while 13% involved haplotypes not identified in this study. The genetic

501 mechanisms which cause this recent gene flow between these lineages remains unclear.
502 McGuire *et al.,* (2004) previously reported such exchanges in *C.parasitica* isolates from the
503 field in North America suggesting that they may be caused by non-meiotic crossing over
504 between vegetatively incompatible genotypes (i.e., parasexual recombination; Pontecorvo,
505 1956, Milgroom *et al.*, 2009). Although we cannot rule out this possibility, other results
506 suggest that regular sexual reproduction may be the main mechanism behind this limited
507 genetic introgression within the lineages for two reasons. First, although a small number of
508 isolates were analyzed in this study, we systematically detected one recombination event
509 which encompassed the mating type locus within four clonal lineages and originated from one
510 of the other French lineages. This genomic region would either be exchanged very regularly
511 between these lineages by an undetermined genetic mechanism or would sometimes be
512 exchanged and maintained in the lineages by regular sexual reproduction within the lineages
513 during European colonization. Second, we expected negative values of Tajima's D in
514 asexually spread lineages (Gladieux *et al.*, 2017), and in the case of population expansion
515 (Aris-Brosou & Excoffier, 1996) as for *C. parasitica* populations in Western Europe (Robin
516 & Heiniger, 2001). But, Tajima's D estimates within each lineages were not significantly
517 negative in all lineages but one (RE019). It is likely that the low polymorphism detected
518 within the lineages and the small number of samples within these lineages decreases the
519 power of the Tajima's D test (Ramos-Onsins & Rozas, 2002). More sequenced isolates are
520 needed to understand how mutations are transmitted in these lineages. For now, the most
521 parsimonious hypothesis is the regular occurrence of genetic recombination among isolates
522 assigned to the same clonal lineage. These homogamic crossings would produce an inverse
523 effect than asexuality and population expansion on the Tajima's D, and prevent us from

524 obtaining a clear signal of rare alleles excess usually found in clonal lineages in other fungi
525 species (Gladieux *et al.*, 2017). The occurrence of preferential crossings within lineages (also
526 called intra-haploid mating, Giraud *et al.,* 2006) should be investigated in the future. It could
527 be an important factor in the success of invasion of *C. parasitica* in Europe. This intra-haploid
528 mating has the advantage of preserving the sequence of adapted haplotypes (with the
529 exception of the mating types and some limited introgressions or mutations), as well as of
530 limiting the accumulation of deleterious mutations, the transmission of virus via asexual
531 reproduction (Day, 1977) and the invasion of transposable elements within genomes (Selker,
532 1990).
533 Genealogical, clustering and haplotypes analyses confirmed the two independent
534 introductions of *C. parasitica* in France previously described in Dutech *et al.,* (2012). Our
535 results showed that the Asian and North American genetic pools do not mix in the areas
536 where they coexist (South-central and northern France). The two lineages RE079 and RE103
537 emerging in the south-central France are associated only with the North American gene pool.
538 A similar result is found in the northern France with lineage H13. However, as a part of their
539 genome was divergent from the other south-eastern clonal lineages, these emergences must
540 have involved at least one other genotype not analyzed in this study. From our samples it is
541 not possible to determine if these genotypes were recently introduced from North America, or
542 if they have been present since the early stages of the introduction in low frequencies in
543 French *C. parasitica* populations. We first hypothesized that the new haplotypes came from
544 the RE092 lineage, which is more diverse than others. However, a rarefaction curve carried
545 out on the genetic diversity of this lineage suggests that most of its diversity has already been
546 sampled in this study (Fig S5). We hypothesize that rare genotypes unrelated to the French

547 clonal lineages, and identified in many southern populations (Dutech *et al.,* 2012), may
548 sometimes cross with the clonal lineages to produce these new emerging lineages. The
549 maintenance of rare genotypes is not consistent with several theoretical studies on expanding
550 populations assumed to lead to rapid fixation of haplotypes along the colonization gradient
551 (Excoffier 2004). However, another theoretical study has showed that the combination of the
552 Allee effect (the decrease of effective reproduction due to the limitation of sexual partners)
553 and unfavorable environmental factors in the early waves of the colonization may sometimes
554 lead to a greater diversity in the colonization wave than in the colonization front (Roques *et*
555 *al.,* 2012). Several factors in Europe, such as a temperature gradient (Robin *et al.,* 2017), or
556 the presence of different lineages of the virus CHV-1 (Feau *et al.,* 2014), may be strong
557 selective pressures for the fungus, thus explaining this pattern of genetic diversity in this
558 invasive species.
559 Haplotype identification and estimates of molecular parameters using BEAST helped us to
560 infer the evolutionary history of *C. parasitica* in Western Europe. We identified a maximum
561 of three different haplotypes for reconstructing 99.5% of the whole-genome haplotype phase
562 of the five lineages associated to the North American genetic pool. This result suggested that
563 a very small founding population was introduced to Italy at the beginning of the 20th century.
564 This result is also supported by an estimated effective population size of 44 individuals from
565 the BEAST analysis. Surprisingly, we also estimated the same number of three haplotypes for
566 reconstructing almost all the haplotype phase (98.7%) of the ten North American genomes
567 analyzed in this study. These results suggest that following the first introduction of Asian
568 strains in North America, a few invasive genotypes may have established and subsequently
569 colonized new areas. Selection for the most invasive genotypes in the first steps of invasion,

570 alongside the colonization of subsequent areas by their progenies (known as the invasive
571 Bridgehead effect; Lombaert *et al.,* 2010), may explain the invasive success of *C. parasitica,*
572 which has such a limited genetic variation between the two invaded continents. We were not
573 able to clearly reconstruct the timing of emergence of the clonal lineages originating from the
574 North American introduction. One scaffold, on which no recombination events were detected
575 between the clonal lineages, suggested dates of emergence between the different lineages that
576 were consistent with the successive colonization of French regions. On other scaffolds, the
577 lack of diversity associated with a recent divergence of the lineages may explain the lack of
578 convergence of the model using BEAST. However, a surprising result from some genomic
579 fragments was a lower number of mutations relative to the historical isolate for some recent
580 isolates than for some of them sampled earlier. This result is inconsistent with an asexual
581 evolution and an accumulation of mutations along the clonal branches. It again suggests the
582 effect of sexual reproduction within these clonal lineages that may eliminate some new
583 mutations, especially if they are slightly deleterious. This hypothesis should be investigated in
584 the future.

585 This study shows that a clonal population structure does not necessarily imply that
586 isolates reproduce only clonaly, as it has been described in other fungal species (Henk *et al.*,
587 2012; Milgroom *et al.*, 2014). First, *Cryphonectria parasitica* clonal lineages have regularly
588 exchanged genomic regions. Second, a part of the genome of new emerging clonal lineages
589 during the colonization are not related to the first emerging lineages. Associated with the
590 presence of the two mating types within each clonal lineage, these results suggest that sexual
591 reproduction may be more frequent than assumed from the simple description of the clonal
592 structure in Europe. In this context, the stability of the lineages over time and through

593 colonization in France, with a limited genetic introgression in their genome, raises the
594 question about potential barriers to gene flow among the French clonal lineages. Pre- or post-
595 zygotic barriers may be involved in these limited crossings between lineages. For example,
596 when comparing the two assembled *C. parasitica* genomes (YVO003 and EP155), we found
597 putative mis-assemblies or genomic rearrangements between them; the latter often being
598 involved in reproductive isolation within species (Brown and O'Neill 2010). We hypothesize
599 that chromosomal rearrangements between clonal lineages may lead to partial reproductive
600 isolation in French populations of C. parasitica. Without ruling out selective processes, a
601 limited introgression after a genetic admixture may also be observed when there is
602 asymmetric gene flow between parental sources (Verdu and Rosenberg 2011). Previous
603 studies have shown the presence of spatial clonal patches of *C. parasitica* within European
604 local chestnut stands (Dutech *et al.,* 2008, Hoegger *et al.,* 2000). If the hybrids resulting from
605 crosses of two clonal lineages are mostly dispersed in one of these patches, conditions of
606 unidirectional back-crosses to one of the parental genotypes can be created, leading to a
607 limited introgression in few generations as observed in this study. Therefore, to disentangle
608 selective or neutral factors explaining the genomic structure of this invasive species, we will
609 now need to analyze allelic combinations from a larger number of different genotypes.

610 **Acknowledgments**

611

612 We thank Donald L. Nuss and the genome portal of the Department of Energy (DOE) Joint
613 Genome Institute (JGI) for making the EP155 genome available; Pierre Gladieux, Christophe
614 Lemaire, Aurélien Tellier, Simone Prospero and Jean-Paul Soularue for their numerous
615 comments and corrections on previous versions of the manuscript; Sophie Siguenza for
616 Mauve alignment of genomes; and Erika Sallet for gene annotations. Ion-torrent sequencing

617 was performed at the Genome Transcriptome Facility of Bordeaux and Illumina at the GeT-
618 PlaGe Facility of Toulouse. This work was supported by the ANR-12-ADAP-0009 (Gandalf
619 project) and an innovative project INRA department EFPA. A. Demené was supported by a
620 PhD fellowship from the Ministère français de l'enseignement supérieur – Université de
621 Bordeaux.
622

References

Aris-Brosou, S., and Excoffier, L., 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Molecular Bioliology and Evolution 13, 494–504. https://doi.org/10.1093/oxfordjournals.molbev.a025610

Barrett, S.C.H., Colautti, R.I., Eckert, C.G., 2008. Plant reproductive systems and evolution during biological invasion. Molecular Ecology 17, 373–383. https://doi.org/10.1111/j.1365-294X.2007.03503.x

Bazin, É., Mathé-Hubert, H., Facon, B., Carlier, J., Ravigné, V., 2014. The effect of mating system on invasiveness: Some genetic load may be advantageous when invading new environments. Biological Invasions 16, 875–886. https://doi.org/10.1007/s10530-013-0544-6

Biraghi, A., 1946. Il cancro del castagno causato da Endothia parasitica. Agric. Ital. 7, 1–9.

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

Brown, J.D., O'Neill, R.J., 2010. Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited. Annual Review of Genomics and Human Genetics 11, 291–316. https://doi.org/10.1146/annurev-genom-082509-141554

Bruford, M.W., Wayne, R.K., 1993. Microsatellites and their application to population genetic studies. Current Opinion in Genetics & Development 3, 939–943. https://doi.org/10.1016/0959-437X(93)90017-J

Bryant, D., Moulton, V., 2003. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Molecular Biology and Evolution 21, 255–265. https://doi.org/10.1093/molbev/msh018

Burt, A., 2000. Perspective: sex, recombination, and the efficacy of selection - Was Waismann right? Evolution 54, 337–351. https://doi.org/10.1111/j.0014-3820.2000.tb00038.x

Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A.-L., Ceppi, M., Conseiller, E., Debuchy, R., Malagnac, F., Goarin, A., Silar, P., Lacoste, S., Sallet, E., Bensimon, A., Giraud, T., Brygoo, Y., 2014. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nature Communications 5. https://doi.org/10.1038/ncomms3876

Cheng, L., Connor, T.R., Siren, J., Aanensen, D.M., Corander, J., 2013. Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. Molecular Biology and Evolution 30, 1224–1228. https://doi.org/10.1093/molbev/mst028

Chuma, I., Isobe, C., Hotta, Y., Ibaragi, K., Futamata, N., Kusaba, M., Yoshida, K., Terauchi, R., Fujita, Y., Nakayashiki, H., Valent, B., Tosa, Y., 2011. Multiple Translocation of the AVR-Pita Effector Gene among Chromosomes of the Rice Blast Fungus Magnaporthe oryzae and Related Species. PLoS Pathogens 7, e1002147–e1002147. https://doi.org/10.1371/journal.ppat.1002147

Corander, J., Marttinen, P., Sirén, J., Tang, J., 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9, 539–539. https://doi.org/10.1186/1471-2105-9-539

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 1000 Genomes Project Analysis, 2011. The variant call format and VCFtools. Bioinformatics (Oxford, England) 27, 2156–8. https://doi.org/10.1093/bioinformatics/btr330

Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements 1394–1403. https://doi.org/10.1101/gr.2289704

Darpoux H, 1949. Le chancre du châtaignier causé par l'Endothia parasitica. Document Phytosanitaire No 7. Paris, France: Ministère de l'Agriculture

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9, 772–772. https://doi.org/10.1038/nmeth.2109

Day, P.R., 1977. Double-stranded RNA in Endothia parasitica. Phytopathology 77, 1393–1393. https://doi.org/10.1094/Phyto-67-1393

Denver, D.R., Morris, K., Lynch, M., Thomas, W.K., 2004. High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Nature 430, 679–682. https://doi.org/10.1038/nature02697

Didelot, X., Wilson, D.J., 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. PLOS Computational Biology 11, e1004041–e1004041. https://doi.org/10.1371/journal.pcbi.1004041

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214–214. https://doi.org/10.1186/1471-2148-7-214

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29, 1969–1973. https://doi.org/10.1093/molbev/mss075

Dutech, C., Rossi, J.-P., Fabreguettes, O., Robin, C., 2008. Geostatistical genetic analysis for inferring the dispersal pattern of a partially clonal species: example of the chestnut blight fungus. Molecular Ecology 17, 4597–4607. https://doi.org/10.1111/j.1365-294X.2008.03941.x

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

Dutech, C., Barrès, B., Bridier, J., Robin, C., Milgroom, M.G., Ravigné, V., 2012. The chestnut blight fungus world tour: Successive introduction events from diverse origins in an invasive plant fungal pathogen. Molecular Ecology 21, 3931–3946. https://doi.org/10.1111/j.1365-294X.2012.05575.x

Excoffier, L., 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Molecular Ecology 13, 853–864. https://doi.org/10.1046/j.1365-294X.2003.02004.x

Feau, N., Dutech, C., Brusini, J., Rigling, D., Robin, C., 2014. Multiple introductions and recombination in Cryphonectria hypovirus 1: Perspective for a sustainable biological control of chestnut blight. Evolutionary Applications 7, 580–596. https://doi.org/10.1111/eva.12157

Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. NATURE REVIEWS | GENETICS 9, 397–405. https://doi.org/10.1038/nrg2337

Flutre, T., Duprat, E., Feuillet, C., Quesneville, H., 2011. Considering Transposable Element Diversification in De Novo Annotation Approaches. PLoS ONE 6, e16526. https://doi.org/10.1371/journal.pone.0016526

Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J.L.N., Nhani, A., Chen, L., Terauchi, R., Lebrun, M.-H., Tharreau, D., Mitchell, T., Pedley, K.F., Valent, B., Talbot, N.J., Farman, M., Fournier, E., 2018. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus *Magnaporthe oryzae*. mBio 9. https://doi.org/10.1128/mBio.01219-17

Gladieux, P., Ravel, S., Rieux, A., Cros-Arteil, S., Adreit, H., 2017. Coexistence ofmultiple endemic and pandemic lineages of the rice blast pathogen. MBio9, e01806-17. https://mbio.asm.org/content/9/2/e01806-17

Gladieux, P., Wilson, B.A., Perraudeau, F., Montoya, L.A., Kowbel, D., Hann-Soden, C., Fischer, M., Sylvain, I., Jacobson, D.J., Taylor, J.W., 2015. Genomic sequencing reveals historical, demographic and selective factors associated with the diversification of the fire-associated fungus *Neurospora discreta*. Molecular Ecology 24, 5657–5675. https://doi.org/10.1111/mec.13417

Goss, E.M., Tabima, J.F., Cooke, D.E.L., Restrepo, S., Fry, W.E., Forbes, G.A., Fieland, V.J., Cardenas, M., Grunwald, N.J., 2014. The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. Proceedings of the National Academy of Sciences 111, 8791–8796. https://doi.org/10.1073/pnas.1401884111

Henk, D.A., Shahar-Golan R., Devi, K.R., Boyce, K.J., Zhan, N., Fedorova, N.D., Nierman, W.C., Hsueh, P.R., Yuen, K.Y., Sieu, T.P.M., Van Kinh, N., Wertheim, H., Baker, S.G., Day, J.N., Vanittanakom, N., Bignell, E.M., Andrianopoulos, A., Fisher, M.C., 2012. Clonality despite sex: the evolution of host-associated sexual neighborhoods in the pathogenic fungus Penicillium marneffei. PLoS Pathogens 8 (10), e1002851. https://doi.org/10.1371/journal.ppat.1002851

Hoegger, P.., Rigling, D.., Heiniger, U., Conedera, M., 2000. Genetic structure of newly established populations of Cryphonectria parasitica. Mycological Research 104, 1108–1116.

Huang, X., 2014. Horizontal transfer generates genetic variation in an asexual pathogen. https://doi.org/10.7717/peerj.650

Hua-Van, A., Le Rouzic, A., Boutin, T.S., Filée, J., Capy, P., 2011. The struggle for life of the genome's selfish architects. Biology Direct 6, 19–19. https://doi.org/10.1186/1745-6150-6-19

Huson, D.H., Bryant, D., 2006. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution 23, 254–267. https://doi.org/10.1093/molbev/msj030

Jamilloux, V., Daron, J., Choulet, F., Quesneville, H., 2017. De Novo Annotation of Transposable Elements: Tackling the Fat Genome Issue. Proceedings of the IEEE 105, 474–481. https://doi.org/10.1109/JPROC.2016.2590833

Kaessmann, H., Vinckenbosch, N., Long, M., 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nature Reviews Genetics 10, 19–31. https://doi.org/10.1038/nrg2487

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Ding, L., 2009. VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25, 2283–2285. https://doi.org/10.1093/bioinformatics/btp373

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research 22, 568–576. https://doi.org/10.1101/gr.129684.111

Lanave, C., Preparata, G., Sacone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. Journal of Molecular Evolution 20, 86–93. https://doi.org/10.1007/BF02101990

Librado, P, Rozas, J, 2009. DnaSP v5: a software for comprehen-sive analysis of DNA polymorphism data. Bioinformatics 25,1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lombaert, E., Guillemaud, T., Cornuet, J.-M., Malausa, T., Facon, B., Estoup, A., 2010. Bridgehead Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. PLoS ONE 5, e9743. https://doi.org/10.1371/journal.pone.0009743

Liu, Y.C., Cortesi, P., Double, M. L., MacDonald, W. L., and Milgroom, M. G. (1996). Diversity and Multilocus Genetic Structure in Populations of *Cryphonectria parasitica*. *Phytopathology*, 86, 1344-1351. link

McGuire, I.C., Marra, R.E., Milgroom, M.G., 2004. Mating-type heterokaryosis and selfing in Cryphonectria parasitica. Fungal Genetics and Biology 41, 521–533. https://doi.org/10.1016/j.fgb.2003.12.007

McGuire, I.C., Marra, R.E., Turgeon, B.G., Milgroom, M.G., 2001. Analysis of Mating-Type Genes in the Chestnut Blight Fungus, Cryphonectria parasitica. Fungal Genetics and Biology 34, 131–144. https://doi.org/10.1006/fgbi.2001.1295

Milgroom, M.G, 1995. Population biology of the chestnut blight fungus, *Cryphonectria parasitica.* Canadian Journal of Botany. 73, 311-319. https://doi.org/10.1139/b95-261

Milgroom, M.G., Cortesi, P., 2004. Biological control of Chestnut Blight with hypovirulence: A Critical Analysis. Annual Review of Phytopathology 42, 311–338. https://doi.org/10.1146/annurev.phyto.42.040803.140325

Milgroom, M.G., Jiménez-Gasco, M. del M., Olivares García, C., Drott, M.T., Jiménez-Díaz, R.M., 2014. Recombination between Clonal Lineages of the Asexual Fungus Verticillium dahliae Detected by Genotyping by Sequencing. PLoS ONE 9, e106740. https://doi.org/10.1371/journal.pone.0106740

Milgroom, M.G., Sotirovski, K., Risteski, M., Brewer, M.T., 2009. Heterokaryons and parasexual recombinants of *Cryphonectria parasitica* in two clonal populations in southeastern Europe. Fungal Genetics and Biology 46, 849–854. https://doi.org/10.1016/j.fgb.2009.07.007

Milgroom, M.G., Sotirovski, K., Spica, D., Davis, J.E., Brewer, M.T., Milev, M., Cortesi, P., 2008. Clonal population structure of the chestnut blight fungus in expanding ranges in southeastern Europe. Molecular Ecology 17, 4446–4458. https://doi.org/10.1111/j.1365-294X.2008.03927.x

Mostowy, R., Croucher, N.J., Andam, C.P., Corander, J., Hanage, W.P., Marttinen, P., 2017. Efficient Inference of Recent and Ancestral Recombination within Bacterial Populations. Molecular Biology and Evolution 34, 1167–1182. https://doi.org/10.1093/molbev/msx066

Muller, H.J., 1964. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1, 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Nieuwenhuis, B.P.S., James, T.Y., 2016. The frequency of sex in fungi. Philosophical Transactions of the Royal Society B: Biological Sciences 371, 20150540–20150540. https://doi.org/10.1098/rstb.2015.0540

Nuss, D.L., 1992. Biological control of chestnut blight: an example of virus-mediated attenuation of fungal pathogenesis. Microbiological reviews 56, 561–76. https://doi.org/10.1099/vir.0.025411-0

Pontecorvo, G., 1956. The parasexual cycle in fungi. Annual Review of Microbiology 10, 393–400. https://doi.org/10.1146/Annurev.Mi.10.100156.002141

Posada, D. and Crandall, K., 2001 Evaluation of methods for detecting recombinationfrom DNA sequences: computer simulations. Proceedings of the National Academy of Sciences 98, 13757–13762. https://doi.org/10.1073/pnas.241370698

Prospero, S., Rigling, D., 2012. Invasion Genetics of the Chestnut Blight Fungus Cryphonectria parasitica in Switzerland 102. https://doi.org/10.1094/PHYTO-02-11-0055

Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., Anxolabehere, D., 2005. Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Computational Biology 1, e22. https://doi.org/10.1371/journal.pcbi.0010022

Raboin, L.-M., Selvi, A., Oliveira, K.M., Paulet, F., Calatayud, C., Zapater, M.-F., Brottier, P., Luzaran, R., Garsmeur, O., Carlier, J., D'Hont, A., 2007. Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen Ustilago scitaminea. Fungal Genetics and Biology 44, 64–76. https://doi.org/10.1016/j.fgb.2006.07.004

Ramos-Onsins, S.E., and Rozas,J., 2002. Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution 19(12), 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034

Rigling, D., and Prospero, S., 2017. Cryphonectria parasitica,the causal agent of chestnut blight: invasion history, popula-tion biology and disease control. Moleuclar Plant Pathology 19(1), 7-20. https://doi.org/10.1111/mpp.12542

Robin, C., Heiniger, U., 2001. Chestnut blight in Europe: diversity of Cryphonectria parasitica, hypovirulence and biocontrol. Forest Snow Landscape Resesearch 76, 361–367. https://pdfs.semanticscholar.org/f6fa/8bde1a9f27c4bbee2d3ad8d1fa9048fd7061.pdf

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

Robin, C., Andanson, A., Saint-Jean, G., Fabreguettes, O., Dutech, C., 2017. What was old is new again: thermal adaptation within clonal lineages during range expansion in a fungal pathogen. Molecular Ecology 26, 1952–1963. https://doi.org/10.1111/mec.14039

Roques, L., Garnier, J., Hamel, F., Klein, E.K., 2012. Allee effect promotes diversity in traveling waves of colonization. Proceedings of the National Academy of Sciences 109, 8828–8833. https://doi.org/10.1073/pnas.1201695109

Sax, D.F., Brown, J.H., 2000. The paradox of invasion. Global Ecology and Biogeography 9, 363–371. https://doi.org/10.1046/j.1365-2699.2000.00217.x

Selker, E.U., 1990. Premeiotic Instability of Repeated Sequences in Neurospora crassa. Annual Review of Genetics 24, 579–613. https://doi.org/10.1146/annurev.ge.24.120190.003051

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steimel, J., Engelbrecht, C.J.B., Harrington, T.C., 2004. Development and characterization of microsatellite markers for the fungus Ceratocystis fimbriata. Molecular Ecology Notes 4, 215–218. https://doi.org/10.1111/j.1471-8286.2004.00621.x

Stukenbrock, E.H., Dutheil, J.Y., 2018. Fine-Scale Recombination Maps of Fungal Plant 208, 1209–1229. https://doi.org/10.1534/genetics.117.300502/-/DC1.1

Thon, M.R., Pan, H., Diener, S., Papalas, J., Taro, A., Mitchell, T.K., Dean, R. a, 2006. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus Magnaporthe oryzae. Genome biology 7, R16–R16. https://doi.org/10.1186/gb-2006-7-2-r16

Travis, J.M.J., Hammershøj, M., Stephenson, C., n.d. Adaptation and propagule pressure determine invasion dynamics: insights from a spatially explicit model for sexually reproducing species 16.

Turner, S.D., 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv 005165–005165. https://doi.org/10.1101/005165

941 Verdu, P., Rosenberg, N.A., 2011. A General Mechanistic Model for Admixture Histories of
942 Hybrid Populations 189, 1413–1426. https://doi.org/10.1534/genetics.111.132787
943
944 Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A.,
945 Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A
946 unified classification system for eukaryotic transposable elements. Nature Reviews Genetics
947 8, 973–982. https://doi.org/10.1038/nrg2165
948
949 Zuckerkandl, E., Pauling, L., 1965. Evolutionary Divergence and Convergence in Proteins, in:
950 Evolving Genes and Proteins. Elsevier, pp. 97–166. https://doi.org/10.1016/B978-1-4832-
951 2734-4.50017-6

Data Accessibility Statement

All DNA sequences are available on NCBI with the accession number SRP162210.
Genome, gene annotation and transposable elements: doi:10.25794/reference/UQ0T2ENU

Author contributions

A. D. carried out the data analyses and interpretation and wrote the manuscript. C. D. conceived and supervised the study and wrote the manuscript. L. L. and J. G. worked on the raw data. R. D. and O. F. performed the high molecular height DNA extraction of YVO003. O. F. did the molecular laboratory work. G. S.-J. carried out the monospore isolations.

Tables and Figures

Tables

Table 1. Summary statistics of genomic variations through sliding and non-overlapping 100kb windows for different subset of isolates. † sample size; ‡ number of bi-allelic sites; § measure of nucleotide diversity; ¶ North America. In bold are Tajima's D values significantly different from zero.

Lineage	n†	SNP‡	TajimaD	Pi§	Singleton Number
Asia	2	18263	-	$3.02E^{-04}$	11204
Asian Introduction	10	18201	0.923	$1.75E^{-04}$	3161
RE053	4	583	-0.870	$4.24E^{-05}$	39
RE053 Nr[a]	2	23	-	$6.67E^{-06}$	23
RE043	4	66	0.621	$1.26E^{-05}$	30
RE043 Nr[a]	3	51	-	$1.21E^{-05}$	17
NA¶	10	13387	0.958	$1.30E^{-04}$	564
NA¶ Introduction	24	12807	1.192	$1.07E^{-04}$	393
RE092	6	5386	0.280	$8.71E^{-05}$	52
RE019	7	405	**-1.766**	$1.36E^{-05}$	58
RE019 Nr[a]	5	120	-	$8.80E^{-06}$	40
RE079	5	135	-1.268	$8.33E^{-06}$	55
RE079 Nr[a]	4	51	-	$4.17E^{-06}$	50
RE103	4	1135	-0.837	$4.24E^{-05}$	116
RE103 Nr[a]	2	17	-	$7.08E^{-06}$	17

971 **Table 2.** Recombination statistics of events detected within five clonal lineages with the three
972 methods used: Nucleotide diversity: method using nucleotide diversity within clonal lineages
973 and maximum likelihood genealogies, FastGEAR (Mostowy *et al.*, 2017) and
974 ClonalFrameML (Didelot and Wilson, 2015). Largest recombinations are larger than 20kb. †
975 kilo base pairs; ‡ 95% confidence interval.
976

Method	Nucleotide diversity	FastGEAR	ClonalFrameML
Recombination count	44	33	42
Largest recombinations count	31	30	28
Size of largest recombinations (kb†)	1190	5550	1190
Mean size of largest recombinations (kb†) ($CI_{95\%}$‡)	230 (76)	753 (483)	220 (93)
Mean size of smallest region (kb†)	10	3	0.011

977

Figure captions

Fig. 1. Distribution of the multilocus genotypes (MLGs) of *Cryphonectria parasitica* isolates sampled in 6 French subpopulations (following Robin *et al.* 2017 classification : South-western France = S1 + S2, South-central France = S3, South-eastern France = S4, North-western France = N1, North-central France = N2 + N3, North-eastern France = N4). Colors represent the most frequent MLGs and white represents other MLGs. Pie charts were constructed using data from 583 isolates genotyped (Southern France, Dutech *et al.*, 2010) and 411 isolates genotyped (Northern France, Robin *et al.*, 2017) for ten microsatellite loci. Stars shows the affiliation to sub-populations and MLGs of each of the 32 French isolates sequenced and analyzed in this study. 1938 is the first official report of *C. parasitica* in Italy (Darpoux, 1949), 1949 is the first mention of *C. parasitica* on *Castanea crenata* trees in the northern coast of Spain (Darpoux, 1949). *C. parasitica* was reported in all but one sample sites in the survey carried out in 1996 and 1997 (de Villebonne 1998, Robin *et al.*, 2017)

Fig. 2. Population subdivision defined by BAPS from the two hierarchical analysis and distance-based tree (Neighbor-joining) from SplitsTree4. Circles represent the partition obtained with different k values in the two BAPS analysis and color of each circle describes the cluster membership of each isolate. K fixed values are shown on the circles. Color patches define the clonal lineages: RE028 in yellow, RE043 in purple, RE053 in green, RE092 in grey, RE019 in blue, RE079 in orange, RE103 in red and H13 in turquoise. Names in dark blue are the ten North American isolates and the historical isolate introduced from North America is in light blue. Names in dark red are the two Asian isolates and the historical isolate introduced from Asia is in light purple.
a) BAPS analysis with all isolates (First analysis) and b) BAPS analysis with only the North American origin isolates (Second analysis).
c) Neighbor-net network of the 46 isolates of *C. parasitica*, based on 38,592 SNPs estimated from SplitsTree4 with uncorrected P distance. Potential recombinations are shown by reticulations. Except for the internal reticulations, nodes showed bootstrap values greater than 0.95 (data not shown for clarity).

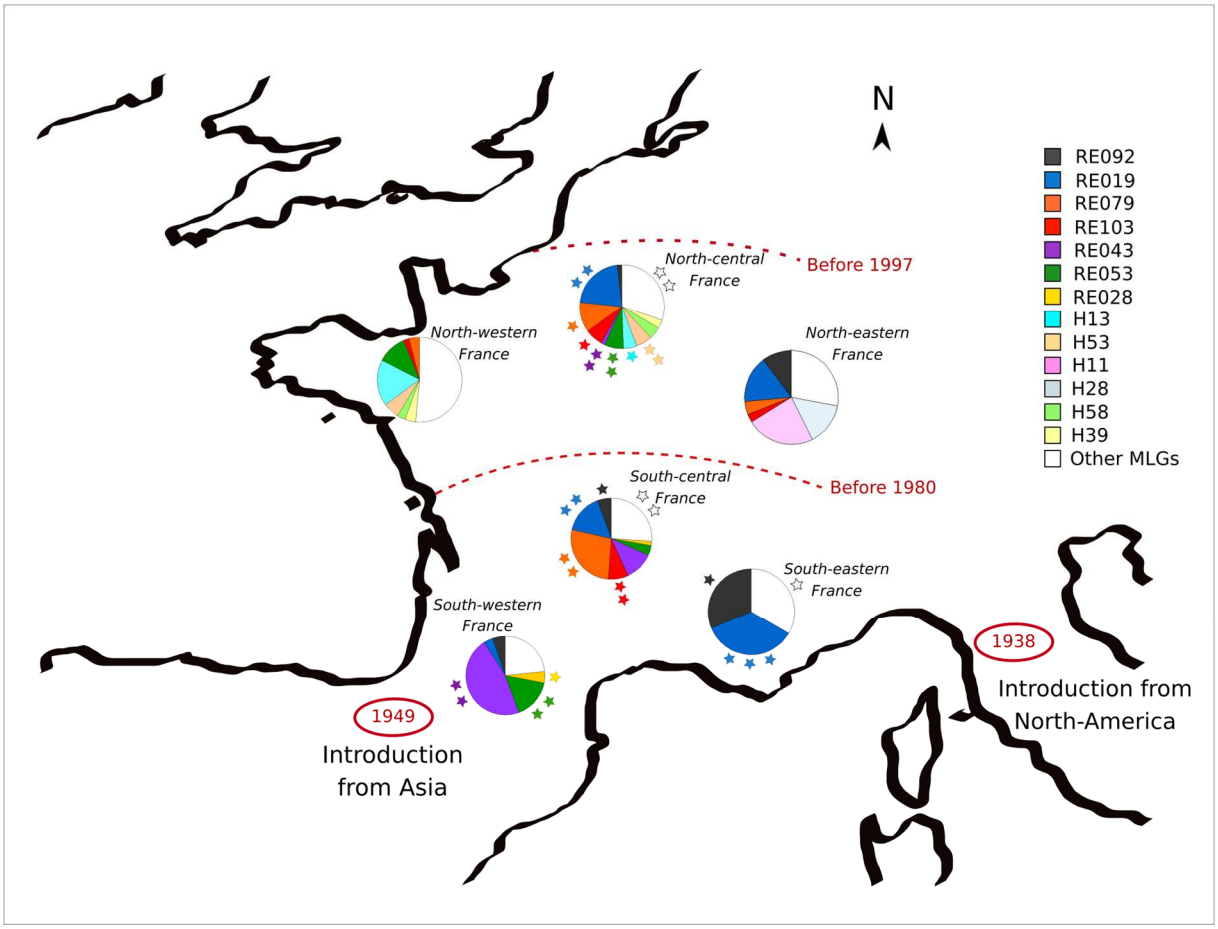
Fig. 3. Log-linear plot of linkage disequilibrium (r^2) according to the distance between nucleotides from 1 to 500kb in three groups of *C. parasitica* isolates: global subset (two Asian isolates, ten North American, five French including the historical Italian isolates from the North American genetic pool) in green, NA subset (ten North American isolates) in purple and FNA (five French including the historical Italian isolates from the North American genetic pool) in blue. Change of the r^2 along the genome is represented using an estimated smoothed curve. Arrows indicate the half-decay values of LD of its maximum estimated value.

Fig. 4. Genomic distribution of the 31 major recombination regions (>20kb) detected using the nucleotide diversity method of detection of recombinations within the five clonal lineages studied here. Colors of the regions of introgression correspond to the donor lineage of the fragment. Unknown origins are in white.

1023 **Fig. 5.** Maximum clade credibility trees of the 2.4Mb alignment located on the scaffold S1 of
1024 *C. parasitica*. Each internal node is labeled with the posterior probability of the robustness of
1025 the corresponding clade. The blue bars illustrate the extent of the 95% highest posterior
1026 density intervals for the node age. The x-axis scale is in years.

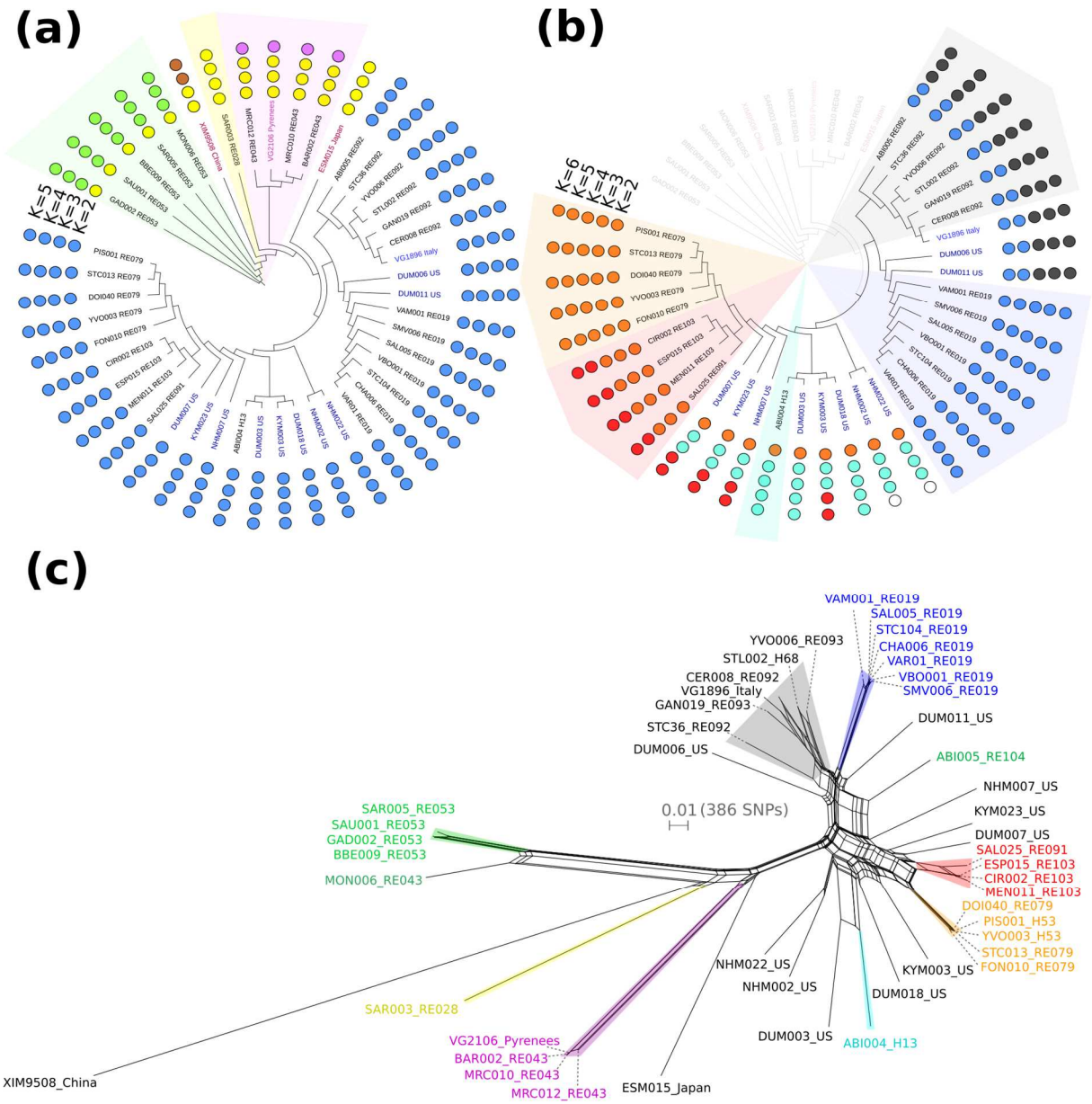
1027 **Fig. 1.** ***(Colors should be used in print)***

1028

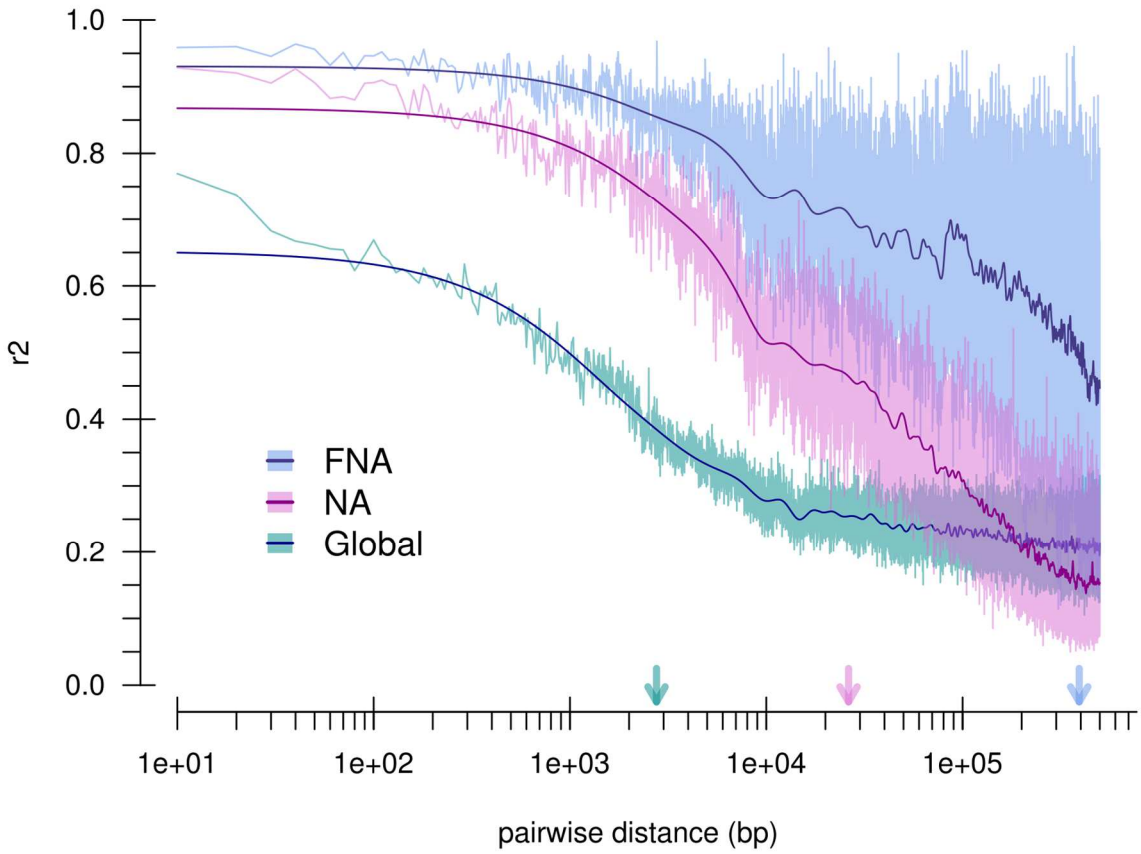


1029

Fig. 2. ***(Colors should be used in print)***

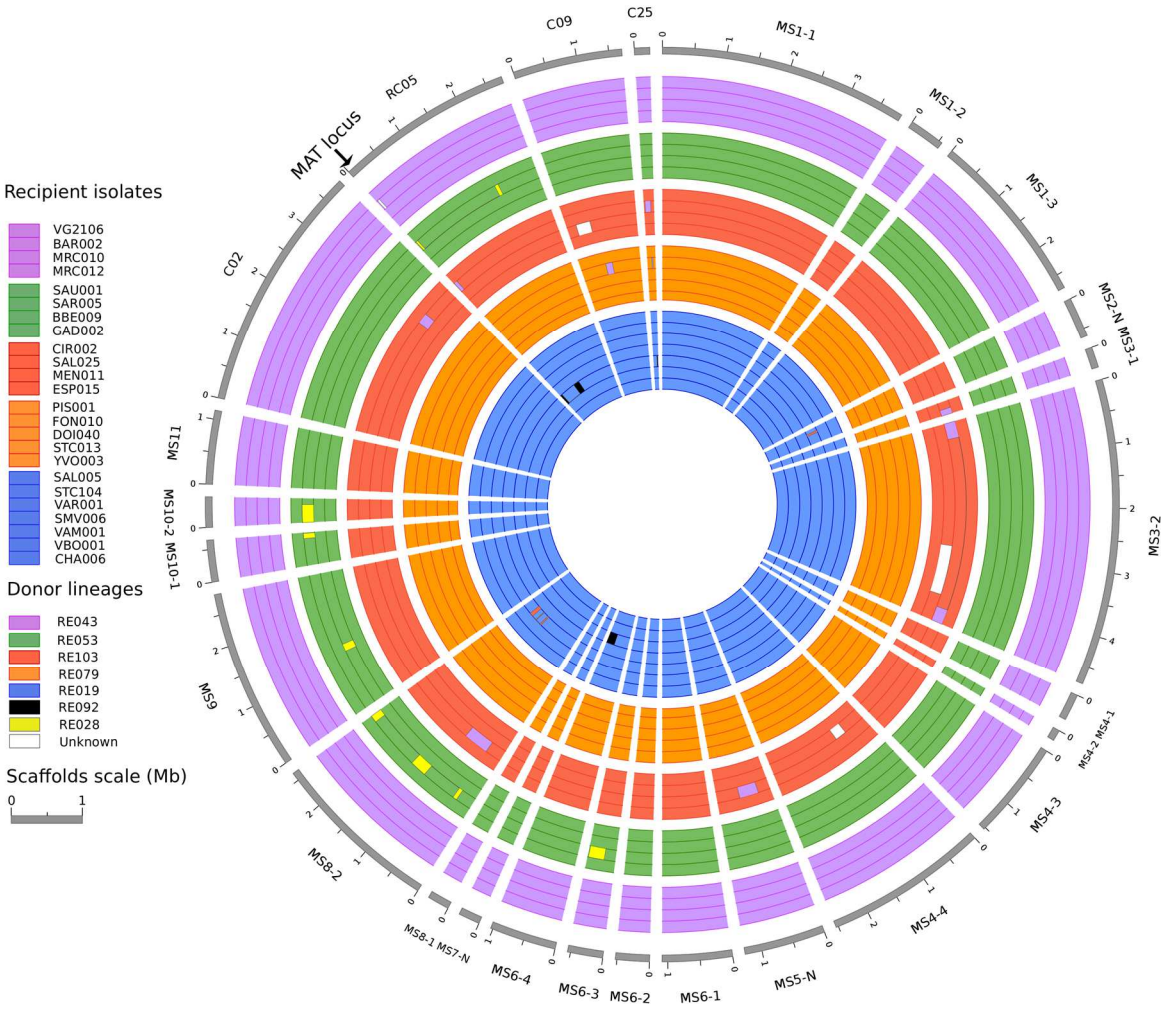


1033 **Fig. 3.** ***(Colors should be used in print)***
1034



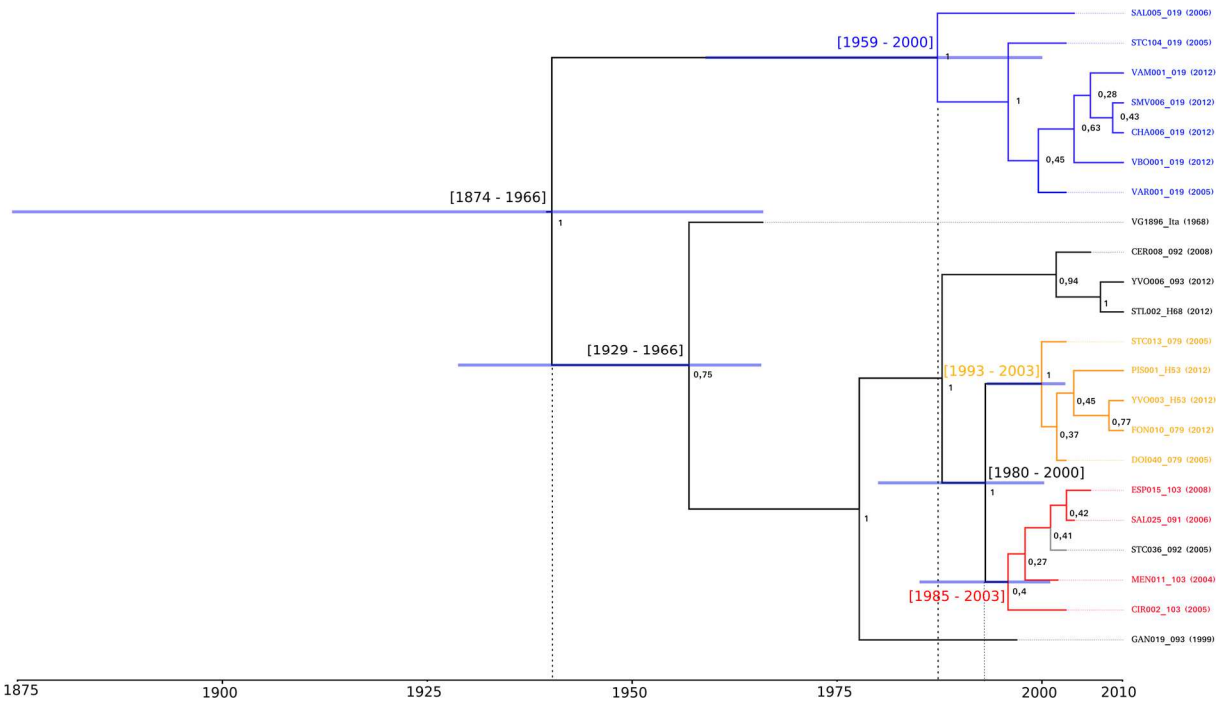
1035

1036 **Fig. 4.** ***(Colors should be used in print)***
1037



1038 **Fig. 5.** ***(Colors should be used in print)***

1039



1040

1041 **Supporting information for online publication**
1042
1043 **Text S1.** Protocol of *C. parasitica* monospore isolation.
1044
1045 **Text S2.** Annotation of protein coding genes
1046
1047 **Table S1.** Sequencing information and mating-type of the 46 *Cryphonectria parasitica*
1048 isolates used in this study. † Number of different microsatellite allele(s) from the closest
1049 French clonal lineage; ‡ Coverage of sequencing in X; § Mating-type allele MAT1-1 and
1050 MAT1-2; ¶ North America.
1051
1052 **Table S2.** Summary results of TEdenovo and TEannot pipeline (REPET Package). REPET
1053 notation refer to the Wicker classification (Wicker 2007): the first letter refers to the initial of
1054 the class (retrotransposons or DNA transposons), the second to the Order (L for long terminal
1055 repeat (LTR), T for terminal inverted repeat (TIR)) and the last to the Super-family. X
1056 signifies an undetermined classification.
1057
1058 **Figure S1.** Comparison of the YVO003 (top) and EP155 (bottom) genomes obtained using
1059 progressive Mauve software (Darling *et al.*, 2004). The colors represent the 35 scaffolds
1060 defined in YVO003 denovo assembly. The potential structural rearrangements are indicated
1061 by the features that link both genome representations. Black lines represent the scaffolding of
1062 the two reference strains. Sizes of the two reference genomes are shown on the right.
1063
1064 **Figure S2.** Plot of the nucleotide diversity (π) calculated by 10kb windows along the 26
1065 major scaffolds within the five studied French clonal lineages: RE019 (in blue), RE103 (in
1066 red), RE053 (in green), RE079 (in orange) and RE043 (in purple). Gray line represents the
1067 global nucleotide diversity (46 isolates).
1068
1069 **Figure S3.** The haplotypic patterns of ten *C. parasitica* French isolates based on 23,240 SNP
1070 (singletons removed) and given by 10kb windows along the genome. Color of each vertical
1071 line defines the frequency of the haplotype for the 10 isolates analyzed, and length of the
1072 sequence is represented by horizontal rectangles. Frequency of each haplotype was defined
1073 distinctly for the two introductions separated by the red line. White haplotype represents the
1074 most frequent haplotype found either in the south-eastern introduction, or in the south-western
1075 introduction. Black ones represent the second most frequent haplotype, blue the third and red
1076 the rarest.
1077
1078 **Figure S4.** RAxML genealogical trees used as starting tree for the BEAST analysis of the
1079 four fragments with no detected recombination using ClonalFrameML. a) MS1 b) MS3 c)
1080 MS4 d) MS8
1081
1082 **Figure S5.** Non-linear regression of the number of SNPs discovered as a function of the
1083 number of isolates of line RE092 considered between 1 and 6. The red line is the asymptote
1084 estimated by R for x[number of RE092 isolates]=+inf; it represents the supposed maximum
1085 genetic diversity that can be found in this line if an infinite number of individuals are
1086 sequenced. The tick named H13 shows the number of SNPs reached when comparing the six

1087 RE092 isolates and the H13 isolate. The red dotted lines represent the confidence intervals at
1088 +2.5% and -2.5%.