UC Berkeley
UC Berkeley Previously Published Works

Title
Genome Sequence of the Chestnut Blight Fungus Cryphonectria parasitica EP155: A Fundamental Resource for an Archetypical Invasive Plant Pathogen.

Permalink
https://escholarship.org/uc/item/2nm816dr

Journal
Phytopathology, 110(6)

ISSN
0031-949X

Authors
Crouch, Jo Anne
Dawe, Angus
Aerts, Andrea
et al.

Publication Date
2020-06-01

DOI
10.1094/phyto-12-19-0478-a

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: A fundamental resource for an archetypical invasive plant pathogen

Jo Anne Crouch[1]*, Angus Dawe[2]*, Andrea Aerts[3], Kerrie Barry[3], Alice C.L. Churchill[4], Jane Grimwood[5], Bradley I. Hillman[6], Michael G. Milgroom[4], Jasmyn Pangilinan[3], Myron Smith[7], Asaf Salamov[3], Jeremy Schmutz[3,5], Jagjit S. Yadav[8], Igor V. Grigoriev[5,9], and Donald L. Nuss[10,11]

Affiliations

[1] Mycology and Nematology Genetic Diversity and Biology Laboratory, United States Department of Agriculture, Agricultural Research Service, 10300 Baltimore Avenue, Building 010A, Beltsville, MD, U.S.A.

[2] Department of Biological Sciences, Mississippi State University, 295 Lee Boulevard, Mississippi State, MS, U.S.A.

[3] United States Department of Energy Joint Genome Institute, Walnut Creek, CA, U.S.A.

[4] School of Integrative Plant Science, Plant Pathology and Plant-Microbe Biology Section, Cornell University, Ithaca, NY, U.S.A.

[5] HudsonAlpha Institute for Biotechnology, Huntsville, AL, U.S.A.

[6] Department of Plant Biology, Rutgers University, 59 Dudley Road, New Brunswick, NJ, U.S.A.

[7] Department of Biology, Carleton University, 1125 Colonel by Drive, Ottawa, ON, Canada.

[8] Environmental Genetics and Molecular Toxicology Division, Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH, U.S.A.

[9] Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, U.S.A.

[10] Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD, U.S.A

[11] Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV, U.S.A

***** Authors for correspondence: Jo Anne Crouch, Mycology and Nematology Genetic Diversity and Biology Laboratory, United States Department of Agriculture, Agricultural Research Service, 10300 Baltimore Avenue, Building 010A, Beltsville, MD, U.S.A, (301) 504-6922, joanne.crouch@usda.gov; Angus Dawe, Department of Biological Sciences, Mississippi State University, 295 Lee Boulevard, Mississippi State, MS 39762, U.S.A. (662) 325-7577, dawe@biology.msstate.edu

48 **Abstract**
49 *Cryphonectria parasitica* is the causal agent of chestnut blight, a fungal
50 disease that almost entirely eliminated mature American chestnut from
51 North America over a 50-year period. Here we formally report the genome of
52 *C. parasitica* EP155 using a Sanger shotgun sequencing approach. After
53 finishing and integration with SSR markers, the assembly was 43.8 Mb in 26
54 scaffolds (L50=5; N50=4.0Mb). Eight chromosomes are predicted: five
55 scaffolds have two telomeres and 12 scaffolds have one telomere sequence.
56 A total of 11,609 gene models were predicted, of which 85% show
57 similarities to other proteins. This genome resource has already increased
58 the utility of a fundamental plant pathogen experimental system through
59 new understanding of the fungal vegetative incompatibility system, with
60 significant implications for enhancing mycovirus-based biological control.
61

62 **Key words**
63 Chestnut blight, forest pathology, mycology, mycoviruses, transposable
64 elements, vegetative incompatibility
65

66 **Introduction**
67

68 Plant disease epidemics have had profound ecological and economic
69 consequences and significantly influenced human history. One of the most
70 momentous disease epidemics was chestnut blight, which completely
71 changed the ecological landscape of the hardwood forests of the eastern
72 United States during the 20th century (reviewed in Anagnostakis, 1988). The
73 causal agent, the ascomycete fungus *Cryphonectria parasitica*, found a very
74 susceptible host in the American chestnut tree, *Castanea dentata*, following
75 its introduction into North America on nursery stock of resistant Asian
76 chestnut tree species. The resulting disease epidemic, first identified in 1903
77 in the Bronx Zoo, spread rapidly, resulting in the destruction of an estimated
78 four billion mature American chestnut trees in the following 50 years
79 (Anagnostakis, 1988). Although the root systems of infected trees often
80 survive and re-sprout, since the new sprouts remain susceptible to endemic
81 *C. parasitica*, the once dominant tree now survives throughout its former
82 natural range primarily as an understory shrub (Dalgleish et al. 2015).
83 Although *C. parasitica* can colonize other members of the family
84 Fagaceae, canker formation and mortality are restricted to North American
85 and European chestnut (*Castanea dentata* and *C. sativa*, respectively) and
86 American chinkapin trees (*C. pumila*; Dallavalle and Zambonelli, 1999). Asian
87 chestnut species (*Castanea crenata* and *C. mollissima*) are resistant to
88 chestnut blight, presumably due to co-adaptation of the two organisms, and
89 provide the basis for a resistance back-cross breeding program
90 (Anagnostakis, 2012). The mechanisms underlying the ability of *C. parasitica*
91 to effectively penetrate defense barriers and rapidly expand in the cambium
92 tissues of susceptible hosts remain ill defined, with no identified role for
93 toxins, specific secondary metabolites or hydrolytic enzymes.

94 The discovery of a group of RNA viruses, now classified in the family
95 Hypoviridae (hypoviruses), that reduce the virulence of *C. parasitica* (causing
96 hypovirulence), stimulated intensive research into the potential of using
97 fungal viruses for the biological control of fungal diseases (reviewed in Dawe
98 and Nuss, 2001; Hillman and Suzuki, 2004; Nuss 1992; 2005). Subsequent
99 epidemiological, population genetic and molecular studies have established
100 the chestnut/*C. parasitica*/hypovirus pathosystem as the textbook example
101 of both the consequences of accidental introduction of an exotic organism
102 and of hypovirulence-mediated biological control of fungal pathogens. The
103 use of hypovirulent *C. parasitica* strains to treat individual disease cankers is
104 highly effective, and hypovirulence provides a level of blight control in some
105 locations in Europe (Heiniger and Rigling, 1994). However, attempts to
106 introduce hypovirulent strains into North American forest ecosystems to
107 control chestnut blight has been problematic, primarily due to the inability of
108 the introduced hypoviruses to spread through the fungal population
109 (Milgroom and Cortesi 2004). The use of severe hypovirus strains that reduce
110 *C. parasitica* ecological fitness, as well as virulence, is one factor that
111 contributes to this poor performance (Milgroom and Cortesi, 2004). A second
112 factor that limits hypovirus spread in North American forests is a diverse self-
113 nonself fungal recognition system termed vegetative incompatibility (*vic*),
114 that regulates the ability of *C. parasitica* strains to undergo anastomosis
115 (Milgroom and Cortesi, 2004). Like mycoviruses in general, hypoviruses lack
116 an extracellular phase to their life cycle and spread to virus-free strains in a
117 population by fusion of the hyphae (anastomosis) (Van Alfen et al. 1975).
118 However, the recent identification, genetic characterization, and systematic
119 disruption of the *C. parasitica* genes that regulate the *vic* system has
120 provided a promising new opportunity for overcoming this major barrier to
121 hypovirus dissemination (Zhang and Nuss 2016; Stauder et al. 2019).
122 In this paper, we report the *C. parasitica* strain EP155 genome
123 sequence as an important resource for elucidating the genomic basis for the
124 selective pathogenicity, niche-associated evolution and virus-based
125 biological control of this classic forest pathogen. This genome resource
126 provides a valuable source of data to enable ongoing and future research of
127 this important plant-associated fungus.
128
129 **Materials and Methods**
130
131 **Genome sequencing and assembly.** *Cryphonectria parasitica* isolate
132 EP155 (ATCC 38755), an orange-pigmented, virulent, hypovirus-free strain
133 (vc type EU-5, *MAT1-2*) was originally isolated in 1977 from a canker on
134 *Castanea dentata* in Bethany, CT, U.S.A. (Anagnostakis and Day 1979). The
135 fungus was grown on potato dextrose agar overlaid with cellophane at room
136 temperature on the laboratory bench for approximately 7 d; mycelium and
137 conidia were scraped from the cellophane using a sterile razor blade, allowed
138 to air dry under a laminar flow hood, then pulverized using liquid nitrogen in
139 a mortar and pestle. DNA was extracted as described by Choi et al. (2012).

140 All sequencing reads were collected with standard Sanger sequencing
141 protocols on ABI 3730XL capillary sequencing machines (ThermoFisher
142 Scientific, Waltham, MA) at the U.S. Department of Energy Joint Genome
143 Institute in Walnut Creek, CA. The genome was sequenced from fosmid and
144 plasmid libraries using a Sanger whole-genome shotgun approach. Three
145 different-sized libraries were used as templates for the plasmid/fosmid
146 subclone sequencing process and both ends were sequenced as follows:
147 332,747 reads from a 2.3 kb-sized plasmid library, 265,247 reads from a 6.8
148 kb sized plasmid library, and 107,327 reads from a 39.3 kb-sized fosmid
149 library. Sequence reads were assembled using a modified version of Arachne
150 v.20071016 (Jaffe et al. 2003) with parameters maxcliq1=100,
151 correct1_passes=0, and BINGE_AND_PURGE=True. This produced 42 scaffold
152 sequences, with N50 of 4.0 Mb, 294 contigs with a 333.9 Kb N50, and a total
153 scaffold size of 44.1 Mb.
154
155 **Finishing and map integration.** The initial whole-genome shotgun
156 assembly was broken down into scaffolds and each scaffold piece was
157 reassembled with Phrap (www.phrap.org) and manually improved and
158 finished using Consed (Gordon 2004). All low-quality regions and gaps were
159 targeted with computationally selected Sanger sequencing reactions
160 completed with 4:1 BigDye terminator: dGTP chemistry (ThermoFisher
161 Scientific). These automated rounds included walking on 2.3kb and 6.8kb
162 plasmid subclones using custom primers (4,526 primers were selected).
163 Following completion of the automated rounds, a trained finisher manually
164 inspected each assembly. Reactions were manually selected to improve the
165 genome. Remaining gaps and hairpin structures were resolved by generating
166 small insert shatter libraries of 6.8kb-spanning clones (Grigoriev et al. 2014).
167 Five fosmid clones were shotgun sequenced and finished to fill large gaps
168 and resolve larger repeats. All these sequencing reactions were generated
169 using Sanger long-read technology. Each assembly was validated by an
170 independent quality assessment. This examination included a visual
171 examination of subclone paired ends and visual inspection of high-quality
172 discrepancies and all remaining low-quality areas. The EP155 assembly was
173 further refined with the aid of the *C. parasitica* genetic linkage map
174 constructed from a cross of Japanese *C. parasitica* isolate JA17 and Italian
175 isolate P17-8 (Kubisiak and Milgroom, 2006) that was upgraded by the
176 addition of 141 single sequence repeat (SSR) markers mined from the EP155
177 genome sequence using Primer3 (Rozen and Skaletsky 2000). Allele data for
178 96 ascospore progeny of the JA17 X P17-8 cross were collected for 60
179 polymorphic EP155-derived SSR markers located at the terminal ends of the
180 EP155 scaffolds. Finished segments (33 scaffolds with 34 contigs) were
181 localized and ordered into pseudomolecules using a 30-marker map with
182 seven joins to form the final assembly. Summary statistics were generated
183 using the JGI Annotation Pipeline, and genome completeness was assessed
184 using BUSCO v3.02 (Simão et al. 2015).
185

186 **Gene models and functional predictions.** EP155 gene models were
187 predicted and annotated using the JGI Annotation Pipeline, which combined
188 homology-based, *ab initio* and transcriptome-based gene predictors
189 predictions (Dawe et al. 2003; Grigoriev et al. 2014, Kuo et al 2014; Shang et
190 al. 2008) using the JGI EST pipeline. Putative protein domains were identified
191 by querying against a local InterProScan database (Jones et al. 2014). The
192 mitochondrial genome was annotated using MITOS v.2 (Bernt et al. 2013).
193 Scaffold regions enriched in genes of similar function were identified
194 using a hypergeometric test ($P > 0.01$, with multiple testing) over a sliding
195 window consisting of 60 ORFs with 10-ORF increments. ORFs were identified
196 in each window that shared the same KOG definition, InterPro label and/or
197 GO label. The probability that ORFs share the same label and co-occur in the
198 same window was calculated, and significantly enriched clusters were
199 identified ($P \leq 0.01$). Tested features included InterPro annotation, Gene
200 Ontology and KOG annotation. For the latter two categorical tests, we tested
201 each category from the leaf to the root node.
202 Putative cytochrome P450 monooxygnases (P450s) were identified
203 through BLAST searches for the conserved P450 signature domains, namely,
204 the oxygen-binding motif and the heme-binding motif. P450s that showed
205 both domains were considered authentic P450s. Incomplete P450 sequences
206 were manually annotated to deduce the full-length sequence and grouped
207 under 'authentic P450s'. The P450s that showed only one of the two
208 signature domains were considered as 'tentative P450s'. Identified P450s
209 were then classified into CYP families and subfamilies based on the existing
210 nomenclature criteria of > 40% nucleotide similarity for assigning a family
211 and > 55% for a subfamily. The P450 families were further grouped into
212 clans. These clan, family, subfamily groupings were based on the
213 classification criteria recommended by the International P450 Superfamily
214 Nomenclature Committee. P450s that could not be assigned to any known
215 clan based on the existing classification scheme were assigned to an
216 appropriate clan(s) based on their relative position in the phylogenetic tree.
217 P450omes of other fungi available at the cytochrome P450 webpage
218 (http://drnelson.uthsc.edu/CytochromeP450.html) were used for a
219 comparative analysis.
220 Protein localizations were predicted using the classical secretion
221 pipeline Secretool (Cortázar et al. 2014). The predicted secretome was
222 evaluated for candidate effector genes using the EffectorP v2.0 prediction
223 tool in combination with EffectorP v1.0, a stringent approach that is
224 recommended to limit false positives to less than 8% (Sperschneider et al.
225 2018).
226 Putative heterokaryon incompatibility genes were identified through a
227 combination of BLASTp searches and syntenic comparisons with the
228 *Neurospora crassa* genome (Galagan et al. 2003). The following
229 heterokaryon incompatibility proteins were used to query the EP155
230 assembly: generic HET domains (PF06985.2), *het* proteins from *Neurospora*
231 *crassa* (*het-c*, *het-6*, *pin-c, tol, un24)* and *het* proteins from *Podospora*

232 *anserina* (*het-c*, *het-D/het-E*, *het-s*). Several motifs are also associated with
233 some *het* genes and may function in programmed cell death; therefore we
234 also searched for NACHT (PF05729.3) and WD40 repeat (PF00400.22)
235 domains, both of which are found associated with HET domains in *P. anserina*
236 *het-D/het-e* (Paoletti et al. 2007). Because of the large number of these
237 motifs associated with other functions, we only searched for them by
238 comparing motifs in regions adjacent to ORFs with HET domains. We also
239 looked for synteny in *C. parasitica* to the genomic regions of *N. crassa* that
240 contain *het* genes. This was done simply using the VISTA tracks in the JGI
241 genome browser by zooming out and looking for similar genes in the two
242 genomes. We especially looked for the pairs of *het* genes that function
243 together in *N. crassa*: *het-6*/*un24* and *het-c/pin-c*.
244 Transposable elements (TEs) were identified using REPET v2.5 (Flutre
245 et al. 2011) as described (Rivera et al. 2018). TE families that contained ≥10
246 sequences (minimum of one sequence ≥300 bp) were assessed for RIP using
247 RIPCAL v2 (Hane and Oliver 2008) using di-nucleotide frequency and
248 alignment-based algorithms. Evidence of RIP mutation was present if di-
249 nucleotide frequencies met the following criteria: (CpA + TpG)/(A/C +
250 GpT)≤1.03 and (TpA/ApT) ≥0.89, and RIPCAL alignments showed peaks for
251 (CA← →TA) + (TG ← → TA) mutations.
252
253 **Data deposition**. The genome assembly and annotations are made
254 available via the JGI fungal genome portal MycoCosm
255 (mycocosm.jgi.doe.gov/fungi; Grigoriev et al. 2014). The data are also
256 deposited at DDBJ/EMBL/GenBank under the following accessions (TO BE
257 PROVIDED UPON ACCEPTANCE). Supplementary tables and figures are
258 available through the National Agricultural Library AgData Commons at
259 DOI(TO BE PROVIDED UPON ACCEPTANCE).
260
261 **Results and discussion**
262
263 **Genome sequencing, assembly and integration with karyotypes and**
264 **genetic linkage map.** The genome of *Cryphonectria parasitica* EP155 was
265 sequenced using the Sanger whole-genome shotgun approach using paired-
266 end sequencing reads of plasmid and fosmid libraries at a coverage of
267 ~8.54X. After trimming for vector and quality, the EP155 genome was
268 originally assembled into 39 main genome scaffolds totaling 43.9 Mb (version
269 1), with the eleven largest scaffolds containing 90% (39,571,974 bp) of the
270 genome sequence. The version 1 assembly was condensed using
271 recombinational linkage data, yielding the final condensed version 2
272 assembly contained in 26 scaffolds with 33 contigs (L50 = 5; N50 = 4.0Mb;
273 43.9 Mb) with an estimated error rate of less than 1 error in 100,000 base
274 pairs. The allele data also allowed placement of the *C. parasitica* mating-type
275 locus (*MAT1*) on Scaffold 2 and vegetative incompatibility loci *vic1*, *vic2*,
276 *vic4*, *vic6* and *vic7* on Scaffolds 5, 7, 4, 3 and 6, respectively. A brief
277 overview of the genome assembly is provided in Table 1; complete summary

278 statistics can be accessed at the JGI website
279 https://genome.jgi.doe.gov/Crypa2/Crypa2.home.html.
280 Overall, there was close correspondence between EP155 karyotypes
281 (Eusebio-Cope et al. 2009) and the 26 scaffolds of the EP155 genome
282 assembly. Estimated *C. parasitica* chromosome sizes based on PFGE ranged
283 from 3.3 Mb to 9.7 Mb with no evidence of mini-chromosomes or accessory
284 chromosomes that are associated with some plant pathogenic fungi
285 (Bertazzoni et al. 2018; Eusebio-Cope et al. 2009). The 16 telomeric
286 sequences identified in the EP155 assembly indicated a minimum of eight
287 chromosomes, in good agreement with cytological and electrophoretic
288 karyotyping datasets that predicted chromosome counts of either seven or
289 nine (Eusebio-Cope et al. 2009; Milgroom et al. 1992). Five of the EP155
290 scaffolds were within the estimated chromosome size range and were
291 complete from telomere to telomere (scaffolds 1-4, 8). Twelve scaffolds had
292 a telomere on one end (scaffolds 5-7, 9-11). The remaining scaffolds were
293 smaller and lacked telomeres.
294 The eleven largest scaffolds ranged in size from 1 Mb to 7.4 Mb, and
295 together comprised 99.2% of the genome sequence (43.3 Mb). Roughly half
296 of the genome was contained in four scaffolds of at least 5.1 Mb in length.
297 Scaffold 8 was closest in size to that predicted for chromosome 9 (3.2 Mb vs
298 3.3 Mb), contained two telomere sequences and contained the beta-tubulin
299 gene that was previously identified as residing on chromosome 9 by
300 Southern analysis (Eusebio-Cope, et al. 2009). The close correspondence
301 between karyotyping results and the draft genome sequence analysis
302 provides a promising platform for further refinement of the sequence
303 assembly to the chromosome level.
304
305 **Gene model predictions.** After filtering for EST support, completeness and
306 homology support, a total of 11,609 genes were structurally and functionally
307 annotated from the EP155 version 2 assembly (Table 1). This number is
308 similar to that predicted for related fungi such as *Neurospora crassa*
309 (10,620), *Magnaporthe oryzae* (12,841) and *Fusarium graminearum*
310 (11,640). Structural features of the predicted *C. parasitica* genes conformed
311 to those reported for the sequenced genomes of other ascomycete fungi in
312 terms of average gene length (1645.62 nt), average size of encoded protein
313 (421.78 aa), exon number (2.9) and intron size (123 nt). Over 85% of
314 predicted proteins show similarities to other proteins from NCBI non-
315 redundant protein database. Over 66% of the predicted proteins contained
316 Pfam domains, with the most highly represented domains including: major
317 facilitator superfamily MFS-1 (PF07690; *n*=244), fungal Zn(20-Cys(6)
318 transcriptional regulatory protein (PF000172; *n*=166), short-chain
319 dehydrogenase/reductase SDR (PF000106; *n*=132), cytochrome P450
320 (PF00067; n=116), and serine /threonine protein kinases (PF00069; *n*=116).
321 Relative to the predicted proteomes of seven other members of the
322 Diaporthales curated by the JGI Mycoportal (Diaporthales MCL.2920), 85.7%
323 of the predicted *C. parasitica* proteins are members of multigene clusters.

324
325 **Mitochondrial genome.** The mitochondria of *C. parasitica* have been
326 extensively studied for their association with virulence attenuation (i.e.
327 hypovirulence), either derived from viral infection, mtDNA mutations or
328 mitochondrial plasmids (Baidyaroy et al. 2000; Monteiro-Vitorello et al. 1995,
329 Polashock and Hillman 1994). Some *C. parasitica* strains harbor
330 mitochondrial plasmids that elicit hypovirulence (Monteiro -Vitorello et al.
331 2000). Similarly, some strains of *C. parasitica* are subject to mitochondrial
332 hypovirulence, a cytoplasmically transmissible form of hypovirulence
333 associated with defects in the mitochondria (Baidyaroy et al. 2000; Monteiro-
334 Vitorello et al. 1995). Although distinct from viral-induced hypovirulence,
335 there are remarkable parallels between the non-viral forms of hypovirulence,
336 including virulence attenuation and the shared alteration of transcript
337 accumulation of over 70 genes (Allen and Nuss 2004; Monteiro-Vitorello et al.
338 1995, 2000). Mitochondria of some *C. parasitica* strains are also shown to
339 harbor small RNA viruses that can reduce fungal virulence (Polashock and
340 Hillman 1994) and can be transmitted to several other fungal species (Shahi
341 et al., 2019). Given that *C. parasitica* EP155 is a virus-free, virulent strain of
342 the fungus, it was not surprising that the mtDNA assembly did not share
343 significant similarity with any of the known indicators of mitochondrial
344 hypovirulence, e.g., the assembly did not contain sequences with similarity
345 to SSU rDNA *InC9* [AF218209], *Cryphonectria parasitica mitovirus 1-NB631*
346 [NC004046], or *pCRY1* [AF031368]. The ratio of UGA (=cytoplasmic
347 terminator) to UGG codons predicted to encode Trp in the *C. parasitica*
348 mitochondrial genome is high compared to other fungi, ~95%, and this
349 correlates positively with the relatively high number of UGA codons predicted
350 to encode Trp in *Cryphonectria* mitochondrial viruses compared to other
351 related mitochondrial viruses (Nibert, 2018).
352 Overall, the mtDNA genome contained a full complement of protein
353 coding genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*,
354 *nad4*, *nad4L*, *nad5*, *nad6*), rRNA and ribosomal proteins (*rrns*, *rnL*, *rps3*), and
355 29 tRNAs. Similar to the mitochondrial genomes of most filamentous
356 ascomycetes, the majority of the tRNAs were clustered together, with 10
357 tRNA genes located side-by-side on each side of the *rrnL* ribosomal gene.
358 Endonuclease ORFs were abundant, with 36 LAGLI-DADG homing
359 endonucleases and 27 GIY-YIG endonucleases predicted to occupy 29% of
360 the mtDNA assembly. Numerically, *C. parasitica* EP155 had a total of 63
361 predicted mitochondrial endonucleases, exhibiting one of the largest overall
362 cohorts of such enzymes identified to date (*Sclerotinia borealis*=61,
363 *Rhizoctonia solani* Rhs1AP=43, *Agaricus bisporus*=46; Mardonov et al. 2014).
364 A physical and genetic map of the *C. parasitica* EP155 mitochondrion
365 was published in 1996 (Bell et al. 1996), with a predicted size of 157 kbp and
366 13 genes mapped. In the current study, genome sequencing recovered the
367 mitochondrial genome within a single 158,902 bp scaffold, consistent with
368 the published physical map. All open reading frames contained within the
369 mtDNA assembly were in the same orientation. Gene order predictions

370 between the physical map and the mtDNA genome sequence only partially
371 matched. Gene ordering in the portion of the physical map containing
372 *atp6*/*nad5*/*nad3*/*cob*/*cox1*/*nad6* agreed with the assembly, but the remainder
373 of the physical map did not correspond with gene ordering derived from
374 sequencing.
375
376 **Secondary metabolites genes and gene clusters with similar**
377 **function.** In fungi, genes involved in complex coordinated functions such as
378 pathogenicity and secondary metabolite (SM) production can occur as co-
379 regulated gene clusters (Nierman et al. 2005; reviewed in Keller et al. 2005;
380 Keller 2019). Altogether, 59 predicted SM genes/gene clusters were
381 identified from scaffolds 1-11 of the *C. parasitica* EP155 genome assembly.
382 All but six of these regions contained either polyketide synthase (PKS;
383 annotated as PKS1-PKS31) and/or non-ribosomal peptide synthase (NPS)
384 genes (annotated as NPS1-10; NPS12-13; PKS/NPS1-PKS/NPS5; ACS1; LYS1;
385 FASA; OAS1). The predicted SMs ranged in size from just a single gene
386 (44.8%) to clusters containing up to 39 genes. The two largest clusters,
387 PKS16 and PKS26 (39 and 32 genes, respectively), each spanned over 100
388 kbp of scaffold 10 and 1, respectively.
389 Enrichment analysis identified 69 clusters ($P \leq 1 \times 10^{-4}$). Overall, these
390 functional gene clusters were more or less evenly distributed across all
391 scaffolds with the exception that six clusters containing a total of 23 MFS
392 transporter ORFs were located on the right half (1.6 Mb) of scaffold 8. ORFs
393 that were most commonly clustered in *C. parasitica* (enriched at $P < 1 \times 10^{-4}$)
394 corresponded with genes present in the predicted SM clusters, and included
395 MFS transporters (163 ORFs distributed across 31 clusters), NAD(P)-binding
396 with reductase or dehydrogenase activity (94 ORFs / 14 clusters), P450s (83
397 ORFs / 32 clusters) and putative PKS genes (23 ORFs / 9 clusters).
398
399 **Cytochrome P450 monooxygenase.** Cytochrome P450 monooxygenases
400 (P450s) are heme-thiolate proteins found across the biological kingdoms.
401 These enzymes perform a wide variety of reactions such as hydroxylation,
402 epoxidation, dealkylation, sulfoxydation, deamination, desulphuration,
403 dehalogenation, and nitric oxide reduction (Sono et al 1996). Fungi in
404 general possess extraordinarily large numbers of P450 genes (P450ome) in
405 their genomes, second only to plants. In the genomes of mushroom-forming
406 fungi, especially white rot fungi, genes encoding P450s are highly expanded
407 in number, facilitating the breakdown of lignins and other complex
408 substances (Syed and Yadav 2012; Suzuki et al. 2012).
409 Our analysis revealed that the *C. parasitica* EP155 genome contains
410 122 P450s (P450ome). Using the Cytochrome P450 (CYP) nomenclature
411 criteria, *C. parasitica* P450s were classified into 76 CYP families and 101 sub-
412 families. The majority of the P450s in the *C. parasitica* P450ome were
413 orphans, with no known function. Fifteen novel sub-families were identified
414 (one under each of the CYP families CYP503, CYP526, CYP548, CYP567,
415 CYP584, CYP614, CYP638, CYP639, CYP660, CYP5091, CYP5093, CYP5111,

416 CYP5129, CYP5168 and CYP5227). In comparison with other euascomycetes,
417 the *C. parasitica* genome has a moderately sized P450ome, comparable in
418 size to the P450omes of *Pyricularia oryzae* (123 P450s), *F. verticillioides* (126
419 P450s) and *A. terreus* (124 P450s). Interestingly, unlike a majority of the
420 euascomycetes, which do not contain basidiomycete-like P450s, the *C.*
421 *parasitica* P450ome revealed five such basidiomycete P450 homologs,
422 namely CYP5053, CYP5227, CYP5090, CYP5093 and CYP5201. The
423 abundance of basidiomycete P450 homologs in *C. parasitica* implies that
424 these P450s may play a key role in the oxidation of wood-derived
425 compounds and tree pathogenesis (Syed and Yadav 2012; Suzuki et al.
426 2012).
427
428 **Vegetative incompatibility.** In *C. parasitica*, six di-allelic loci controlling
429 vegetative incompatibility in European *C. parasitica* populations have been
430 identified by classical genetics (Cortesi and Milgroom, 1998), five of which
431 function in preventing heterokaryon formation (Choi et al. 2012; Smith et al.
432 2006; Zhang et al. 2014). At least two more *vic* loci are thought to function in
433 natural populations of *C. parasitica* (Liu and Milgroom, 2007; Robin et al.
434 2000). These loci have historically been referred to in *C. parasitica* as
435 vegetative incompatibility (*vic*) loci (Anagnostakis, 1988), but are assumed
436 to share characteristics with heterokaryon incompatibility (*het*) loci in other
437 fungi. Incompatibility genes at six of these *vic* loci were identified in *C.*
438 *parasitica* by a combination of linkage mapping and comparative genomics
439 that made use of the EP155 genome sequence (Choi et al. 2012; Zhang et al.
440 2014). Systematic disruption of these *vic* genes demonstrated their role in
441 allorecognition and their ability to restrict virus transmission (Choi et al.
442 2012; Zhang et al. 2014; Zhang and Nuss 2016).
443 From fungi other than *C. parasitica*, thirteen genes with heterokaryon
444 incompatibility function have been characterized at the molecular level:
445 seven from *N. crassa* and six from *P. anserina* (Glass and Dementhon, 2006;
446 Paoletti and Saupe, 2009). As in *N. crassa* and *P. anserina*, *vic* genes in *C.*
447 *parasitica* are characterized by significant allelic polymorphisms, several are
448 idiomorphic, and two encode proteins with a HET domain. The HET domain is
449 defined by three conserved motifs of about 18, 36 and 10 amino acids in
450 length, arranged in a specific order within a ~200 amino acid region, which is
451 also a feature of six of the *het* genes characterized in *P. anserina* and *N.*
452 *crassa* (Paoletti and Clavé, 2007; Smith et al. 2000). BLASTp analysis of the
453 *C. parasitica* EP155 genome sequence identified 94 proteins with homologs
454 from other ascomycete *het* genes. *C. parasitica* protein 88866 (annotated
455 *hch1*, for *het-c* homolog) is homologous to *N. crassa het-c* (PF07217.2); in *C.*
456 *parasitica* the region containing *hch1* is syntenic to the region of the *N.*
457 *crassa* genome containing *het-c*, but we found no homolog to *pin-c* in *C.*
458 *parasitica*, which is the linked interacting partner to *het-c* in *N. crassa*.
459 Several *C. parasitica*-encoded proteins were found with high levels of
460 similarity to *het* genes from *P. anserina*. A homolog of the *P. anserina het-D/E*
461 genes, *C. parasitica* protein 84049 (annotated *pdh1*, for *Podospora het-D*

462 homolog), clearly contains conserved NACHT, WD-repeats and HET domains.
463 Protein 106535 (annotated *pch1*, for *Podospora het-c* homolog) is highly
464 similar to *P. anserina het-C*. None of these *het* gene homologs in *C. parasitica*
465 map to regions associated with known vegetative incompatibility function.
466 Overall, there are 124 genes annotated in *C. parasitica* that contain the
467 HET domain. This number is among the highest found yet in any ascomycete
468 genome. *Podospora anserina* was previously described as containing the
469 most recorded HET domains with 120 (Paoletti et al. 2007), *N. crassa* has 55
470 and *Aspergillus oryzae* has 38 (Federova et al. 2005). We found several ORFs
471 that putatively encode HET domains with high similarity to *pin-c*, *tol* or *het-6*
472 from *N. crassa*, but none was found in regions syntenic with the *N. crassa*
473 homologs and therefore we did not name these specifically as homologs.
474 Genes with HET domains that were not clearly homologous to known *het*
475 genes were considered to have HET-like domains. Considerably more ORFs
476 that have HET-domains occur in filamentous ascomycete genomes than
477 known functional *het* genes (Federova et al. 2005). Similarly, homologs of
478 known *het* genes are not necessarily functional *het* genes in other species.
479 The overall lack of synteny among HET domain genes between *C. parasitica*
480 and other ascomycete species, and the dispersed repetitive distribution, both
481 intra- and intergenomically, supports the view that the HET domain
482 represents a component of a mobile genetic element (Paoletti and Saupe,
483 2009).
484
485 **Transposable elements.** Transposable elements (TEs) representing both
486 Class I (retrotransposons) and Class II transposons (transposons with DNA
487 transposition intermediates) were present in the genome of *C. parasitica*
488 EP155 (Table 1). The TE load, ~14% of the total genome sequence, was
489 largely due to the presence of 2,716 Class I retroelements, comprising
490 almost 5.0 Mb total. Class I elements in the family *Metaviridae* (*Gypsy/Ty3*
491 elements) were the most abundant group of retroelements, with 2,040
492 elements comprising over 4 Mb of the genome and making up over 75% of
493 all TEs. No copy of a *Metaviridae* retrotransposon containing an intact coding
494 sequence was identified. *Metaviridae* elements were commonly located in
495 TE-rich clusters. Of note was the region surrounding the *MAT1* locus on
496 scaffold 5, where there are numerous retrotransposon fragments on either
497 side of the *MAT1-2* gene. Overall, the 4-Mb scaffold where *MAT1* resides
498 (scaffold 5) was ~17% transposon-derived. The presence of TEs surrounding
499 the *Mat1* locus was consistent with mapping studies performed by Kubisiak
500 and Milgroom (2006), where significant recombination suppression and high
501 levels of heterogeneity in the region surrounding the *MAT1* locus was
502 documented.
503 Seventeen full-length TEs with intact coding sequences were identified
504 in the EP155 genome. Nine of the 17 intact transposons were copies of the
505 *hAT*-family Class II transposon *Crypt1*, the only *C. parasitica* element that has
506 been shown experimentally to be active (Linder-Basso et al. 2001).

507 Repeat-induced point mutation (RIP) is a fungal genome defense
508 mechanism that may mutate repeated sequences such as TEs, most
509 commonly leading to sequences with a reduced GC content and C→T
510 transition mutations. RIP is well defined and extremely efficient in
511 *Neurospora crassa* (Cambareri et al. 1998) and has been documented at a
512 much less efficient level in several other filamentous fungi (reviewed in
513 Galagan and Selker, 2004; Clutterbuck, 2011). Duplicate contiguous
514 sequences of greater than ~400 bases within a given genome are detected
515 by an unknown mechanism and then disabled by methylation of cytosine
516 bases in either copy of the duplicated sequence, followed by subsequent
517 deamination of the methylated cytosines to thymine. The only gene known
518 to be required for RIP encodes a DNA methyltransferase called *rid* (RIP-
519 defective; Freitag et al. 2002), and this gene is present in the *C. parasitica*
520 EP155 genome. However, RIPCAL analyses and dinucleotide frequencies
521 showed little evidence for RIP mutation across the *C. parasitica* EP155
522 genome. Our detections of RIP were limited to DIRS elements (*n*=537),
523 helitrons (*n*=24), and an unidentified Class II element (*n*=23). Our analysis
524 did not detect a signature of RIP mutation from *Metaviridae* elements,
525 although using a de-RIP approach, Clutterbuck (2011) identified 10 *Gypsy*
526 elements with dinucleotide ratios consistent with RIP mutation.
527
528 **Conclusions**
529 *Cryphonectria parasitica* has great importance as a plant pathogen both
530 historically and contemporarily. The fungus caused the pandemic that
531 reshaped forests in North America and is still abundant in the environment,
532 suppressing chestnut populations. Thus, information about the fungal
533 genome is vital to understanding host/pathogen interactions and supports
534 traditional breeding and transgenic approaches to develop resistance against
535 the disease. As important scientifically, *C. parasitica* is a model system for
536 examining virus/fungus interactions at the molecular and population levels.
537 These interactions have provided a powerful and exploitable platform for
538 identifying cellular mechanisms important for fungal pathogenesis, and for
539 examining the potential of viruses for biological control of a fungal pathogen.
540 Understanding this fungal genome is therefore critical to understanding the
541 fungus as a pathogen, as a host for its own pathogens, and to explore the
542 potential of using mycoviruses to productively alter fungal phenotype.
543
544 **Acknowledgements**
545 This work was supported by the U.S. Department of Energy (DOE) Joint
546 Genome Institute (JGI), a DOE Office of Science User Facility, which is
547 supported by the Office of Science of the DOE under Contract No. DE-AC02-
548 05CH11231 awarded to DLN, ACLC and MGM, Lawrence Livermore National
549 Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos
550 National Laboratory under contract No. DE-AC02-06NA25396. The project
551 has been supported since its inception through the USDA National Institute of
552 Food and Agriculture (USDA-NIFA) Hatch Multistate Research Program,

553 initially through project No. NE140, currently project No. NE1833. JAC is
554 supported by USDA-ARS project 8042-22000-298-00-D. ALD and DLN were
555 supported by an NSF collaborative award, MCB-1051453 and MCB-1051331,
556 respectively. BIH was supported by USDA-NIFA Hatch and McIntire-Stennis
557 Research Programs and by the New Jersey Agricultural Experiment Station.
558 MLS is supported by an NSERC Discovery Grant. The P450 work was
559 supported by the University of Cincinnati funds to JSY.
560
561 Mention of trade names or commercial products in this publication is solely
562 for the purpose of providing specific information and does not imply
563 recommendation or endorsement by the USDA. The USDA is an equal
564 opportunity provider and employer.

565 **References**

566 Allen, T.D., and Nuss, D.L. 2004. Linkage between mitochondrial
567 hypovirulence and viral hypovirulence in the chestnut blight fungus
568 revealed by cDNA microarray analysis. Eukaryot. Cell 3:1227-1232.
569 Anagnostakis, S.L. 1988. *Cryphonectria parasitica*, cause of chestnut blight.
570 Adv. Plant Pathol. 6:123-136.
571 Anagnostakis, S.L. 2012. Chestnut breeding in the United States for disease
572 and insect resistance Plant Dis. 96:1392-1403
573 Anagnostakis, S.L., and Day, P.R. 1979. Hypovirulence conversion in
574 *Endothia parasitica*. Phytopathology 69:1226-1229.
575 Baidyaroy, D., Huber, D.H., Fulbright, D.W., and Bertrand, H. 2000.
576 Transmissable mitochondrial hypovirulence in a natural population of
577 *Cryphonectria parasitica*. Mol. Plant Microbe Interact. 13:88-95.
578 Bertazzoni, S., Williams, A.H., Jones, D.A., Syme, R.A., Tan, K.-C., and Hane,
579 J.K. 2018. Accessories make the outfit: Accessory chromosomes and
580 other dispensable DNA regions in plant pathogenic fungi. Mol. Plant
581 Microb. Int. 31:779-788.
582 Bell, J.A., Monteiro-Vitorello, C.B., Hausner, G., Fulbright, D.W., and Bertrand,
583 H. 1996. Physical and genetic map of the mitochondrial genome of
584 *Cryphonectria parasitica*. Current Genet. 30:34-43.
585 Cambareri, E.B., Aisner, R., and Carbon, J. 1998. Structure of the
586 chromosome VII centromere region in *Neurospora crassa:* degenerate
587 transposons and single repeats. Mol. Cell Biol. 18:5465-5477.
588 Choi, G.H., Dawe, A.L., Churbanov, A., Smith, M.L., Milgroom, M.G., and Nuss,
589 D.L. 2012. Molecular characterization of the vegetative incompatability
590 genes that restrict hypovirus transmission in the chestnut blight fungus
591 *Cryphonectria parasitica*. Genetics 190:113-127.
592 Clutterbuck, A.J. 2011. Genomic evidence of Repeat-Induced Point mutation
593 (RIP) in filamentous ascomycetes. Fungal Genet. Biol. 48:306-326.
594 Cortázar, A.R., Aranasay, A.M., Alfaro, M., Oguiza, J.A., and Lavín, J.L. 2014.
595 SECRETOOL: integrated secretome analysis tool for fungi. Amino Acids
596 46:471-473.
597 Cortesi, P., McCulloch, C.E., Song, H., Lin, H., and Milgroom, M.G. 2001.
598 Genetic control of horizontal virus transmission in the chestnut blight
599 fungus, *Cryphonectria parasitica*. Genetics 159:107-118.
600 Cortesi, P., and Milgroom, M.G. 1998. Genetics of vegetative incompatibility
601 in *Cryphonectria parasitica*. Appl. Environ. Microbiol. 64:2988-2994.
602 Dalgleish, H.J., Nelson, C.D., Scrivani, J.A., and Jacobs, D.F. 2015.
603 Consequences of shifts in abundance and distribution of American
604 chestnut for restoration of a foundation forest tree. Forests 7:4-9.
605 Dalavalle, E., and Zambonelli, A. 1999. Epidemiological role of strains of
606 *Cryphonectria parasitica* isolated from hosts other than chestnut.
607 Europ. J. Forest Pathol. 29:97-102.
608 Dawe, A.L., and Nuss, D.L. 2001. Hypoviruses and chestnut blight: exploiting
609 viruses to understand and modulate fungal pathogenesis. Ann. Rev.
610 Genetics 35:1-29.

611 Dawe, A.L., McMains, V.C., Panglao, M., Kasahara, S., Chen, B., and Nuss, D.L.
612 2003. An ordered collection of expressed sequences from
613 *Cryphonectria parasitica* and evidence of genomic microsynteny with
614 *Neurospora crassa* and *Magnaporthe grisea*. Microbiology 149:2373-
615 2384.
616 Eusebio-Cope, A., Suzuki, N., Sadeghi-Garmaroodi, H., and Taga, M. 2009.
617 Cytological and electrophoretic karyotyping of the chestnut blight
618 fungus *Cryphonectria parasitica*. Fungal Genet. Biol. 46:342-351.
619 Fedorova, N.D., Badger, J.H., Robson, G.D., Wortman, J.R., and Nierman, W.C.
620 2005. Comparative analysis of programmed cell death pathways in
621 filamentous fungi. BMC Genomics 6:177.
622 Flutre, T., Duprat, E., Feuillet, C., and Quenesville, H. 2011. Considering
623 transposable element diversification in *de novo* annotation
624 approaches. PLoS One 6:e16526.
625 Freitag, M., Williams, R.L., Kothe, G.O., and Selker, E.U. 2002. A cytosine
626 methyltransferase homologue is essential for repeat-induced point
627 mutations in *Neurospora crassa*. Proc. Natl. Acad. Sci. U.S.A. 99:8802-
628 8807.
629 Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D.,
630 FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T.,
631 Engels, R., Wang, S., Nielsen, C.B., Butler, J., Endrizzi, M., Qui, D.,
632 Ianakiev, P., Bell-Pedersen, D., Nelson, M.A., Werner-Washburne, M.,
633 Selitrennikoff, C.P., Kinsey, J.A., Braun, E.L., Zelter, A., Schulte, U.,
634 Kothe, G.O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg,
635 D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R.,
636 Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S.,
637 Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R.L., Perkins,
638 D.D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt,
639 R.J., Osmani, S.A., DeSouza, C.P., Glass, L., Orbach, M.J., Berglund, J.A.,
640 Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A.,
641 Aramayo, R., Natvig, D.O., Alex, L.A., Mannhaupt, G., Ebbole, D.J.,
642 Freitag, M., Paulsen, I., Sachs, M.S., Lander, E.S., Nusbaum, C., and
643 Birren, B. 2003. The genome sequence of the filamentous fungus
644 *Neurospora crassa*. Nature 422:859-68.
645 Galagan, J.E., and Selker, E.U. 2004. RIP: the evolutionary cost of genome
646 defense. Trends Genet. 20:417-423.
647 Glass, N.L., and Dementhon K. 2006. Non-self recognition and programmed
648 cell death in filamentous fungi. Curr. Opin. Microbiol. 9:553-558.
649 Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R.,
650 Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H.,
651 Dubchak, I., and Shabalov, I. 2014. MycoCosm portal: gearing up for
652 1000 fungal genomes. Nucleic Acids Res. 42:D699-704.
653 Gordon, D. 2003. Viewing and editing assembled sequences using Consed.
654 Curr. Protocols Bioinfor. 2:11.2.1-11.2.43.

655 Hane, J. K., and Oliver, R. P. 2008. RIPCAL: a tool for alignment-based
656 analysis of repeat-induced point mutations in fungal genomic
657 sequences. BMC Bioinformatics 9:478.
658 Heiniger, U., and Rigling, D. 1994. Biological control of chestnut blight in
659 Europe. Ann. Rev. Phytopathol. 32:581-599.
660 Hillman, B.I., and Suzuki, N. 2004. Viruses of the chestnut blight fungus,
661 *Cryphonectria parasitica*. Adv. Virus Res. 63:423-472.
662 Jaffe, D.B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov, J.P.,
663 Zody, M.C., and Lander, E.S. 2003. Whole-genome sequence assembly
664 for mammalian genomes: Arachne 2. Genome Res. 13:91-96.
665 Keller, N.P., Turner, G., and Bennett, J.W. 2005. Fungal secondary
666 metabolism - from biochemistry to genomics. Nat. Rev. Microbiol.
667 3:937-947.
668 Keller, N.P. 2019. Fungal secondary metabolism: regulation, function and
669 drug discovery. Nat. Rev. Microbiol. 17:167-180.
670 Kubisiak, T.L., and Milgroom, M.G. 2006. Markers linked to vegetative
671 incompatibility (vic) genes and a region of high heterogeneity and
672 reduced recombination near the mating type locus (MAT) in
673 *Cryphonectria parasitica*. Fungal Genet. Biol. 43:453-463.
674 Kuo, A., Bushnell, B., Grigoriev, I.V. (2014). Fungal genomics: Sequencing
675 and annotation. Adv. Bot. Res. 70:1-52.
676 Linder-Basso, D., Foglia, R., Zhu, P., and Hillman, B.I. 2001. *Crypt1*, an active
677 *Ac*-like transposon from the chestnut blight fungus, *Cryphonectria*
678 *parasitica*. Mol. Genet. Genomic 265:730-738.
679 Liu, Y.C. and Milgroom, M.G. 2007. High diversity of vegetative compatibility
680 types in *Cryphonectria parasitica* in Japan and China. Mycologia
681 99:279-284.
682 Mardonov, A.V., Beletsky, A.V., Kadnikov, V.V., Ignatov, A.N., and Ravin, N.V.
683 2014. The 203 kbp mitochondrial genome of the phytopathogenic
684 fungus *Sclerotinia borealis* reveals multiple invasions of introns and
685 genomic duplications. PLoS One 9:e107536.
686 Milgroom, M.G., and Cortesi, P. 2004. Biological control of chestnut blight
687 with hypovirulence: a critical analysis. Ann. Rev. Phytopathol. 42:311-
688 338.
689 Milgroom, M.G., Lipari, S.E., and Powell, W.A. 1992. DNA fingerprinting and
690 analysis of population structure in the chestnut blight fungus,
691 *Cryphonectria parasitica*. Genetics 131: 297-306.
692 Monteiro-Vitorello, C.B., Baidyaroy, D., Bell, J.A., Hausner, G., Fulbright, D.W.,
693 and Bertrand, H. 2000. A circular mitochondrial plasmid incites
694 hypovirulence in some strains of *Cryphonectria parasitica*. Curr. Genet.
695 37:242-256.
696 Monteiro-Vitorello, C.B., Bell, J.A. Fulbright, D.W., and Bertrand, H.A. 1995. A
697 cytoplasmically transmissible hypovirulence phenotype associated with
698 mitochondrial DNA mutations in the chestnut blight fungus
699 *Cryphonectria parasitica*. Proc. Natl. Acad. Sci. U.S.A. 92:5935-5939.

700 Nibert, M.L. 2018. Mitovirus UGA(Trp) codon usage parallels that of host
701 mitochondria. Virology 507:96-100.
702 Nuss, D.L. 1992. Biological control of chestnut blight: an example of virus-
703 mediated attenuation of fungal pathogenesis. Microbiol. Rev. 56:561-
704 576.
705 Nuss, D.L. 2005. Hypovirulence: mycoviruses at the fungal-plant interface.
706 Nat. Rev. Microbiol. 3:632-642.
707 Paoletti, M., and Clavé, C. 2007. The fungus-specific HET domain mediates
708 programmed cell death in *Podospora anserina*. Eukaryot. Cell 6:2001-
709 2008.
710 Paoletti, M., and Saupe, S.J. 2009. Fungal incompatibility: Evolutionary origin
711 and pathogen defense? BioEssays 31:1201-1210.
712 Paoletti, M., Saupe, S.J., and Clavé C. 2007. Genesis of a fungal non-self
713 recognition repertoire. PLoS One 2:e283.
714 Polashock, J.J., and Hillman, B.I. 1994. A small mitochondrial double-stranded
715 (ds) RNA element associated with a hypovirulent strain of the chestnut
716 blight fungus and ancestrally related to yeast cytoplasmic T and W
717 dsRNAs. Proc. Natl. Acad. Sci. U.S.A. 91:8680-8684.
718 Rivera, Y., Salgado-Salazar, C., Veltri, D., Malapi-Wight, M., Crouch, JA. 2018.
719 Genome analysis of the ubiquitous boxwood pathogen *Pseudonectria*
720 *foliicola*. PeerJ 6:e5401.
721 Robin, C., Anziani, C., and Cortesi, P. 2000. Relationship between biological
722 control, incidence of hypovirulence, and diversity of vegetative
723 incompatibility types of *Cryphonectria parasitica* in France.
724 Phytopathology 90:730-737.
725 Rozen, S., and Skaletsky, H. 2000. Primer3 on the WWW for general users
726 and for biologist programmers. Methods Mol. Biol. 132:365-386.
727 Shahi, S., Eusebio-Cope, A., Kondo, H., Hillman, B.I., and Suzuki N. 2019.
728 Investigation of host range and host defense against a mitochondrially
729 replicating mitovirus. J. Virol. 93: e01503-18.
730 Shang, J., Wu, X., Lan, X., Fan, Y., Dong, H., Deng, Y., Nuss, D.L., and Chen, B.
731 2008. Large-scale expressed sequence tag analysis for the chestnut
732 blight fungus *Cryphonectria parasitica*. Fungal Genet. Biol. 45:319-327.
733 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov,
734 E.M. 2015. BUSCO: assessing genome assembly and annotation
735 completeness with single-copy orthologs. Bioinformatics 31:3210-3212.
736 Smith, M.L., Micali, O.C., Hubert, S.P., Mir-Rashed, N., Jacobson, D.J., and
737 Glass, N.L. 2000. Vegetative incompatibility in the *het-6* region of
738 *Neurospora crassa* is mediated by two linked genes. Genetics
739 155:1095-1104.
740 Smith, M.L., Gibbs, C.C., and Milgroom, M.G. 2006. Heterokaryon
741 incompatibility function of barrage-associated vegetative
742 incompatibility genes (*vic*) in *Cryphonectria parasitica*. Mycologia
743 98:43-50.
744 Sono, M., Roach, M.P., Coulter, E.D., and Dawson, J.H. 1996. Heme-containing
745 oxygenases. Chem. Rev. 96:2841-2888.

746 Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B., and Taylor, J.M.
747 2018. Improved prediction of fungal effector proteins from secretomes
748 with EffectorP 2.0. Mol. Plant Pathol. 19:2094-2110.
749 Stauder, C.M., Nuss, D.L., Zhang, D.-X., Double, M.L., MacDonald, W.L.,
750 Matheny, A.M., and Kasson, M.T. 2019. Enhanced hypovirus
751 transmission by engineered super donor strains of the chestnut blight
752 fungus, *Cryphonectria parasitica*, into a natural population of strains
753 exhibiting diverse vegetative compatibility genotypes. Virology 528:1-
754 6.
755 Suzuki H., MacDonald, J. Syed, K., Salamov, A., Chiaki, H., Aerts, A.,
756 Henrisaat, B., Wiebenga, A., Vankuyk, P.A., Barry, K., Lindquist E.,
757 LaButti, K., Lapidus, A., Lucas, S., Coutinho, P., Gong, Y., Samejima, M.,
758 Mahadevan, R., Abou-Zaid, M., deVaries, R.P. Igarashi, K., Yadav, J.S.,
759 Grigoriev, I.V., and Master, E.R. 2012. Comparative genomics of the
760 white-rot fungi, *Phanerochaete carnosa* and *P. chrysosporium*, to
761 elucidate the genetic basis of the distinct wood types they colonize.
762 BMC Genomics 13:444.
763 Syed, K., and Yadav, J.S. 2012. P450 monooxygenases (P450ome) of the
764 model white rot fungus *Phanerochaete chrysosporium*. Crit. Rev.
765 Microbiol. 38:339-363.
766 Zhang, D.-X., and Nuss, D.L. 2016. Engineering super mycovirus donor
767 strains of chestnut blight fungus by systematic disruption of multilocus
768 vic genes. Proc. Nat. Acad. Sci. U.S.A. 113:2062-2067.
769 Zhang. D.-X., Spiering, M.J., Dawe, A.L., and Nuss, D.L. 2014. Vegetative
770 incompatibility loci with dedicated roles in allorecognition restrict
771 mycovirus transmission in chestnut blight fungus. Genetics 197:701-
772 714.
773

774 775 **Table 1.** Summary features of the *Cryphonectria parasitica* EP155 genome assembly.

Assembly data	Scaffold count	26
	Contig count	33
	Scaffold assembly length	43.9 Mb
	Estimated % sequence bases in gaps	0.2%
	Scaffold N50/L50	4.0/5.1 Mb
	Contig N50/L50	5.0/4.0 Mb
	Average scaffold length (Mb)	1.69
	Maximum scaffold size (Mb)	7.44
	Number of scaffolds >5.0 kb	20
	Number of scaffolds >50.0 kb	13
	% of assembly in scaffolds >50.0 kb	99.6%
	No. Ns per 100 kb	160
	GC content	50.8%
Annotation	Gene Models	11,609
	BUSCO eukaryote orthologs	94.4%
	BUSCO fungal orthologs	98.6%
	Gene density (genes/Mb scaffold)	264.4
	Average gene length (bp)	1648.6
	Average protein length (aa)	422.9
	Average exon frequency	2.91 exons/gene
	Average exon length	487.0
	Average intron length	122.6
	% complete gene models (with start and stop codons)	84%
	% genes with homology support	85%
	% genes with Pfam domains	66%
Gene clusters	Multigene clusters	10,803
	Average multigene cluster size (bp)	5.32
	Singletons	6,605
No. secondary metabolite clusters (46 total)		
	DMAT[1]	2
	NRPS[2] and NRPS-like	14
	PKS[3] and PKS-like	22
	Hybrid PKS/NRPS	4
	TC	4
No. CAZyme families (548 total)		
	Auxillary activities	66
	Carbohydrate-binding module	36
	Carbohydrate esterase	40
	Distantly related to plant expansins	6
	Glycoside hydrolase	291
	Glycosyl transferases	100
	Polysaccharide lyase	9
Secretome and effector predictions		
	Secretome (3.4% of genes in proteome predicted as secreted)	397
	Candidate effectors (8% of secretome)	32
Cytochrome P450 monooxygenases		122
No. transposable elements (2,955 total elements, spanning 5.3 Mb)		
	2,040 LTR[4] (Class I)	4,029,976
	1 LINE[5] (Class I)	102

537 DIRS (Class I)	705,541
192 TIR[6] (Class II)	256,785
24 Helitron (Class II)	96,326
127 Unknown Order (Class I)	220,656
23 MITE[7]	4,568

776 [1] DMAT: dimethylallyl transferase
777 [2] NRPS: nonribosomal peptide synthetase
778 [3] PKS: polyketide synthase
779 [4] LTR: long terminal repeat
780 [5] LINE: long interspersed nuclear element
781 [6] TIR: tandem inverted repeat
782 [7] MITE: miniature inverted repeat