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 Jasmyne 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}

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.

Abstract

Cryphonectria parasitica is the causal agent of chestnut blight, a fungal disease that almost entirely eliminated mature American chestnut from North America over a 50-year period. Here, we formally report the genome of *C. parasitica* EP155 using a Sanger shotgun sequencing approach. After finishing and integration with simple-sequence repeat markers, the assembly was 43.8 Mb in 26 scaffolds ($L_{50} = 5$; $N_{50} = 4.0$ Mb). Eight chromosomes are predicted: five scaffolds have two telomeres and six scaffolds have one telomere sequence. In total, 11,609 gene models were predicted, of which 85% show similarities to other proteins. This genome resource has already increased the utility of a fundamental plant pathogen experimental system through new understanding of the fungal vegetative incompatibility system, with significant implications for enhancing mycovirus-based biological control.

Plant disease epidemics have had profound ecological and economic consequences and have significantly influenced human history. One of the most momentous disease epidemics was chestnut blight, which completely changed the ecological landscape of the hardwood forests of the eastern United States during the 20th century (Anagnostakis 1988). The causal agent, the ascomycete fungus *Cryphonectria parasitica*, found a very susceptible host in the American chestnut tree, *Castanea dentata*, following its introduction into North America on nursery stock of resistant Asian chestnut tree species. The resulting

Funding

This work was supported by the U.S. Department of Energy (DOE) Joint Genome Institute (JGI), a DOE Office of Science User Facility, which is supported by the Office of Science of the DOE under contract number DE-AC02-05CH11231 awarded to D. L. Nuss, A. C. L. Churchill, and M. G. Milgroom, Lawrence Livermore National Laboratory under contract number DE-AC52-07NA27344, and Los Alamos National Laboratory under contract number DE-AC02-06NA25396. The project has been supported since its inception through the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Hatch Multistate Research Program, initially through project number NE140, currently project number NE1833. J. A. Crouch is supported by USDA Agricultural Research Service project 8042-22000-298-00-D. A. Dawe and D. L. Nuss were supported by National Science Foundation collaborative awards MCB-1051453 and MCB-1051331, respectively. B. I. Hillman was supported by USDA-NIFA Hatch and McIntire-Stennis Research Programs and by the New Jersey Agricultural Experiment Station. M. Smith is supported by an NSERC Discovery Grant. The P450 work was supported by the University of Cincinnati funds to J. S. Yadav.

Keywords

chestnut blight, forest pathology, mycology, mycoviruses, transposable elements, vegetative incompatibility

[†]Corresponding authors: J. A. Crouch; joanne.crouch@usda.gov; and A. Dawe; dawe@biology.msstate.edu

Funding: U.S. Department of Energy (DOE) Joint Genome Institute (JGI), grant number DE-AC02-05CH11231.

The author(s) declare no conflict of interest.

Accepted for publication 10 March 2020.

disease epidemic, first identified in 1903 in the Bronx Zoo, spread rapidly, resulting in the destruction of an estimated four billion mature American chestnut trees in the following 50 years (Anagnostakis 1988). Although the root systems of infected trees often survive and resprout, because the new sprouts remain susceptible to endemic *Cryphonectria parasitica*, the once dominant tree now survives throughout its former natural range primarily as an understory shrub (Dalglish et al. 2015).

The mechanisms underlying the ability of *C. parasitica* to effectively penetrate defense barriers and rapidly expand in the cambium tissues of susceptible hosts remain ill defined, with no identified role for toxins, specific secondary metabolites (SM), or hydrolytic enzymes. In this article, we report the *C. parasitica* strain EP155 genome sequence as an important resource for elucidating the genomic basis for the selective pathogenicity, niche-associated evolution, and virus-based biological control of this classic forest pathogen. *C. parasitica* EP155 (ATCC 38755) is an orange-pigmented, virulent, hypovirus-free strain (vc type EU-5, *MAT1-2*) originally isolated in 1977 from a canker on *Castanea dentata* in Bethany, CT, U.S.A. (Anagnostakis and Day 1979). The fungus was grown on potato dextrose agar overlaid with cellophane at room temperature on the laboratory bench for approximately 7 days; mycelium and conidia were scraped from the cellophane using a sterile razor blade, allowed to air dry under a laminar flow hood, then pulverized using liquid nitrogen in a mortar and pestle. DNA was extracted as described (Choi et al. 2012).

Genome Sequencing and Assembly

All sequencing reads were collected with standard Sanger sequencing protocols on ABI 3730XL capillary sequencing machines (Thermo Fisher Scientific, Waltham, MA, U.S.A.) at the United States Department of Energy Joint Genome Institute (JGI) in Walnut Creek, CA. The *Cryphonectria parasitica* EP155 genome was sequenced using a Sanger whole-genome shotgun approach from paired-end sequencing reads of plasmid and fosmid libraries at a coverage of approximately 8.54x. Three different-sized libraries were used as templates for plasmid and fosmid subclone sequencing, with both ends sequenced as follows: 332,747 reads from a 2.3-kb-sized plasmid library, 265,247 reads from a 6.8-kb-sized plasmid library, and 107,327 reads from a 39.3-kb-sized fosmid library. Reads were assembled using a modified version of Arachne v.20071016 (Jaffe et al. 2003) with parameters maxcliq1 = 100, correct1_passes = 0, and BINGE_AND_PURGE = True. After trimming for vector and quality, the genome was assembled into 39 main scaffolds spanning 43.9 Mb (version 1), with the 11 largest scaffolds containing 90% of the genome sequence.

Genome Finishing and Map Integration

The initial whole-genome shotgun assembly was broken down into scaffolds and each scaffold piece was reassembled with Phrap (www.phrap.org), then manually improved and finished using Consed (Gordon 2003). All low-quality regions and gaps were targeted with computationally selected Sanger sequencing reactions completed with 4:1 BigDye terminator: dGTP chemistry (Thermo Fisher Scientific). These automated rounds included walking on 2.3- and 6.8-kb plasmid subclones using 4,526 custom primers. Following the automated rounds, a trained finisher manually inspected each assembly. Remaining gaps and hairpin structures were resolved by generating small insert shatter libraries of 6.8-kb-spanning clones (Grigoriev et al. 2014). Five fosmid clones were shotgun sequenced using Sanger long-read technology and finished to fill large gaps and resolve larger repeats. Each assembly was validated by an independent quality assessment, which included visual examinations of subclone paired ends, high-quality discrepancies, and any low-quality areas. The assembly was further refined with the aid of the *C. parasitica* genetic linkage map constructed from a cross of Japanese *C. parasitica* isolate JA17 and Italian isolate P17-8 (Kubisiak and Milgroom 2006) that was upgraded by the addition of 141 simple-sequence repeat (SSR) markers mined from the EP155 genome sequence using Primer3 (Rozen and Skaletsky 2000). Allele data for 96 ascospore progeny of the JA17 × P17-8 cross were collected for 60 polymorphic EP155-derived SSR markers located at the terminal ends of the EP155 scaffolds. Finished segments (33 scaffolds with 34 contigs) were localized and ordered into pseudomolecules using a 30-marker map with seven joins to form the final assembly. The version 1 assembly was condensed using the recombinational linkage data, yielding the final version 2 assembly contained in 26 scaffolds with 33 contigs ($L_{50} = 5$; $N_{50} = 4.0$ Mb; 43.9 Mb) with an estimated error rate of less than 1 error in 100,000 bp. The allele data also

Table 1. Summary features of the *Cryphonectria parasitica* EP155 genome assembly^a

Features ^b	Data
Assembly data	
Scaffold count	26
Contig count	33
Scaffold assembly length	43.9 Mb
Estimated percent sequence bases in gaps	0.2%
Scaffold N ₅₀ /L ₅₀	4.0/5.1 Mb
Contig N ₅₀ /L ₅₀	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
Percentage of assembly in scaffolds > 50.0 kb	99.6%
Number of 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)	1,648.6
Average protein length (aa)	422.9
Average exon frequency	2.91 exons/gene
Average exon length	487.0
Average intron length	122.6
Percentage of complete gene models (with start and stop codons)	84%
Percentage of genes with homology support	85%
Percentage of genes with Pfam domains	66%
Gene family clusters ^c	
Shared gene family clusters	10,803
Average multigene family size	5.32
Singletons	6,605
Number of secondary metabolite clusters (46 total)	
Dimethylallyl transferase	2
NRPS and NRPS-like	14
PKS and PKS-like	22
Hybrid PKS/NRPS	4
Terpene cyclase	4
Number of CAZyme families (548 total)	
Auxiliary 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 ^d	
Secretome (3.4% of genes in proteome predicted as secreted)	397
Candidate effectors (8% of secretome)	32
Cytochrome P450 monooxygenases	122
Number of transposable elements (2,955 total elements, spanning 5.3 Mb) ^e	
2,040 LTR (class I)	4,029,976
1 LINE (class I)	102
537 DIRS (class I)	705,541
192 TIR (class II)	256,785
24 Helitron (class II)	96,326
127 Unknown order (class I)	220,656
23 MITE	4,568

^a Data were generated using the Joint Genome Institute (JGI) annotation pipeline unless otherwise specified.

^b NRPS = nonribosomal peptide synthetase, PKS = polyketide synthase, CAZyme = carbohydrate active enzymes, LTR = long terminal repeat, LINE = long interspersed nuclear element, TIR = tandem inverted repeat, and MITE = miniature inverted repeat.

^c Gene family clusters were predicted using TRIBE-MCL as part of the JGI annotation pipeline (<https://mycocosm.jgi.doe.gov/mycocosm/help/clustering/index.creole>).

^d Candidate effectors were predicted using a EffectorP v.2.0 (using the stringent combination of EffectorP 2.0 and 1.0) (<http://effectorp.csiro.au/>).

^e Transposable elements predicted using REPET v2.5 (Flutre et al. 2011).

allowed placement of the *C. parasitica* mating-type locus (*MAT1*) on scaffold 2 and vegetative incompatibility loci *vic1*, *vic2*, *vic4*, *vic6*, and *vic7* on scaffolds 5, 7, 4, 3 and 6, respectively. Summary statistics were generated using the JGI Annotation Pipeline, and genome completeness was assessed using BUSCO v3.02 (Simão et al. 2015). A brief overview of the genome assembly is provided in Table 1; complete summary statistics can be accessed at the JGI website (<https://mycocosm.jgi.doe.gov/Crypa2/Crypa2.home.html>), and supplementary information is available at AgData Commons (Crouch et al. 2020). An overview of the genome assembly and summary of some key groups of genes is detailed in the following paragraphs.

Overall, there was close correspondence between EP155 karyotypes (Eusebio-Cope et al. 2009) and the 26 scaffolds of the EP155 genome assembly. Estimated *C. parasitica* chromosome sizes based on pulsed-field gel electrophoresis ranged from 3.3 to 9.7 Mb with no evidence of mini-chromosomes or accessory chromosomes that are associated with some plant pathogenic fungi (Bertazzoni et al. 2018; Eusebio-Cope et al. 2009). The 16 telomeric sequences—identified in the EP155 assembly through manual inspection of scaffold ends for the telomeric repeat sequence (T/C)TAGGG—indicated a minimum of eight chromosomes, in good agreement with cytological and electrophoretic karyotyping datasets that predicted chromosome counts of either seven or nine (Eusebio-Cope et al. 2009; Milgroom et al. 1992). Five of the EP155 scaffolds were within the estimated chromosome size range and were complete from telomere to telomere (scaffolds 1 to 4 and 8). Six scaffolds had a telomere on one end (scaffolds 5 to 7 and 9 to 11). The remaining scaffolds were smaller and lacked telomeres.

The 11 largest scaffolds ranged in size from 1.0 to 7.4 Mb and, together, comprised 99.2% of the genome assembly. Roughly half of the genome was contained in four scaffolds of at least 5.1 Mb in length. Scaffold 8 was closest in size to that predicted for chromosome 9 (3.2 versus 3.3 Mb) and contained the β -tubulin gene that was previously identified as residing on chromosome 9 by Southern analysis (Eusebio-Cope, et al. 2009). The close correspondence between karyotyping results and the draft genome sequence analysis provides a promising platform for further refinement of the sequence assembly to the chromosome level.

Gene Models and Functional Predictions

EP155 gene models were predicted and annotated using the JGI Annotation Pipeline, which combined homology-based, ab initio, and transcriptome-based gene predictors (Dawe et al. 2003; Grigoriev et al. 2014, Kuo et al. 2014; Shang et al. 2008) using the JGI expressed sequence tag (EST) pipeline. Putative protein domains were identified by querying against a local InterProScan database. After filtering for EST support, completeness, and homology support, 11,609 genes were structurally and functionally annotated from the EP155 version 2 assembly (Table 1). This number is similar to that predicted for related fungi such as *Neurospora crassa* (10,620), *Magnaporthe oryzae* (12,841), and *Fusarium graminearum* (11,640).

Functional predictions were performed using the JGI Annotation Pipeline (<https://mycocosm.jgi.doe.gov/mycocosm/home>) unless otherwise specified. Over 85% of the predicted proteins showed similarities to other proteins from the NCBI nonredundant protein database. Over 66% of the predicted proteins contained Pfam domains, with the most highly represented domains including major facilitator superfamily MFS-1 (PF07690; $n = 244$), fungal Zn(20-Cys(6) transcriptional regulatory protein (PF000172; $n = 166$), short-chain dehydrogenase/reductase SDR (PF000106; $n = 132$), cytochrome monooxygenase P450 (P450) (PF00067; $n = 116$), and serine/threonine protein kinases (PF00069; $n = 116$). As calculated using TRIBE-MCL (Enright et al. 2002), relative to the predicted proteomes of seven other members of the order Diaporthales curated by the JGI Mycoportal (Diaporthales MCL.2920), 85.7% of the predicted *C. parasitica* proteins were members of multigene family clusters.

Sm Genes and Gene Clusters

In fungi, genes involved in complex coordinated functions such as pathogenicity and SM production can occur as coregulated gene clusters, where genes are physically organized together (Keller 2019; Keller et al. 2005). Altogether, 59 predicted SM genes or gene clusters were identified from scaffolds 1 to 11 of the *C. parasitica* EP155 genome assembly. All but six of these regions contained either polyketide synthase (PKS; annotated as PKS1 to PKS31) or nonribosomal peptide synthase (NPS) genes (annotated as NPS1 to 10, NPS12 and NPS13, PKS/NPS1 to PKS/NPS5, ACS1, LYS1, FASA, and OAS1) or both. The predicted SM clusters ranged in size from just a single gene (44.8%) to clusters containing up to 39 genes, and

included genes with functions typically associated with SM clusters (e.g., transporters and transcription factors). The two largest clusters, PKS16 and PKS26 (39 and 32 genes, respectively), each spanned over 100 kb of scaffolds 10 and 1, respectively.

Cytochrome P450S

Cytochrome P450s perform a wide variety of reactions such as hydroxylation, epoxidation, dealkylation, sulfoxidation, deamination, desulfuration, dehalogenation, and nitric oxide reduction (Sono et al. 1996). Fungi in general possess extraordinarily large numbers of P450 genes, second only to plants. Putative cytochrome P450s were identified from the *C. parasitica* EP155 assembly through BLAST searches for the two conserved P450 signature domains (<http://drnelson.uthsc.edu/CytochromeP450.html>); namely, the oxygen-binding and heme-binding motifs. P450s that showed both domains were considered authentic P450s. Identified P450s were then classified into families and subfamilies based on the existing nomenclature criteria of >40% nucleotide similarity for assigning a family and >55% for a subfamily based on the classification criteria recommended by the International P450 Superfamily Nomenclature Committee. P450s that could not be assigned to any known clan based on the existing classification scheme were assigned to an appropriate clan or clans based on their relative position in the phylogenetic tree.

The *C. parasitica* EP155 genome contained 122 P450s (P450ome) classified into 76 CYP families and 101 subfamilies. The majority of the P450s in the *C. parasitica* P450 genome were orphans, with no known function. Fifteen novel subfamilies were identified (one under each of the CYP families CYP503, CYP526, CYP548, CYP567, CYP584, CYP614, CYP638, CYP639, CYP660, CYP5091, CYP5093, CYP5111, CYP5129, CYP5168, and CYP5227). In comparison with other euscomycetes, the *C. parasitica* genome had a moderately sized P450ome, comparable in size with the P450omes of *Aspergillus terreus*, *F. verticillioides*, and *M. oryzae* (123 to 126 P450s). Interestingly, unlike the majority of the euscomycetes, which do not contain basidiomycete-like P450s, the *C. parasitica* genome contained five such basidiomycete P450 homologs; namely CYP5053, CYP5227, CYP5090, CYP5093, and CYP5201. The abundance of basidiomycete P450 homologs in *C. parasitica* implies that these P450s may play a key role in the oxidation of wood-derived compounds and tree pathogenesis (Suzuki et al. 2012; Syed and Yadav 2012).

Mitochondrial Genome

The mitochondria of *C. parasitica* have been extensively studied for their association with virulence attenuation (i.e., hypovirulence), derived from either viral infection, mtDNA mutations, or mitochondrial plasmids (Baidyaroy et al. 2000; Monteiro-Vitorello et al. 1995; Polashock and Hillman 1994). Some *C. parasitica* strains harbor mitochondrial plasmids that elicit hypovirulence (Monteiro-Vitorello et al. 2000). Similarly, some strains of *C. parasitica* are subject to mitochondrial hypovirulence, a cytoplasmically transmissible form of hypovirulence associated with defects in the mitochondria (Baidyaroy et al. 2000; Monteiro-Vitorello et al. 1995). Although distinct from viral-induced hypovirulence, there are remarkable parallels between the nonviral forms of hypovirulence, including virulence attenuation and the shared alteration of transcript accumulation of over 70 genes (Allen and Nuss 2004; Monteiro-Vitorello et al. 1995, 2000). Mitochondria of some *C. parasitica* strains are also shown to harbor small RNA viruses that can reduce fungal virulence (Polashock and Hillman 1994) and can be transmitted to several other fungal species (Shahi et al. 2019). Given that *C. parasitica* EP155 is a virus-free, virulent strain of the fungus, it was not surprising that the mtDNA assembly did not share significant similarity with any of the known indicators of mitochondrial hypovirulence; for example, the assembly did not contain sequences with similarity to small subunit ribosomal DNA *InC9* (AF218209), *Cryphonectria parasitica mitovirus 1-NB631* [NC004046], or *pCRY1* [AF031368]. The ratio of UGA (= cytoplasmic terminator) to UGG codons predicted to encode Trp in the *C. parasitica* mitochondrial genome is high compared with other fungi, approximately 95%, and this correlates positively with the relatively high number of UGA codons predicted to encode Trp in *Cryphonectria* mitochondrial viruses compared with other related mitochondrial viruses (Nibert 2017).

The *C. parasitica* EP155 mitochondrial genome was annotated using MITOS v.2 (Bernt et al. 2013). Genome sequencing recovered the mitochondrial genome within a single 158,902-bp scaffold, consistent with the published physical map (Bell et al. 1996). Overall, the mtDNA

genome contained a full complement of protein-coding genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*), ribosomal RNA and ribosomal proteins (*rrns*, *rnl*, and *rps3*), and 29 tRNAs. Similar to the mitochondrial genomes of most filamentous ascomycetes, the majority of the tRNAs are clustered together, with 10 tRNA genes located side by side on each side of the *rnl* ribosomal gene. Endonuclease open reading frames (ORFs) were abundant, with 36 LAGLI-DADG homing endonucleases and 27 GIY-YIG endonucleases predicted to occupy 29% of the mtDNA assembly. Numerically, *C. parasitica* EP155 had a total of 63 predicted mitochondrial endonucleases, exhibiting one of the largest overall cohorts of such enzymes identified to date (*Sclerotinia borealis* = 61, *Rhizoctonia solani* Rhs1AP = 43, *Agaricus bisporus* = 46) (Mardanov et al. 2014).

Vegetative Incompatibility

Attempts to introduce hypovirulent strains of *C. parasitica* into North American forest ecosystems to control chestnut blight have been hindered by the difficulty of introducing hypoviruses (family Hypoviridae) into the resident fungal populations (Milgroom and Cortesi 2004). The transfer of hypoviruses from hypovirulent strains to virus-free natural populations of *C. parasitica* is dependent on hyphal anastomosis and the subsequent transfer of cytoplasmic hypoviruses from a donor strain to the virus-free recipient strain (Dawe and Nuss 2001; Hillman and Suzuki 2004; Nuss 1992, 2005; Smith et al. 2000; Van Alfen et al. 1975). This process is regulated by a diverse self-nonsel self vegetative incompatibility (*vic*) system (Milgroom and Cortesi 2004). In *C. parasitica*, six di-allelic loci controlling vegetative incompatibility in European *C. parasitica* populations have been identified by classical genetics (Cortesi and Milgroom 1998), five of which function in preventing heterokaryon formation (Choi et al. 2012; Smith et al. 2006; Zhang et al. 2014). At least two more *vic* loci are thought to function in natural populations of *C. parasitica* (Liu and Milgroom 2007; Robin et al. 2000). The recent identification, genetic characterization, and systematic disruption of incompatibility genes at six of these *vic* loci has provided a promising new opportunity for overcoming this major barrier to hypovirus dissemination across *C. parasitica* populations (Choi et al. 2012; Stauder et al. 2019; Zhang and Nuss 2016; Zhang et al. 2014).

Although *C. parasitica* vegetative incompatibility loci have historically been referred to as *vic* loci (Anagnostakis 1988), they are assumed to share characteristics with heterokaryon incompatibility (*het*) loci in other fungi (Glass and Dementhon 2006; Paoletti and Saupe 2009). Putative *het* genes in the *C. parasitica* EP155 genome were identified through a combination of BLASTp searches and syntenic comparisons with the *N. crassa* genome (Galagan et al. 2003). The following *het* proteins were used to query the EP155 assembly: generic HET domains (PF06985.2), *het* proteins from *N. crassa* (*het-c*, *het-6*, *pin-c*, *tol*, and *un24*) and *het* proteins from *Podospora anserina* (*het-c*, *het-D/het-E*, and *het-s*). Several motifs are also associated with some *het* genes and may function in programmed cell death; therefore, we also searched for NACHT (PF05729.3) and WD40 repeat (PF00400.22) domains, both of which are found associated with HET domains in *P. anserina* *het-D/het-e* (Paoletti and Clavé 2007). Because of the large number of these motifs associated with other functions, we only searched for them by comparing motifs in regions adjacent to ORFs with HET domains. We also looked for synteny in *C. parasitica* to the genomic regions of *N. crassa* that contain *het* genes. This was done simply using the VISTA tracks in the JGI genome browser by zooming out and looking for similar genes in the two genomes. We especially looked for the pairs of *het* genes that function together in *N. crassa*: *het-6/un24* and *het-c/pin-c*.

BLASTp searches of the *C. parasitica* EP155 genome sequence with other ascomycete *het* genes identified 94 homologous proteins. *C. parasitica* protein 88866 (annotated *hch1*, for *het-c* homolog) is homologous to *N. crassa* *het-c* (PF07217.2); in *C. parasitica*, the region containing *hch1* is syntenic to the region of the *N. crassa* genome containing *het-c* but we found no homolog to *pin-c* in *C. parasitica*, which is the linked interacting partner to *het-c* in *N. crassa*. Several *C. parasitica*-encoded proteins were found with high levels of similarity to *het* genes from *P. anserina*. A homolog of the *P. anserina* *het-D/E* genes, *C. parasitica* protein 84049, clearly contains conserved NACHT, WD-repeats, and HET domains. Protein 355540 is highly similar to *P. anserina* *het-C*. None of these *het* gene homologs in *C. parasitica* map to regions associated with known vegetative incompatibility function. Therefore, the genomic locations and make up of allorecognition factors appear to be distinct among species, as is the distribution of HET domain genes.

Overall, our analysis showed that there are 124 genes annotated in *C. parasitica* that contain the HET domain. This number is among the highest found yet in any ascomycete genome. *P. anserina* was previously described as containing the most recorded HET domains with 120 (Paoletti and Clavé 2007), *N. crassa* has 55, and *A. oryzae* has 38 (Fedorova et al. 2005). We found several ORFs that putatively encode HET domains with high similarity to *pin-c*, *tol*, or *het-6* from *N. crassa* but none was found in regions syntenic with the *N. crassa* homologs and, therefore, we did not name these specifically as homologs. Considerably more ORFs that have HET domains occur in filamentous ascomycete genomes than known functional *het* genes (Fedorova et al. 2005). Similarly, homologs of known *het* genes are not necessarily functional *het* genes in other species. The overall lack of synteny among HET domain genes between *C. parasitica* and other ascomycete species, and the dispersed repetitive distribution both intra- and intergenomically, supports the view that the HET domain represents a component of a mobile genetic element (Micali and Smith 2006; Paoletti and Saupe 2009). We speculate that, within species lineages, independent movement of HET domain elements into distinct genomic locations led to the evolution of unique allorecognition loci in different filamentous fungi.

Transposable Elements

Transposable elements (TEs) were identified using REPET v2.5 (Flutre et al. 2011) as described (Rivera et al. 2018). TEs representing both class I transposons (retrotransposons) and class II transposons (transposons with DNA transposition intermediates) were present in the genome of *C. parasitica* EP155 (Table 1). The TE load, approximately 14% of the total genome sequence, was largely due to the presence of 2,716 class I retroelements, comprising almost 5.0 Mb in total. Class I elements in the family Metaviridae (*Gypsy/Ty3* elements) were the most abundant group of retroelements, with 2,040 elements composing over 4 Mb of the genome and making up over 75% of all TEs. No copy of a Metaviridae retrotransposon containing an intact coding sequence was identified. Metaviridae elements were commonly located in TE-rich clusters. Of note was the region surrounding the *MAT1* locus on scaffold 5, where there are numerous retrotransposon fragments on either side of the *MAT1-2* gene. Overall, the 4-Mb scaffold where *MAT1* resides (scaffold 5) was approximately 17% transposon-derived. The presence of TEs surrounding the *Mat1* locus was consistent with mapping studies performed by Kubisiak and Milgroom (2006), where significant recombination suppression and high levels of heterogeneity in the region surrounding the *MAT1* locus was documented.

Seventeen full-length TEs with intact coding sequences were identified in the EP155 genome. Of the 17 intact transposons, nine were copies of the *hAT*-family class II transposon *Crypt1*, the only *C. parasitica* element that has been shown experimentally to be active (Linder-Basso et al. 2001).

RIP is a fungal genome defense mechanism that may mutate repeated sequences such as TEs, most commonly leading to sequences with a reduced GC content and C→T transition mutations. RIP is well defined and extremely efficient in *N. crassa* (Cambareri et al. 1998) and has been documented at a much less efficient level in several other filamentous fungi (Clutterbuck 2011; Galagan and Selker 2004). Duplicate contiguous sequences of greater than approximately 400 bases within a given genome are detected by an unknown mechanism and then disabled by methylation of cytosine bases in either copy of the duplicated sequence, followed by subsequent deamination of the methylated cytosines to thymine. *C. parasitica* EP155 TE families that contained ≥10 sequences (minimum of 1 sequence ≥300 bp) were assessed for repeat-induced point mutation (RIP) using RIPCAL v2 (Hane and Oliver 2008) using di-nucleotide frequency and alignment-based algorithms. Evidence of RIP mutation was present if di-nucleotide frequencies met the following criteria: $(\text{CpA} + \text{TpG})/(\text{A/C} + \text{GpT}) \leq 1.03$ and $(\text{TpA/ApT}) \geq 0.89$, and RIPCAL alignments showed peaks for $(\text{CA} \leftrightarrow \text{TA}) + (\text{TG} \leftrightarrow \text{AC})$ mutations. The only gene known to be required for RIP encodes a DNA methyltransferase called RIP-defective (*rid*) (Freitag et al. 2002), and this gene is present in the *C. parasitica* EP155 genome. However, RIPCAL analyses and dinucleotide frequencies showed little evidence for RIP mutation across the *C. parasitica* EP155 genome. Our detections of RIP were limited to DIRS elements ($n = 537$), helitrons ($n = 24$), and an unidentified class II element ($n = 23$). Our analysis did not detect a signature of RIP mutation from Metaviridae elements, although using a de-RIP approach, Clutterbuck (2011) identified 10 *Gypsy* elements with dinucleotide ratios consistent with RIP mutation.

Data Deposition

The genome assembly and annotations are made available via the JGI fungal genome portal MycoCosm (<https://mycocosm.jgi.doe.gov/mycocosm/home>) (Grigoriev et al. 2014). The data are also deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under accession WHUS000000000. The version described in this paper is version WHUS01000000. Supplementary tables and figures are available through the National Agricultural Library AgData Commons at <https://doi.org/10.15482/USDA.ADC/1518485>.

Acknowledgments

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer.

Literature Cited

- Allen, T. D., and Nuss, D. L. 2004. Linkage between mitochondrial hypovirulence and viral hypovirulence in the chestnut blight fungus revealed by cDNA microarray analysis. *Eukaryot. Cell* 3:1227-1232.
- Anagnostakis, S. L. 1988. *Cryphonectria parasitica*, cause of chestnut blight. *Adv. Plant Pathol.* 6:123-136.
- Anagnostakis, S. L., and Day, P. R. 1979. Hypovirulence conversion in *Endothia parasitica*. *Phytopathology* 69:1226-1229.
- Baidyaroy, D., Huber, D. H., Fulbright, D. W., and Bertrand, H. 2000. Transmissible mitochondrial hypovirulence in a natural population of *Cryphonectria parasitica*. *Mol. Plant-Microbe Interact.* 13:88-95.
- Bell, J. A., Monteiro-Vitorello, C. B., Hausner, G., Fulbright, D. W., and Bertrand, H. 1996. Physical and genetic map of the mitochondrial genome of *Cryphonectria parasitica*. *Curr. Genet.* 30:34-43.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., Pütz, J., Middendorf, M., and Stadler, P. F. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phyl. Evol.* 69:313-319.
- Bertazzoni, S., Williams, A. H., Jones, D. A., Syme, R. A., Tan, K.-C., and Hane, J. K. 2018. Accessories make the outfit: Accessory chromosomes and other dispensable DNA regions in plant pathogenic fungi. *Mol. Plant-Microbe Interact.* 31: 779-788.
- Cambarelli, E. B., Aisner, R., and Carbon, J. 1998. Structure of the chromosome VII centromere region in *Neurospora crassa*: Degenerate transposons and single repeats. *Mol. Cell. Biol.* 18:5465-5477.
- Choi, G. H., Dawe, A. L., Churbanov, A., Smith, M. L., Milgroom, M. G., and Nuss, D. L. 2012. Molecular characterization of the vegetative incompatibility genes that restrict hypovirus transmission in the chestnut blight fungus *Cryphonectria parasitica*. *Genetics* 190:113-127.
- Clutterbuck, A. J. 2011. Genomic evidence of Repeat-Induced Point mutation (RIP) in filamentous ascomycetes. *Fungal Genet. Biol.* 48:306-326.
- Cortesi, P., and Milgroom, M. G. 1998. Genetics of vegetative incompatibility in *Cryphonectria parasitica*. *Appl. Environ. Microbiol.* 64:2988-2994.
- Crouch, J. A., Dawe, A., Aerts, A., Barry, K., Churchill, A. C. L., Grimwood, J., Hillman, B. I., Milgroom, M. G., Pangilinan, J., Smith, M., Salamov, A., Schmutz, J., Yadav, J. S., Grigoriev, I. V., and Nuss, D. L. 2020. Data from: Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: A fundamental resource for an archetypical invasive plant pathogen. AgData Commons. <https://doi.org/10.15482/USDA.ADC/1518485>.
- Dalgleish, H. J., Nelson, C. D., Scrivani, J. A., and Jacobs, D. F. 2015. Consequences of shifts in abundance and distribution of American chestnut for restoration of a foundation forest tree. *Forests* 7:4-9.
- Dawe, A. L., McMains, V. C., Panglao, M., Kasahara, S., Chen, B., and Nuss, D. L. 2003. An ordered collection of expressed sequences from *Cryphonectria parasitica* and evidence of genomic microsynteny with *Neurospora crassa* and *Magnaporthe grisea*. *Microbiology* 149:2373-2384.
- Dawe, A. L., and Nuss, D. L. 2001. Hypoviruses and chestnut blight: Exploiting viruses to understand and modulate fungal pathogenesis. *Annu. Rev. Genet.* 35: 1-29.
- Enright, A. J., Van Dongen, S., and Ouzounis, C. A. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30: 1575-1584.
- Eusebio-Cope, A., Suzuki, N., Sadeghi-Garmaroodi, H., and Taga, M. 2009. Cytological and electrophoretic karyotyping of the chestnut blight fungus *Cryphonectria parasitica*. *Fungal Genet. Biol.* 46:342-351.
- Fedorova, N. D., Badger, J. H., Robson, G. D., Wortman, J. R., and Nierman, W. C. 2005. Comparative analysis of programmed cell death pathways in filamentous fungi. *BMC Genomics* 6:177.
- Flutre, T., Duprat, E., Feuillet, C., and Quenesville, H. 2011. Considering transposable element diversification in *de novo* annotation approaches. *PLoS One* 6:e16526.
- Freitag, M., Williams, R. L., Kothe, G. O., and Selker, E. U. 2002. A cytosine methyltransferase homologue is essential for repeat-induced point mutations in *Neurospora crassa*. *Proc. Natl. Acad. Sci. U.S.A.* 99:8802-8807.
- Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L. J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C. B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M. A., Werner-Washburne, M., Selitrennikoff, C. P., Kinsey, J. A., Braun, E. L., Zelter, A., Schulte, U., Kothe, G. O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamysseis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzner, R. L., Perkins, D. D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R. J., Osmani, S. A., DeSouza, C. P., Glass, L., Orbach, M. J., Berglund, J. A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D. O., Alex, L. A., Mannhaupt, G., Ebbold, D. J., Freitag, M., Paulsen, I., Sachs, M. S., Lander, E. S., Nusbaum, C., and Birren, B. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422:859-868.
- Galagan, J. E., and Selker, E. U. 2004. RIP: The evolutionary cost of genome defense. *Trends Genet.* 20:417-423.
- Glass, N. L., and Dementhon, K. 2006. Non-self recognition and programmed cell death in filamentous fungi. *Curr. Opin. Microbiol.* 9:553-558.
- Gordon, D. 2003. Viewing and editing assembled sequences using Consed. *Curr. Protoc. Bioinf.* 2:11.2.1-11.2.43.
- Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., and Shabalov, I. 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42:D699-D704.
- Hane, J. K., and Oliver, R. P. 2008. RIPCAL: A tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinf.* 9:478.
- Hillman, B. I., and Suzuki, N. 2004. Viruses of the chestnut blight fungus, *Cryphonectria parasitica*. *Adv. Virus Res.* 63:423-472.
- Jaffe, D. B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov, J. P., Zody, M. C., and Lander, E. S. 2003. Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* 13:91-96.
- Keller, N. P. 2019. Fungal secondary metabolism: Regulation, function and drug discovery. *Nat. Rev. Microbiol.* 17:167-180.
- Keller, N. P., Turner, G., and Bennett, J. W. 2005. Fungal secondary metabolism—From biochemistry to genomics. *Nat. Rev. Microbiol.* 3:937-947.
- Kubisiak, T. L., and Milgroom, M. G. 2006. Markers linked to vegetative incompatibility (*vic*) genes and a region of high heterogeneity and reduced recombination near the mating type locus (*MAT*) in *Cryphonectria parasitica*. *Fungal Genet. Biol.* 43:453-463.
- Kuo, A., Bushnell, B., and Grigoriev, I. V. 2014. Fungal genomics: Sequencing and annotation. *Adv. Bot. Res.* 70:1-52.

- Linder-Basso, D., Foglia, R., Zhu, P., and Hillman, B. I. 2001. *Crypt1*, an active *Ac*-like transposon from the chestnut blight fungus, *Cryphonectria parasitica*. *Mol. Genet. Genomics* 265:730-738.
- Liu, Y. C., and Milgroom, M. G. 2007. High diversity of vegetative compatibility types in *Cryphonectria parasitica* in Japan and China. *Mycologia* 99:279-284.
- Mardanov, A. V., Beletsky, A. V., Kadnikov, V. V., Ignatov, A. N., and Ravin, N. V. 2014. The 203 kbp mitochondrial genome of the phytopathogenic fungus *Sclerotinia borealis* reveals multiple invasions of introns and genomic duplications. *PLoS One* 9:e107536.
- Micali, C. O., and Smith, M. L. 2006. A nonself recognition gene complex in *Neurospora crassa*. *Genetics* 173:1991-2004.
- Milgroom, M. G., and Cortesi, P. 2004. Biological control of chestnut blight with hypovirulence: A critical analysis. *Annu. Rev. Phytopathol.* 42:311-338.
- Milgroom, M. G., Lipari, S. E., and Powell, W. A. 1992. DNA fingerprinting and analysis of population structure in the chestnut blight fungus, *Cryphonectria parasitica*. *Genetics* 131:297-306.
- Monteiro-Vitorello, C. B., Baidyary, D., Bell, J. A., Hausner, G., Fulbright, D. W., and Bertrand, H. 2000. A circular mitochondrial plasmid incites hypovirulence in some strains of *Cryphonectria parasitica*. *Curr. Genet.* 37:242-256.
- Monteiro-Vitorello, C. B., Bell, J. A., Fulbright, D. W., and Bertrand, H. A. 1995. A cytoplasmically transmissible hypovirulence phenotype associated with mitochondrial DNA mutations in the chestnut blight fungus *Cryphonectria parasitica*. *Proc. Natl. Acad. Sci. U.S.A.* 92:5935-5939.
- Nibert, M. L. 2017. Mitovirus UGA(Trp) codon usage parallels that of host mitochondria. *Virology* 507:96-100.
- Nuss, D. L. 1992. Biological control of chestnut blight: An example of virus-mediated attenuation of fungal pathogenesis. *Microbiol. Rev.* 56:561-576.
- Nuss, D. L. 2005. Hypovirulence: Mycoviruses at the fungal-plant interface. *Nat. Rev. Microbiol.* 3:632-642.
- Paoletti, M., and Clavé, C. 2007. The fungus-specific HET domain mediates programmed cell death in *Podospora anserina*. *Eukaryot. Cell* 6:2001-2008.
- Paoletti, M., and Saupé, S. J. 2009. Fungal incompatibility: Evolutionary origin and pathogen defense? *BioEssays* 31:1201-1210.
- Polashock, J. J., and Hillman, B. I. 1994. A small mitochondrial double-stranded (ds) RNA element associated with a hypovirulent strain of the chestnut blight fungus and ancestrally related to yeast cytoplasmic T and W dsRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 91:8680-8684.
- Rivera, Y., Salgado-Salazar, C., Veltri, D., Malapi-Wight, M., and Crouch, J. A. 2018. Genome analysis of the ubiquitous boxwood pathogen *Pseudonectria follicola*. *PeerJ* 6:e5401.
- Robin, C., Anziani, C., and Cortesi, P. 2000. Relationship between biological control, incidence of hypovirulence, and diversity of vegetative incompatibility types of *Cryphonectria parasitica* in France. *Phytopathology* 90:730-737.
- Rozen, S., and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132:365-386.
- Shahi, S., Eusebio-Cope, A., Kondo, H., Hillman, B. I., and Suzuki, N. 2019. Investigation of host range and host defense against a mitochondrially replicating mitovirus. *J. Virol.* 93:e01503-18.
- Shang, J., Wu, X., Lan, X., Fan, Y., Dong, H., Deng, Y., Nuss, D. L., and Chen, B. 2008. Large-scale expressed sequence tag analysis for the chestnut blight fungus *Cryphonectria parasitica*. *Fungal Genet. Biol.* 45:319-327.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Smith, M. L., Gibbs, C. C., and Milgroom, M. G. 2006. Heterokaryon incompatibility function of barrage-associated vegetative incompatibility genes (*vic*) in *Cryphonectria parasitica*. *Mycologia* 98:43-50.
- Smith, M. L., Micali, O. C., Hubert, S. P., Mir-Rashed, N., Jacobson, D. J., and Glass, N. L. 2000. Vegetative incompatibility in the *het-6* region of *Neurospora crassa* is mediated by two linked genes. *Genetics* 155:1095-1104.
- Sono, M., Roach, M. P., Coulter, E. D., and Dawson, J. H. 1996. Heme-containing oxygenases. *Chem. Rev.* 96:2841-2888.
- Stauder, C. M., Nuss, D. L., Zhang, D.-X., Double, M. L., MacDonald, W. L., Matheny, A. M., and Kasson, M. T. 2019. Enhanced hypovirus transmission by engineered super donor strains of the chestnut blight fungus, *Cryphonectria parasitica*, into a natural population of strains exhibiting diverse vegetative compatibility genotypes. *Virology* 528:1-6.
- Suzuki, H., MacDonald, J., Syed, K., Salamov, A., Chiaki, H., Aerts, A., Henrisaat, B., Wiebenga, A., Vankuyk, P. A., Barry, K., Lindquist, E., LaButti, K., Lapidus, A., Lucas, S., Coutinho, P., Gong, Y., Samejima, M., Mahadevan, R., Abou-Zaid, M., deVaries, R. P., Igarashi, K., Yadav, J. S., Grigoriev, I. V., and Master, E. R. 2012. Comparative genomics of the white-rot fungi, *Phanerochaete carmosa* and *P. chrysosporium*, to elucidate the genetic basis of the distinct wood types they colonize. *BMC Genomics* 13:444.
- Syed, K., and Yadav, J. S. 2012. P450 monooxygenases (P450ome) of the model white rot fungus *Phanerochaete chrysosporium*. *Crit. Rev. Microbiol.* 38:339-363.
- Van Alfen, N. K., Jaynes, R. A., Anagnostakis, S. L., and Day, P. R. 1975. Chestnut blight: Biological control by transmissible hypovirulence in *Endothia parasitica*. *Science* 189:890-891.
- Zhang, D.-X., and Nuss, D. L. 2016. Engineering super mycovirus donor strains of chestnut blight fungus by systematic disruption of multilocus *vic* genes. *Proc. Natl. Acad. Sci. U.S.A.* 113:2062-2067.
- Zhang, D.-X., Spiering, M. J., Dawe, A. L., and Nuss, D. L. 2014. Vegetative incompatibility loci with dedicated roles in allorecognition restrict mycovirus transmission in chestnut blight fungus. *Genetics* 197:701-714.