First Draft Genome Resource for the Tomato Black Leaf Mold Pathogen *Pseudocercospora fuligena*

Alex Z. Zaccaron and Ioannis Stergiopoulos[†]

Department of Plant Pathology, University of California Davis, One Shields Avenue, Davis, CA 95616-8751, U.S.A.

Abstract

Pseudocercospora fuligena is a fungus that causes black leaf mold, an important disease of tomato in tropical and subtropical regions of the world. Despite its economic importance, genomic resources for this pathogen are scarce and no reference genome was available thus far. Here, we report a 50.6-Mb genome assembly for *P. fuligena*, consisting of 348 contigs with an N₅₀ value of 0.407 Mb. In total, 13,764 protein-coding genes were predicted with an estimated BUSCO completeness of 98%. Among the predicted genes there were 179 candidate effectors, 445 carbohydrate-active enzymes, and 30 secondary metabolite gene clusters. The resources presented in this study will allow genome-wide comparative analyses and population genomic studies of this pathogen, ultimately improving management strategies for black leaf mold of tomato.

Funding

This work was supported by the National Science Foundation award number 1557995 and by the United States Department of Agriculture National Institute of Food and Agriculture Hatch Project CA-D-PPA-2185-H.

Keywords

black leaf mold, Dothideomycetes, effectors, fungi, genomics, Mycosphaerellaceae, pathogen, RNAseq

Genome Resource Announcement

Pseudocercospora fuligena (Roldan) Deighton (Deighton 1976; Roldan 1938) is a fungal pathogen that causes black leaf mold of tomato (*Solanum lycopersicum*). The disease is prevalent in tropical and subtropical regions of the world, where hot and humid conditions favor infections from the fungus (Phengsintham et al. 2012). Phylogenetically, *P. fuligena* is placed in Mycosphaerellaceae, a large family in the order Capnodiales (Ascomycetes; Dothideomycetes) that includes several notorious plant pathogens, including *Pseudocercospora fijiensis*, the causal agent of the black sigatoka disease of banana (Churchill 2011); *Cercospora beticola* that causes Cercospora leaf spot in beets (Weiland and Koch 2004); *Zymoseptoria tritici* that causes Septoria leaf blotch of wheat (Orton et al. 2011); and many others (Ohm et al. 2012).

The fungus was first described in 1938 in the Philippines (Roldan 1938) but, since then, it has been reported in many countries across the tropics and subtropics (Phengsintham et al. 2012). In the United States, it was noted as early as 1974 in Florida (Blazquez and Alfieri 1974) but tomato infections from the fungus were recently reported in Ohio (Subedi et al. 2015) and North Carolina (Lookabaugh et al. 2018) as well. Disease symptoms are characterized by irregularly shaped, chlorotic, pale-yellow to light-green spots on infected leaves which, at late stages, enlarge and coalesce, thus causing wilt and premature defoliation

[†]Corresponding author: I. Stergiopoulos; istergiopoulos@ucdavis.edu

*The *e-Xtra* logo stands for “electronic extra” and indicates that supplementary materials are published online.

The author(s) declare no conflict of interest.

Accepted for publication 2 September 2020.

Table 1. Genome assembly and gene prediction statistics of *Pseudocercospora fuligena*

Characteristics	Statistics
Genome assembly	
Assembly size (bp)	50,525,800
Number of contigs	348
Longest contig (Mb)	1.46
N ₅₀ (Mb)	0.407
L ₅₀	41
PacBio genome coverage	37x
Illumina genome coverage	124x
Genome annotation	
Number of protein-coding genes	13,764
Average gene length (bp)	1,521
Average intron length (bp)	82
Average protein length (amino-acids)	472
Gene density (genes/Mb)	272
BUSCO completeness (%)	98
GC content (%): Coding DNA	52.3
GC content (%): Overall	46.3

(Wang et al. 1995). As a result, infected tomato plants produce fewer or smaller fruit, leading to yield losses of up to 30% (Hartman 1992).

Despite its importance, little information is currently available regarding the molecular basis of pathogenesis in this fungus. An exemption is *PfAvr4*, a functional ortholog of the *Avr4* effector protein from the tomato pathogen *Cladosporium fulvum*. *PfAvr4* is shown to bind to chitin in the fungal cell walls to protect it against host chitinases during infections (Kohler et al. 2016). However, apart from these studies, other information on effectors or other virulence factors of this pathogen is lacking, partially due to the absence of genomic resources. Therefore, in order to facilitate further research on the pathogenesis of this fungus, we generated a high-quality draft genome assembly, the first for *P. fuligena*.

P. fuligena strain CBS109729 was used for genomic DNA and RNA extractions. The strain was obtained from the Westerdijk Fungal Biodiversity Institute in the Netherlands and was originally collected from an infected tomato plant in Taiwan. Genomic DNA was isolated using the UltraClean Microbial DNA isolation kit (MoBio Laboratories Inc.) according to the manufacturer's instructions. Nucleic acid sequencing was conducted at the University of California-Davis Genome Sequencing Core facility on the Illumina HiSeq2500 platform (PE 150 × 150) and the PacBio RSII system (average read length = 2.9 kb), which generated 7.4 and 2.4 Gbp of data, respectively. For transcriptome sequencing, RNA was obtained from in vitro cultures of the fungus grown in poor (KH₂PO₄ at 1 g/liter, KNO₃ at 1 g/liter, MgSO₄ · 7H₂O at 0.5 g/liter, KCl at 0.5 g/liter, sucrose at 0.5 g/liter, and glucose at 0.5 g/liter) and rich (yeast extract at 10 g/liter and glucose at 30 g/liter) liquid media at 25°C, with constant shaking at 100 rpm. Total RNA was extracted using the Qiagen RNeasy Plant Mini kit, according to the manufacturer's instructions. Sequencing of the RNAseq libraries was performed on the Illumina HiSeq2500 platform (PE 100 × 100), which generated 5.3 Gbp of data.

The absence of bacterial contamination in the reads was verified prior to genome assembly via exact *k*-mer matching (*k* = 29) performed with the *bbduk.sh* script from BBMap v38.59, utilizing the NCBI 16S ribosomal database (update 8 August 2020) as reference. No Illumina read matched a 16S ribosomal sequence, while 0.23% of the PacBio reads had a 29-mer matching bacterial sequences. Upon further inspection, 98% of these reads were identified to contain *P. fuligena* mitochondrial DNA or 18S ribosomal sequence, indicating that the reads were essentially free of bacterial contamination.

Canu v1.9 (Koren et al. 2017) was chosen to assemble the *P. fuligena* genome. Canu's built-in steps of read correction, trimming, and assembly make it a good choice for error-prone PacBio reads. In a recent genome assembly evaluation for eukaryotic organisms, Canu was considered the best among 10 PacBio assemblers due to fewer assembly errors combined with good balance of computational requirements (Jayakumar and Sakakibara 2019). To assemble the *P. fuligena* genome, the expected genome size was set to 55 Mb based on the size of genomes from closely related species available in the NCBI database. Other parameters were utilized as default. Canu's default parameters are adjusted for read depths

between 30x and 60x. By default, it also requires a minimum read overlap length of 500 bp during assembly, and reads shorter than 1 kb are not utilized. This default behavior fits well for the sequenced *P. fuligena* PacBio reads at 37x coverage and average length of 2.9 kb.

For genome polishing, Illumina reads were quality trimmed with fastp v0.20 (Chen et al. 2018) and mapped to the assembled contigs with BWA-MEM v0.7.17-r1188 (Li and Durbin 2009). The aligned reads were then used to polish the assembly three times with Pilon v1.23 (Walker et al. 2014). The resulting assembly contained 50.6 Mb organized into 348 contigs ranging from 1 kb to 1.46 Mb in size (Table 1). A custom repetitive DNA library was generated with RepeatModeler v1.0.11 (Smit and Hubley 2015) and fed to RepeatMasker v4.0.8 (Smit et al. 2015), which eventually masked 18.5% of the bases. Based on the repeat masking, 17% of the *P. fuligena* genome was covered by interspersed repeats, with 13.1% classified as retrotransposons, 1.3% as DNA transposons, and 2.6% were unclassified. The remaining 1.5% corresponded to low-complexity regions or simple repeats.

For gene prediction, RNAseq reads were mapped to the assembly with HISAT2 v2.1.0 (Kim et al. 2015) and transcripts were assembled with Trinity v2.9.1 (Grabherr et al. 2011) in genome-guided mode. The assembled transcripts along with protein sequences of *Cercospora beticola* (GCF_002742065.1), *Z. tritici* (GCF_000219625.1), and *P. fijiensis* (GCF_000340215.1) were used as evidence to predict genes with Maker v2.31.10 (Cantarel et al. 2008). An initial round of predictions was carried out with settings adjusted to find genes directly from transcript and protein evidence. In total, 1,500 gene models with all of their splicing sites and exons supported by either transcript or protein alignments were extracted. These were next used to train the *ab initio* predictors Augustus (Stanke et al. 2006) and SNAP (Korf 2004), which were subsequently imported in Maker for a new round of gene predictions. Functional annotations were carried out by querying the predicted proteins against the SwissProt/Uniprot database using BLASTP with an *e*-value < 1e-5 and searching for conserved domains with InterProScan v5 (Jones et al. 2014). Eukaryotic clusters of orthologous groups (KOGs) were identified with EggNOG mapper v2 (Huerta-Cepas et al. 2017). Secreted proteins were identified with SignalP v5.0 (Emanuelsson et al. 2007) and effectors were predicted with EffectorP v2.0 (Sperschneider et al. 2016). Genes encoding carbohydrate-active enzymes (CAZymes) were identified with dbCAN2 web server (Zhang et al. 2018) and secondary metabolite (SM) gene clusters were identified with antiSMASH v5.1.2 (Blin et al. 2019).

In total, 13,764 protein-coding genes were predicted in the *P. fuligena* genome, with an average length of 1,521 bp. Gene completeness estimated with BUSCO v4.0.5 (Simão et al. 2015), based on 1,706 conserved genes among Ascomycetes, revealed 98% completeness, with 0.4% duplication and 0.6% fragmentation. Among the predicted genes, 8,671 proteins had a homologous sequence in the SwissProt/Uniprot database with an *e*-value < 1e-5, and 9,278 proteins had a conserved Pfam domain. In total, 2,043, 1,488, and 3,819 proteins were attributed to KOG categories related to cellular processes and signaling, information storage and processing, and metabolism, respectively. Approximately 10% of all proteins (1,317) were predicted to be secreted, of which 179 were classified as candidate effectors. Included among these was the PfAvr4 effector that was previously characterized from this fungus (Kohler et al. 2016). In total, 445 genes encoding CAZymes were identified as well, which included 467 modules from the six main CAZyme families; that is, 257 glycoside hydrolases, 90 glycosyltransferases, 18 carbohydrate esterases, 24 carbohydrate-binding modules, 8 polysaccharide lyases, and 70 auxiliary activities enzymes. Finally, 30 putative SM gene clusters were identified, including 11 polyketide synthases (PKS), 11 nonribosomal peptide synthetases (NRPS), 2 hybrid PKS-NRPS, and 6 terpene synthases.

The protein-coding genes of *P. fuligena* were also used for estimating its phylogenetic relationship to other species of Mycosphaerellaceae. For this purpose, protein sequences from 12 other species of Mycosphaerellaceae with annotated sequenced genomes available at NCBI database were obtained and, along with *P. fuligena* proteins, were organized into 14,941 orthogroups with OrthoFinder v2.3.12 (Emms and Kelly 2019). OrthoFinder subsequently utilized STAG (Emms and Kelly 2018) to infer a species phylogenetic tree based on the identified orthogroups. As expected, the resulting tree supported the grouping of *P. fuligena* with other *Pseudocercospora* spp. However, although *P. fuligena* is a tomato pathogen, it surprisingly grouped with *P. eumusae* and *P. musae* that, together with *P. fijiensis*, constitute the notorious sigatoka disease complex of banana (Supplementary Fig. S1). The three banana pathogens are thought to have emerged from a recent common

ancestor (Chang et al. 2016) and, thus, the phylogenetic placing of *P. fuligena* within their clade suggests a host jump from banana to tomato.

The *P. fuligena* genome assembly presented in this study represents a valuable resource for comparative and population genomic analyses of this pathogen, and will be instrumental for further research on the molecular mechanisms of pathogenesis of this fungus on its tomato host. The Whole Genome Shotgun project has been deposited at DNA Data Bank of Japan/European Nucleotide Archive/GenBank under the accession JABCIY000000000. The version described in this article is version JABCIY010000000. The sequenced strain is stored at Westerdijk Fungal Biodiversity Institute in The Netherlands.

Acknowledgments

We thank P. W. Crous at the Westerdijk Fungal Biodiversity Institute in The Netherlands for providing us with a culture of the sequenced strain of *P. fuligena*.

Author-Recommended Internet Resources

BBMap v38.59: <https://sourceforge.net/projects/bbmap/>
dbCAN: <http://bcb.unl.edu/dbCAN2/blast.php>
EffectorP v2.0: <http://effectorp.csiro.au>
EggNOG-mapper v2.0: <http://eggno-mapper.embl.de>
NCBI 16S ribosomal database: <ftp.ncbi.nlm.nih.gov/blast/db>
SignalP v5.0: <http://www.cbs.dtu.dk/services/SignalP>
Westerdijk Fungal Biodiversity Institute: <https://wi.knaw.nl>

Literature Cited

- Blazquez, C. H., and Alfieri, S. A. 1974. Cercospora leaf mold of tomato. *Phytopathology* 64:443-445.
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., and Weber, T. 2019. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47:W81-W87.
- Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18:188-196.
- Chang, T.-C., Salvucci, A., Crous, P. W., and Stergiopoulos, I. 2016. Comparative genomics of the Sigatoka disease complex on banana suggests a link between parallel evolutionary changes in *Pseudocercospora fijiensis* and *Pseudocercospora eumusae* and increased virulence on the banana host. *PLoS Genet.* 12:e1005904.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884-i890.
- Churchill, A. C. L. 2011. *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: Progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Mol. Plant Pathol.* 12: 307-328.
- Deighton, F. C. 1976. Studies on Cercospora and allied genera: Pseudocercospora Speg., Pantospora Cif. and Cercoseptoria Petr. Page 168 in: *Mycological Papers*, Vol. 140. Commonwealth Mycological Institute, Kew, U.K.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2: 953-971.
- Emms, D. M., and Kelly, S. 2018. STAG: Species tree inference from all genes. *bioRxiv*. doi:10.1101/267914
- Emms, D. M., and Kelly, S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644-652.
- Hartman, G. L. 1992. Black leaf mold development and its effect on tomato yield. *Plant Dis.* 76:462-465.
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., and Bork, P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34:2115-2122.
- Jayakumar, V., and Sakakibara, Y. 2019. Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. *Brief. Bioinf.* 20:866-876.
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., and Hunter, S. 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30:1236-1240.
- Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12:357-360.
- Kohler, A. C., Chen, L. H., Hurlburt, N., Salvucci, A., Schwessinger, B., Fisher, A. J., and Stergiopoulos, I. 2016. Structural analysis of an Avr4 effector ortholog offers insight into chitin binding and recognition by the Cf-4 receptor. *Plant Cell* 28: 1945-1965.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722-736.
- Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinf.* 5:59.
- Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
- Lookabaugh, E. C., Thomas, A., Shew, B. B., Butler, S. C., and Louws, F. J. 2018. First report of black leaf mold of tomato caused by *Pseudocercospora fuligena* in North Carolina. *Plant Dis.* 102:442.
- Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., Condon, B. J., Copeland, A. C., Dhillon, B., Glaser, F., Hesse, C. N., Kosti, I., LaButti, K., Lindquist, E. A., Lucas, S., Salamov, A. A., Bradshaw, R. E., Ciuffetti, L., Hamelin, R. C., Kema, G. H. J., Lawrence, C., Scott, J. A., Spatafora, J. W., Turgeon, B. G., de Wit, P. J. G. M., Zhong, S., Goodwin, S. B., and Grigoriev, I. V. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog.* 8:e1003037.
- Orton, E. S., Deller, S., and Brown, J. K. M. 2011. *Mycosphaerella graminicola*: From genomics to disease control. *Mol. Plant Pathol.* 12:413-424.
- Phengsintham, P., Chuksatiro, E., McKenzie, E., Hyde, K., and Braun, U. 2012. Tropical phytopathogens 2: *Pseudocercospora fuligena*. *Plant Pathol. Quar.* 2: 57-62.
- Roldan, E. 1938. New or noteworthy lower fungi of the Philippine Islands, II. *Philipp. J. Sci.* 66:1-7.

- 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.
- Smit, A. F. A., and Hubley, R. 2015. RepeatModeler Open-1.0. Institute for Systems Biology, Seattle, WA, U.S.A. <http://www.repeatmasker.org/RepeatModeler/>
- Smit, A. F. A., Hubley, R., and Green, P. 2015. RepeatMasker Open-4.0. Institute for Systems Biology, Seattle, WA, U.S.A. <http://www.repeatmasker.org/>
- Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *New Phytol.* 210: 743-761.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. 2006. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34: W435-W439.
- Subedi, N., Testen, A. L., Baysal-Gurel, F., and Miller, S. A. 2015. First report of black leaf mold of tomato caused by *Pseudocercospora fuligena* in Ohio. *Plant Dis.* 99:285.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
- Wang, T. C., Hartman, G. L., Hsieh, W. H., and Black, L. L. 1995. Reactions of Solanaceous species to *Pseudocercospora fuligena*, the causal agent of tomato black leaf mold. *Plant Dis.* 79:661-665.
- Weiland, J., and Koch, G. 2004. Sugarbeet leaf spot disease (*Cercospora beticola* Sacc.). *Mol. Plant Pathol.* 5:157-166.
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46:W95-W101.