Draft Genome Assembly of *Passalora sequoiae* a Needle Blight Pathogen on Leyland Cypress

Warren Copes (✉ warren.copes@usda.gov)

USDA Agricultural Research Service   https://orcid.org/0000-0001-7360-7506

Jorge Ibarra Caballero

Colorado State University

Ebrahiem Babiker

USDA-ARS Southeast Area

Jane Stewart

Colorado State University

Valerie Orner

USDA-ARS Southeast Area

Alan Windham

University of Tennessee Knoxville

Renee Arias

USDA-ARS Southeast Area

Abstract

Objective

Passalora sequoiae (family Mycosphaerellaceae) causes a twig blight on Leyland cypress that requires numerous fungicide applications annually to minimize economic losses for ornamental plant nursery and Christmas tree producers. The objective was to generate a high-quality draft assembly of the whole genome as a resource for primer development and to investigate genotype diversity.

Data description:

We report here the genome sequence of *P. sequoiae* 9LC2 that was isolated from Leyland cypress 'Leighton Green' in 2017 in southern Mississippi, USA. The draft genome was obtained using Pacific Biosciences (PacBio) SMRT and Illumina HiSeq 2500 sequencing. Illumina reads were mapped to PacBio assembled contigs to determine base call consistency. Based on a total of 44 contigs with 722 kilobase (kb) average length (range 9.4 kb to 3.4 Mb), the whole genome size was estimated at 31,768,716 bp. Mapping of Illumina reads to PacBio contigs resulted in a 1000 x coverage and were used to confirm accuracy of the consensus sequences.

Objective

Passalora sequoiae (Ellis & Everh.) Y.L. Guo & W.H. Hsieh (teleomorph *Mycosphaerella*) [syn. *Cercosporidium sequoiae* (Ellis and Everh.) Baker and Partridge; *Asperisporium sequoiae* (Ellis and Everh.) Sutton and Hodges; *Cercospora sequoiae* (Ellis and Everh.)] is a fungus that causes needle blight on genera in the Cupressaceae, mainly Leyland cypress (x *Cupressocyparis leylandii*) [1, 2]. Disease symptoms of brown to grayish needles appear during the spring and progressively appear throughout the tree canopy to result in unmarketable trees (Fig. 1). Annual fungicide costs and crop loss inflict significant costs on the ornamental tree and Christmas tree industries [3, 4, 5].

The objective of this work was to sequence the whole genome of *P. sequoiae* using PacBio and Illumina to assemble contigs. A lack of genetic information for this fungus prevents utilization of genetic tools to determine genetic diversity of isolates, potential differences in virulence, and ultimately the development of control practices. Currently, only two sequence entries are listed in GenBank (NCBI), both correspond to the 18S rDNA gene of this fungus, a total of 5,476 base pairs (bp).

A problem in sampling *P. sequoiae* populations is that numerous dematiaceous hyphomycetes with morphologically similar conidia and conidioma are found in many regions (Figs. 2 and 3). Proper identification of these organisms is further complicated by the numerous name revisions over the last two decades [1, 6, 7, 8, 9, 10, 11]. Worldwide, Crous and Braun [9] recognized 659 names of *Cercospora* and approximately 550 names of *Passalora*, plus multiple other genera in a checklist containing 5,720 taxa and taxonomic reallocations in the Mycosphaerellaceae (Ascomycota). A constraining factor is only a small number of dematiaceous hyphomycetes have been included in genetic phylogenies using DNA loci, mRNA and proteins [7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19].

Data Description

A single spore isolate of *P. sequoiae* 9LC2 was recovered from a Christmas tree near Hattiesburg, MS, USA. DNA was extracted from freeze-dried mycelium [20] and sheared to approximately 20 kb fragments. SMRTbell library was prepared, then sequenced on a PacBio Sequel sequencer using one SMRT cell with P6-C4 sequencing chemistry at USDA-ARS Genomics and Bioinformatics Research Unit, Stoneville, MS, USA. Bam files were processed using Finishing Module 20.0 of CLC_Bio Workbench v.12 (Qiagen LLC, Hilden, Germany). A total of 519,499 subreads with a total of 6,612,712,889 nucleotides (nt), average length 14,247 nt, N50 21,720, were generated. Subreads were corrected and *de novo* assembled.

The 19 contigs were manually examined and split when necessary, rendering 44 contigs of 722,016 nt average and 44 x coverage and an overall GC content of 48.3%. A total of 244,368,646 reads with an average length of 148 nt after trimming were obtained from Illumina sequencing. These reads were mapped to the PacBio contigs assembled by the Finishing Module and resulted in an average of 1,011 x coverage. A small percentage of gaps, 2–4 nt in length, approximately 2–3 gaps every 150,000 nt were observed in the assembly of Illumina reads compared to the PacBio assembly, and they corresponded to microsatellites; thus, in all cases, the PacBio assembly was chosen.

Basic Local Alignment Search Tool (BLAST) [21] was used to analyze fragments of the contigs, and these corresponded to matches with the genera *Pseudocercospora*, *Cercospora* and *Passalora*. The 9360 nt contig uploaded to NCBI as "P.sequoiae_contig_43" contained the 18S rDNA gene and internal transcribed spacers of *P. sequoiae* isolate 9LC2. Alignment of this contig to the 5476 nt NCBI entry GU214667.1 from *Passalora sequoiae* had a 99.65% match, 5457/5476 nt, confirming identity of isolate 9LC2.

Structural annotation of the genome assembly was determined using MAKER v.2.31.8 [22]. The MAKER pipeline included programs 1) RepeatMasker v.4.0.6 [23] to mask interspersed repeats and low complexity DNA sequences; 2) three gene predictors: GeneMark-ES [24]; SNAP [25], trained with Sordariomycetidae proteins from the Uniprot database; and Augustus [26]; and 3) tRNAscan [27] to identify tRNA genes in the genomic sequence. The total number of genes identified by Maker was 10,657. Of those, 10,576 genes were predicted to have proteins ≥ 50 amino acids. Maker also identified 81 tRNA and 3.42% of the genome corresponded to short repetitive sequences.

DbCAN2 [27] identified 331 predicted proteins that had signatures as carbohydrate active enzymes (CAZymes). Of those 52, 9, 186, 3, 79 and 9 corresponded to auxiliary activity enzymes, carbohydrate esterases, glycoside hydrolases, polysaccharide lyases, glycosyl transferases and carbohydrate binding modules, respectively. Thirty-four proteins had blast hits to the phi-database [28].

This whole-genome project has been deposited in DDBJ/ENA/GenBank under the accession number WSQC00000000 [29]. The version described in this paper is the first version, WSQC01000000.

Table 1
Overview of data files/data sets.

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data File 1	ALL_CONTIGS_Passalora_sequoiae_RenamedDec12_2019.fsa	FASTA (.fsa)	GenBank Accession: WSQC00000000 [29]
Data set 1	Figure 1 Leyland cypress blight symptoms	.jpg	DOI: 10.15482/USDA.ADC/1518905 [30]
Data set 1	Figure 2 Passalora sporulation	.png	DOI: 10.15482/USDA.ADC/1518905 [30]
Data set 1	Figure 3 Passalora conidia	.png	DOI: 10.15482/USDA.ADC/1518905 [30]
Data set 1	Methodology	WORD (.docx)	DOI: 10.15482/USDA.ADC/1518905 [30]

Limitations

The genome sequence of a single isolate of *P. sequoiae* is being reported; thus, sequences of additional isolates would be needed to perform comparative genomics. Mapping of the Illumina sequences to PacBio contigs resulted in small gaps of low frequency; therefore, no serious limitation of data quality was evident. Reconstruction of whole chromosomes showing predicted genes and their annotation would provide characterization of the structural and functional levels.

Abbreviations

NCBI National Center for Biotechnology/Information; PacBio Pacific Biosciences;

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank [29]. The version described in this paper is version WSQC01000000. Given size limitations for uploading, raw data are available from Renee.Arias@usda.gov upon reasonable request. Due to the extremely slow growth and nutritional requirements of this organism, the type strain has been stored at USDA-ARS Thad Cochran Southern Horticultural Laboratory, Poplarville, MS. The dataset of figures and the full methodology is available in the Ag Data Commons repository maintained by the United States Department of Agriculture [30].

Competing interests

The authors declare that they have no competing interests.

Funding

Author's contributions

WEC recovered and stored the isolate, purchased sequencing services and drafted the manuscript; EB extracted DNA from lyophilized tissue and submitted samples for sequencing; RSA performed de-novo assembly, blasting and molecular analysis and led the project conceptualization; VAO submitted the genome to GenBank; JIC performed a structural bioinformatic analysis in the laboratory of JES; and ASW provided high quality photographs. WEC, EB, RSA, JIC and JES contributed to reviewing and editing. All authors agree to the publication policies of the BMC Research Notes.

Acknowledgements

Authors' information

[1] USDA ARS Thad Cochran Southern Horticultural Laboratory, Poplarville, MS, USA. [2] Colorado State University, Fort Collins, CO, USA. [3] USDA ARS National Peanut Research Laboratory, Dawson, Georgia, USA. [4] University of Tennessee, Nashville, TN.

References

[1]   Braun U, Nakashima C, Crous PW. Cercosporoid fungi (*Mycosphaerellaceae*) 1. Species on other fungi, *Pteridophyta* and *Gymnospermae*. IMA Fungus. 2013;4:265-345.

[2]   Reddy MR, Pandey P. Cercospora needle blight of Radiata pine in India. Indian Forester. 1973;99:308-309.

[3]   Leahy RM. Cercosporidium blight of Leyland cypress and related conifers. Florida Dept Agric and Consumer Serv. Plant Pathol Circ. 2000; No 397.

[4]   Williams-Woodward JL, Windham AS. Chapter 54. Leyland cypress diseases. In: Jones RK, Benson DM, editors. Diseases of woody ornamentals and trees diseases. APS press: St. Paul, MN; 2001. p. 212-215.

[5]   Little, E. L. 2017 Georgia plant disease loss estimates. UGA Coop Ext Ann Pub. 102-10; 2019. https://extension.uga.edu/publications/detail.html?number=AP102-10. Accessed 27 Feb 2020.

[6]   Baker WA, Partridge EC, Morgan-Jones G. Notes on Hyphomycetes. LXXVII. *Asperisporium sequoiae*, the causal organism of conifer needle blight, reclassified in *Cercosporidium*, with comments on the status of the genus. Mycotaxon. 2000;76:247-256.

[7]   Bensch K, Braun U, Groenewald JZ, Crous PW. The genus *Cladosporium*. Studies Mycol. 2012;72:1-401.

[8]   Braun U, Crous PW. Additions and corrections to names published in *Cercospora* and *Passalora*. Mycotaxon. 2005;92:395-416.

[9]   Crous PW, Braun U. *Mycosphaerella* and its anamorphs. 1. Names published in *Cercospora* and *Passalora*. CBS Biodiversity Series. 2003;1:1-571.

[10]   Crous PW, Summerell BA, Carnegie AJ, Wingfield MJ, Hunter GC, Burgess TI, Andjic V, Barber PA, Groenewald JZ. Unraveling *Mycosphaerella*: do you believe in genera? Persoonia. 2009;23:99-118.

[11]   Groenewald JZ, Nakashima C, Nishikawa J, Shin HD, Park JH, Jama AN, Groenewald M, Braun U, Crous PW. Species concepts in *Cercospora*: spotting the weeds among the roses. Studies Mycol. 2013;75:115-170.

[12]   Crous PW, Groenewald JZ, Mansilla JP, Hunter GC, Wingfield MJ. Phylogenetic reassessment of *Mycosphaerella* spp. and their anamorphs occurring on Eucalyptus. Studies Mycol. 2004;50:195-214.

[13]   Crous PW, Schoch CL, Hyde KD, Wood AR, Gueidan C, de Hoog GS, Groenewald JZ. Phylogenetic lineages in the *Capnodiales*. Studies In Mycol. 2009a;64:17-47.

[14]   Crous PW, Summerell BA, Shivas RG, Burgess TI, Decock CA, Dreyer LL, et al. Fungal planet description sheets: 107-127. Persoonia. 2012;28:138-182.

[15]   Crous PW, Wingfield MJ, Burgess TI, Carnegie AJ, Hardy GEStJ, Smith D, et al. Fungal planet description sheets: 625–715. Persoonia. 2017;39:270-467.

[16]   de Wit PJGM, van der Burgt A, Ökmen B, Stergiopoulos I, Abd-Elsalam KA, Aerts AL, et al. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but

also signatures of common ancestry. PLoS Genet. 2012;8(11): e1003088; doi:10.1371/journal.pgen.1003088.

[17] Goodwin SB, Dunkle LD, Zismann VL. Phylogenetic analysis of *Cercospora* and *Mycosphaerella* based on the internal transcribed spacer region of ribosomal DNA. Phytopathology. 2001;91:648-658.

[18] Luo M, Dang P, Bausher MG, Holbrook CC, Lee RD, Lynch RE, Guo BZ. Identification of transcripts involved in resistance responses to leaf spot disease caused by *Cercosporidium personatum* in peanut (*Arachis hypogaea*). Phytopathology 2005;95:381-387.

[19] Meghvansi MK, Khan MH, Gupta R, Veer V. Identification of a new species of *Cercospora* causing leaf spot disease in *Capsicum assamicum* in northeastern India. Res Microbiol. 2013;164:894-902.

[20] Babiker, EM, Hulbert, SH, and Paulitz TC. *Hyaloperonospora camelinae* on *Camelina sativa* in Washington state: Detection, seed transmission, and chemical control. Plant Dis. 2012; 96:1670-1674.

[21] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–410; doi: 10.1016/S0022-2836(05)80360-2.

[21] Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18(1):188-196.

[22] Smit AFA, Hubley R, Green P. RepeatMasker 1996. http://repeatmasker.org. Accessed 10 May 2019.

[23] Ter-Hovhannisyan V, Lomsadze A, Chernoff Y, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 2008;18:1979-1990.

[24] Zaharia M, Bolosky WJ, Curtis K, Fox A, Patterson D, Shenker S, Stoica I, Karp RM, Sittler T. Faster and more accurate sequence alignment with SNAP. 2011;arXiv:1111.5572

[25] Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 2011;27(6):757-763.

[26] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955-964.

[27] Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018; 46:95-101.

[28] Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, Sadanadan V, Khamari L, Billal S, Mohanty S, Hammond-Kosack KE. PHI-base: A new interface and further additions for the multi-species pathogen–host interactions database. Nucleic Acids Res. 2017; 45:D604–D610.

[29] Copes WE, Babiker E, Orner VA, Arias RS. *Passalora sequoiae* isolate 9LC2, whole genome shotgun sequencing project. National Center for Biotechnology Information. 2020. https://www.ncbi.nlm.nih.gov/nuccore/WSQC01000000. Accessed 24 June 2020.

[30] Copes WE, Caballero JI; Babiker E, Stewart JE, Orner VA, Windham AS, Arias RS. Data from: Draft genome assembly of *Passalora sequoiae* a needle blight pathogen on Leyland cypress. Ag Data Commons. 2020. https://doi.org/10.15482/USDA.ADC/1518905. Accessed 24 June 2020.
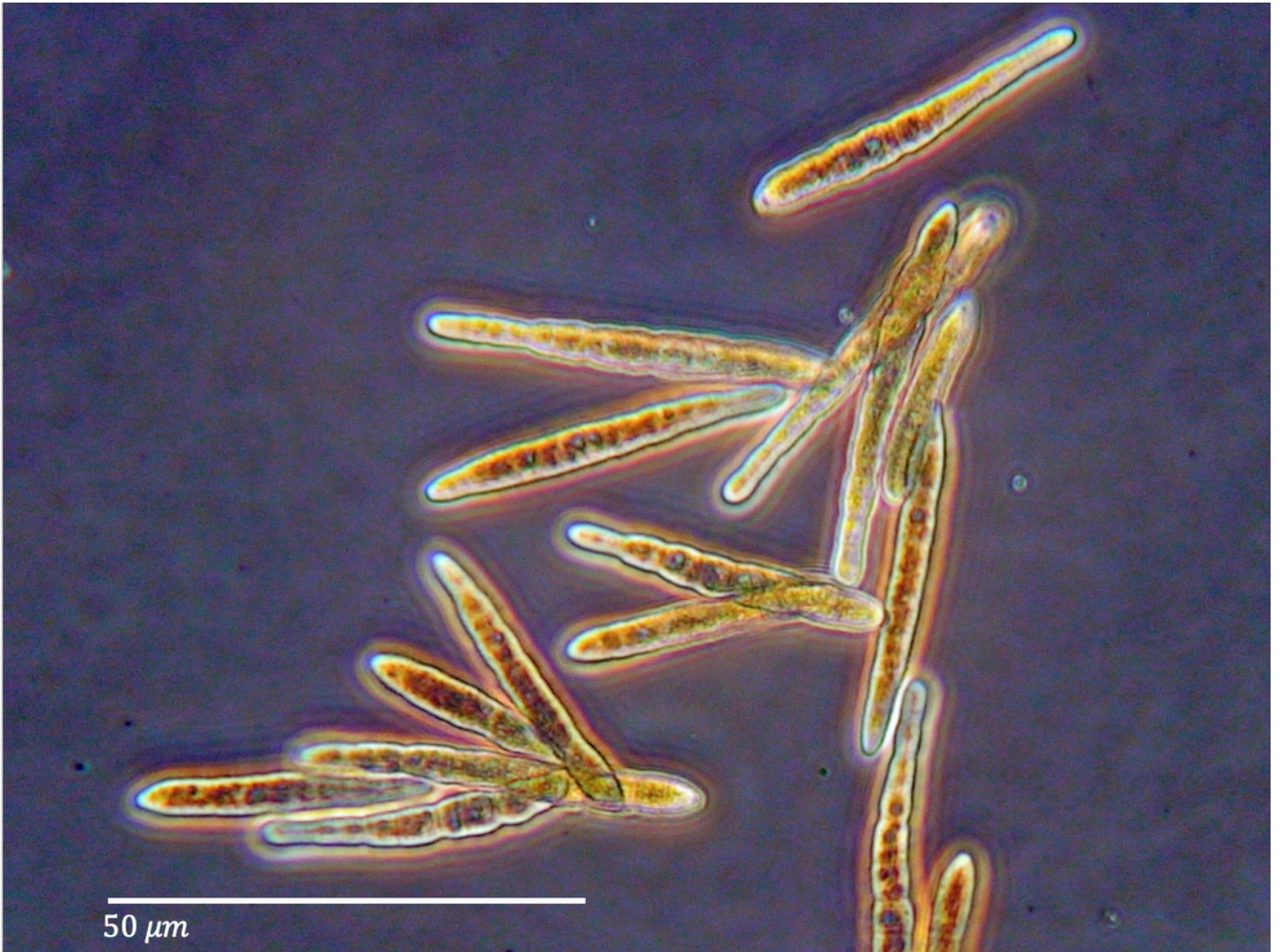
Figures

Figure 1

Leyland cypress tree showing Passalora twig blight symptoms.

Figure 2

Infected Leyland cypress leaf with sporulating conidioma of Passalora sequoiae.

Figure 3

Conidia of Passalora sequoiae.