BMC Research Notes

DATA NOTE

Open Access

Draft genome assembly of *Passalora sequoiae* a needle blight pathogen on Leyland cypress

Warren E. Copes[1*], Jorge Ibarra Caballero[2], Ebrahiem Babiker[1], Jane E. Stewart[2], Valerie A. Orner[3], Alan S. Windham[4] and Renee S. Arias[3]

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 genome of *P. sequoiae* as a resource for primer development 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 × coverage and were used to confirm accuracy of the consensus sequences.

Keywords: *Cupressocyparis leylandii*, Genome annotation, Illumina, Leyland cypress, Needle blight, PacBio

Objective

Passalora sequoiae (Ellis & Everh.) Y.L. Guo & W.H. Hsieh (syn. *Cercosporidium sequoiae* (Ellis and Everh.) Baker and Partridge) 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 gray needles appear during the spring and progressively appear throughout the tree canopy to result in unmarketable trees (Fig. 1). Annual fungicide application and crop loss inflict significant costs on the ornamental tree and Christmas tree industries [3–5].

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 three entries are listed for *Passalora* spp. in GenBank (NCBI), corresponding to the 18S rDNA gene of this fungus, a total of 5476 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, 6–12]. A further constraining factor is that only a small number of dematiaceous hyphomycetes have been included in genetic phylogenies using DNA loci, mRNA and proteins [7, 10–20]. *Mycosphaerellaceae* was recently narrowed to 120 genera based on phylogenetic data [12].

*Correspondence: Warren.Copes@usda.gov
[1] USDA ARS Thad Cochran Southern Horticultural Laboratory, Poplarville, MS 39470, USA
Full list of author information is available at the end of the article

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Fig. 1 Leyland cypress tree showing Passalora twig blight symptoms

Fig. 2 Infected Leyland cypress leaf with sporulating conidioma of *Passalora sequoiae*

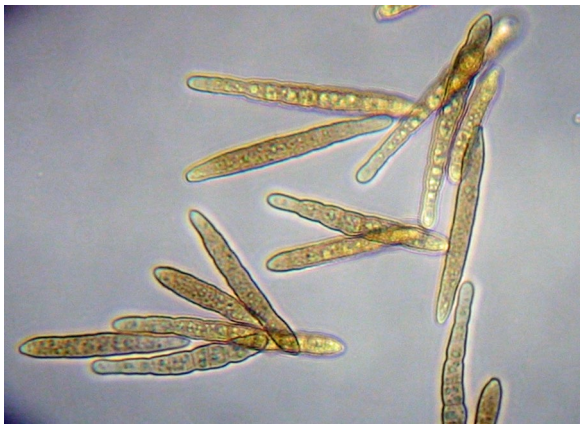

Fig. 3 Conidia of *Passalora sequoiae*

Data description

A single spore isolate of *P. sequoiae* 9LC2 was recovered from a Christmas tree near Hattiesburg, MS, USA. DNA was extracted [21] and sheared to approximately 20 kb fragments. SMRTbell library was prepared, then sequenced on a PacBio Sequel sequencer at USDA-ARS, 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 6,612,712,889 nucleotides (nt) total, average length 14,247 nt, N50 21,720, were generated. Subreads were corrected and de novo assembled. The initial 19 contigs were manually split when necessary, rendering 44 contigs of 722,016 nt average and 44 x coverage. 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 assembled contigs resulting in 1011 x average coverage. A small percentage of gaps, 2–4 nt in length, approximately 2–3 gaps every 150,000 nt were observed using Illumina reads on the PacBio assembly, and they corresponded to microsatellites; thus, in all cases, the PacBio assembly was chosen (Table 1).

Basic Local Alignment Search Tool (BLAST) [22] of a 9360 nt contig containing the 18S rDNA gene and internal transcribed spacers of *P. sequoiae* isolate 9LC2 showed a 99.65% identity with the 5476 nt NCBI entry *Passalora sequoiae* GU214667.1 [10]. The 5476 bp region of 9LC2 was used to retrieve 20 closely related sequences with 100% coverage. A Neighbor Joining [23] phylogenetic radial tree was constructed [24] using CLC Genomics Workbench 20.0 (Fig. 4), using NCBI accessions: GU214655.1; GU214656.1; GU214658.1; GU214661.1; GU214662.1; GU214664.1; GU214665.1; GU214666.1; GU214667.1; GU214668.1; GU214670.1; GU214671.1; GU214673.1; GU214678.1; GU214684.1; GU214686.1; GU214688.1; GU214697.1; GU214698.1; GU214699.1. *Passalora sequoiae* 9LC2 showed 99.7% identity to *P. sequoiae* CPC 11258, and 99.2 identity to *P. brachycarpa* CBS 115124. Though the taxonomy of *Passalora* is still being debated [12], *P. sequoiae* 9LC2 grouped with previously reported *Passalora* spp.

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_Renamed-Dec12_2019.fsa	FASTA (.fsa)	GenBank Accession: https://identifiers.org/ncbi/insdc:WSQC01000000 [33]
Data set 1	Fig. 1 LC blight symptoms	.jpg	DOI: https://www.doi.org/10.15482/USDA.ADC/1518905 [34]
Data set 1	Fig. 2 Passalora sporulation	.png	DOI: https://www.doi.org/10.15482/USDA.ADC/1518905 [34]
Data set 1	Fig. 3 Passalora conidia	.png	DOI: https://www.doi.org/10.15482/USDA.ADC/1518905 [34]
Data set 1	Fig. 4 Passalora 9LC2 phylogeny	.pdf	DOI: https://www.doi.org/10.15482/USDA.ADC/1518905 [34]
Data set 1	Methodology	WORD (.docx)	DOI: https://www.doi.org/10.15482/USDA.ADC/1518905 [34]

Structural annotation of the genome assembly was determined using MAKER v.2.31.8 [25]. The MAKER pipeline included programs 1) RepeatMasker v.4.0.6 [26] to mask interspersed repeats and low complexity DNA sequences; 2) three gene predictors: GeneMark-ES [27]; SNAP [28], trained with Sordariomycetidae proteins from the Uniprot database; and Augustus [29]; and 3) tRNAscan [30] 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 [31] 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 [32].

This whole-genome project has been deposited in DDBJ/ENA/GenBank under the accession number WSQC00000000 [33]. The version described in this paper is the first version, WSQC01000000.

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.

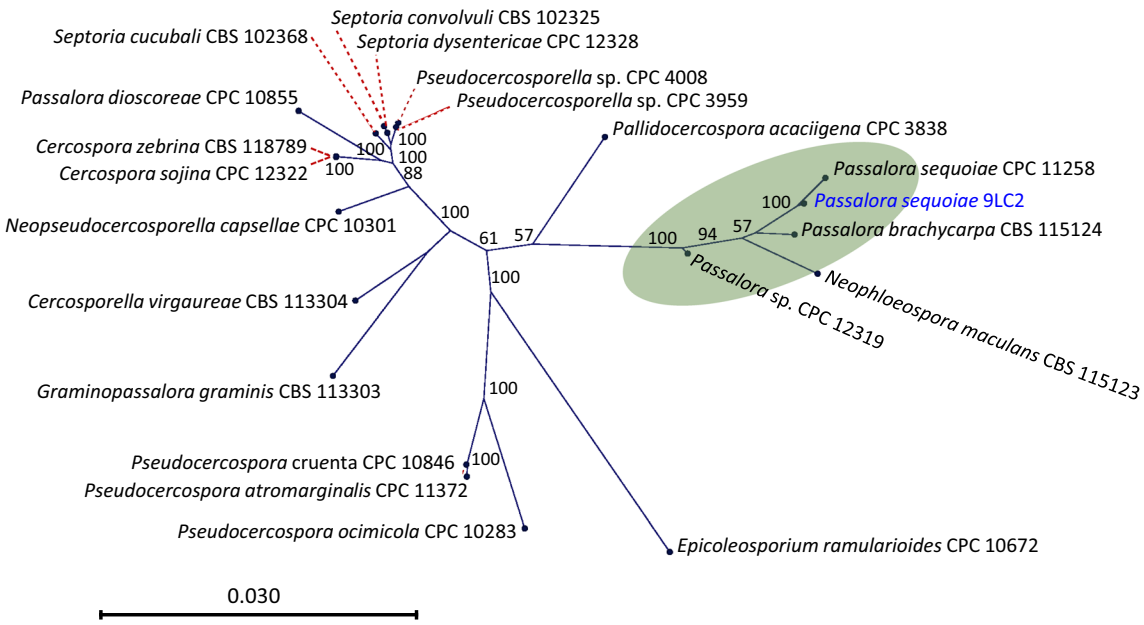

Fig. 4 Phylogeny of *Passalora sequoiae* 9LC2 and closely related species based on Neighbor-Joining analysis of 5465 nt of 18S ribosomal RNA (rRNA) gene, Internal transcribed spacer (ITS) 1, 5.8S rRNA gene, ITS2 and 28S ribosomal RNA gene partial sequence. Bootstrap of 100 resampling are shown at the nodes; scale is nucleotide substitution rate

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; BLAST: Basic Local Alignment Search Tool.

Acknowledgements
The PacBio sequencing was performed by Miss Sheron A. Simpson at USDA-ARS, Genomics and Bioinformatics Research Unit, Stoneville, MS, USA. The Illumina sequencing was performed by LC Sciences, LLC, Houston, TX, USA. We thank John T. Dobbs for assisting in structural genome annotation.

Authors' 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. All authors read and approved the final manuscript.

Funding
This project was funded by the USDA-ARS NP305 CRIS Project 6404-21430-001-00D.

Availability of data and materials
This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank [33]. The version described in this paper is version https://identifiers.org/ncbi/insdc: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 [34].

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] USDA ARS Thad Cochran Southern Horticultural Laboratory, Poplarville, MS 39470, USA. [2] Dept. of Agricultural Biology, Colorado State University, Fort Collins, CO 80523, USA. [3] USDA ARS National Peanut Research Laboratory, Dawson, GA 39842, USA. [4] Soil, Plant and Pest Center, University of Tennessee, Nashville, TN 37211, USA.

Received: 31 May 2020 Accepted: 9 October 2020
Published online: 04 November 2020

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;99:308–9.
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. St. Paul, MN: APS press; 2001. p. 212-5.
5. Little, E. L. 2017 Georgia plant disease loss estimates. UGA Coop Ext Ann Pub. 2019. p. 102-10. 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. 76:247-56.
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;75:115–70.
12. Videira SIR, Groenewald JZ, Nakashima C, Braun U, Barreto RW, de Wit PJGM, Crous PW. Mycosphaerellaceae—Chaos or clarity? Stud Mycol. 2017;87:257–421.
13. 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.
14. Crous PW, Schoch CL, Hyde KD, Wood AR, Gueidan C, de Hoog GS, Groenewald JZ. Phylogenetic lineages in the *Capnodiales*. Studies Mycol. 2009;64:17–47.
15. Crous PW, Summerell BA, Shivas RG, Burgess TI, Decock CA, Dreyer LL, et al. Fungal planet description sheets: 107-127. Persoonia. 2012;28:138–82.
16. 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.
17. 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. https://doi.org/10.1371/journal.pgen.1003088.
18. Goodwin SB, Dunkle LD, Zismann VL. Phylogenetic analysis of *Cercospora* and *Mycosphaerella* based on the internal transcribed spacer region of ribosomal DNA. Phytopathology. 2001;91:648–58.
19. 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;95:381–7.
20. 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.
21. Babiker EM, Hulbert SH, Paulitz TC. *Hyaloperonospora camelinae* on *Camelina sativa* in Washington state: detection, seed transmission, and chemical control. Plant Dis. 2012;96:1670–4.
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.
23. Saitu N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
24. Jukes TH, Cantor CR. Chapter 24. Evolution of Protein Molecules. In: Munro HN, editor. Mammalian Protein Metabolism. New York: Academic

Press; 1969. p. 21-132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7.
25. 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;18:188–96.
26. Smit AFA, Hubley R, Green P. RepeatMasker 1996. http://repeatmasker.org. Accessed 10 May 2019.
27. 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;18:1979–90.
28. 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.
29. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 2011;27:757–63.
30. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
31. 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.
32. 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:D604–10.
33. Copes WE, Babiker E, Orner VA, Arias RS. *Passalora sequoiae* isolate 9LC2, whole genome shotgun sequencing project. National Center for Biotechnology Information. 2020. https://identifiers.org/ncbi/insdc:WSQC01000000. Accessed 24 June 2020.
34. 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://www.doi.org/10.15482/USDA.ADC/1518905. Accessed 24 June 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC