



Draft Genome Sequences of Two *Cadophora* Strains Isolated from Water and a Nonalcoholic Beverage Ingredient

 Ewelina Stefanovic,^a Colm Scully,^a Jimmy Roder,^c Finbarr Murphy,^a Jos Houbraken,^b Duong Vu^b

^aAnalytical Services Laboratory, European Refreshments, Ballina, County Mayo, Ireland

^bWesterdijk Fungal Biodiversity Institute, Utrecht, the Netherlands

^cR&D Operation Brussels, S.A. Coca-Cola Services N.V., Brussels, Belgium

ABSTRACT Members of the fungal genus *Cadophora* are isolated from a variety of habitats, including plants, soil, water, food, and indoor environments. Here, we report the draft genome sequences of two strains, *Cadophora malorum* M34 and *Cadophora* sp. strain M221.

The genus *Cadophora* was introduced by Lageberg in 1927 with *Cadophora fastigiata* as the type species (1). Based on phialide morphology similarity, the genus was synonymized with *Phialophora* (2). The initial proposal to reclassify *Phialophora*-like anamorphs according to morphological characteristics (3) was confirmed by the phylogenetic analysis of internal transcribed spacer (ITS) and 28S ribosomal DNA (rDNA) sequences and showed that *Cadophora* represents a separate genus in *Ploettnerulaceae* (*Helotiales*, *Leotiomyces*) (4). Currently, *Cadophora* encompasses over 20 species (5). *Cadophora* species belong to dark septate endophytes, a polyphyletic group of fungi that produce melanized, septate hyphae (6) associated with plants, are present in soil (6–8), on decaying wood (7, 8), and in water (9), and can survive in extreme habitats, such as Antarctica (10). The two *Cadophora* strains analyzed in this study were isolated from water (*Cadophora malorum* M34 [=CBS 147375]) and a low-pH ingredient used in a nonalcoholic beverage (*Cadophora* sp. strain M221 [=CBS 147358]) and were deposited in the CBS culture collection at the Westerdijk Fungal Biodiversity Institute, the Netherlands.

Both strains were grown on acidified malt extract agar (5 days at 25°C for M34 and 7 days at 20°C for M221). DNA was isolated with a DNeasy UltraClean microbial kit (Qiagen GmbH, Germany), and genomic libraries were prepared with the Nextera XT DNA library preparation kit (Illumina, Inc., USA). The 250-bp paired-end read sequencing was performed on an Illumina HiSeq platform (MicrobesNG, UK). The *de novo* assembly was performed on samples using SPAdes v3.7 (11), and the quality of the genome assemblies was assessed by QUAST (26). Funannotate v1.7.4 (12) was used for gene prediction and annotation with the protein evidence from a publicly available genome of a *Cadophora* species (13). Specifically, the protein evidence was aligned to the genome using DIAMOND v0.9.24 (14) and Exonerate v2.20 (15). The gene predictions were performed using GeneMark-ES (16), BUSCO 2 (17), AUGUSTUS v3.3.2 (18), and EvidenceModeler (19). The gene models were functionally annotated with Pfam (20), a DIAMOND search on UniProt v2020_24, InterProScan 5 (21), eggNOG v5.0 (22), MEROPS (23), CAZyme (24), BUSCO 2 (dikarya_odb9), and SignalP v5.0 (25).

Sequence assemblies indicated coverages of 203× and 57× for the strains M34 and M221, respectively. The length of the *Cadophora malorum* M34 genome was 48,945,047 bp, consisting of 945 nonoverlapping contigs, with a contig N_{50} value of 121,946 bp and a maximum contig size of 931,894 bp. A total of 16,400 protein-coding genes were identified, along with 107 RNAs. In the case of *Cadophora* sp. M221, the assembly yielded a genome sequence of 60,032,319 bp, consisting of 6,664

Citation Stefanovic E, Scully C, Roder J, Murphy F, Houbraken J, Vu D. 2021. Draft genome sequences of two *Cadophora* strains isolated from water and a nonalcoholic beverage ingredient. *Microbiol Resour Announc* 10:e00368-21. <https://doi.org/10.1128/MRA.00368-21>.

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2021 Stefanovic et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ewelina Stefanovic, estefanovic@coca-cola.com.

Received 30 April 2021

Accepted 6 June 2021

Published 1 July 2021

contigs. The maximum contig size was 316,907 bp, and the contig N_{50} value was 47,756 bp, while a total of 17,654 protein-coding genes and 124 tRNAs were identified. The GC content of the genomes was 48.01% for M34 and 46.46% for M221. A summary of genome completeness using BUSCO 2 (dikarya_odb9) revealed 98.6% complete, 98.2% complete and single-copy, 0.4% complete and duplicated, 0.8% fragmented, and 0.6% missing orthologs for M34. For M221, these values were 98.1%, 97.7%, 0.4%, 1.4%, and 0.5%, respectively.

These sequencing data present a valuable addition to the set of currently available *Cadophora* genomes and enable deeper understanding of the features and further comparative genome analysis of strains of this genus.

Data availability. The whole-genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession numbers [JAFJYH000000000](https://doi.org/10.2307/3754512) for M34 and [JAFJYG000000000](https://doi.org/10.1101/gr.081612.108) for M221. The raw reads have been deposited under the SRA accession numbers [SRR14354866](https://doi.org/10.1007/978-1-4939-9173-0_14) and [SRR14354867](https://doi.org/10.1007/978-1-4939-9173-0_14) for M34 and [SRR14354808](https://doi.org/10.1007/978-1-4939-9173-0_14) and [SRR14354809](https://doi.org/10.1007/978-1-4939-9173-0_14) for M221.

ACKNOWLEDGMENT

We are grateful to Tilen Kranjc for his valuable help during the submission process.

REFERENCES

- Lagerberg T. 1927. Biological and practical researches into blueing in pine and spruce. *Sven Skogsvardsforen Tidskr* 25:145–272.
- Conant NF. 1937. The occurrence of a human pathogenic fungus as a saprophyte in nature. *Mycologia* 29:597–598. <https://doi.org/10.2307/3754512>.
- Gams W. 2000. *Phialophora* and some similar morphologically little-differentiated anamorphs of divergent ascomycetes. *Stud Mycol* 45:187–199.
- Harrington TC, McNew DL. 2003. Phylogenetic analysis places the *Phialophora*-like anamorph genus *Cadophora* in the *Helotiales*. *Mycotaxon* 87:e151.
- Maciá-Vicente JG, Piepenbring M, Koukol O. 2020. Brassicaceous roots as an unexpected diversity hot-spot of helotialean endophytes. *IMA Fungus* 11:1–23. <https://doi.org/10.1186/s43008-020-00036-w>.
- Walsh E, Duan W, Mehdi M, Naphri K, Khiste S, Scalera A, Zhang N. 2018. *Cadophora meredithiae* and *C. interclivum*, new species from roots of sedge and spruce in a western Canada subalpine forest. *Mycologia* 110:201–214. <https://doi.org/10.1080/00275514.2017.1406748>.
- Travadon R, Lawrence DP, Rooney-Latham S, Gubler WD, Wilcox WF, Rolshausen PE, Baumgartner K. 2015. *Cadophora* species associated with wood-decay of grapevine in North America. *Fungal Biol* 119:53–66. <https://doi.org/10.1016/j.funbio.2014.11.002>.
- Bien S, Damm U. 2020. *Arboricolonus simplex* gen. et sp. nov. and novelties in *Cadophora*, *Minutiella* and *Proliferodiscus* from *Prunus* wood in Germany. *MycKeys* 63:119–161. <https://doi.org/10.3897/mycokeys.63.46836>.
- Novak Babic M, Gunde-Cimerman N. 2021. Water-transmitted fungi are involved in degradation of concrete drinking water storage tanks. *Microorganisms* 9:160. <https://doi.org/10.3390/microorganisms9010160>.
- Arenz BE, Held BW, Jurgens JA, Farrell RL, Blanchette RA. 2006. Fungal diversity in soils and historic wood from the Ross Sea region of Antarctica. *Soil Biol Biochem* 38:3057–3064. <https://doi.org/10.1016/j.soilbio.2006.01.016>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Palmer J, Stajich J. 2016. Funannotate: a fungal genome annotation and comparative genomics pipeline (v1. 5.2). Zenodo <https://doi.org/10.5281/zenodo3354704>.
- Knapp DG, Németh JB, Barry K, Hainaut M, Henrissat B, Johnson J, Kuo A, Lim JHP, Lipzen A, Nolan M, Ohm RA, Tamás L, Grigoriev IV, Spatafora JW, Nagy LG, Kovács GM. 2018. Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. *Sci Rep* 8:1–13. <https://doi.org/10.1038/s41598-018-24686-4>.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. <https://doi.org/10.1186/1471-2105-6-31>.
- Ter-Hovhannisyán V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res* 18:1979–1990. <https://doi.org/10.1101/gr.081612.108>.
- Seppy M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol* 1962:227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res* 34:W435–W439. <https://doi.org/10.1093/nar/gkl200>.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7–R22. <https://doi.org/10.1186/gb-2008-9-1-r7>.
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res* 49: D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
- Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
- Rawlings ND, Waller M, Barrett AJ, Bateman A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 42:D503–D509. <https://doi.org/10.1093/nar/gkt953>.
- Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, Yi H, Yang Z, Yin Y. 2018. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res* 46:D516–D521. <https://doi.org/10.1093/nar/gkx894>.
- Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.