

Draft genome sequence of a black yeast fungus *Exophiala xenobiotica* isolated from La Brea Tar Pits

Tania Kurbessoian,¹ Danielle Stevenson,² Renata Haro,¹ Samantha Ying,² Jason E. Stajich¹

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT A 30.28 Mb draft genome sequence was assembled and annotated for the melanized ascomycetous fungus *Exophiala xenobiotica* NRRL_64630 (Pezizomycotina; Chaetothyriales) isolated from La Brea Tar Pits, Los Angeles, California. Species identification was made by phylogenetic assessment of the Internal Transcribed Spacer. This is the first isolated fungal species from this historic space.

KEYWORDS La Brea Tar Pits, extremophilic, black yeast, exophiala

Popular references to tar pits entertain vivid images of large mastodons and saber-tooth cats ensnared in a sticky substance that oozes from the ground (1). Asphalt and tar are similar hydrocarbon substrates, asphalt is released into seeps and collected in large viscous pools while tar is man-made (2). Seeps are found in many parts of the world, from Peru's Talara Tar Pits to the La Brea Tar Pits (LBTP) (3). Culturally, the indigenous Los Angeles Native Gabrieleño-Tongva used tar as an adhesive on their boats and tools as far back as 10,000 years ago (3).

The role microbes and fungi can play in the remediation of tar-soaked soils remains an important opportunity for new technology. Fossils were collected from the LBTP in the early 1900s by UC Berkeley, LA High School, and the Southern California Academy of Sciences eventually leading to LBTP in the 1970s (4).

First, 5 mm of soil was collected from the La Brea site "Project 23" neighboring the Los Angeles County Museum of Art parking structure Box 5B (now found within the museum grounds). Cultures of fungi were obtained by plating dilutions of soils on nutrient-rich Malt Extract Yeast Extract (MEYE) media. Culturing followed steps enriching for melanized fungi and the addition of antibiotics to exclude bacteria (5). Growth of one black yeast strain was isolated and subcultured on MEYE at room temperature twice for axenic confirmation and designated TK_68. Genomic DNA was extracted from yeast colonies grown for 1 week on MEYE plates using a cetyltrimethylammonium bromide (CTAB) protocol (6). Genomic DNA was measured by Nanodrop and Qubit and diluted to ~28 ng/L. DNA sequencing libraries were prepared in the Institute for Integrative Genome Biology Genomics Core at University of California, Riverside (Riverside, CA) with Nextera DNA Flex Library kit and sequenced with 2 × 150 bp on Illumina NovoSeq 6000 in the QB3 Genomics Facility at the University of California, Berkeley.

The TK_68 strain was identified as (100% identity) *Exophiala xenobiotica*, accession number [KX426972.1](#), by PCR amplifying the Internal Transcribed Spacer (ITS) region (28S + ITS1 + 5.8S + ITS2 + 18S) using ITS1F and ITS4 primers and amplification protocols (7), sequenced with Sanger sequencing, and BLASTN search to the NCBI refseq ITS database. To further confirm species identity, ITS regions were extracted with ITSx (v.1.1.3) (8) from public *Exophiala* genomes in GenBank and outgroup *Aspergillus fumigatus* ATCC 1022. A multiple sequence alignment of the ITS region was constructed with MUSCLE (v.5.1) (9), trimmed with ClipKit (v.1.3.0) (10), and the phylogenetic relationships inferred with

Editor Leighton Pritchard, University of Strathclyde, Glasgow, Scotland, United Kingdom

Address correspondence to Jason E. Stajich, jason.stajich@ucr.edu.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 28 September 2023

Accepted 10 December 2023

Published 5 January 2024

Copyright © 2024 Kurbessoian et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

TABLE 1 Strain and species designation, isolation source, sequencing read, assembly, and annotation statistics^a

Species	Location	No. of read pairs	Coverage	No. of contigs	Genome size	Contig L ₅₀	Contig N ₅₀	G + C content (%)	Genome completion (BUSCO %)	Genome duplication (BUSCO %)	No. of genes	Telomeres found (forward, reverse, T2T)
<i>Exophiala xenobiotica</i> NRRL 64630	La Brea Tar Pits, Los Angeles, California	9,065,778	60.43×	682	30.28 Mb	15	676 kb	51.86	97.9%	0.0%	11,317	7 F 8 R, 0 T2T (both)

^aThe species identification number for NRRL is listed, as is the location where the soil was collected from. The number of reads was used to help determine the coverage value. Genome assembly calculations include a number of contigs, genome size, N₅₀, L₅₀, and G + C content, while genome annotation results include a number of genes predicted and annotated. BUSCO completion statistics and comparisons were determined using the *sordariomycetes_odb10* database with 3817 genes. Telomeres were calculated using the *find_telomere.py* script (Hiltunen et al. 2021).

IQTree2 (v.2.2.2.6) (11). The strain was deposited in the USDA-ARS NRRL collection as accession NRRL 64630 (Fig. 1).

The NRRL 64630 genome was assembled using the AATF pipeline (v.0.2.3) (12–16) performing filtering and trimming steps for data quality, relying on SPAdes (3.15.4) (17) for assembly, AATF tools for contaminant removal, and assembly polishing. Telomeres were calculated using the *find_telomere.py* script (18). BUSCO (v.5.4.4) *ascomycota_odb10* database (19) was used to assess the completeness of the assembly. Genome annotation was performed with Funannotate (v.1.8.10) (20–38). Default parameters for the underlying tools were applied throughout. Genome sequencing, assembly, and protein-coding gene annotation statistics are summarized in Table 1.

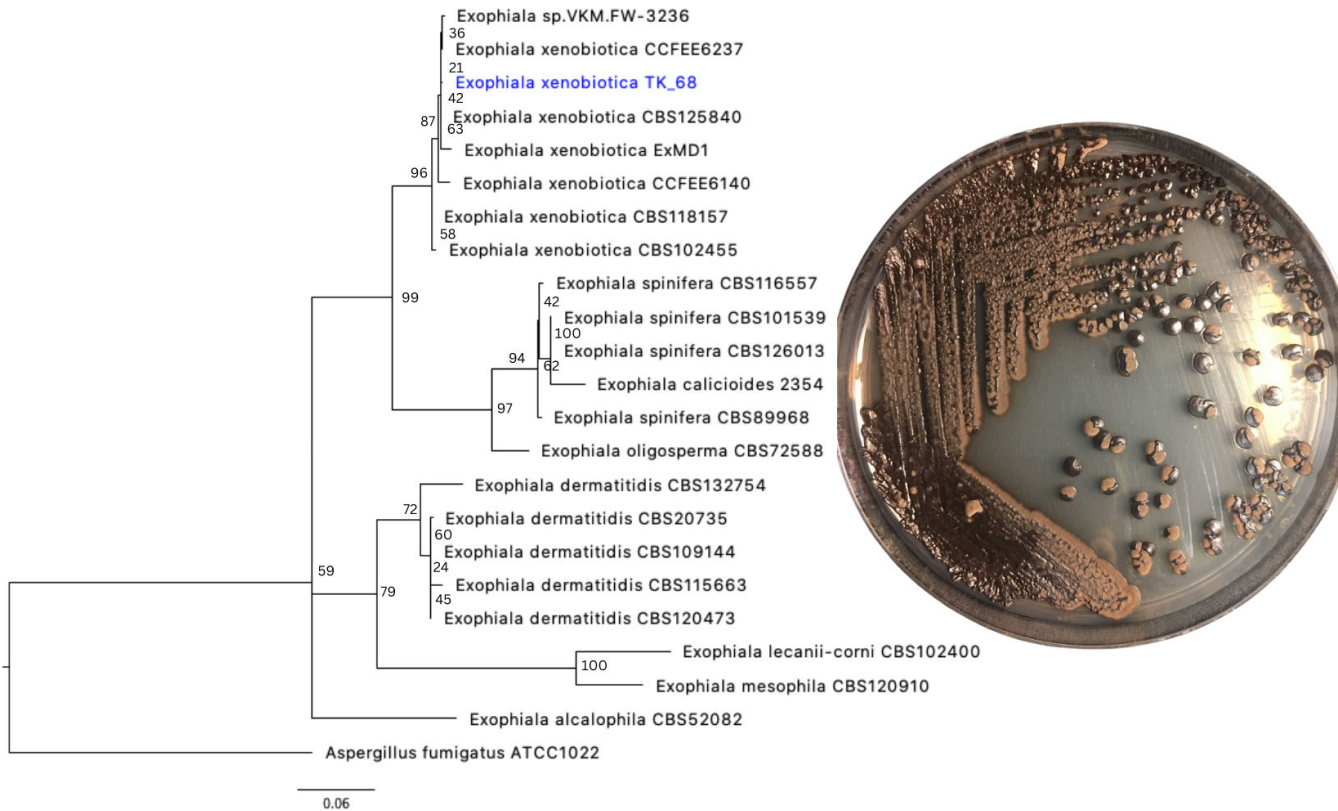


FIG 1 Agar culture of *Exophiala xenobiotica*, axenically isolated from the soil collected in box 5B-P23 C2A, and a phylogenetic tree describing our species along with respective 22 NCBI isolates. (A) Species identity was inferred from the phylogenetic tree constructed from ITS region sequences for our species from this study (in blue), 21 reference sequences from *Exophiala* genera, and rooted with one sequence of outgroup taxa.

ACKNOWLEDGMENTS

We thank the staff at the La Brea Tar Pits who assisted in sample collection and permitting. Genome assembly and annotation were performed on the IIGB High-Performance Computing Cluster supported by NSF DBI-1429826, DBI-2215705, and NIH S10-OD016290 grants. J.E.S. is a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities and supported by the U.S. Department of Agriculture, National Institute of Food and Agriculture Hatch projects CA-R-PPA-211-5062-H.

AUTHOR AFFILIATIONS

¹Department of Microbiology and Plant Pathology, Institute for Integrative Genome Biology, University of California, Riverside, California, USA

²Department of Environmental Studies, University of California, Riverside, California, USA

AUTHOR ORCID*s*

Tania Kurbessoian  <http://orcid.org/0000-0003-3946-0867>

Jason E. Stajich  <http://orcid.org/0000-0002-7591-0020>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	DBI-1429826	Jason E. Stajich
National Science Foundation (NSF)	DBI-2215705	Jason E. Stajich
HHS National Institutes of Health (NIH)	S10-OD016290	Jason E. Stajich

AUTHOR CONTRIBUTIONS

Tania Kurbessoian, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft | Danielle Stevenson, Data curation, Validation | Renata Haro, Data curation, Validation | Samantha Ying, Validation | Jason E. Stajich, Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Validation, Writing – review and editing

DATA AVAILABILITY

This Whole Genome project has been deposited at DDBJ/ENA/GenBank under the accession [JAPDRM000000000.1](https://www.ncbi.nlm.nih.gov/nuclseq/JAPDRM000000000.1) and SRA accession of [SRR22028093](https://www.ncbi.nlm.nih.gov/sra/SRR22028093). Genome assembly, annotation and phylogenetic assessment pipeline and related code are archived at DOI [10.5281/zenodo.8021900](https://doi.org/10.5281/zenodo.8021900) (39). The culture has been deposited in the USDA ARS Culture Collection (NRRL) under the strain accession NRRL 64630.

REFERENCES

1. Wright T. 1987. Geological setting of the rancho La Brea tar pits:87–91.

2. Murali Krishnan J, Rajagopal KR. 2003. Review of the uses and modeling of bitumen from ancient to modern times. *Appl Mech Rev* 56:149–214. <https://doi.org/10.1115/1.1529658>

3. O'Reilly K. 2021. *Asphalt: A history*. U of Nebraska Press.

4. "La Brea tar pits". University of California museum of Paleontology. Retrieved 26 Apr 2023. <https://ucmp.berkeley.edu/quaternary/labrea.html>.

5. Kurbessoian T. 2019. Cultivating melanized fungi from biological soil crust and rock surfaces v1. Available from: <https://doi.org/10.17504/protocols.io.3fxgjpjn>

6. Carter-House D, Stajich JE, Unruh S, Kurbessoian T. 2020. Fungal CTAB DNA extraction. *Protocols io*. Available from: <https://doi.org/10.17504/protocols.io.bhx8j7rw>

7. White TJ, Bruns T, Lee S, Taylor J. 1990. 38 - Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. *PCR Protocols*. San Diego: Academic Press. p. 315–322.

8. Bengtsson - Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez - García M, Ebersberger I, de Sousa F, Amend A, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand YJK, Sanli K, Eriksson KM, Vik U, Veldre V, Nilsson RH, Bunce M. 2013. Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing . *Methods Ecol Evol* 4:914–919. <https://doi.org/10.1111/2041-210X.12073>

9. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <https://doi.org/10.1186/1471-2105-5-113>

10. Steenwyk JL, Buida TJ 3rd, Li Y, Shen X-X, Rokas A, Hejnal A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol* 18:e3001007. <https://doi.org/10.1371/journal.pbio.3001007>

11. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa131>
12. Palmer JM, Stajich JE. 2022. Automatic Assembly for the fungi (AAFTF): genome assembly pipeline. Available from: <https://zenodo.org/record/6326242>
13. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
14. Bushnell B. 2014. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. Joint Genome Institute.
15. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>
16. Titus Brown C, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. *J Open Source Softw* 1:27. <https://doi.org/10.21105/joss.00027>
17. 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>
18. Hiltunen M, Ament-Velázquez SL, Johannesson H. 2021. The assembled and annotated genome of the fairy-ring fungus *Marasmius oreades*. *Genome Biol Evol* 13:evab126. <https://doi.org/10.1093/gbe/evab126>
19. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>
20. Palmer JM, Stajich J. 2020. Funannotate V1.8.1: eukaryotic genome annotation. Available from: <https://zenodo.org/record/4054262>
21. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* 117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>
22. Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11. <https://doi.org/10.1186/s13100-015-0041-9>
23. Smit AFA. 2004. Repeat-masker open-3.0. Available from: <http://www.repeatmasker.org>
24. Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>
25. 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>
26. Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform* 2:lqaa026. <https://doi.org/10.1093/nargab/lqaa026>
27. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20:2878–2879. <https://doi.org/10.1093/bioinformatics/bth315>
28. 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>
29. 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>
30. 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. <https://doi.org/10.1186/gb-2008-9-1-r7>
31. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–W57. <https://doi.org/10.1093/nar/gkw413>
32. Jones P, Binns D, Chang H-Y, 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>
33. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222–30. <https://doi.org/10.1093/nar/gkt1223>
34. 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>
35. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>
36. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* 46:D624–D632. <https://doi.org/10.1093/nar/gkx1134>
37. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. *Nucleic Acids Res* 46:W200–W204. <https://doi.org/10.1093/nar/gky448>
38. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>
39. kurbessoian T. 2023. Tania-k/LBTP_Exophiala_xenobiotica: Version1 Exophiala Xenobiotica. Available from: <https://doi.org/10.5281/zenodo.8021900>