AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Environmental Microbiology | Announcement

Draft genome sequences of related *Paeniglutamicibacter* sp. isolates from two disparate cave systems

Earl Middlebrook,[1] Jennifer Hathaway,[2] Anastasia Pittis,[2] George Abernathy,[3] Karen Davenport,[1] Cheryl Gleasner,[1] Kimberly Mcmurry,[1] Julia Kelliher,[1] Andrew W. Bartlow,[1] Migun Shakya,[1] Michaeline Albright,[4] Anand Kumar,[3] Blake Hovde,[1] Brett Youtsey,[1] Buck Hanson,[5] Aaron Robinson,[1] Mark Flynn,[1] Shannon Johnson,[1] Diana Northup,[2] Armand E. K. Dichosa[1]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We present the genome assemblies of two similar *Paeniglutamicibacter* strains, ORCA_105 and MACA_103, isolated from Mammoth and Oregon Cave systems, respectively. These closely related, but distinct genomes will provide a resource for those studying genomic adaptation to caves.

KEYWORDS cave, draft genome, bacterial isolates, 16S sequencing, whole genome sequencing, bioinformatics

Many bacterial lineages thrive in extreme terrestrial cave systems (1, 2) and have yet to be identified. Here, we present draft genomes of two *Paeniglutamicibacter* isolates from Oregon and Kentucky caves with identical 16S rRNA genes.

Isolate ORCA_105 was collected on 3 October 2018 from the floor of Oregon Caves National Monument (ORCA) 68 m from the cave entrance (Oregon, USA; UTM NAD83 10T 466241, 4661543) using a sterile swab and inoculated onto a ½ R2A agar plate with nystatin (150,000 U/L). Isolate MACA_103 was collected on 9 April 2019 from Mammoth Cave National Park (MACA) ceiling (Kentucky, USA; UTM NAD83 16S 581551, 4109203) 400 m from the cave entrance and inoculated onto ⅕ R2A agar with 5 g/L of pulverized cave rocks and nystatin (150,000 U/L). Isolates with unique morphologies were subcultured using streak isolation and incubation at 8°C until the isolate appeared pure, at which time DNA extraction was performed from the lawn using a 1.0 µL loop.

DNA extractions were performed using the Biospec BeadBeater (Bartlesville, OK) for 1.5 min at medium speed and DNeasy UltraClean Microbial Kit (Qiagen Germantown, MD). Genomic libraries were generated with NEBNext Ultra DNA II Library Kit (New England Biolabs, Cat. #E7645L) and sequenced on an Illumina NextSeq 500 with the Mid Output Kit v2.5 (Illumina, Cat. #20024905) generating 2 × 151 bp reads. Sequence analysis was performed with the EDGE bioinformatics UI platform (v2.4.0) (3). Reads were first trimmed/filtered using *faQC* (v2.08) (4) with three bases clipped from each end and removing reads below 20 average quality and/or 50 bp length, then assembled using *IDBA* (v1.1.1) (5) with options "--pre_correction --mink 31—maxk 121—step 20 min_contig 200." Contigs with less than 100× coverage, identified by *BWA* (v0.7.12) (6), were removed. *CheckM* (v1.2.2) (7) was used to estimate completeness and contamination, whereas *Prokka* (v1.14.5) (8) was used to predict CDSs, tRNAs, and rRNAs. Assembly details are provided in Table 1.

The end-trimmed alignment of ORCA_105 and MACA_103 16S rRNA gene sequences shows that they are identical. By BLAST search (9–11), full-length ORCA_105 and MACA_103 16S gene sequences show very high similarity to the 16S rRNA gene of *Paeniglutamicibacter sulfureus* (NR_026237.1) at 99.47% and 99.85%, respectively, and *P. antarcticus* (NR_115079.1) at 99.47% and 99.70%, respectively. Using CEANIA (https://github.com/eamiddlebrook/CEANIA), these two strains have an average nucleotide

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Armand E. K. Dichosa, armand@lanl.gov.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 7 February 2025
Accepted 5 April 2025
Published 29 April 2025

TABLE 1 Genome assembly statistics, BLAST results, and accession numbers

Characteristic	ORCA_105	MACA_103
No. of reads	32,351,542	32,156,656
Assembly size (MBP)	4.42	4.80
No. of contigs	118	147
N50 values (BP)	89,925	88,696
Max contig len	237,673	202,090
Avg. coverage	1062	979
GC content (%)	0.64968	0.65791
Best hit taxon	*P. sulfureus* and *P. antarcticus*	*P. sulfureus*
Best hit accn.	NR_026237.1 and NR_115079.1	NR_026237.1
Est. complete (%)	99.77	99.54
Est. contamination (%)	1.87	2.29
No. coding sequences[a]	3,934	4,257
No. rRNA[a]	7	4
No. tRNA[a]	62	62
GenBank assembly accn.[b]	GCF_045798545.1	GCF_045798555.1
NCBI BioSample accn.	SAMN44524352	SAMN44524351
NCBI SRA accn.	SRX27112100	SRX27112101

[a] Annotations provided by Prokka (v1.14.5) (8).

[b] Annotations of GenBank entries were provided by PGAP (12).

identity of 0.903 across their core genes. A phylogenomic tree with related bacteria shows a close association between ORCA_105 and MACA_103 and *P. sulfureus* As4PL and *P. quisquiliarum* ABSL32-1, respectively (Fig. 1).

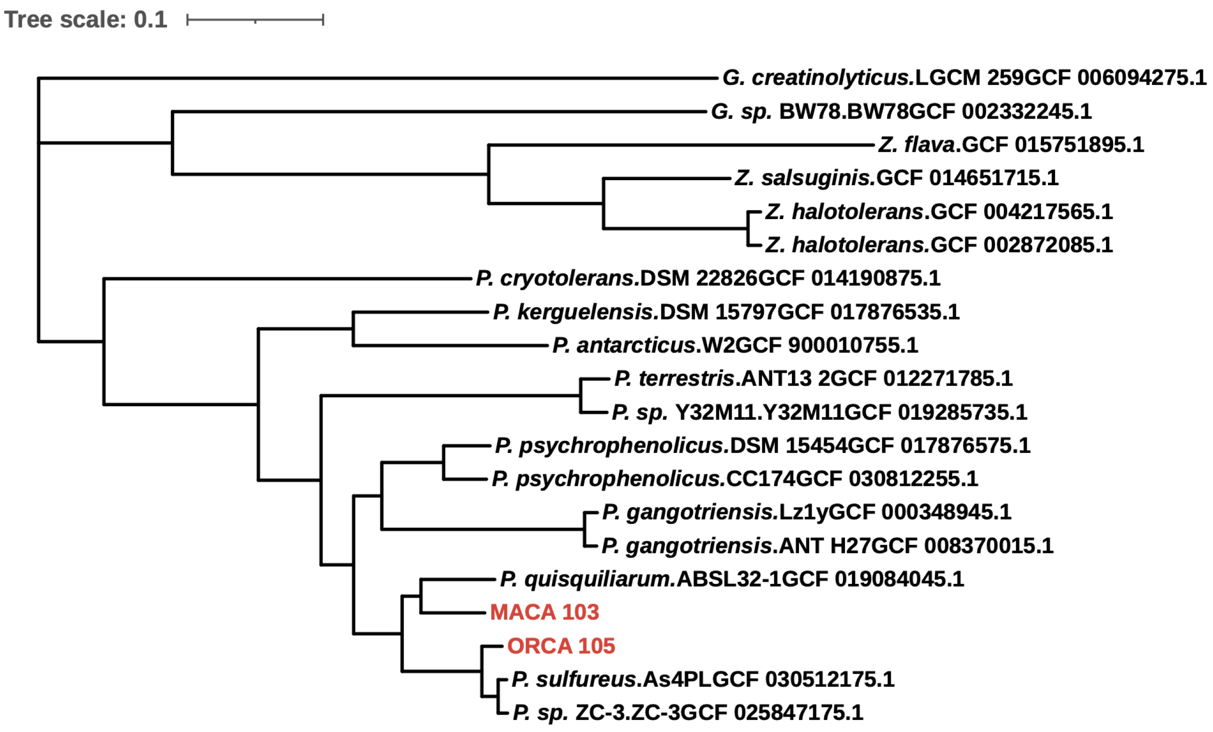

FIG 1 Phylogenomic tree of ORCA_105 and MACA_103 along with 18 closely related isolates. The tree was inferred with the *OrthoPhyl* pipeline (v1.0) (12). Briefly, *OrthoFinder* (v2.5.4) (13, 14) was used to infer 1,122 strict single-copy orthologs that were converted to codon alignments with *PAL2NAL* (15). *IQ-TREE* (v2.2.0.3) (16–18) was used to infer the final tree from the concatenated alignments. The tree was visualized with iTOL (19). Only bootstrap supports (calculated from 100 replicates) that are less than 100 are shown on the tree. NCBI accessions are shown at leaves. G: *Glutamicibacter*, *P*: *Paeniglutamicibacter,* and Z: *Zhihengliuella.*

These assemblies from cave-adapted *Paeniglutamicibacter* strains will be valuable additions to the growing number of genomic sequences from cave environments for identifying genomic signatures of cave adaptation.

ACKNOWLEDGMENTS

Sample collection was supported through the Cooperative Ecosystem Studies Unit, cooperative agreement P18AC01346. We also extend our thanks to the Northup Lab, UNM Department of Biology, and the Molecular Biology Facility. Research reported in this publication was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under award number P30 GM110907. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Genomic sequencing was supported by the Oak Ridge Institute for Science and Education through the Joint Science and Technology Institute -West STEM program (Project Code: SCG1). We thank Christopher Gallegos, Shane Johnson, Alice Chung-MacCoubrey, and Rickard Toomey for their work on this project.

AUTHOR AFFILIATIONS

[1]B-GEN: Genomics and Bioanalytics, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA
[2]Department of Biology, The University of New Mexico, Albuquerque, New Mexico, USA
[3]B-TEK: Biochemistry and Biotechnology, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA
[4]Allonnia, Boston, Massachusetts, USA
[5]B-IOME: Microbial and Biome Sciences, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

AUTHOR ORCIDs

Earl Middlebrook http://orcid.org/0000-0003-0508-455X
Andrew W. Bartlow http://orcid.org/0000-0002-1254-4481
Anand Kumar https://orcid.org/0000-0001-9290-0927
Blake Hovde http://orcid.org/0000-0002-7112-3310
Shannon Johnson http://orcid.org/0000-0002-3972-9208
Armand E. K. Dichosa http://orcid.org/0000-0003-0640-6629

FUNDING

Funder	Grant(s)	Author(s)
Cooperative Ecosystems Studies Unit, Cooperative Agreement	P18AC01346	Diana Northup
HHS | NIH | National Institute of General Medical Sciences (NIGMS)	P30 GM110907	Diana Northup
DOE | Oak Ridge Institute for Science and Education (ORISE)	SCG1	Armand E. K. Dichosa

AUTHOR CONTRIBUTIONS

Earl Middlebrook, Conceptualization, Formal analysis, Visualization, Writing – review and editing | Jennifer Hathaway, Conceptualization, Data curation, Resources, Writing – review and editing | Anastasia Pittis, Data curation, Resources | George Abernathy, Conceptualization, Investigation, Methodology, Writing – review and editing | Karen Davenport, Conceptualization, Formal analysis, Software | Cheryl Gleasner, Investigation, Methodology | Kimberly Mcmurry, Investigation, Methodology | Julia Kelliher, Conceptualization, Formal analysis, Investigation | Andrew W. Bartlow, Conceptualization, Formal

analysis, Investigation, Writing – review and editing | Migun Shakya, Conceptualization, Formal analysis, Investigation | Michaeline Albright, Conceptualization, Formal analysis, Investigation | Anand Kumar, Conceptualization, Formal analysis, Investigation | Blake Hovde, Conceptualization, Formal analysis, Investigation | Brett Youtsey, Conceptualization, Formal analysis, Investigation | Buck Hanson, Conceptualization, Formal analysis, Investigation | Aaron Robinson, Conceptualization, Formal analysis, Methodology | Mark Flynn, Conceptualization, Software | Shannon Johnson, Conceptualization, Formal analysis, Investigation | Diana Northup, Conceptualization, Data curation, Formal analysis, Writing – review and editing | Armand E. K. Dichosa, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft

DATA AVAILABILITY

These whole-genome projects have been deposited in NCBI's GenBank under the BioProject PRJNA1180589 and BioSample accessions SAMN44524352 (ORCA_105) and SAMN44524351 (MACA_103). Raw reads have been deposited in NCBI's SRA with accessions SRX27112100 (ORCA_105) and SRX27112101 (MACA_103). Assemblies are available with accessions GCF_045798545.1 (ORCA_105) and GCF_045798555.1 (MACA_103).

REFERENCES

1. Lange-Enyedi NT, Németh P, Borsodi AK, Spötl C, Makk J. 2024. Calcium carbonate precipitating extremophilic bacteria in an alpine ice cave. Sci Rep 14:2710. https://doi.org/10.1038/s41598-024-53131-y
2. Kosznik-Kwaśnicka K, Golec P, Jaroszewicz W, Lubomska D, Piechowicz L. 2022. Into the unknown: microbial communities in caves, their role, and potential use. Microorganisms 10:222. https://doi.org/10.3390/microorganisms10020222
3. Li P-E, Lo C-C, Anderson JJ, Davenport KW, Bishop-Lilly KA, Xu Y, Ahmed S, Feng S, Mokashi VP, Chain PSG. 2017. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. Nucleic Acids Res 45:67–80. https://doi.org/10.1093/nar/gkw1027
4. Lo C-C, Chain PSG. 2014. Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics 15:366. https://doi.org/10.1186/s12859-014-0366-2
5. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174
6. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303:3997. https://doi.org/10.48550/arXiv.1303.3997.
7. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
8. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
10. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214. https://doi.org/10.1089/10665270050081478
11. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. 2008. Database indexing for production MegaBLAST searches. Bioinformatics 24:1757–1764. https://doi.org/10.1093/bioinformatics/btn322
12. Middlebrook EA, Katani R, Fair JM. 2023. OrthoPhyl – a turn-key solution for large scale whole genome bacterial phylogenomics. bioRxiv. https://doi.org/10.1101/2023.06.27.546815.
13. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157. https://doi.org/10.1186/s13059-015-0721-2
14. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238. https://doi.org/10.1186/s13059-019-1832-y
15. Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–12. https://doi.org/10.1093/nar/gkl315
16. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
17. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285
18. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281
19. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301