AMERICAN SOCIETY FOR MICROBIOLOGY
Microbiology Resource Announcements

GENOME SEQUENCES

Basaltic Lava Tube Hosts a Putative Novel Genus in the Family *Solirubrobacteraceae*

C. B. Fishman,[a] J. G. Bevilacqua,[a] O. Gadson,[a] A. S. Hahn,[b] A. C. McAdam,[c] J. Bleacher,[d] S. S. Johnson[e]

[a]Department of Biology, Georgetown University, Washington, DC, USA
[b]Koonkie Inc., Menlo Park, California, USA
[c]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[d]NASA Headquarters, Washington, DC, USA
[e]Science, Technology, and International Affairs Program, Georgetown University, Washington, DC, USA

ABSTRACT We report the draft genome sequence of a putative new genus and species, *Siliceabacter maunaloa*, in the family *Solirubrobacteraceae*. The members of this family of *Actinobacteria* are generally Gram positive and mesophilic. Found within a Hawaiian lava tube, this microbe illuminates the types of prokaryotes inhabiting secondary minerals in subsurface basaltic environments.

Beneath Mauna Loa, the largest active volcano on the planet, lies a maze of basaltic lava tubes (1, 2). A sample was collected in August 2019 (19°36.392′N, 155°28.993′W) from a pink, gypsum-rich mineral deposit in a lava tube with sunlight exposure. The sample was transported on ice and frozen at −80°C. Then, DNA was extracted using the ZymoBIOMICS DNA/RNA miniprep kit. A DNA library was prepared using the Illumina DNA prep kit, and shotgun sequencing was performed using the NovaSeq 6000 platform.

The reads were processed, assembled, and binned following the Atlas workflow with default parameters (3). Duplicate reads were excluded using Clumpify v38.86 (4). Adapters were removed, and the reads were filtered based on quality scores greater than 10 and lengths greater than 51 bp using BBTools v38.86 (4). The merged, paired-end reads were assembled using metaSPAdes v3.15 and mapped back to the assembly using BBMap v38.86 and SAMtools v1.9 (5–7). Metagenome-assembled genomes (MAGs) were binned using MetaBAT v2 and MaxBin v2.0, and the bins were combined using DAS Tool v1.0 (8–10). Each sample's unique set of genomes was found using dRep v2.2.2, and the genomes were clustered into a set of nonredundant MAGs using Mash v1.1.1 and MUMmer v3.23 (11–13).

This is a high-quality genome, according to the Genomic Standards Consortium, predicted to be 96.01% complete and 0.43% contaminated, as assessed using CheckM v1.1.2 (14, 15) (Fig. 1). It represents a putative novel species in the *Solirubrobacteraceae* family, as assigned using GTDB-Tk v1.5 (17). Forty-six tRNA genes and three rRNA genes were identified, for which similarity searches were performed using online BLAST-based homology against the RefSeq database (18–22). The closest 16S rRNA gene in the database is 97.94% similar and annotated as an uncultured member of *Actinobacteria*, the phylum that contains *Solirubrobacteraceae*. The 23S and 5S rRNA genes are respectively 92.38% and 98.20% similar to those of *Conexibacter woesei*, a species in the same order (*Solirubrobacterales*) (23). This genetic distance suggests that this species belongs to a new genus in the family *Solirubrobacteraceae*.

The ORFs were annotated using MetaPathways v2.5 with default settings (24, 25). The sequences were queried against the following reference databases: MetaCyc (accessed 10 August 2020) (26, 27), CAZy (accessed 1 June 2020) (28), RefSeq-NR

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine

Address correspondence to S. S. Johnson, sarah.johnson@georgetown.edu.

The authors declare no conflict of interest.

Received 4 June 2022
Accepted 19 September 2022
Published 3 October 2022

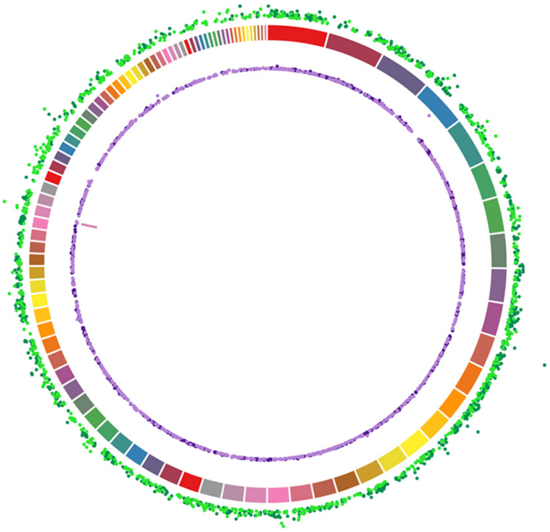

FIG 1 Visualization of the MAG in this study, with 6,042,029 reads mapped, 120 contigs, or overlapping DNA segments, a median coverage of 245×, and 72.14% GC content. The genome is 4.64 Mbp with an N_{50} value of 62,108 bp. It contains 3,770 open reading frames (ORFs), of which 2,161 are annotated and 1,609 are hypothetical proteins, as found using Prodigal v2.6.3 with default settings (16). The dark purple and dark green dots are ORFs with functional annotations, while the light purple and light green dots are hypothetical proteins. The dark pink line denotes the rRNA genes. The colored blocks represent contigs, which are ordered by size. Without a complete assembly, the structure of the genome remains unknown.

(accessed 12 August 2020) (22), and UniProtKB/Swiss-Prot (accessed 12 August 2020) (29).

ORFs for central enzymes of the reductive pentose phosphate pathway, including RubisCO, suggest an ability to fix CO_2 for energy storage and incorporation (30–32). An ORF encoding carbon monoxide dehydrogenase suggests CO oxidation ability (33, 34). Multiple ORFs for hydrogenase development and for [NiFe] hydrogenase and molybdopterin oxidoreductase suggest that this microbe can perform hydrogen oxidation I (aerobic) (35–38). This microbe is likely a facultative lithotroph, which have rarely been reported living mixotrophically via aerobic hydrogen and CO oxidation and CO_2 reduction (30). An ORF encoding a carotenoid may be responsible for the pink hue of the sample, as actinobacteria are able to produce a wide range of pigments, including pink (39).

Using Pathway Tools v24.5 with default settings, 133 metabolic pathways were predicted for biosynthesis (54.89%), degradation (17.29%), energy metabolism (5.26%), detoxification (2.26%), and other pathways (20.30%) (40). The predicted pathways confirmed aerobic respiration and included arsenate detoxification, which may be critical for living in an arsenic-rich volcanic environment (41).

The taxonomic and genomic profile of a novel facultative lithotroph such as this one informs us of the types of microbes living in undercharacterized volcanic habitats and how they may evolve functional potential as a consequence of the limited nutrient availability.

Data availability. Sequencing data and information regarding this MAG are available under BioProject accession number PRJNA818798. Information about the cave metagenome sample is available under BioSample accession number SAMN29624104. The raw whole-genome sequencing (WGS) reads are available under Sequence Read Archive accession number SRR19994649. The WGS assembly is available under genome accession number JANDLV000000000. Information about this MAG is available under BioSample accession number SAMN26879537. The ORFs (https://doi.org/10.6084/m9.figshare.20387157), assembly (https://doi.org/10.6084/m9.figshare.20387154), and annotations (https://doi.org/10.6084/m9.figshare.20387145) are also available on Figshare.

ACKNOWLEDGMENTS

This research was supported by NASA's PSD Research Program through the GSFC ISFM Goddard Instrument Field Team program and the GSFC ISFM Fundamental Laboratory Research program, as well as the NASA DC Space Grant Consortium.

REFERENCES

1. Trusdell FA. 2012. Mauna Loa—history, hazards, and risk of living with the world's largest volcano: U.S. Geological Survey fact sheet 2012–3104. USGS, Denver, CO.
2. Greeley R. 1987. The role of lava tubes in Hawaiian volcanoes, p 1589–1602. *In* Decker RW, Wright TL, Stauffer PH (ed), U.S. Geological Survey professional paper 1350. USGS, Denver, CO.
3. Kieser S, Brown J, Zdobnov EM, Trajkovski M, McCue LA. 2020. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. BMC Bioinformatics 21:257. https://doi.org/10.1186/s12859-020-03585-4.
4. Bushnell B. 2018. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. Joint Genome Institute.
5. Nurk S, Meleshko D, Korobeynikov A, Pevzner P. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
6. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner.
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
8. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359. https://doi.org/10.7717/peerj.7359.
9. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.
10. Sieber C. 2017. Dereplication, aggregation and scoring tool (DAS Tool) v1.0. Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA.
11. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126.
12. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132. https://doi.org/10.1186/s13059-016-0997-x.
13. Delcher AL, Salzberg SL, Phillippy AM. 2003. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics 00:10.3.1–10.3.18. https://doi.org/10.1002/0471250953.bi1003s00.
14. 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.
15. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Genome Standards Consortium, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.
16. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
17. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
18. Seemann T. 2013. Barrnap: basic rapid ribosomal RNA predictor.
19. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.
20. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences, p 1–14. *In* Kollmar M (ed), Gene prediction, vol 1962. Springer, New York, NY.
21. 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.
22. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745. https://doi.org/10.1093/nar/gkv1189.
23. Albuquerque L, da Costa MS. 2014. The families Conexibacteraceae, Patulibacteraceae and Solirubrobacteraceae, p 185–200. *In* Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (ed), The prokaryotes: Actinobacteria. Springer, Berlin, Germany.
24. Konwar KM, Hanson NW, Pagé AP, Hallam SJ. 2013. MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. BMC Bioinformatics 14:202. https://doi.org/10.1186/1471-2105-14-202.
25. Konwar KM, Hanson NW, Bhatia MP, Kim D, Wu S-J, Hahn AS, Morgan-Lang C, Cheung HK, Hallam SJ. 2015. MetaPathways v2.5: quantitative functional, taxonomic and usability improvements. Bioinformatics 31:3345–3347. https://doi.org/10.1093/bioinformatics/btv361.
26. Karp PD, Riley M, Paley SM, Pellegrini-Toole A. 2002. The MetaCyc database. Nucleic Acids Res 30:59–61. https://doi.org/10.1093/nar/30.1.59.
27. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S. 2020. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. Nucleic Acids Res 48:D445–D453. https://doi.org/10.1093/nar/gkz862.
28. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The Carbohydrate-Active Enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
29. The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049.
30. Nanba K, King GM, Dunfield K. 2004. Analysis of facultative lithotroph distribution and diversity on volcanic deposits by use of the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase. Appl Environ Microbiol 70:2245–2253. https://doi.org/10.1128/AEM.70.4.2245-2253.2004.
31. Grostern A, Alvarez-Cohen L. 2013. RubisCO-based CO_2 fixation and C_1 metabolism in the actinobacterium Pseudonocardia dioxanivorans CB1190. Environ Microbiol 15:3040–3053. https://doi.org/10.1111/1462-2920.12144.
32. Ivanovsky RN, Fal YI, Berg IA, Ugolkova NV, Krasilnikova EN, Keppen OI, Zakharchuc LM, Zyakun AM. 1999. Evidence for the presence of the reductive pentose phosphate cycle in a filamentous anoxygenic photosynthetic bacterium, Oscillochloris trichoides strain DG-6. Microbiology (Reading) 145:1743–1748. https://doi.org/10.1099/13500872-145-7-1743.
33. Cunliffe M. 2013. Physiological and metabolic effects of carbon monoxide oxidation in the model marine bacterioplankton Ruegeria pomeroyi DSS-3. Appl Environ Microbiol 79:738–740. https://doi.org/10.1128/AEM.02466-12.
34. Wawrousek K, Noble S, Korlach J, Chen J, Eckert C, Yu J, Maness PC. 2014. Genome annotation provides insight into carbon monoxide and

hydrogen metabolism in Rubrivivax gelatinosus. PLoS One 9:e114551. https://doi.org/10.1371/journal.pone.0114551.

35. Li X, Luo Q, Wofford NQ, Keller KL, McInerney MJ, Wall JD, Krumholz LR. 2009. A molybdopterin oxidoreductase is involved in H_2 oxidation in Desulfovibrio desulfuricans G20. J Bacteriol 191:2675–2682. https://doi.org/10.1128/JB.01814-08.
36. Hedrich S, Johnson DB. 2013. Aerobic and anaerobic oxidation of hydrogen by acidophilic bacteria. FEMS Microbiol Lett 349:40–45. https://doi.org/10.1111/1574-6968.12290.
37. Watanabe S, Kawashima T, Nishitani Y, Kanai T, Wada T, Inaba K, Atomi H, Imanaka T, Miki K. 2015. Structural basis of a Ni acquisition cycle for [NiFe] hydrogenase by Ni-metallochaperone HypA and its enhancer. Proc Natl Acad Sci U S A 112:7701–7706. https://doi.org/10.1073/pnas.1503102112.
38. Garg RP, Menon AL, Jacobs K, Robson RM, Robson RL. 1994. The hypE gene completes the gene cluster for H2-oxidation in Azotobacter vinelandii. J Mol Biol 236:390–396. https://doi.org/10.1006/jmbi.1994.1149.
39. Sandmann G. 2021. Carotenoid biosynthesis in the phylum Actinobacteria. Adv Exp Med Biol 1261:175–181. https://doi.org/10.1007/978-981-15-7360-6_14.
40. Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, Latendresse M, Ong WK, Subhraveti P, Caspi R, Fulcher C, Keseler IM, Paley SM. 2021. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Brief Bioinform 22:109–126. https://doi.org/10.1093/bib/bbz104.
41. Masuda H. 2018. Arsenic cycling in the Earth's crust and hydrosphere: interaction between naturally occurring arsenic and human activities. Prog Earth Planet Sci 5:68. https://doi.org/10.1186/s40645-018-0224-3.