AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences of Nine "*Candidatus* Nanosynbacter sp. HMT-352" Strains Cultured from the Human Oral Cavity

Jett Liu,[a] Daniel R. Utter,[b] Jie Nie,[c] Kristopher A. Kerns,[d] Eleanor I. Lamont,[d] Erik L. Hendrickson,[d] Xiaoyan Wang,[c] Jeffrey S. McLean,[d] Xuesong He,[a,e] Batbileg Bor[a,e]

[a]The Forsyth Institute, Cambridge, Massachusetts, USA
[b]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA
[c]Department of Cariology and Endodontology, Peking University School of Stomatology, Beijing, China
[d]Department of Periodontics, University of Washington, Seattle, Washington, USA
[e]Department of Oral Medicine, Infection, and Immunity, Harvard School of Dental Medicine, Boston, Massachusetts, USA

Jie Nie and Daniel R. Utter contributed equally to this article. Author order was determined alphabetically by first name.

ABSTRACT Here, we report draft genome sequences for nine strains of "*Candidatus* Nanosynbacter sp. HMT-352." These strains and their sequences were used to interrogate strain-level variations in host range, gene content, and growth dynamics among the phylum "*Candidatus* Saccharibacteria."

The nine "*Candidatus* Nanosynbacter sp. HMT-352" (hereafter, HMT-352) strains reported here (Table 1) were recently isolated from the human oral cavity (1) and are the first members of the phylum "*Candidatus* Saccharibacteria," a major lineage of the Candidate Phyla Radiation (CPR) (2), to be characterized at the strain level.

These nine HMT-352 strains were isolated from human saliva using a previously described "baiting" method (1, 3). Briefly, saliva samples were centrifuged, filtered through a 0.45-μM filter, and cocultured in brain heart infusion medium (catalog number 237500; BD, NJ, USA) with potential basibionts (bacterial hosts). The cocultures were incubated at 37°C and passaged at a dilution of 1:10 every 2 days into fresh medium. A previously described modified MasterPure DNA isolation kit (catalog number MGP04100; Epicentre, WI, USA) protocol (4) was used to isolate genomic DNA (gDNA) from both filter-isolated HMT-352 cells and the HMT-352-basibiont cocultures. Briefly, bacterial cultures were mixed with glass beads (catalog number G8772; Sigma, St. Louis, MO) and disrupted using a bead-beating homogenizer. gDNA isolation was then performed according to the manufacturer's protocol. The gDNA was randomly fragmented by sonication and then end-polished, A-tailed, and ligated with full-length Illumina adapters. The library constructs were purified using the AMPure XP system (Beckman Coulter, IN, USA) and checked for size distribution using a 2100 Bioanalyzer (Agilent Technologies, CA, USA). The libraries were then sequenced on an Illumina NovaSeq instrument (paired-end [PE] 150-bp reads).

Default parameters were used for the computational analyses except where otherwise noted. The reads were quality controlled using iu-filter-quality-minoche from illumina-utils v2.12 (5). For each new HMT-352 strain, genomes were assembled using a previously described Anvi'o v7.1 workflow (1, 6) that employed both the isolate and coculture genomic libraries. Briefly, for each strain, libraries from the isolated HMT-352 were individually assembled using metaSPAdes v3.15.3 (7) and binned using MaxBin2 v2.2.4-1 (8). The bins were then manually refined and reassembled using both the isolate and coculture libraries. Genes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.3 (9). All genomes were less than 5% redundant, between 83% and 85% complete, and contained between 732 and 801 genes (Table 1).

Editor Irene L. G. Newton, Indiana University, Bloomington

Address correspondence to Batbileg Bor, bbor@forsyth.org.

The authors declare no conflict of interest.

Received 25 April 2022
Accepted 11 July 2022
Published 27 July 2022

TABLE 1 Summary information for the nine "*Ca*. Nanosynbacter sp. HMT-352" strains cultured from the human oral cavity

Strain name	BioSample accession no.	SRA accession no. for:		Total length (bp)	No. of contigs	N_{50} (bp)	GC content (%)	Completion (%)	Redundancy (%)	No. of genes	Assembly coverage (×) for:	
		Monoculture	Coculture								Monoculture	Coculture
TM7-001	SAMN23492223	SRR18278454	SRR18278446	771,807	1	771,807	43.25	83.10	2.82	803	3,276	432
TM7-008	SAMN23492221	SRR18278459	SRR18278451	725,580	2	548,424	43.17	84.51	0.00	747	3,063	2,387
TM7-053	SAMN23492224	SRR18278461	SRR18278463	755,984	2	546,117	43.18	84.51	4.23	775	1,403	1,097
TM7-057	SAMN23492218	SRR18278457	SRR18278449	758,480	6	490,402	42.99	84.51	2.82	804	728	1,156
TM7-072	SAMN23492220	SRR18278453	SRR18278445	733,210	1	733,210	43.31	84.51	1.41	742	2,911	1,235
TM7-075	SAMN23492222	SRR18278460	SRR18278462	730,938	2	448,955	43.15	84.51	4.23	755	3,484	1,767
TM7-076	SAMN23492217	SRR18278456	SRR18278448	756,098	4	170,979	43.19	83.10	0.00	791	3,519	1,213
TM7-087	SAMN23492219	SRR18278458	SRR18278450	741,912	3	536,841	43.42	84.51	0.00	771	3,259	2,736
TM7-037	SAMN23492216	SRR18278452	SRR18278444	718,283	1	718,283	43.26	83.10	2.82	735	651	390

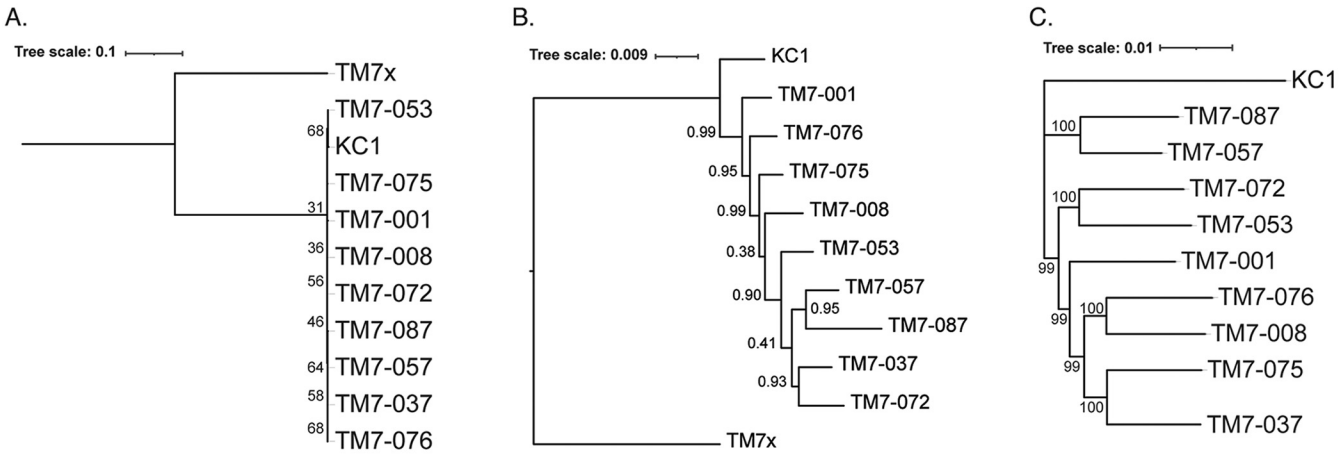

FIG 1 16S rRNA, select marker gene, and single-copy core gene trees of the reported HMT-352 strains. The previously reported HMT-352 strain KC1 (15) is additionally included. The node labels represent bootstrap support. (A) Maximum-likelihood tree based on full-length 16S rRNA sequences constructed using IQ-TREE v2.1.4-beta with ultrafast bootstrap (-bb 1500) (16). *Nanosynbacter lyticus* strain TM7x (17) is included as an outgroup. (B) Phylogenomic tree constructed using FastTree 2 v2.1.11-1 (18) with 60 concatenated core protein amino acid sequences and TM7x (HMT-952) included as an outgroup. (C) Maximum-likelihood tree inferred using 523 concatenated single-copy core gene amino acid sequences found in all strains constructed using IQ-TREE with ultrafast bootstrap (-bb 1500).

The HMT-352 strains were compared to representative *Candidatus* Saccharibacteria from eHOMD v15.22 (10) using full-length 16S rRNA sequences aligned with MAFFT v7.490 (11). All strains had more than 98% homology to the closest eHOMD species. The average nucleotide identity (ANI) values over all alignable genome fractions, however, ranged between 93% and 95%, at or below the extreme end of the accepted range for intraspecies variation (12–14). Such substantial intraspecies genetic diversity is additionally apparent in Fig. 1, which provides a comparison of the phylogenetic differences between the 16S rRNA, select marker gene, and single-copy core gene trees. The unexpectedly high nucleotide diversity among these strains warrants further investigation and accentuates that broad phylogenetic characterization of the CPR is the next step in understanding these bacteria.

Data availability. Cultures of these strains are available upon request. The sequence data have been deposited at NCBI under the BioProject accession number PRJNA784561. The BioSample and SRA accession numbers are listed in Table 1. All code used to assemble and analyze the genomes is available at https://www.borlab.org/resources.

ACKNOWLEDGMENTS

We were supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under awards 1R01DE023810, 1R01DE020102, and 1R01DE026186 (to X.H. and J.S.M.) and 5K99DE027719-02 and 5R00DE027719-04 (to B.B.). Additional support was provided to D.R.U. by Harvard University's Department of Organismic and Evolutionary Biology program and the National Science Foundation Graduate Research Fellowship Program under grant DGE1745303.

REFERENCES

1. Nie J, Utter DR, Kerns KA, Lamont EI, Hendrickson EL, Liu J, Wu T, He X, McLean J, Bor B. 2022. Strain-level variation and diverse host bacterial responses in episymbiotic saccharibacteria. mSystems 7:e01488-21. https://doi.org/10.1128/msystems.01488-21.
2. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48.
3. Bor B, Collins AJ, Murugkar PP, Balasubramanian S, To TT, Hendrickson EL, Bedree JK, Bidlack FB, Johnston CD, Shi W, McLean JS, He X, Dewhirst FE. 2020. Insights obtained by culturing Saccharibacteria with their bacterial hosts. J Dent Res 99:685–694. https://doi.org/10.1177/0022034520905792.
4. Chipashvili O, Utter DR, Bedree JK, Ma Y, Schulte F, Mascarin G, Alayyoubi Y, Chouhan D, Hardt M, Bidlack F, Hasturk H, He X, McLean JS, Bor B. 2021. Episymbiotic Saccharibacteria suppresses gingival inflammation and bone loss in mice through host bacterial modulation. Cell Host Microbe 29:1649–1662.e7. https://doi.org/10.1016/j.chom.2021.09.009.
5. Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using Illumina paired-end technology. PLoS One 8:e66643. https://doi.org/10.1371/journal.pone.0066643.
6. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.
7. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
8. 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.
9. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
10. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. 2018. New insights into human nostril microbiome from the expanded human oral microbiome database (eHOMD): a resource for the microbiome of the human aerodigestive tract. mSystems 3:e00187-18. https://doi.org/10.1128/mSystems.00187-18.
11. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.
12. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.
13. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Carnevali PBM, Banfield JF. 2020. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. mSystems 5:e00731-19. https://doi.org/10.1128/mSystems.00731-19.
14. Rodríguez-R L, Konstantinidis K. 2014. Bypassing cultivation to identify bacterial species: culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. Microbe Wash DC 9:111–118. https://doi.org/10.1128/microbe.9.111.1.
15. Cross KL, Klingeman DM, Podar M. 2022. Complete genome sequence of human oral saccharibacterium "*Candidatus* Nanosynbacter sp. HMT352" strain KC1. Microbiol Resour Announc 11:e01205-21. https://doi.org/10.1128/mra.01205-21.
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. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu S-Y, Dorrestein PC, Esquenazi E, Hunter RC, Cheng G, Nelson KE, Lux R, Shi W. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci U S A 112:244–249. https://doi.org/10.1073/pnas.1419038112.
18. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.