



# Complete Genome Sequence of Strain JB001, a Member of *Saccharibacteria* Clade G6 (“*Candidatus* Nanogingivalaceae”)

 Jonathon L. Baker<sup>a</sup>

<sup>a</sup>Genomic Medicine Group, J. Craig Venter Institute, La Jolla, California, USA

**ABSTRACT** At least 6 highly diverse clades of *Saccharibacteria* inhabit the human oral cavity. However, all oral *Saccharibacteria* strains with currently available complete genome sequences or cultured isolates belong to clade G1, leaving clades G2 through G6 poorly understood. Here, a complete genome sequence of JB001, a clade G6 (“*Candidatus* Nanogingivalaceae”) *Saccharibacteria* strain, is reported.

*Saccharibacteria* (formerly TM7) have reduced genomes, a small cell size, and appear to have a parasitic lifestyle, dependent on host bacteria (1–3). At least 6 major clades of *Saccharibacteria* (G1 through G6) inhabit the human oral cavity; however, all currently available complete genome sequences and cultured isolates belong to G1, leaving G2 through G6 quite poorly understood (4). Recent studies provided the first draft genome sequences from clades G3, G5, and G6 (5–8), which displayed major differences in encoded functional pathways, suggesting that the lifestyle and host dependency of the clades may be distinct (5, 6, 8). *Saccharibacteria* frequently lack what are considered “essential” core genes, which are typically relied upon to estimate the completion of draft genome sequences. Indeed, the sequence for the chromosome of TM7x, the first *Saccharibacteria* strain to be cultivated, had a completeness of 65%, according to CheckM (5). Therefore, obtaining complete genome sequences of *Saccharibacteria* is of special importance. In this study, Nanopore sequencing was used to deliver the first complete genome sequence of an oral *Saccharibacteria* strain outside clade G1, the G6 (proposed name, “*Candidatus* Nanogingivalaceae”) taxon JB001.

The draft assembly of JB001, *Candidatus*\_Nanogingivalaceae\_FGB1\_strain\_JCVI\_27\_bin.3, reported in 2021, was obtained from human saliva in Los Angeles, CA, USA, and fragmented into 67 contigs (6). Here, high-molecular-weight genomic DNA was extracted from the same saliva sample as used to obtain the original draft genome, using a phenol-chloroform-based protocol (9), and was examined for purity, size, and concentration using the TapeStation system (Agilent Technologies). The DNA was not sheared or size selected. A long-read library was prepared using a ligation sequencing kit (Oxford Nanopore Technologies) and sequenced on a GridION using an R9.4.1 flow cell (Oxford Nanopore Technologies). Base calling, quality control, and adapter trimming were performed using Guppy v4.0.11/MinKNOW v20.06.9 (Oxford Nanopore Technologies), resulting in 3,199,915 reads ( $N_{50}$ , 13,719 bp). Two independent methods generated improved draft assemblies. (i) Human reads were removed using minimap2 v2.17-r941 (10), and the remaining long reads were assembled using meta-flye v2.8-b1674 (11). MegaBLAST v2.2.26 (12) was used to identify the circular JB001 contig within the metagenome assembly. (ii) Long reads mapping to the draft genome of JB001 were extracted using minimap2. These long reads, along with the short reads used to generate the original JB001 draft assembly, were used by Unicycler v0.4.8 (13) to obtain a draft genome of 4 contigs. Three short contigs were removed based on disparate GC content, coverage, and BLAST hits to other organisms (Anvi’o v7-dev [14]), leaving one circular contig. Tricycler v0.3.0 (<https://github.com/rrwick/Tricycler>) was used to develop a consensus assembly from the two draft assemblies. The resulting assembly was polished using Medaka

**Citation** Baker JL. 2021. Complete genome sequence of strain JB001, a member of *Saccharibacteria* clade G6 (“*Candidatus* Nanogingivalaceae”). *Microbiol Resour Announc* 10:e00517-21. <https://doi.org/10.1128/MRA.00517-21>.

**Editor** Steven R. Gill, University of Rochester School of Medicine and Dentistry

**Copyright** © 2021 Baker. 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 [jobaker@jcvl.org](mailto:jobaker@jcvl.org).

**Received** 18 May 2021

**Accepted** 19 May 2021

**Published** 10 June 2021

v1.0.3 (<https://github.com/nanoporetech/medaka>), then Pilon v1.23 (15). Circulator v1.5.5 (16) was used to rotate the genome sequence start to *dnaA*. The scripts, parameters, and versions of the software tools used are available at [https://github.com/jonbakerlab/JB001\\_genome\\_completion](https://github.com/jonbakerlab/JB001_genome_completion). Default parameters were used unless otherwise noted. JB001 was annotated using the NCBI Prokaryotic Genome Annotation Pipeline v5.1. The resulting chromosome was 662,051 bp with a GC content of 36.4% and is predicted to carry 687 genes. This resource will provide valuable information regarding the lifestyle and evolution of G6 *Saccharibacteria*.

**Data availability.** The complete genome sequence of JB001 has been deposited in GenBank under the accession number [CP072208](https://www.ncbi.nlm.nih.gov/nuccore/CP072208). The BioProject accession number for the genome is [PRJNA624185](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA624185). The raw Nanopore read library has been deposited in the Sequence Read Archive (SRA) database under the accession number [SRX10387815](https://www.ncbi.nlm.nih.gov/sra/SRX10387815). The short reads used to generate the original JB001 draft assembly are available in the SRA database under the accession number [SRX4318838](https://www.ncbi.nlm.nih.gov/sra/SRX4318838).

## ACKNOWLEDGMENTS

I thank Karrie Goglin-Almeida, Jelena Jablanovic, and Kara Riggsbee for performing the library preparation and sequencing and Jeffrey S. McLean for helpful discussions.

This research was supported by NIH/NIDCR K99-DE029228.

## REFERENCES

- Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523:208–211. <https://doi.org/10.1038/nature14486>.
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hemsdorf 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>.
- He X, McLean JS, Edlund A, Yooshep 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>.
- Bor B, Bedree JK, Shi W, McLean JS, He X. 2019. *Saccharibacteria* (TM7) in the human oral microbiome. *J Dent Res* 98:500–509. <https://doi.org/10.1177/0022034519831671>.
- McLean JS, Bor B, Kerns KA, Liu Q, To TT, Solden L, Hendrickson EL, Wrighton K, Shi W, He X. 2020. Acquisition and adaptation of ultra-small parasitic reduced genome bacteria to mammalian hosts. *Cell Rep* 32:107939. <https://doi.org/10.1016/j.celrep.2020.107939>.
- Baker JL, Morton JT, Dinis M, Alvarez R, Tran NC, Knight R, Edlund A. 2021. Deep metagenomics examines the oral microbiome during dental caries, revealing novel taxa and co-occurrences with host molecules. *Genome Res* 31:64–74. <https://doi.org/10.1101/gr.265645.120>.
- Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper SJ, Griffen A, Heaton M, Joshi S, Klingeman D, Leys E, Yang Z, Parks JM, Podar M. 2019. Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol* 37:1314–1321. <https://doi.org/10.1038/s41587-019-0260-6>.
- Shaiber A, Willis AD, Delmont TO, Roux S, Chen L-X, Schmid AC, Yousef M, Watson AR, Lolans K, Esen OC, Lee STM, Downey N, Morrison HG, Dewhirst FE, Mark Welch JL, Eren AM. 2020. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol* 21:292. <https://doi.org/10.1186/s13059-020-02195-w>.
- Baker JL, Edlund A. 2020. Composite long- and short-read sequencing delivers a complete genome sequence of B04Sm5, a reutericyclin- and mutanocyclin-producing strain of *Streptococcus mutans*. *Microbiol Resour Announc* 9:e01067-20. <https://doi.org/10.1128/MRA.01067-20>.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Kolmogorov M, Bickhart DM, Behsas B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 17:1103–1110. <https://doi.org/10.1038/s41592-020-00971-x>.
- 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>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Eren AM, Esen OC, 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>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circulator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.