Composite Long- and Short-Read Sequencing Delivers a Complete Genome Sequence of B04Sm5, a Reutericyclin- and Mutanocyclin-Producing Strain of *Streptococcus mutans*

Jonathon L. Baker,[a] Anna Edlund[a,b]

aGenomic Medicine Group, J. Craig Venter Institute, La Jolla, California, USA
bDepartment of Pediatrics, University of California at San Diego, La Jolla, California, USA

ABSTRACT *Streptococcus mutans* strain B04Sm5 was recently shown to inhibit the growth of neighboring commensal bacteria using reutericyclin, an acylated tetramic acid produced by the *muc* biosynthetic gene cluster. Here, a complete genome sequence of B04Sm5 is reported.

*S*treptococcus mutans is considered a major etiologic agent of dental caries, which is globally the most common chronic infectious disease (1). As it is not typically considered a pioneer colonizer of the tooth, *S. mutans* must be able to outcompete its already-established bacterial neighbors (which are typically health-associated commensals) to successfully institute itself as a member of the dental plaque microbiota and cause disease (2, 3). A subset of *S. mutans* strains (but not the well-characterized type strain, UA159) encode a hybrid nonribosomal peptide synthetase/polyketide synthase (NRPS/PKS) biosynthetic gene cluster (BGC), *muc*, which has significant homology (48% to 69%) to cognates involved in the biosynthesis of reutericyclin by *Lactobacillus* (4–7). Indeed, the *S. mutans* B04Sm5 *muc* BGC was recently shown to produce not only reutericyclin but also two reutericyclin analogs, as well as mutanocyclin, an unacylated reutericyclin derivative (6). *S. mutans* B04Sm5 employed reutericyclin to inhibit the growth of the adjacent competing commensal organisms, *Streptococcus sanguinis*, *Streptococcus mitis*, and *Streptococcus gordonii* (6). This suggests that the carriage of *muc* by *S. mutans* strains may impact the virulence of this pathogen.

S. mutans B04Sm5, a serotype c strain, was originally isolated from a carious lesion in a child with rampant caries by Argimon and Caufield, as described in reference 8. A draft assembly of the genome of B04Sm5, fragmented into 81 contigs, was published in 2014 (9). Here, a combination of Oxford Nanopore and Illumina sequencing was used to obtain a complete, circular genome sequence. B04Sm5 was grown in brain heart infusion (BHI) medium at 37°C under 5% CO_2/95% air. High-molecular-weight genomic DNA (gDNA) was extracted using a phenol-chloroform-based protocol (10) with the following modifications: (i) 300 U/ml of mutanolysin was used in lysis; (ii) following lysis in SDS, the sample was homogenized using 0.1-mm glass beads for 30 s in a FastPrep-24 homogenizer (MP Biomedicals); and (iii) an additional final ethanol precipitation was performed. The resulting gDNA was examined for purity, size, and concentration using the TapeStation system (Agilent Technologies) and a Qubit fluorometer (Thermo Fisher Scientific). A long-read library was prepared using a ligation sequencing kit (Oxford Nanopore Technologies) and sequenced on the GridION system using an R9.4.1 flow cell (Oxford Nanopore Technologies). Base calling was performed using Guppy v.4.8.11, resulting in 780,164 reads (N_{50}, 14,642 bp). A short-read library was prepared using a TruSeq DNA PCR-free library prep kit (Illumina) and sequenced on a MiSeq instrument (Illumina), generating 486,071 paired-end 150-bp reads. Quality control was performed on the short reads using KneadData v.0.5.4 (https://github

.com/biobakery/kneaddata) and on the long reads by Filtlong v.0.2.0 (https://github .com/rrwick/Filtlong). Draft assemblies were produced using a combination of the long- and short-read libraries and the Flye v.2.8-b1674 (11), hybridSPAdes v.3.14.0 (12), and Unicycler v.0.4.8 (13) assemblers. Trycycler v.0.3.0 (https://github.com/rrwick/ Trycycler) was used to develop a circular consensus assembly from the draft assemblies. The final assembly was polished using the full long-read set with medaka v.1.0.3 (https://github.com/nanoporetech/medaka) and then the short-read set using Pilon v.1.23 (14). The scripts and parameters of the software tools used are available at https://github.com/jaybake5/B04Sm5_genome_completion. The genome of B04Sm5 is a single chromosome of 2,011,542 bp, encoding 1,964 genes with a GC content of 36.9%. This resource will provide valuable information regarding the acquisition/evolution of reutericyclin and mutanocyclin production in *S. mutans*.

Data availability. The complete genome sequence of *S. mutans* B04Sm5 has been deposited in GenBank under the accession number CP061071. The BioProject accession number for the genome is PRJNA661123. The raw read libraries (both Oxford Nanopore and Illumina) have been deposited in the Sequence Read Archive (SRA) database under the accession number PRJNA661123.

REFERENCES

1. Pitts NB, Zero DT, Marsh PD, Ekstrand K, Weintraub JA, Ramos-Gomez F, Tagami J, Twetman S, Tsakos G, Ismail A. 2017. Dental caries. Nat Rev Dis Primers 3:17030. https://doi.org/10.1038/nrdp.2017.30.
2. Bowen WH, Burne RA, Wu H, Koo H. 2018. Oral biofilms: pathogens, matrix, and polymicrobial interactions in microenvironments. Trends Microbiol 26:229–242. https://doi.org/10.1016/j.tim.2017.09.008.
3. Lamont RJ, Koo H, Hajishengallis G. 2018. The oral microbiota: dynamic communities and host interactions. Nat Rev Microbiol 16:745–759. https://doi.org/10.1038/s41579-018-0089-x.
4. Liu L, Hao T, Xie Z, Horsman GP, Chen Y. 2016. Genome mining unveils widespread natural product biosynthetic capacity in human oral microbe *Streptococcus mutans*. Sci Rep 6:37479. https://doi.org/10.1038/srep37479.
5. Aleti G, Baker JL, Tang X, Alvarez R, Dinis M, Tran NC, Melnik AV, Zhong C, Ernst M, Dorrestein PC, Edlund A. 2019. Identification of the bacterial biosynthetic gene clusters of the oral microbiome illuminates the unexplored social language of bacteria during health and disease. mBio 10:e00321-19. https://doi.org/10.1128/mBio.00321-19.
6. Tang X, Kudo Y, Baker JL, LaBonte S, Jordan PA, McKinnie SMK, Guo J, Huan T, Moore BS, Edlund A. 2020. Cariogenic *Streptococcus mutans* produces tetramic acid strain-specific antibiotics that impair commensal colonization. ACS Infect Dis 6:563–571. https://doi.org/10.1021/acsinfecdis.9b00365.
7. Hao T, Xie Z, Wang M, Liu L, Zhang Y, Wang W, Zhang Z, Zhao X, Li P, Guo Z, Gao S, Lou C, Zhang G, Merritt J, Horsman GP, Chen Y. 2019. An anaerobic bacterium host system for heterologous expression of natural product biosynthetic gene clusters. Nat Commun 10:3665. https://doi.org/10.1038/s41467-019-11673-0.
8. Argimon S, Caufield PW. 2011. Distribution of putative virulence genes in *Streptococcus mutans* strains does not correlate with caries experience. J Clin Microbiol 49:984–992. https://doi.org/10.1128/JCM.01993-10.
9. Argimon S, Konganti K, Chen H, Alekseyenko AV, Brown S, Caufield PW. 2014. Comparative genomics of oral isolates of *Streptococcus mutans* by *in silico* genome subtraction does not reveal accessory DNA associated with severe early childhood caries. Infect Genet Evol 21:269–278. https://doi.org/10.1016/j.meegid.2013.11.003.
10. Chen YY, Clancy KA, Burne RA. 1996. *Streptococcus salivarius* urease: genetic and biochemical characterization and expression in a dental plaque streptococcus. Infect Immun 64:585–592. https://doi.org/10.1128/IAI.64.2.585-592.1996.
11. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
12. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32:1009–1015. https://doi.org/10.1093/bioinformatics/btv688.
13. 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.
14. 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.