Complete Genome Sequence of Human Oral Phylogroup 1 *Treponema* sp. Strain OMZ 804 (ATCC 700766), Originally Isolated from Periodontitis Dental Plaque

Yuki Chan,[a] Yong-Biao Huo,[a*] Xiaolin Yu,[a*] Huihui Zeng,[a] Wai Keung Leung,[a] Rory M. Watt[a]

[a]Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China

ABSTRACT Host-associated treponeme bacteria play etiological roles in human and animal soft tissue infections. *Treponema* sp. strain OMZ 804 (ATCC 700766) was isolated from dental plaque sampled from a patient with periodontitis in Switzerland in 1994. We report here the complete genome sequence of its 2.98-Mb circular chromosome.

The host-associated treponeme bacteria *Treponema medium* and *Treponema vincentii* play etiological roles in oral infectious diseases in humans and animals, especially periodontal diseases, and in various animal soft tissue infections (1–5). There are many uncertainties in treponeme taxonomy, with *T. medium*/*T. vincentii*-like taxa frequently identified within infected human and animal niches (4–9). *Treponema* sp. strain OMZ 804 (MH1F1, ATCC 700766) was originally isolated from the oral cavity of a periodontitis patient in Switzerland in 1994 (10). A previous multilocus sequence analysis (MLSA) study placed *Treponema* sp. strain OMZ 804 within a distinct species-level phylotype (*Treponema* sp. strain IC) that was phylogenetically related to *T. medium* and *T. vincentii* (9). Here, we report the complete genome sequence of *Treponema* sp. strain OMZ 804, obtained directly from the depositor, Chris Wyss.

Axenic culture of OMZ 804 was maintained anaerobically at 37°C in supplemented tryptone-yeast extract-gelatin-volatile fatty acids-serum (TYGVS) medium as described previously (9). Genomic DNA was extracted using a QIAamp DNA mini extraction kit (Qiagen, Germany) following the manufacturer's protocol.

Long-read sequencing was performed on an Oxford Nanopore Technologies (ONT) minION Mk1B device with an R9.4 flow cell (FLO-MIN106). The whole-genome sequencing library was prepared using the ONT 1D genomic DNA ligation sequencing kit (SQK-LSK108) and barcoding kit (EXP-NBD103) according to the manufacturer's protocol (version NBE_9006_v103_revP_21Dec2016), excluding the optional genomic DNA fragmentation step. The DNA was repaired using the NEBNext formalin-fixed, paraffin-embedded (FFPE) repair mix (New England Biolabs), cleaned with AMPure XP beads (Beckman Coulter), and dA tailed using the NEBNext end repair/dA-tailing module; then, sequencing adapters were ligated onto the prepared ends. ONT reads were base called with Albacore v2.3.3 (11), and adapters were removed from the sequence reads using Porechop v0.2.3 (12). The ONT sequencing generated a total of 13,606 reads that consisted of 169,920,109 bases, and the N_{50} value was 24,668.

Short-read sequencing was performed on the Illumina HiSeq 2500 platform at the Centre for PanorOmic Sciences (CPOS; The University of Hong Kong). Shotgun libraries for 101-bp paired-end reads (PE100) were prepared using the TruSeq DNA sample prep kit according to the manufacturer's protocol (part number 15026486 Rev.C). The Illumina sequences were quality filtered to remove reads containing >5% of unknown bases, >50% of bases with a quality value of ≤10, and bases with a read length of

Address correspondence to Rory M. Watt, rmwatt@hku.hk.

* Present address: Yong-Biao Huo, Zhujiang New Town Dental Clinic, Guanghua School of Stomatology, Hospital of Stomatology, Guangdong Provincial Key Laboratory of Stomatology, Sun Yat-sen University, Guangzhou, China; Xiaolin Yu, Department of Oral Implantology, Guanghua School of Stomatology, Hospital of Stomatology, Guangdong Provincial Key Laboratory of Stomatology, Sun Yat-sen University, Guangzhou, China.

<20 bp. Adaptor sequences were trimmed using Trimmomatic v0.39 (13). This yielded 5,974,798 reads and 603,431,232 bases (99.91%) after quality filtering.

Hybrid genome assembly was performed using Canu v1.8 (minReadlength = 5000) (14), with sequence polishing using Racon v1.3.1 (15) and Pilon v1.22 (16) correction using the Illumina reads. The completeness and circularity of the genome assembly were confirmed through mapping the Illumina and ONT reads using GraphMap v0.5.2 (17), placing the *dnaA* gene at the beginning of the genome assembly. Default parameters were used for all software unless otherwise specified. This resulted in one circular complete sequence (2.98 Mb; coverage, 260×) with a GC content of 44.3%.

The genome was annotated using Prokka v1.3 (18). A total of 2,891 protein-coding genes (CDSs), 49 tRNAs, and 6 rRNAs were annotated in the chromosome. Genome similarity calculated using orthologous average nucleotide identity (OrthoANI) (19) showed 84.7% identity to *T. vincentii* ATCC 35580 and F0403, 89.8% to *T. medium* ATCC 700293, and 90.0% to *Treponema* sp. strain OMZ 838 (20). The presence of PTK2 focal adhesion kinase 1 (EC 2.7.10.2), bacitracin transport system ATP-binding protein BcrA, and NADP-reducing hydrogenase subunit HndC (EC 1.12.1.3) are unique to this strain in comparison to other genome-sequenced phylogroup 1 treponeme species/phylotypes.

Data availability. The complete genome sequence has been deposited in GenBank under the accession number CP048020. The raw sequencing reads were deposited in the NCBI Sequence Read Archive (SRA) under the accession numbers SRR10954731 and SRR10954730 and BioProject accession number PRJNA602574.

REFERENCES

1. Visser MB, Ellen RP. 2011. New insights into the emerging role of oral spirochaetes in periodontal disease. Clin Microbiol Infect 17:502–512. https://doi.org/10.1111/j.1469-0691.2011.03460.x.
2. Ellen RP, Galimanas VB. 2005. Spirochetes at the forefront of periodontal infections. Periodontology 2000 38:13–32. https://doi.org/10.1111/j.1600-0757.2005.00108.x.
3. Dewhirst FE, Tamer MA, Ericson RE, Lau CN, Levanos VA, Boches SK, Galvin JL, Paster BJ. 2000. The diversity of periodontal spirochetes by 16S rRNA analysis. Oral Microbiol Immunol 15:196–202. https://doi.org/10.1034/j.1399-302x.2000.150308.x.
4. Chan ECS, McLaughlin R. 2000. Taxonomy and virulence of oral spirochetes. Oral Microbiol Immunol 15:1–9. https://doi.org/10.1034/j.1399-302x.2000.150101.x.
5. Edwards AM, Dymock D, Jenkinson HF. 2003. From tooth to hoof: treponemes in tissue-destructive diseases. J Appl Microbiol 94:767–780. https://doi.org/10.1046/j.1365-2672.2003.01901.x.
6. Paster BJ. 2001. Phylum XV. Spirochaetes, p 471–566. *In* Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Brown D, Parte A (ed), Bergey's manual of systematic bacteriology, 2nd ed, vol. 4. Springer, New York, NY.
7. Clegg SR, Carter SD, Birtles RJ, Brown JM, Hart CA, Evans NJ. 2016. Multilocus sequence typing of pathogenic treponemes isolated from cloven-hoofed animals and comparison to treponemes isolated from humans. Appl Environ Microbiol 82:4523–4536. https://doi.org/10.1128/AEM.00025-16.
8. You M, Mo S, Leung WK, Watt RM. 2013. Comparative analysis of oral treponemes associated with periodontal health and disease. BMC Infect Dis 13:174. https://doi.org/10.1186/1471-2334-13-174.
9. Huo Y-B, Chan Y, Lacap-Bugler DC, Mo S, Woo PCY, Leung WK, Watt RM. 2017. Multilocus sequence analysis of phylogroup 1 and 2 oral treponeme strains. Appl Environ Microbiol 83:e02499-16. https://doi.org/10.1128/AEM.02499-16.
10. Wyss C. 1998. Flagellins, but not endoflagellar sheath proteins, of *Treponema pallidum* and of pathogen-related oral spirochetes are glycosylated. Infect Immun 66:5751–5754. https://doi.org/10.1128/IAI.66.12.5751-5754.1998.
11. Sahoo N. 2017. Sequence base-calling through Albacore software: a part of the Oxford Nanopore Technology. MSc thesis. Orissa University of Agriculture and Technology, Bhubaneswar, India. http://krishikosh.egranth.ac.in/handle/1/5810031407.
12. Wick RR. 2017. Porechop. Github. https://github.com/rrwick/Porechop.
13. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
14. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
15. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
16. 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.
17. Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. 2016. Fast and

sensitive mapping of Nanopore sequencing reads with GraphMap. Nat Commun 7:11307. https://doi.org/10.1038/ncomms11307.

18. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

19. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.

20. Chan Y, Ma APY, Lacap-Bugler DC, Huo Y-B, Keung Leung W, Leung FC, Watt RM. 2014. Complete genome sequence for Treponema sp. OMZ 838 (ATCC 700772, DSM 16789), isolated from a necrotizing ulcerative gingivitis lesion. Genome Announc 2:e01333-14. https://doi.org/10.1128/genomeA.01333-14.