Complete Genome Sequences of Three *Phocaeicola vulgatus* Strains Isolated from a Healthy Japanese Individual

Hanh Vu,[a] Yoshinori Muto,[b] Masahiro Hayashi,[b,c] Hideki Noguchi,[d] Kaori Tanaka,[a,b,c] Yoshimasa Yamamoto[a]

aUnited Graduate School of Drug Discovery and Medical Information Sciences, Gifu University, Gifu, Japan
bInstitute for Glyco-core Research (iGCORE), Gifu University, Gifu, Japan
cLife Science Research Center, Gifu University, Gifu, Japan
dJoint Support-Center for Data Science Research, Research Organization of Information and Systems, Mishima, Japan

ABSTRACT *Phocaeicola vulgatus* (formerly *Bacteroides vulgatus*) is a pathogenic anaerobic bacterium frequently involved in human infections. We present the complete genome sequences of three *Phocaeicola vulgatus* strains isolated from the same healthy person, determined by hybrid assembly using Nanopore long-read sequencing and DNBseq short-read sequencing.

*P*hocaeicola vulgatus, one of the most numerically predominant *Bacteroides* species, has been extensively studied for its critical roles in infectious diseases as well as in other aspects of human health (1–4). Despite their clinical importance, *Bacteroides* species are most likely underrepresented in public genome databases, especially in terms of complete genomes of strains isolated from humans. Thus, there exists an urgent need for high-quality, well-established genome data to support studies on microbiome diversity and function. Hence, we sequenced and assembled the complete genomes of three *P. vulgatus* strains isolated from a healthy Japanese volunteer.

The fecal sample was collected from a volunteer who had normal bowel activity and no history of antibiotic use during the 3 months prior to the study. Samples were collected with Puritan fecal Opti-Swab CB-206 and cultured in an anaerobic atmosphere on *Bacteroides* bile esculin (BBE) agar (Kyokuto) and BBE with ceftazidime (30 mg/L) for 48 h at 37°C. Colonies of more than 1-mm diameter were subcultured on Gifu anaerobic medium agar (Nissui) under the same conditions. The isolates were identified using matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (MALDI Biotyper [MBT]). *P. vulgatus* strains were incubated in ABCM broth (Eiken Chemical) in an anaerobic chamber (Hirasawa) at 35°C for 12 h. Bacterial pellets were obtained by centrifugation. Total DNA was extracted using a NucleoBond high-molecular-weight (HMW) DNA kit (Macherey-Nagel). The DNA was quantified using Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kits (Thermo Fisher) and qualified with a NanoDrop instrument (Thermo Fisher) at an optical density of 260/280 nm ($OD_{260/280}$), and fragments were checked by electrophoresis. The same qualified DNA templates were used for both long-read and short-read sequencing.

Genome assembly was conducted using a hybrid approach combining Nanopore long-read sequencing and DNBseq short-read sequencing, similar to a previously described method (5). For long-read sequencing, a library was constructed using a ligation sequencing kit (SQK-LSK-109; Oxford Nanopore Technologies [ONT]), and sequencing was performed using a GridION X5 system (ONT) on a FLO-MIN106 flow cell. Long-read sequence data were base called using Guppy v.4.2.3. The raw reads were subjected to trimming and quality filtering using Porechop v.0.2.4 (https://github.com/rrwick/Porechop) and Filtlong v.0.2.0 (minimum length 100 bp) (https://github.com/rrwick/Filtlong). For short-read sequencing, the MGIEasy FS DNA library prep set (MGI Tech) was used for library construction. Subsequently, 2 × 150-bp paired-end sequencing was performed using the DNBSEQ-G400 platform (MGI Tech). The raw

Vu et al.

Microbiology Resource Announcements

TABLE 1 Information of the complete genome sequences of three *P. vulgatus* strains isolated from a healthy Japanese individual

Parameter[a]	Data for strain:		
	MG01-03	MG01-07	MG01-10
DNBseq sequencing			
No. of reads	32,198,982	17,338,628	28,459,232
Size (kb)	4,829,847	2,600,794	4,268,884
Avg coverage ($\times$)	911	490	805
DRA accession no.	DRR328570	DRR328571	DRR328572
ONT sequencing			
No. of reads	133,757	175,618	123,967
Size (kb)	1,129,693	1,170,228	1,148,433
Avg read length (bp)	8,445	6,663	9,264
Avg coverage ($\times$)	213	221	217
DRA accession no.	DRR328574	DRR328575	DRR328573
Assembly			
Genome structure	1 chromosome and 2 plasmids	1 chromosome and 4 plasmids	1 chromosome and 4 plasmids
DDBJ/GenBank accession no. (chromosome/plasmid name)	AP025232 (GAIMETA0103) AP025233 (pMG01-03_1) AP025234 (pMG01-03_2)	AP025235 (GAIMETA0107) AP025236 (pMG01-07_1) AP025237 (pMG01-07_2) AP025238 (pMG01-07_3) AP025239 (pMG01-07_4)	AP025240 (GAIMETA0110) AP025241 (pMG01-10_1) AP025242 (pMG01-07_2) AP025243 (pMG01-10_3) AP025244 (pMG01-10_4)
Genome size (bp) (chromosome/plasmid name)	5,073,274 (GAIMETA0103) 5,594 (pMG01-03_1) 4,306 (pMG01-03_2)	4,985,319 (GAIMETA0107) 7,659 (pMG01-07_1) 5,594 (pMG01-07_2) 4,306 (pMG01-07_3) 2,784 (pMG01-07_4)	4,957,169 (GAIMETA0110) 7,659 (pMG01-10_1) 5,594 (pMG01-10_2) 4,306 (pMG01-10_3) 2,784 (pMG01-10_4)
G+C content (%) (chromosome/plasmid name)	42.4 (GAIMETA0103) 39.6 (pMG01-03_1) 42.9 (pMG01-03_2)	42.3 (GAIMETA0107) 37.8 (pMG01-07_1) 39.6 (pMG01-07_2) 42.9 (pMG01-07_3) 41.5 (pMG01-07_4)	42.2 (GAIMETA0110) 37.8 (pMG01-10_1) 39.6 (pMG01-10_2) 42.9 (pMG01-10_3) 41.5 (pMG01-10_4)
No. of coding sequences[b]	4,552	4,484	4,454
No. of RNAs[b]	102	108	108

[a] DRA, DDBJ Sequence Read Archive.
[b] DFAST, DDBJ Fast Annotation and Submission Tool.

sequencing reads were processed using fastp v.0.20.1 (6) for trimming adapters and low-quality data, and approximately 3.5 million read pairs (1.0 Gbp) were sampled using SeqKit v.0.16.1 (7). High-quality short-read and long-read sequences were assembled using Unicycler v.0.4.8 (8) with default settings. The assembled contig graph was confirmed using Bandage v.0.8.1 (9), and the integrity of the assembled genomic data was confirmed using CheckM v.1.1.3 (10). Annotation of the assembled genomes was performed using the DDBJ Fast Annotation and Submission Tool (DFAST) (https://dfast.nig.ac.jp/). Default parameters were used for all software unless otherwise specified.

The genome information is summarized in Table 1. According to CheckM, all three of the obtained genomes were 99.25% complete with no contamination.

Data availability. The complete genome sequences of the three *P. vulgatus* strains are available in DDBJ/EMBL/GenBank under the accession numbers listed in Table 1.

ACKNOWLEDGMENTS

This work was supported by the Japan Society for the Promotion of Science KAKENHI (grant 20H00561) and the Collaboration Program at ROIS Joint Support-Center for Data Science Research (grant 012RP2021).

The study was approved by the Ethics Committee of Gifu University (Gifu, Japan; approval number 2019-164), and the study participant provided written informed consent.

Volume 11 Issue 2 e01124-21

mra.asm.org 2

REFERENCES

1. Caesar R, Fak F, Backhed F. 2010. Effects of gut microbiota on obesity and atherosclerosis via modulation of inflammation and lipid metabolism. J Intern Med 268:320–328. https://doi.org/10.1111/j.1365-2796.2010.02270.x.
2. Zou R, Xu F, Wang Y, Duan M, Guo M, Zhang Q, Zhao H, Zheng H. 2020. Changes in the gut microbiota of children with autism spectrum disorder. Autism Res 13:1614–1625. https://doi.org/10.1002/aur.2358.
3. Yoshida N, Emoto T, Yamashita T, Watanabe H, Hayashi T, Tabata T, Hoshi N, Hatano N, Ozawa G, Sasaki N, Mizoguchi T, Amin HZ, Hirota Y, Ogawa W, Yamada T, Hirata KI. 2018. Bacteroides vulgatus and Bacteroides dorei reduce gut microbial lipopolysaccharide production and inhibit atherosclerosis. Circulation 138:2486–2498. https://doi.org/10.1161/CIRCULATIONAHA.118.033714.
4. Usyk M, Pandey A, Hayes RB, Moran U, Pavlick A, Osman I, Weber JS, Ahn J. 2021. Bacteroides vulgatus and Bacteroides dorei predict immune-related adverse events in immune checkpoint blockade treatment of metastatic melanoma. Genome Med 13:160. https://doi.org/10.1186/s13073-021-00974-z.
5. Sydenham TV, Overballe-Petersen S, Hasman H, Wexler H, Kemp M, Justesen US. 2019. Complete hybrid genome assembly of clinical multidrug-resistant Bacteroides fragilis isolates enables comprehensive identification of antimicrobial-resistance genes and plasmids. Microb Genom 5:e000312. https://doi.org/10.1099/mgen.0.000312.
6. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
7. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 11:e0163962. https://doi.org/10.1371/journal.pone.0163962.
8. 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.
9. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
10. 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.