AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology® Resource Announcements

Complete Genome Sequence of *Collinsella aerofaciens* JCM 10188ᵀ

Dieter M. Tourlousse,ᵃ Mitsuo Sakamoto,ᵇ Takamasa Miura,ᶜ Koji Narita,ᵈ Akiko Ohashi,ᵃ Yoshihito Uchino,ᶜ Atsushi Yamazoe,ᶜ Keishi Kameyama,ᵈ Jun Terauchi,ᵈ Moriya Ohkuma,ᵇ Hiroko Kawasaki,ᶜ Yuji Sekiguchiᵃ

ᵃBiomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan
ᵇMicrobe Division/Japan Collection of Microorganisms, RIKEN BioResource Research Center, Tsukuba, Ibaraki, Japan
ᶜBiological Resource Center, National Institute of Technology and Evaluation (NITE), Kisarazu, Chiba, Japan
ᵈJapan Microbiome Consortium (JMBC), Osaka, Osaka, Japan

ABSTRACT We report a complete genome sequence of *Collinsella aerofaciens* JCM 10188ᵀ (=VPI 1003ᵀ). The genome consists of a circular chromosome (2,428,218 bp with 60.6% G+C content) and two extrachromosomal elements. The genome was predicted to contain 5 sets of rRNA genes, 58 tRNA genes, and 2,079 protein-encoding sequences.

Members of the genus *Collinsella* represent an important group of bacteria in the human gut. *Collinsella* spp. are considered pathobionts, being associated with diseases such as irritable bowel syndrome (1), psoriatic arthritis (2), nonalcoholic steatohepatitis (3), and symptomatic atherosclerosis (4). Elevated abundance of *Collinsella* was also correlated with rheumatoid arthritis (5), with *Collinsella aerofaciens* being experimentally demonstrated to increase gut permeability and augmented arthritis severity in mice. To allow better understanding of the bacterium's role in the gut, this study provides a complete genome sequence of *C. aerofaciens* JCM 10188ᵀ (=VPI 1003ᵀ), the authentic type strain of this species (6).

Cells were obtained from the Japan Collection of Microorganisms and cultured under an N_2 atmosphere in modified Gifu anaerobic medium (GAM) broth with 1% glucose. DNA was purified using the EZ1 DNA tissue kit (Qiagen) from cell lysates prepared by bead beating and lysozyme/proteinase K treatment. Short-read libraries were prepared with the TruSeq Nano DNA kit and sequenced on a MiSeq instrument (2 × 251-bp reads) at a coverage of ~320×. Long-read libraries were constructed with the ligation sequencing kit (SQK-LSK109; Oxford Nanopore Technologies [ONT]) and native barcoding expansion pack (EXP-NBD104; ONT); sequencing was performed on an R9.4.1 flow cell (FLO-MIN106) using the ONT MinION device. All software for read processing and assembly was run with default settings unless indicated otherwise. Base calling was performed with Guppy v3.1.5 (ONT) in high-accuracy mode with concurrent library demultiplexing and barcode trimming. To extract high-quality reads, ONT reads were filtered using NanoFilt v2.5.0 (7) (length, >1,000 bp; quality, ≥9) and then compared to the Illumina reads and quality controlled using Trimmomatic v0.38 (8) with Filtlong v0.2.0 (https://github.com/rrwick/Filtlong); the poorest 10% of read bases were discarded. A total of 68,855 ONT reads (N_{50}, 6,088 bp; ~140× coverage) were used to generate a long-read assembly using Flye v2.5 (9). Hybrid assembly was performed with Unicycler v0.4.7 (10) using the Flye assembly and 3,478,246 Illumina reads as inputs. Annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline v4.11 (11).

The genome of *C. aerofaciens* JCM 10188ᵀ contains a circular chromosome of 2,428,218 bp with a G+C content of 60.6%. The assembly also contained two extra-

chromosomal elements of 23,044 bp (57.3% G+C content) and 4,066 bp (59.3% G+C content), identified as circular by Flye/Unicycler. A BLAST search against NCBI's nucleotide database indicated that the 23-kbp element had similarity (91% identity and 64% query coverage) to an unidentified plasmid sequence from human feces (GenBank accession number CP021588), whereas the top hit of the 4-kbp element was to a plasmid sequence previously reported in the rat gut metamobilome (GenBank accession number LN852765). Further analyses would, however, be needed to substantiate that both sequences represent genuine extrachromosomal elements and to resolve their function. As a whole, the genome was predicted to contain 5 rRNA operons, 58 tRNA genes, and 2,079 protein-coding sequences.

Data availability. The genome sequence has been deposited in DDBJ/EMBL/GenBank under the accession numbers CP048433, CP048434, and CP048435. The raw ONT and Illumina sequencing reads are available in the Sequence Read Archive (SRA) under the accession numbers SRR10968455 and SRR10968458, respectively.

ACKNOWLEDGMENT

REFERENCES

1. Kassinen A, Krogius-Kurikka L, Mäkivuokko H, Rinttilä T, Paulin L, Corander J, Malinen E, Apajalahti J, Palva A. 2007. The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. Gastroenterology 133:24–33. https://doi.org/10.1053/j.gastro.2007.04.005.
2. Shapiro J, Cohen NA, Shalev V, Uzan A, Koren O, Maharshak N. 2019. Psoriatic patients have a distinct structural and functional fecal microbiota compared with controls. J Dermatol 46:595–603. https://doi.org/10.1111/1346-8138.14933.
3. Astbury S, Atallah E, Vijay A, Aithal GP, Grove JI, Valdes AM. 2019. Lower gut microbiome diversity and higher abundance of proinflammatory genus *Collinsella* are associated with biopsy-proven nonalcoholic steatohepatitis. Gut Microbes 7: 1–2. https://doi.org/10.1080/19490976.2019.1681861.
4. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, Bäckhed F, Nielsen J. 2012. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun 3:1245. https://doi.org/10.1038/ncomms2266.
5. Chen J, Wright K, Davis JM, Jeraldo P, Marietta EV, Murray J, Nelson H, Matteson EL, Taneja V. 2016. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. Genome Med 8:43. https://doi.org/10.1186/s13073-016-0299-7.
6. Kageyama A, Benno Y, Nakase T. 1999. Phylogenetic and phenotypic evidence for the transfer of *Eubacterium lentum* to the genus *Collinsella* as *Collinsella aerofaciens* gen. nov., comb. nov. Int J Syst Bacteriol 49:557–565. https://doi.org/10.1099/00207713-49-2-557.
7. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
8. 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.
9. 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.
10. 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.
11. 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.