Draft Genome Sequence of *Herminiimonas contaminans* Strain CCM 7991^T, a Biopharmaceutical Contaminant

Luis Johnson Kangale,[a,b] Rita Zgheib,[a,b] ⓘ Eric Ghigo,[b,e] ⓘ Didier Raoult,[b,c,d] ⓘ Pierre-Edouard Fournier[a,b]

[a]Aix-Marseille University, IRD, AP-HM, SSA, VITROME, Marseille, France
[b]IHU Méditerranée Infection, Marseille, France
[c]Aix-Marseille University, IRD, AP-HM, MEPHI, Marseille, France
[d]Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia
[e]Techno-Jouvence, Marseille, France

ABSTRACT *Herminiimonas contaminans* was described as a new bacterial species, a contaminant isolated from a biopharmaceutical production process in Sweden. Since the genome sequence was not available, we performed draft genome sequencing. The genome of strain CCM 7991^T (=CCUG 53591^T = DSM 28178^T = Marseille-Q4544^T) was 4,038,814 bp long, with a G+C content of 53.9%; a total of 3,860 genes were identified, along with 3 rRNAs, 44 tRNAs, and 4 noncoding RNAs (ncRNAs).

The genus *Herminiimonas* was first described by Fernandes et al. in 2005, with *Herminiimonas fonticola* as the type species (1). Strain CCM 7991 is a contaminant isolated from a biopharmaceutical production process in Sweden and was described as the type strain of *Herminiimonas contaminans* (2). Working on the planarian microbiota, we identified a new bacterial species. The characterization of this species required a comparison with *H. contaminans*, but no genome sequence was available for *H. contaminans* at the time. We acquired *H. contaminans* DSM 28178^T (=CCM 7991^T) from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). *H. contaminans* DSM 28178^T was then deposited in the Collection des Souches de l'Unité des Rickettsies (CSUR) culture collection under the number Marseille-Q4544^T. However, because no genome sequence was available for *H. contaminans*, we performed whole-genome sequencing for this species. In our laboratory, strain Marseille-Q4544^T was grown on Columbia agar supplemented with 5% sheep blood (bioMérieux, Marcy l'Etoile, France), at 28°C for 24 h. To extract bacterial genomic DNA (gDNA) from strain Marseille-Q4544^T, a mechanical treatment was first performed on a single colony by glass beads acid washed (Sigma-Aldrich Chimie, Saint-Quentin-Fallavier, France) using a FastPrep-24 5G grinder (MP Biomedicals, Illkirch, France) at maximum speed (6.5) for 90 sec. Then, after a 30-min lysozyme incubation at 37°C, DNA was extracted on an EZ1 BioRobot system with the EZ1 DNA tissue kit (Qiagen, Hilden, Germany). The gDNA was sequenced using MiSeq technology (Illumina, Inc., San Diego, CA, USA) (3) with the Nextera XT DNA sample prep kit (Illumina) and mate pair strategy. The gDNA was fragmented and then amplified by 12 cycles of PCR, followed by the addition of tag adapters and the introduction of dual-index barcodes. The libraries were purified using AMPure XP beads (Beckman Coulter, Inc., Fullerton, CA, USA), normalized using specific beads following the Nextera XT protocol, and pooled into a single library for sequencing. After 39 h, automated cluster generation and paired-end sequencing with dual-index reads were performed in a single run in 2 × 250-bp format. Total information of 5.51 Gb was obtained from a 578,000/mm^2 cluster density with clusters passing quality control filters at 93.8%. Within this run, the index representation

for strain Marseille-Q4544^T was determined to be 6.2%. The 11,232,685 paired-end reads of the MiSeq run were quality checked using FastQC version 0.11.8 with default parameters (4). We used Shovill on the Galaxy Australia server (www.usegalaxy.org.au) (Galaxy version 1.0.4+galaxy0) with the following options: Trimmomatic version 0.39 for trimming the reads and SPAdes version 3.14 for genomic assembly, with default parameters (5, 6). Strain Marseille-Q4544^T was assembled into 56 contigs (N_{50}, 201,954 bp; L_{50}, 6; coverage, 46.95×) for a total size of 4,038,814 bp, with a G+C content of 53.9%. Genomic annotation was obtained using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.13 (7). A total of 3,860 genes were identified, along with 3 rRNAs, 44 tRNAs, and 4 noncoding RNAs (ncRNAs). A BLASTN search of complete 16S rRNA gene sequences on NCBI showed 100% identity with *Herminiimonas contaminans* (16S rRNA GenBank accession number NR_108871.1).

Data availability. The genome and read sequences of *Herminiimonas contaminans* strain Marseille-Q4544^T (BioProject PRJNA677876 and BioSample SAMN16775933) were deposited in GenBank under the accession numbers JADOEL010000001 through JADOEL010000056 and SRR13040143, respectively. *Herminiimonas contaminans* strain Marseille-Q4544^T is available in the CSUR, CCM, CCUG, and DSMZ culture collections under the accession numbers CSUR Q4544^T, CCM 7991^T, CCUG 53591^T, and DSM 28178^T, respectively.

REFERENCES

1. Fernandes C, Rainey FA, Nobre MF, Pinhal I, Folhas F, da Costa MS. 2005. Herminiimonas fonticola gen. nov., sp. nov., a betaproteobacterium isolated from a source of bottled mineral water. Syst Appl Microbiol 28:596–603. https://doi.org/10.1016/j.syapm.2005.03.013.

2. Kämpfer P, Glaeser SP, Lodders N, Busse H-J, Falsen E. 2013. Herminiimonas contaminans sp. nov., isolated as a contaminant of biopharmaceuticals. Int J Syst Evol Microbiol 63:412–417. https://doi.org/10.1099/ijs.0.039073-0.

3. Ravi RK, Walton K, Khosroheidari M. 2018. MiSeq: a next generation sequencing platform for genomic analysis. Methods Mol Biol 1706:223–232. https://doi.org/10.1007/978-1-4939-7471-9_12.

4. Wingett SW, Andrews S. 2018. FastQ Screen: a tool for multi-genome mapping and quality control. F1000Res 7:1338. https://doi.org/10.12688/f1000research.15931.2.

5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

6. 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.

7. 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.