



Complete Genome Sequences of the Two Strains *Methylobacterium extorquens* NBC_00036 and NBC_00404

Faurdal, David; Jørgensen, Tue S.; Alvarez-Arevalo, Maria; Mourched, Anna-Sophie; Sterndorff, Eva B.; Charusanti, Pep; Weber, Tilmann

Published in:
Microbiology Resource Announcements

Link to article, DOI:
[10.1128/mra.00115-23](https://doi.org/10.1128/mra.00115-23)

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Faurdal, D., Jørgensen, T. S., Alvarez-Arevalo, M., Mourched, A-S., Sterndorff, E. B., Charusanti, P., & Weber, T. (2023). Complete Genome Sequences of the Two Strains *Methylobacterium extorquens* NBC_00036 and NBC_00404. *Microbiology Resource Announcements*, 12(7), e0011523. <https://doi.org/10.1128/mra.00115-23>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Complete Genome Sequences of the Two Strains *Methylorubrum extorquens* NBC_00036 and NBC_00404

David Faurdal,^a  Tue S. Jørgensen,^a Maria Alvarez-Arevalo,^a Anna-Sophie Mourched,^a Eva B. Sterndorff,^a Pep Charusanti,^a  Tilmann Weber^a

^aThe Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

ABSTRACT Here, we report the complete genome sequences of *Methylorubrum extorquens* NBC_00036 and *Methylorubrum extorquens* NBC_00404. The genomes were sequenced using the Oxford Nanopore Technologies MinION and Illumina NovaSeq systems. Both genomes are circular, with sizes of 5,661,342 bp and 5,869,086 bp, respectively.

Soil-derived bacteria have served as a source of many important secondary metabolites. Yet, methylotrophs, common in soil, have not been widely studied for their biosynthetic gene clusters (BGCs) (1). Here, we present the complete genome sequences and biosynthetic potential of two strains of *Methylorubrum extorquens*, NBC_00036 and NBC_00404.

In 2016, *Methylorubrum extorquens* NBC_00036 and NBC_00404 were isolated from Danish and Spanish soil on ISP-2 medium (2) by mixing 1 g of soil with 3 mL of 6% yeast extract, diluting this mixture 1:49 in sterile water and spreading the dilution onto International *Streptomyces* Project 2 (ISP-2) medium. Single colonies were then picked and used to make glycerol stocks. The strains were inoculated from the glycerol stock directly into liquid ISP-2 and grown at 30°C and 140 rpm for genomic DNA purification using a modified Qiagen Genomic-tip 20/G kit protocol, as described in (3).

Libraries were prepared for Nanopore sequencing using the rapid barcoding 96 kit (SQK-RBK110.96; Oxford Nanopore Technologies, Inc.) and then sequenced on an R9.4.1 flow cell using a MinION device. Barcode 23 was used for NBC_00036, and barcodes 46 and 87 were used for NBC_00404.

The data were basecalled and demultiplexed using Guppy (5.0.17 + 99baa5b, client-server API V7.0.0) applying the high-accuracy model, excluding reads smaller than 1 kb. This yielded 15,270 raw reads for NBC_00036, with an N_{50} value of 26,081 bp and a coverage of 34×, and 118,436 raw reads for NBC_00404, with an N_{50} value of 11,876 bp and a coverage of 163×.

Next-generation sequencing (NGS) DNA prep set (Novogene) libraries with size selection from 250 to 400 bp were used to generate Illumina NovaSeq data. This yielded 6,310,138 and 7,033,369 2 × 150-nucleotide (nt) read pairs for NBC_00036 and NBC_00404, providing coverages of 324× and 374×, respectively (as reported by Polypolish). The reads were trimmed using Trim Galore with Cutadapt V2.10 to a minimum length of 100 bp with a quality of at least 20 (4).

The Nanopore reads were assembled *de novo* using Flye V2.9 (5), applying --nano-raw and 5 iterations of polishing. The assemblies were then polished with the Illumina reads using Polypolish version 0.5 (6), leading to 1,744 and 6,044 changed positions from NBC_00036 and NBC_00404.

Finally, the assemblies were polished using POLCA (7) (part of MaSuRCA V4.0.5). BUSCO V5.1.2 (8) was used to check the quality of assemblies with the alphaproteobacteria_odb10 data set (8). The strains were identified as members of *Methylobacterium extorquens* (called *Methylorubrum extorquens* at NCBI) using GTDB-tk V1.7.0 (9), with the closest match being the genome found under GenBank accession number [GCF_900234795.1](https://www.ncbi.nlm.nih.gov/nuccore/GCF_900234795.1) (ANI, 96.88% and

Editor David A. Baltus, The University of Arizona

Copyright © 2023 Faurdal et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Tue S. Jørgensen, tuesparholt@gmail.com, or Tilmann Weber, tiwe@biosustain.dtu.dk.

The authors declare no conflict of interest.

Received 10 March 2023

Accepted 1 June 2023

Published 20 June 2023

TABLE 1 Attributes of *Methylorubrum extorquens* NBC_00036 and NBC_00404

| Attribute | Data for strain: | |
|---|------------------|-----------|
| | NBC_00036 | NBC_00404 |
| Mean Nanopore Coverage (×) | 34 | 163 |
| Size (bp) | 5,661,342 | 5,869,086 |
| GC content (%) | 68.37 | 68.29 |
| No. of coding sequences | 5,197 | 5,489 |
| No. of rRNAs | 15 | 15 |
| No. of antiSMASH-predicted BGCs | | |
| Homo serine lactone clusters | 1 | 1 |
| Nonalpha poly-amino acid-like ϵ -polylysine clusters | 1 | 1 |
| Nonribosomal peptide synthetase clusters | 1 | 1 |
| Redox-cofactors | 1 | 1 |
| RiPP-like factors | 1 | 0 |
| Siderophore clusters | 1 | 1 |
| Type I polyketide synthases | 1 | 1 |
| Terpene clusters | 3 | 3 |
| No. of BUSCOs | | |
| Complete | 431 | 431 |
| Single-copy complete | 428 | 430 |
| Duplicated complete | 3 | 1 |
| Fragmented | 0 | 0 |
| Missing | 1 | 1 |
| Total | 432 | 432 |

96.65% for NBC_00036 and NBC_00404, respectively). The genomes were annotated using PGAP V6.3 (10). Finally, the BGCs encoded by the strains were predicted using antiSMASH V6.2 (11), with ClusterBlast, KnownClusterBlast, and SubClusterBlast applied.

The relevant genome statistics, attributes, and predicted BGCs of the two strains are provided in Table 1. The two strains have an average nucleotide identity of 99.33%, determined using FastANI (12). This is also reflected in the fact that the predicted BGCs are almost identical, except for the RiPP-like factor encoded by NBC_00036.

Data availability. The data are available under NCBI BioProject accession number [PRJNA884983](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA884983). The raw reads were deposited in the SRA under BioSample accession numbers [SAMN31331780](https://www.ncbi.nlm.nih.gov/biosample/SAMN31331780) (NBC_00036) and [SAMN31331782](https://www.ncbi.nlm.nih.gov/biosample/SAMN31331782) (NBC_00404). The genome sequences can be found under GenBank accession numbers [CP110131](https://www.ncbi.nlm.nih.gov/genbank/CP110131) and [CP110130](https://www.ncbi.nlm.nih.gov/genbank/CP110130).

ACKNOWLEDGMENTS

We thank Oliwia Vuksanovic for her invaluable technical assistance.

This work was funded by grants from the Novo Nordisk Foundation (NNF20CC0035580 and NNF16OC0021746).

REFERENCES

- Puri AW. 2019. Specialized metabolites from methylotrophic proteobacteria. *Curr Issues Mol Biol* 33:211–224. <https://doi.org/10.21775/cimb.033.211>.
- Shirling EB, Gottlieb D. 1966. Methods for characterization of Streptomyces species. *Int J Syst Evol Microbiol* 16:313–340. <https://doi.org/10.1099/00207713-16-3-313>.
- Alvarez-Arevalo M, Sterndorff EB, Faurdal D, Jørgensen TS, Mourched A-S, Vuksanovic O, Saha S, Weber T. 2023. Extraction and Oxford Nanopore sequencing of genomic DNA from filamentous Actinobacteria. *STAR Protoc* 4:101955. <https://doi.org/10.1016/j.xpro.2022.101955>.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
- 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>.
- Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* 18:e1009802. <https://doi.org/10.1371/journal.pcbi.1009802>.
- Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput Biol* 16:e1007981. <https://doi.org/10.1371/journal.pcbi.1007981>.
- Manni M, Berkeley MR, Seppely M, Zdobnov EM. 2021. BUSCO: assessing genomic data quality and beyond. *Curr Protoc* 1:e323. <https://doi.org/10.1002/cpz1.323>.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a tool-kit to classify genomes with the genome taxonomy database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
- 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>.
- Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. AntiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.