



Complete, circular genome sequence of a *Bosea* sp. isolate from soil

Alvarez-Arevalo, Maria; Sterndorff, Eva Baggesgaard; Faurdal, David; Mourched, Anna Sophie; Charusanti, Pep; Jørgensen, Tue Sparholt; Weber, Tilmann

Published in:
Microbiology Resource Announcements

Link to article, DOI:
[10.1128/MRA.00360-23](https://doi.org/10.1128/MRA.00360-23)

Publication date:
2023

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

[Link back to DTU Orbit](#)

Citation (APA):
Alvarez-Arevalo, M., Sterndorff, E. B., Faurdal, D., Mourched, A. S., Charusanti, P., Jørgensen, T. S., & Weber, T. (2023). Complete, circular genome sequence of a *Bosea* sp. isolate from soil. *Microbiology Resource Announcements*, 12(9), 1-3. <https://doi.org/10.1128/MRA.00360-23>

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, circular genome sequence of a *Bosea* sp. isolate from soil

Maria Alvarez-Arevalo,¹ Eva Baggesgaard Sterndorff,¹ David Faurdal,¹ Anna-Sophie Mourched,¹ Pep Charusanti,¹ Tue Sparholt Jørgensen,¹ Tilmann Weber¹

AUTHOR AFFILIATION See affiliation list on p. 3.

ABSTRACT Here, we report the complete, circular genome sequence of a potential novel species from the underexplored Alphaproteobacterial genus *Bosea*. *Bosea* sp. NBC_00550 was isolated from a soil sample collected in Lyngby, Denmark. We explore the biosynthetic potential of *Bosea* sp. NBC_00550 and compare it with that of other *Bosea* species.

KEYWORDS *Bosea*, rare soil microbes, Alphaproteobacteria, nanopore

Among the Alphaproteobacteria living in soils, *Bosea* is underexplored, and at the time of writing, only seven complete genomes had been deposited at NCBI. In this paper, a strain divergent from previously sequenced isolates is presented to increase diversity to the genome database of this genus.

Bosea sp. NBC_00550 was isolated from a soil sample collected in Lyngby, Denmark (SAMN29888037, 55.798056 N 12.541667 E) at 3–5 cm depth as described in reference (1). The strain was isolated by plating on solid ISP4 medium and incubating at 30°C (2). A single colony was picked, grown in liquid medium, and frozen in 25% vol/vol glycerol. The frozen isolate was inoculated directly in liquid ISP2 medium and incubated for 5 days at 30°C, 140 rpm before DNA extraction using a modified QIAGEN Genomic-tip 20 protocol according to references (3, 4). Default parameters were used for all software and protocols unless otherwise specified.

A NEBNext Ultra II DNA library with six PCR cycles yielded 10,661,381 read pairs from an Illumina NovaSeq machine using a 2 × 150 bp strategy. Illumina reads were trimmed using Trim Galore (5) [v. 0.6.4_dev, running Cutadapt v. 2.10 (6)] with a minimum length of 100 bp and minimum quality of Q20.

A Nanopore library of *Bosea* sp. NBC_00550 was constructed using DNA from the same extraction as was used to generate the Illumina data. A MinION 9.4.1 flowcell and the SQK-RBK110.96 kit were used to generate Nanopore data, which were basecalled using the high-accuracy model in Guppy (v. 5.0.17 + 99baa5b). Barcodes were trimmed in the MinKNOW software. As a result, 1,164,240,587 bp with an N₅₀ of 10,514 bp was used for assembly.

A Flye assembly (v.2.9-b1768, --iterations 5) (7) was polished with Illumina data using Polypolish (v.0.5.0) (8) and POLCA (v.4.0.5) (9). The genome integrity was evaluated with Bandage (v.0.8.1) (10), and the core gene content was evaluated with BUSCO (v.5.1.2, database alphaproteobacteria_odb10) (11). The low number of bases in the assembly not covered by the Illumina reads indicates a very high assembly quality (Table 1, reported by Polypolish). The taxonomy was determined with GTDBtk (v.1.7.0) (12), and NCBI PGAP (v.6.3) (13) was used for automated annotation. AntiSMASH (v. 6.1.1) (14) was used to identify biosynthetic gene clusters (BGCs), and Clinker (v.0.0.24) (15) was used to compare the gene clusters to *Bosea* complete genomes from the NCBI database.

Editor J. Cameron Thrash, University of Southern California, Los Angeles, California, USA

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

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 22 May 2023

Accepted 26 June 2023

Published 21 August 2023

Copyright © 2023 Alvarez-Arevalo 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/).

TABLE 1 Accession numbers and assembly characteristics of the sequenced *Bosea* strain

<i>Bosea</i> sp. NBC_00550		
Assembly length (bp)	6,267,852	
Total BUSCO groups searched: 432	Complete and single copy	430
	Complete and duplicated	2
	Fragmented	0
	Missing	0
PGAP ^a annotation	No. of CDS ^b	5,908
	No. of tRNAs	50
	No. of rRNAs	2, 2, 2 (5S, 16S, 23S)
Detected BGCs by antiSMASH (14)	Three regions: terpene, NAPAA, NRPS-T1PKS hybrid	
BioProject	PRJNA861150	
BioSample	SAMN29888037	
Assembly accession number	GCA_026020075.1	
GenBank accession no.	CP102772	CP102773
Locus name	NBC_00550	Plasmid
	chromosome	pNBC550
Topology	Circular	Circular
Length (bp)	5,687,190	580,662
Nanopore coverage (x)	194	116
Illumina coverage (x)	500.1	372.5
Uncovered bases (bp)	2	2
Closest GTDBtk hit (genus, species, accession number, %ANI)	<i>Bosea</i> sp. F3-2, GCF_008253865.1 88.05 %ANI	
Complete <i>Bosea</i> genomes from NCBI (accession numbers)	GCA_001741865 , GCA_008253865 , GCA_003952665 , GCA_002220095 , GCA_001562255 , GCA_011764485 , GCA_001713455	

^aPGAP, Prokaryotic Genome Annotation Pipeline.

^bCDS, coding sequences.

The assembly produced a circular chromosome of 5.7 Mbp with 66% GC and a large circular plasmid of 581 kbp with 63% GC (Table 1). The Nanopore and Illumina coverages were 194 and 500, respectively, and 432 (100%) BUSCO genes were complete. *Bosea* sp. NBC_00550 carries a non-alpha poly-amino acid (NAPAA) BGC with the specificity-conferring code DLedLgTVvK, which is also found in all seven other complete *Bosea* genomes (Table 1), and a terpene cluster is also found in all other seven *Bosea* genomes. The complete *Bosea* genomes carry between three and nine regions encoding BGCs, and in addition to the shared NAPAA and terpene BGCs, *Bosea* sp. NBC_00550 carries a hybrid NRPS-T1PKS cluster, which is similar to a hybrid NRPS-T1PKS BGC found in *Bosea vaviloviae* ([GCA_001741865](#)). AntiSMASH predicts that the NRPS-T1PKS hybrid has glycosyltransferases and polysaccharide biosynthesis genes near the small core NRPS-T1PKS gene.

ACKNOWLEDGMENTS

We thank Oliwia Vuksanovic and Alexandra Hoffmeyer for their invaluable technical assistance.

This work was supported by the Novo Nordisk Foundation grants NNF16OC0021746 and NNF20CC0035580.

AUTHOR AFFILIATION

¹The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

AUTHOR ORCID*s*

Tue Sparholt Jørgensen  <http://orcid.org/0000-0002-2437-7086>

Tilmann Weber  <http://orcid.org/0000-0002-8260-5120>

FUNDING

Funder	Grant(s)	Author(s)
Novo Nordisk Fonden (NNF)	NNF16OC0021746	Tilmann Weber
Novo Nordisk Fonden (NNF)	NNF20CC0035580	Tilmann Weber

AUTHOR CONTRIBUTIONS

Maria Alvarez-Arevalo, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft | Eva Baggesgaard Sterndorff, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review and editing | David Faurdal, Conceptualization, Data curation, Formal analysis, Investigation, Writing – review and editing, Resources | Anna-Sophie Mourched, Conceptualization, Methodology, Writing – review and editing, Resources | Pep Charusanti, Resources | Tue Sparholt Jørgensen, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review and editing, Resources, Software, Supervision | Tilmann Weber, Conceptualization, Investigation, Writing – review and editing, Supervision, Funding acquisition, Project administration

DATA AVAILABILITY

Data are available as BioProject [PRJNA861150](https://ncbi.nlm.nih.gov/bioproject/PRJNA861150). SRA accession numbers for raw reads are [SRR20852345](https://ncbi.nlm.nih.gov/sra/SRR20852345) and [SRR20852344](https://ncbi.nlm.nih.gov/sra/SRR20852344). GenBank accession numbers are listed in Table 1.

REFERENCES

- Faurdal D, Jørgensen TS, Alvarez-Arevalo M, Mourched A-S, Sterndorff EB, Charusanti P, Weber T. 2023. Complete genome sequences of the two strains *Methylorubrum Exorquens* Nbc_00036 and Nbc_00404. *Microbiol Resour Announc* 12:e0011523. <https://doi.org/10.1128/mra.00115-23>
- Nielsen JB, Gren T, Mohite OS, Jørgensen TS, Klitgaard AK, Mourched AS, Blin K, Oves-Costales D, Genilloud O, Larsen TO, Tanner D, Weber T, Gotfredsen CH, Charusanti P. 2022. Identification of the biosynthetic gene cluster for pyracrimycin A, an antibiotic produced by *Streptomyces* sp. *ACS Chem Biol* 17:2411–2417. <https://doi.org/10.1021/acscchembio.2c00480>
- QIAGEN Genomic DNA Handbook. 2015. QIAGEN. <https://www.qiagen.com/us/resources/download.aspx?id=d2b85b26-16dd-4259-a3a7-a08cbd2a08a3&lang=en>.
- Alvarez-Arevalo M, Sterndorff EB, Faurdal D, Jørgensen TS, Mourched A-S, Vuksanovic O, Saha S, Weber T. 2022. Extraction and Oxford nanopore sequencing of genomic DNA from filamentous actinobacteria. *STAR Protoc* 4:101955. <https://doi.org/10.1016/j.xpro.2022.101955>
- Krueger F. 2023. Trim Galore! a wrapper around Cutadapt and Fastqc to consistently apply adapter and quality trimming to Fastq files, with extra Functionality for RRBS data. Available from: <https://github.com/FelixKrueger/TrimGalore>
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10. <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, Schneidman-Duhovny D. 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, Ouzounis CA. 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>
- Wick R.R, 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>
- Manni M, Berkeley MR, Seppely M, Zdobnov EM. 2021. BUSCO: Assessing Genomic data quality and beyond. *Current Protocols* 1:e323. <https://doi.org/10.1002/cpz1.323>
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit 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>
- Gilchrist CLM, Chooi YH. 2021. Clinker & clustermapJs: automatic generation of gene cluster comparison figures. *Bioinformatics* 37:2473–2475. <https://doi.org/10.1093/bioinformatics/btab007>