



UNIVERSITY OF HELSINKI

<https://helda.helsinki.fi>

Metagenome-assembled genomes from mineral tundra soils in Rásttigáisá, northern Norway

Pessi, Igor S.; Rutanen, Aino; Hultman, Jenni

2024-01-23

Microbiology Society

<http://hdl.handle.net/10138/592796>

Pessi, I S, Rutanen, A & Hultman, J 2024, 'Metagenome-assembled genomes from mineral tundra soils in Rásttigáisá, northern Norway', *Access microbiology*, vol. 6, no. 1, 000655.v3. <https://doi.org/10.1099/acmi.0.000655.v3>

Downloaded from Helda, University of Helsinki institutional repository. <https://helda.helsinki.fi>
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Metagenome-assembled genomes from mineral tundra soils in Rásttigáisá, northern Norway

Igor S. Pessi^{1,2,3,*}, Aino Rutanen¹ and Jenni Hultman^{1,2,4}

Abstract

Microbial communities in tundra soils remain largely unknown despite their important roles in the cycling of greenhouse gases. Here, we report 59 non-redundant metagenome-assembled genomes (MAGs) recovered from mineral tundra soils in Rásttigáisá, northern Norway. The MAGs were obtained by clustering contigs according to tetranucleotide frequency and differential coverage and were manually curated to remove contigs with outlying GC content and/or mean coverage. Most MAGs were assigned to the bacterial phyla *Candidatus* Dormibacterota ($n=12$), Verrucomicrobiota ($n=10$), and Acidobacteriota ($n=9$). All archaeal MAGs ($n=4$) belong to the genus *Candidatus* Nitrosopolaris (phylum Thermoproteota). The 59 Rásttigáisá MAGs expand our knowledge of the diversity and ecological roles of tundra microbiomes.

DATA SUMMARY

The data generated in this study have been submitted to the European Nucleotide Archive (ENA) under the project PRJEB49283. Raw reads are available under accessions ERR11584940–ERR11584949, and accession numbers for the assembled genomes are listed in Table S1, available in the online version of this article.

INTRODUCTION

Human activities such as unsustainable industrial and agricultural practices are driving irreversible changes in tundra soils and other polar ecosystems. Greenhouse gases (GHGs) play a major role in climate change and, if emissions are not curbed drastically, atmospheric temperatures in the Arctic may be 7°C warmer by the end of the century [1]. Microbial communities in the tundra play key roles in both the production and consumption of greenhouse gases and are thus a crucial component of the global climate system [2, 3]. Despite harsh conditions, tundra soils harbour highly diverse and specialized micro-organisms, most of which still remain unknown [4–8]. Fuelled by ever-increasing sequencing and computational power, recent culture-independent metagenome investigations have revealed that uncultured micro-organisms – the so-called ‘microbial dark matter’ – play critical roles in the carbon [4–6] and nitrogen [6–8] cycles in the tundra. More studies on the diversity and functional capacity of tundra microbiomes are thus needed to understand better their contribution to global biogeochemical cycles and the GHG budget. Here, we report 59 metagenome-assembled genomes (MAGs) recovered from tundra soils in northern Norway.

DESCRIPTION OF THE DATASET

We sampled soils across ten sites in an area of alpine tundra in the Rásttigáisá Fell (69°59' N, 26°15' E, 700 m.a.s.l), Finnmark, Norway. In each site, one sample was taken from the mineral layer (10–15 cm depth) with a soil corer and transferred immediately to dry ice. Soil water content ranged from 9.6–17.3%, organic matter from 1.2–7.6% and pH from 4.5 to 5.3. Metagenomic DNA was extracted with the DNeasy PowerSoil kit (QIAGEN, Hilden, Germany) and libraries were prepared with the Nextera XT kit

Received 27 June 2023; Accepted 12 January 2024; Published 23 January 2024

Author affiliations: ¹University of Helsinki, Helsinki, Finland; ²Helsinki Institute of Sustainability Science (HELSUS), Helsinki, Finland; ³Finnish Environment Institute (Syke), Helsinki, Finland; ⁴Natural Resources Institute Finland (Luke), Helsinki, Finland.

***Correspondence:** Igor S. Pessi, igor.pessi@helsinki.fi

Keywords: *Candidatus* Dormibacterota; metagenomics; tundra soils.

Abbreviations: ENA, European Nucleotide Archive; GC, guanine-cytosine; GHG, greenhouse gas; GTDB, genome taxonomy database; MAG, metagenome-assembled genome.

One supplementary figure and one supplementary table are available with the online version of this article.

000655.v3 © 2024 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

(Illumina, San Diego, CA, USA) following the manufacturers' instructions. Paired-end sequencing was done with the Illumina NextSeq500 platform (forward reads: 170 bp, reverse reads: 140 bp) at the Institute of Biotechnology, University of Helsinki. A total of 114950383 paired-end reads was obtained (35.6 Gb), ranging from 7696757 to 14621365 reads per sample (2.4–4.5 Gb).

Sequencing data was processed using a genome-resolved metagenomics approach. First, *Cutadapt* v1.10 [9] was used to remove adapters and base calls with Phred score <28, and the quality of the data was checked with *FastQC* v0.11.5 [10]. Sequences were then assembled into larger contiguous sequences with *MEGAHIT* v1.1.1 [11]. Assemblies were done for each sample individually and as one co-assembly consisting of pooled data from all samples. For each assembly, contigs ≥2500 bp were binned with *anvi'o* v6.0 [12] according to Pessi et al. [8]. Briefly, MAGs were obtained by manually grouping the contigs according to tetranucleotide frequency and differential coverage using the *anvi-interactive* interface of *anvi'o*. MAGs were visually inspected to remove outlying contigs, and only MAGs with homogeneous sequence composition and coverage were kept (see Fig. S1 and merenlab.org/2015/

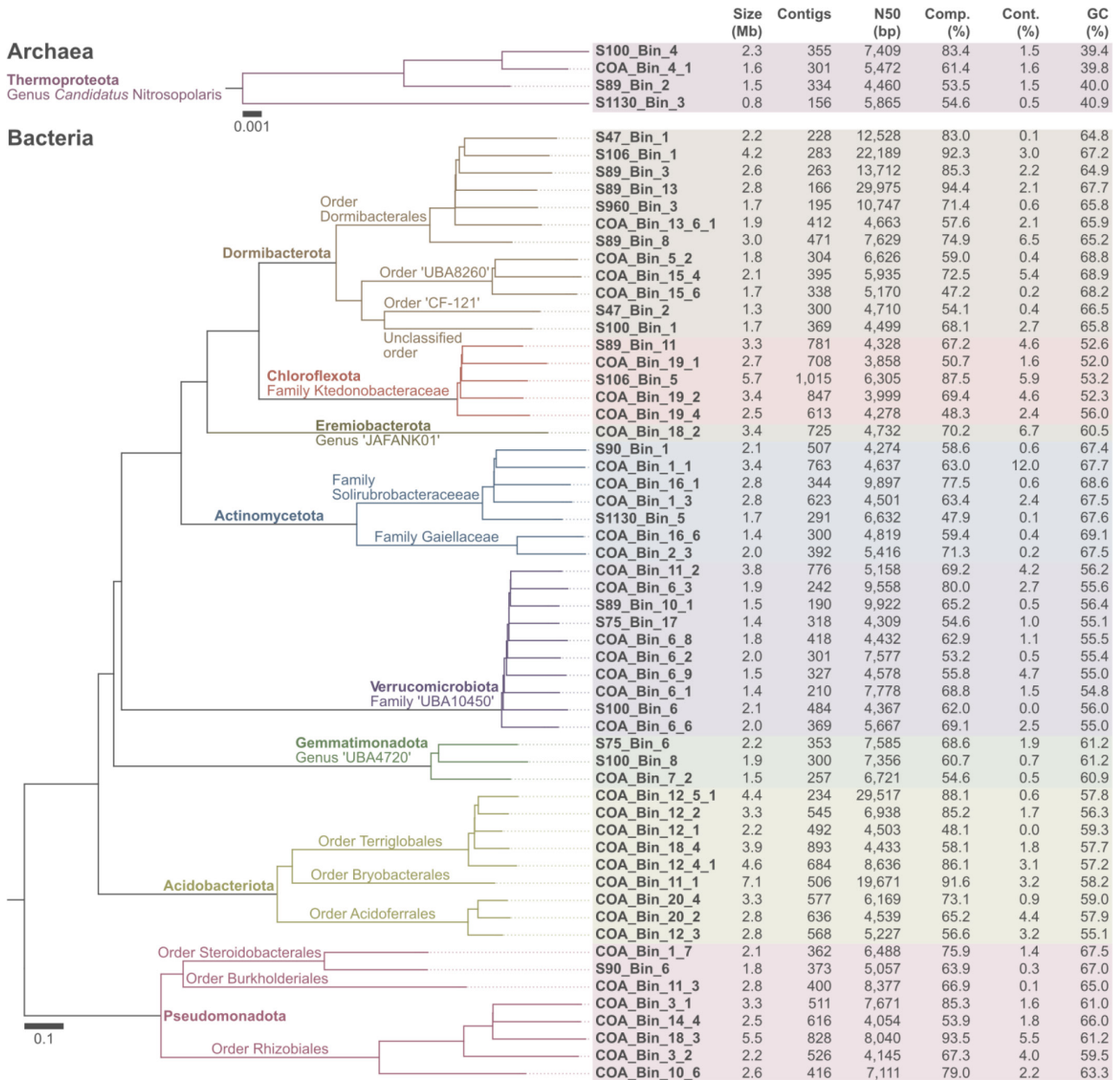


Fig. 1. Fifty-nine metagenome-assembled genomes (MAGs) from mineral tundra soils in Råsttiggåisa, northern Norway. Maximum-likelihood trees obtained with *GTDB-Tk* v2.3.0 based on 53 archaeal and 120 bacterial single-copy genes. Completion and redundancy values were obtained with *Checkm2* v1.0.1.

05/11/anvi-refine for more information). MAGs from all assemblies were combined and *dRep* v2.3.2 [13] was used to yield a set of 59 non-redundant MAGs ($\geq 99\%$ average nucleotide identity).

The 59 non-redundant MAGs range from 0.8 to 7.1 Mb and comprise 156–1015 contigs with N_{50} values of 3858–29 975 bp (Fig. 1, Table S1). Completeness and contamination levels estimated with *CheckM2* v1.0.1 [14] are of 47.2–94.4% and 0.0–12.0%, respectively. These estimates are based on machine-learning models trained on simulated genomes, which seem to give more reliable estimates than previous single-copy gene-based approaches [14]. GC content ranges from 40.0–69.0%. Taxonomic classification with *GTDB-Tk* v2.3.0 [15] and the Genome Taxonomy Database (GTDB) release R214 [16] placed the MAGs within the bacterial phyla *Candidatus* Dormibacterota ($n=12$), Verrucomicrobiota ($n=10$), Acidobacteriota ($n=9$), Pseudomonadota ($n=8$), Actinomycetota ($n=7$), Chloroflexota ($n=5$), Gemmatimonadota ($n=3$) and Eremiobacterota ($n=1$). The remaining MAGs ($n=4$) were assigned to the archaeal phylum Thermoproteota. These have been characterized elsewhere as part of the recently proposed genus *Candidatus* Nitrosopolaris [8].

FUTURE OUTLOOK

The 59 Rásttigáisá MAGs obtained here represent medium- and high-quality genomes from microbial populations in mineral tundra soils. Among other applications, these population genomes can be used in studies of microbial phylogenomics, biogeography and comparative genomics looking at adaptations to life under cold, oligotrophic conditions. An example of this is the study by Pessi et al. [8], in which four of the MAGs reported here were included in the description of a novel genus of putative ammonia-oxidizing archaea that appears to be restricted to polar and alpine environments. Our preliminary analysis of the Rásttigáisá MAGs suggests that other novel taxa might be represented in this dataset (Table S1), but further studies are needed for a better characterization of this potentially novel diversity. The prevalence of uncultured taxa has also been observed in other metagenomic investigations of tundra soils in Alaska [4, 5], Finland [6, 7], Sweden [17], as well as Antarctic soils [18, 19]. In particular, analysis of the MAGs assigned to *Candidatus* Dormibacterota, which is the most abundant phylum in the dataset, can cast light into the ecological roles of this quite enigmatic group of micro-organisms in cold and oligotrophic soils [18, 19]. Further functional annotation and metabolic reconstruction of the Rásttigáisá MAGs can provide clues about the potential roles of these populations in the cycling of GHGs [7, 17]. Despite the small scale when considered individually, this and other similar studies [20–22] contribute to increasing the coverage of the polar microbiome census.

Funding information

This study was funded by the Academy of Finland (project 335354) and the University of Helsinki.

Acknowledgements

We thank Miska Luoto and the BioGeoClimate Modelling Lab at the University of Helsinki for providing the samples and the CSC – IT Center for Science for the computing resources.

Author contributions

I.S.P.: data curation, supervision, validation, writing – original draft, writing – review & editing. A.R.: formal analysis, writing – review & editing. J.H.: conceptualization, funding acquisition, project administration, supervision, writing – review & editing.

Conflicts of interest

The author(s) declare that there are no conflicts of interest.

References

- Post E, Alley RB, Christensen TR, Macias-Fauria M, Forbes BC, et al. The polar regions in a 2°C warmer world. *Sci Adv* 2019;5:eaaw9883.
- Hugelius G, Loisel J, Chadburn S, Jackson RB, Jones M, et al. Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proc Natl Acad Sci U S A* 2020;117:20438–20446.
- Voigt C, Marushchak ME, Abbott BW, Biasi C, Elberling B, et al. Nitrous oxide emissions from permafrost-affected soils. *Nat Rev Earth Environ* 2020;1:420–434.
- Hultman J, Waldrop MP, Mackelprang R, David MM, McFarland J, et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 2015;521:208–212.
- Taş N, Prestat E, Wang S, Wu Y, Ulrich C, et al. Landscape topography structures the soil microbiome in arctic polygonal tundra. *Nat Commun* 2018;9:777.
- Viitamäki S, Pessi IS, Virkkala A-M, Niittynen P, Kemppinen J, et al. The activity and functions of soil microbial communities in the finnish sub-arctic vary across vegetation types. *FEMS Microbiol Ecol* 2022;98:fiac079.
- Pessi IS, Viitamäki S, Virkkala A-M, Eronen-Rasimus E, Delmont TO, et al. In-depth characterization of denitrifier communities across different soil ecosystems in the tundra. *Environ Microbiome* 2022;17:30.
- Pessi IS, Rutanen A, Hultman J. *Candidatus* Nitrosopolaris, a genus of putative ammonia-oxidizing archaea with a polar/alpine distribution. *FEMS Microbes* 2022;3:xtac019.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 2011;17:10.
- Andrews S. FastQC: a quality control tool for high throughput sequence data; 2019. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31:1674–1676.
- Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 2021;6:3–6.

13. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 2017;11:2864–2868.
14. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023;20:1203–1212.
15. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022;38:5315–5316.
16. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–D794.
17. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* 2018;560:49–54.
18. Ji M, Greening C, Vanwonterghem I, Carere CR, Bay SK, et al. Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* 2017;552:400–403.
19. Montgomery K, Williams TJ, Brettle M, Berengut JF, Zhang E, et al. Persistence and resistance: survival mechanisms of *Candidatus* Dormibacterota from nutrient-poor Antarctic soils. *Environ Microbiol* 2022;24:4491.
20. Rusley C, Onstott TC, Vishnivetskaya TA, Layton A, Chauhan A, et al. Metagenome-assembled genome of USC α AHI, a potential high-affinity methanotroph from Axel Heiberg Island, Canadian high arctic. *Microbiol Resour Announc* 2019;8:e01178-19.
21. Castro-Nallar E, Berríos-Farías V, Díez B, Guajardo-Leiva S, Thrash JC. Spatially and temporally explicit metagenomes and metagenome-assembled genomes from the Comau Fjord (42°S), Patagonia. *Microbiol Resour Announc* 2023;12:e00059-23.
22. Sun EW-H, Hajirezaie S, Dooner M, Vishnivetskaya TA, Layton A, et al. Thaumarchaea genome sequences from a high Arctic active layer. *Microbiol Resour Announc* 2020;9:e00326-20.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org

Peer review history

VERSION 2

Editor recommendation and comments

<https://doi.org/10.1099/acmi.0.000655.v2.3>

© 2023 Efthimiou G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

Georgios Efthimiou; University of Hull, Biomedical Science, Hardy Building, Cottingham Road, UNITED KINGDOM, Hull

Date report received: 12 December 2023

Recommendation: Accept

Comments: The work presented is clear and the arguments well formed. This study would be a valuable contribution to the existing literature. This is a study that would be of interest to the field and community. All comments by the reviewers were satisfactorily addressed.

SciScore report

<https://doi.org/10.1099/acmi.0.000655.v2.1>

© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

iThenticate report

<https://doi.org/10.1099/acmi.0.000655.v2.2>

© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

Author response to reviewers to Version 1

This is an interesting data paper about the sequencing of soil bacteria in Norway. It is well-written, with sound methodology and several good relevant references.

We thank the reviewer for reviewing our manuscript.

Few minor corrections:

Abstract: Provide a few sentences in the beginning about soil microbes and your methods.

Abstract: At the end, explain why are your findings important for ecology/medicine/industry.

These suggestions have been implemented (page 1; lines 10–11, 12–15, 17–19):

Future outlook: Expand about similar applications and studies and compare results with theirs.

Future outlook: Highlight the importance of your findings for ecology/medicine/industry.

Future outlook: Are there any tundra soil studies in other countries?

These suggestions have been implemented (pages 3–4; lines 87–89, 92–96):

The authors present metagenome sequencing results from mineral tundra soils from northern Norway. The results include 59 individual genomes assembled from the metagenome data. These genomes have briefly been characterised both in content and taxonomically. Similar to genome announcement papers, the major contribution here is highlighting the availability of the genomic data and contextualising it rather than providing significant in-depth analysis. This said the authors already have provide value from this in another paper highlighted with in. Minor comments are below. No significant changes are recommended beyond expansions and this reviewer congratulates the authors on their work.

We thank the reviewer for reviewing our manuscript.

Very short abstract (<50 words) but then, short paper, I would expand it somewhat but leave this to editors' discretion.

Indeed, we find that a short abstract is commensurate with this type of article. However, the abstract has been expanded according to the comments from Reviewer 1 (page 1; lines 10–11, 12–15, 17–19):

Data summary

Data is accessible via links.

Nothing to comment.

Introduction

I would encourage expanding the introduction, specifically expanding on the contents of references 4-8, it would be helpful for readers, even in a few brief lines to know what these references truly represent (like a mini review).

This suggestion has been implemented (page 2; lines 33–36):

Description of the data set

Line 38 - (170+140 bp) - ? I'm not sure what this means and unsure if this is a typo suggesting paired end reads?

This has been clarified (page 2; lines 47–48):

Line 49 - Is it possible to include an example of an excluded MAG and an included MAG. To provide readers with an example and perhaps a sense of what the limits of this selection process. Ultimately, it's very subjective. I leave it to the editor discretion as other journals have published similarly but I always find this lacks robustness in its reproducibility. This is an important criterion for ACMI.

We thank the reviewer for pointing this out. A figure portraying a “good” and a “bad” bin has been added (page 2; lines 62–63):

Figure S1. Diagram showing examples of a good- and a bad-quality metagenomic bin. Made with *anvi'o* v6.0 (Erenet *al.*, 2021).

Line 54 - I would expand on what Completeness and contamination level estimates mean within the manuscript. GC content for example is relatively self-explanatory. The others require additional knowledge.

This issue has been clarified (page 3, lines 67–69):

VERSION 1

Editor recommendation and comments

<https://doi.org/10.1099/acmi.0.000655.v1.5>

© 2023 Efthimiou G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

Georgios Efthimiou; University of Hull, Biomedical Science, Hardy Building, Cottingham Road, UNITED KINGDOM, Hull

Date report received: 25 October 2023

Recommendation: Minor Amendment

Comments: The work presented is clear and the arguments well formed. This study would be a valuable contribution to the existing literature. This is a study that would be of interest to the field and community. The reviewers have highlighted minor concerns with the work presented. Please ensure that you address their comments.

Reviewer 2 recommendation and comments

<https://doi.org/10.1099/acmi.0.000655.v1.4>

© 2023 Munnoch J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

John Munnoch; University of Strathclyde, SIPBS, Glasgow, UNITED KINGDOM

<https://orcid.org/0000-0002-9018-6026>

Date report received: 22 November 2023

Recommendation: Minor Amendment

Comments: The authors present metagenome sequencing results from mineral tundra soils from northern Norway. The results include 59 individual genomes assembled from the metagenome data. These genomes have briefly been characterised both in content and taxonomically. Similar to genome announcement papers, the major contribution here is highlighting the availability of the genomic data and contextualising it rather than providing significant in-depth analysis. This said the authors already have provide value from this in another paper highlighted with in. Minor comments are below. No significant changes are recommended beyond expansions and this reviewer congratulates the authors on their work. Very short abstract (

Please rate the manuscript for methodological rigour

Satisfactory

Please rate the quality of the presentation and structure of the manuscript

Good

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?

No

Is there a potential financial or other conflict of interest between yourself and the author(s)?

No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?

Yes

Reviewer 1 recommendation and comments

<https://doi.org/10.1099/acmi.0.000655.v1.3>

© 2023 Miller G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License.

Gilbert Miller; University of Cape Town, SOUTH AFRICA

Date report received: 22 November 2023

Recommendation: Minor Amendment

Comments: This is an interesting data paper about the sequencing of soil bacteria in Norway. It is well-written, with sound methodology and several good relevant references. Few minor corrections: Abstract: Provide a few sentences in the beginning about soil microbes and your methods. Abstract: At the end, explain why are your findings important for ecology/medicine/industry Future outlook: Expand about similar applications and studies and compare your results with theirs. Future outlook: Highlight the importance of your findings for ecology/medicine/industry Future outlook: Are there any tundra soil studies in other countries?

Please rate the manuscript for methodological rigour

Good

Please rate the quality of the presentation and structure of the manuscript

Good

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?

No

Is there a potential financial or other conflict of interest between yourself and the author(s)?

No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?

Yes

SciScore report

<https://doi.org/10.1099/acmi.0.000655.v1.1>

© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.

iThenticate report

<https://doi.org/10.1099/acmi.0.000655.v1.2>

© 2023 The Authors. This is an open-access article report distributed under the terms of the Creative Commons License.