

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Metagenomics analysis scripts are publicly available at https://github.com/GreeningLab/AntarcticevolutionManuscript .
Data analysis	Databases used to analyse data or against which data was compared to include: - UniProtKB (https://www.uniprot.org/help/uniprotkb) - Curated metabolic marker databases (https://bridges.monash.edu/collections/_/5230745) - GTDB (release 220; https://gtdb.ecogenomic.org/) - EggNOG (v5.0; http://egglog5.embl.de/) - proMGE (http://promge.embl.de/) Data were analysed using the following published software: - BBTools v38.80 - metaSPAdes v3.13.1 - metaFlye v2.8 - MetaBAT2 v2.15 - MaxBin2 v2.1.1 - CONCOCT v1.1.0 - VAMB v3.0.2 - DAS_Tool v1.1.6 - dRep v3.4.3 - CheckM v1.1.2 - GTDB-Tk v2.4.0 - IQ-TREE v2.3.6

- Prodigal v.2.6.3
- ETE Toolkit v.3
- DIAMOND v2.1.10
- AlphaFold v2.0
- Foldseek v8.ef4e960
- UCSF Chimera v1.19
- MetaCHIP v1.10.13
- eggNOG-mapper v2.1.9
- CoverM v0.6.0
- Anvi'o v8.0
- HyPhy v2.5.29

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The metagenomic data and MAGs used in this study are available through NCBI BioProject PRJNA630822 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA630822/>). The protein amino acid sequences, protein structures and phylogenetic trees of hydrogenases are available at figshare (<https://doi.org/10.6084/m9.figshare.30000151>). All other study data are included in the article and/or supporting information. Source Data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Research sample

Sampling strategy	Sampling was performed via targeted collection of proglacial and mountainous soils selected to represent diverse terrestrial soil habitats within the McMurdo Dry Valleys, one of the largest ice-free regions in Antarctica. No formal statistical methods were used to predetermine the sample size. Instead, the sample size (n = 16 independent soil sites) was determined based on environmental logistics, seasonal field accessibility, and geographic representation. These samples generated 676 metagenome-assembled genomes (MAGs), providing the basis for the downstream HGT and polymorphism analyses.
Data collection	Raw sequencing reads were obtained from Ortiz et al., 2021, as referenced in the text, with additional high-depth short-read (DNBseq and NextSeq500) and long-read (Oxford Nanopore PromethION) sequencing generated for the representative sample MS6-5.
Timing and spatial scale	Soil samples were collected from 14 to 22 January 2015. The study included 16 independent soil samples collected from geographically distinct sites north of the Mackay Glacier region, South Victoria Land, Antarctica.
Data exclusions	No data were excluded from the analyses.
Reproducibility	All analyses were computational and can be reproduced according to the described methods and available code.
Randomization	Randomization was not applicable because this was an observational study analyzing all available Antarctic soil metagenomes without experimental allocation or randomized sampling.
Blinding	Blinding was not applicable as this study involved standardized computational processing of environmental sequencing data without subjective assessment or experimental intervention.
Did the study involve field work?	☒ Yes ☐ No

Field work, collection and transport

Field conditions	The soil samples were collected between 11 and 22 January 2015 from 16 ice-free sites in the Mackay Glacier region, South Victoria Land, Antarctica. Fieldwork was conducted during the austral summer under typical Antarctic field conditions.
Location	Sixteen surface soil samples were collected from independent ice-free sites in the Mackay Glacier region, South Victoria Land, Antarctica, between 11 and 22 January 2015. Samples were collected from proglacial, mountainous and other ice-free habitats at elevations ranging from 157 to 1,109 m above sea level. Surface soils (0-5 cm depth) were collected from approximately 1 m ² at each site. Detailed site-specific metadata, including GPS coordinates, sampling time and other metadata of soil samples are available publicly under the corresponding NCBI accession numbers associated with this study, as well as within the original data source publications.
Access & import/export	This study uses sequencing data generated from archival soil samples originally collected during historical field expeditions (2014/2015) by collaborating institutions. The original field work was performed in strict compliance with the Antarctic Treaty and national environmental protection regulations governing research in South Victoria Land. The collected samples were legally transported to the University of Pretoria (South Africa) and subsequently transferred to Monash University (Australia) under strict biosecurity protocols, including mandatory 50 kGy gamma irradiation at a quarantine-approved facility for compliance with Australian Department of Agriculture regulations. All molecular data generated have been legally deposited into the NCBI database.
Disturbance	Sampling caused minimal environmental disturbance. Approximately 50 g of surface soil (0-5 cm depth) was collected aseptically from an approximately 1 m ² area at each site using sterile techniques, and sampling locations were left with minimal disturbance.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study	n/a	Involved in the study
☒	☐ Antibodies	☒	☐ ChIP-seq
☒	☐ Eukaryotic cell lines	☒	☐ Flow cytometry
☒	☐ Palaeontology and archaeology	☒	☐ MRI-based neuroimaging
☐	☒ Animals and other organisms		
☒	☐ Clinical data		
☒	☐ Dual use research of concern		
☒	☐ Plants		

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	The study did not involve laboratory animals.
Wild animals	The study did not involve wild animals.
Reporting on sex	The study did not involve macro-organisms or sex-based analysis; the biological material examined was the total soil microbiome.
Field-collected samples	Metagenomic data were derived from 16 terrestrial soil samples collected from the north of the Mackay Glacier region in South Victoria Land, Antarctica. Samples were collected using sterile 50 mL Falcon tubes, maintained below 0°C during field transport, and stored at -80°C upon arrival at the laboratory prior to DNA extraction. No live laboratory housing, maintenance, or controlled experimental photoperiods were applicable, as samples were frozen immediately to preserve the in situ microbial community structure.
Ethics oversight	No ethical approval or oversight was required for this study, as it did not involve any human subjects, vertebrate animals, or regulated multicellular organisms. Sampling in Antarctica was conducted in accordance with the relevant national Antarctic research program guidelines and environmental protection protocols.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>