

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	We searched for the presence of near-identical species to Antarctic geothermal MAGs in the Sequence Read Archive (SRA) using the branchwater CLI client v. 0.4.0, with default parameters (SRA was accessed 09-04-2026).
Data analysis	Data were analysed with the following software. Branchwater hits (containment ≥0.3) were evaluated with YACHT v1.3.1 (k=21, minimum coverage 1.0), with metagenomes retrieved via fasterq-dump (SRA Toolkit v3.0.2); spatial consolidation of hits used the sf package v1.1.0 in R, and the hydrothermal-proportion vs. location-count relationship was fitted with nls in base R Stats v4.5.2. A phylogenomic tree of Antarctic geothermal MAGs was built with GToTree v1.8.16 (Prodigal v2.6.3, HMMER3 v3.4, Muscle v5.1, trimAl v1.5.rev1, FastTree2 v2.1.11) and visualised in iTOL. Isolate genomes were quality-trimmed with Trimmomatic v0.39, assembled with SPAdes v4.0.0, assessed with QUAST v5.2.0, and matched to MAGs using the EZBioCloud ANI calculator. Fissure 8 MAG presence in Hawaiian metagenomes was assessed with sourmash gather (k=21), and thermophilic MAGs were selected using GenomeSPOT v1.0.1. MAGs were characterised with CheckM2 v1.01 (genome properties), Tome v1.0.0 (optimal growth temperature), gRodon2 v2.7.2 (maximum growth rate), GenomeSPOT (temperature, pH and salinity tolerance), run_dbcan v4.1.4 (CAZymes), and COGclassifier v1.0.5 (COG categories, including UV-resistance and dormancy gene searches by COG ID). Metabolic marker genes were identified with DIAMOND v2.1.6 against the custom Greening Lab databases (HydDB 2022 for hydrogenases). 16S rRNA amplicon data (ice core, Antarctic soils, air samples) were demultiplexed with idemp, primer-trimmed with Cutadapt v2.1, and processed with DADA2 (v1.26 for ice core and soils; air-sample ASVs inferred as for the ice core data) using SILVA v138.2 for taxonomy (50% bootstrap threshold). Contaminants were removed with Decontam (prevalence method; v1.32.0). Downstream analyses were run in R v4.3.3 using mctoolsr, vegan (richness via specnumber), phyloseq (v1.50.0 for air data), indicpecies (MultiPatt, r.g function, α=0.05), rstatix and base R wilcox.test, with plots produced in ggplot2 and cowplot. Custom code used for DADA2 processing followed the fiererlab/dada2_fiererlab pipeline.

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](#)

All sequence data associated with this project is available online. Raw 16S rRNA gene sequence data for the air samples and controls are on the Sequence Read Archive (SRA) under BioProject PRJNA1456287. Raw 16S rRNA gene sequencing data for the ice cores and controls are on the Sequence Read Archive (SRA) under BioProject PRJNA1460789. Genome sequences for the isolates have been deposited into GenBank, under the accessions JCKWB000000000 (Thermus), JCKWD000000000 (Calidithermus), and JCKWE000000000 (Geobacillus thermoleovorans). The versions of these genomes described in this paper are versions JCKWX010000000. Antarctic geothermal MAG sequences have been deposited into Genbank, Biosamples SAMN57478730 - SAMN57478840. All scripts for processing the data and recreating the figures can be found on Github, <https://github.com/ThermophileResearchUnit/Global-thermophiles-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

Multi-part descriptive and experimental study of whether thermophiles from Antarctic geothermal sites are dispersed. Components: (1) k-mer searches (branchwater, then YACHT-calibrated cANI thresholds) of public SRA metagenomes for near-identical species to Antarctic geothermal MAGs, with habitat classification and spatial consolidation of hits; (2) phylogenomic placement and trait/metabolic-marker characterisation of MAGs; (3) isolation of three thermophiles (Geobacillus, Thermus, Calidithermus) with laboratory survival assays (UV; freezing/starvation/desiccation) and growth-temperature assays for the two Thermaceae; (4) 16S rRNA amplicon profiling of a Fang Glacier ice core, reprocessing of 200 published Antarctic soil datasets, and a preliminary air-sampling survey on Mt Erebus; (5) searching Fissure 8 (Hawai'i) thermophilic MAGs against Hawaiian metagenomes. There are no treatment factors or interactions except exposure time (UV: 0, 2, 5, 8, 12, 18 min; freezing: multiple time points) and, for Geobacillus, culture age (overnight vs 7 days, to vary spore content). Survival assays used triplicate dishes per time point. Ice core: 25 samples, 5 extraction blanks, 3 no-template controls. Air: 3 samples, no replication.

Research sample

Geothermally heated soils/MAGs from Tramway Ridge and Western Crater (Mt Erebus), Cryptogram Ridge (Mt Melbourne) and Mt Rittmann (described in ref 30); Fissure 8 steam-vent biofilms, Big Island, Hawai'i (19 thermophilic MAGs from soils >50 °C, PRJNA1459608); the 12.5 m Fang Glacier ice core (Mt Erebus); three isolates from Tramway Ridge material, chosen to match cosmopolitan MAGs or (for Geobacillus) to match known, widely distributed thermophile; 200 Antarctic soil 16S datasets (PRJNA1214840; 166 retained); SRA metagenomes (accessed 09-04-2026) and PRJEB52128 Hawaiian metagenomes.

Sampling strategy

Convenience/opportunistic sampling from existing projects and public databases; isolates were obtained by enrichment/plating. Sample size calculation was prevented by logistical constraints of Antarctic access, exhaustive use of available MAGs.

Data collection

Soil: top 10 cm removed with a sterile scoop, collected with a sterile spatula into 50 mL tubes, temperature measured with a

Data collection	Checktemp 1C probe. Glass wool incubated in a borehole as a passive fumarolic-steam sampler. Air: SASS3100 sampler, peroxide-cleaned filter, 300 L/min, tripod 1.5 m, downwind edge of geothermal area. Ice: cores cut into 1 m lengths; tephra sections subsampled under contamination controls, outer 0.5–1 cm scraped, melted, filtered (0.2 µm Sterivex). Sequencing: Illumina HiSeq/ NovaSeq X Plus (isolate genomes), MiSeq (16S; Univ. Colorado Boulder, Massey Univ.). Survival data: CFU plate counts. Isotope data: Los Gatos analyser. Samples for culturing were collected by S.E.N., M.B.S., and I.R.M. Air samples collected by I.R.M. and M.B.S. Ice cores collected by P.K.
Timing and spatial scale	Isolate soils: November 2023 or 2024, 10 cm depth. Glass wool borehole incubation: Nov 2023–Nov 2024. Air sampling: November 2023 (≥4 h per sample; Western Crater duration unknown because of battery failure). Ice cores drilled November 2007 (processed later, in 2022). Survival assays and growth assays (7 days) were conducted over several months from Aug. 2025 - Mar. 2026. Spatial scale: individual geothermal sites on three Antarctic volcanoes (Ross Sea region), one Hawaiian site, a glacier cirque on Erebus, 200 soils across Antarctica, and global SRA metagenomes.
Data exclusions	Branchwater hits with containment <0.3 were excluded. Synthetic metagenomes and metagenomes with no assignable habitat class were removed. Four MAGs were omitted from the tree for too few single-copy gene hits: Erebus-WC_bin.38.permissive_final, Erebus-WC_bin.59.orig_final, Mel_bin.5.orig_final, Ritt_bin.10.strict_final. Fissure 8 MAGs were restricted to predicted thermophiles (GenomeSPOT). The 15 m ice core was not used because it was visibly dirty. Ice core 16S: 10 chloroplast and 11 mitochondrial ASVs removed, plus Decontam-identified contaminants. Antarctic soils: 1,064 non-prokaryotic or unassigned ASVs removed; samples with <2,000 reads excluded. Air 16S: 567 contaminant ASVs (including Scott Base-unique ASVs), 41 ASVs with 1 read total, and 5 chloroplast/mitochondrial/ eukaryote ASVs removed, leaving 78 ASVs.
Reproducibility	Survival assays used triplicate dishes per time point, which were the result of splitting one batch culture into separate dishes. For survival assays, a t least one preliminary experiment was used for each isolate to gauge the dilutions needed for CFU counting. Growth experiments were triplicate cultures. Air data are n = 1 per site and so unreplicated and preliminary. 16S PCRs were run in duplicate; extraction blanks and no-template controls were included. Computational steps list software versions, and DADA2 followed the fiererlab pipeline.
Randomization	Not applicable: observational/descriptive sampling, with no allocation of units to groups. In the Geobacillus spore comparison, groups were defined by culture age, not randomly assigned.
Blinding	Blinding was not relevant because sequence analyses were automated and no group allocation was involved. CFU counting was not blinded.
Did the study involve field work?	☒ Yes ☐ No

Field work, collection and transport

Field conditions	Tramway Ridge soil ~65 °C at sampling; borehole 62 °C during incubation; air sampling at the downwind edge of the heated ground in November 2023. Air temperature for all fieldwork was below -20C and above -35C. Wind speed and direction were not recorded for any fieldwork.
Location	Tramway Ridge and Western Crater, Mt Erebus; Cryptogram Ridge, Mt Melbourne; Mt Rittmann; Fang Glacier cirque, northern flank of Mt Erebus; sea ice in front of Scott Base (air control); Fissure 8, Big Island, Hawai'i. Sample locations specifics: Mt Erebus, Western Crater (WC-50; WC air) -77.531962 167.117093 2019 (soil); 2023 (air) Mt Erebus, Tramway Ridge (TR; TR air) -77.518333 167.111389 2009 (soil); 2023 (air) Mt Erebus, Fang Glacier (Fang1-25) -77.4954 167.2311 2007 Fissure 8, Hawai'i (F8-8, 9, 11, 14, 15-22) 19.4615 -154.90906 2022
Access & import/export	Samples collected in Antarctica were collected using permits from the New Zealand Ministry of Foreign Affairs and Trade (MFAT), which authorizes field sampling in Antarctica under the Protocol on Environmental Protection to the Antarctic Treaty. Samples collected in Tramway Ridge were collected within an Antarctic Specially Protected Area (ASPA), with the appropriate approval and ASPA permits obtained from MFAT prior to sampling, as well as subsequent reporting to MFAT. All soil and culture samples were then imported to New Zealand under a Restricted Import Permit for bringing research samples into New Zealand.
Disturbance	Tramway Ridge: soil: small surface (top 10 cm) samples taken with sterile tools. A borehole was used for the one-year glass wool incubation; during installation, walking on and disturbance to warm soils were minimized. Tyvek suits were worn while on site at Tramway Ridge, with all equipment and suits and nitrile gloves sterilized using 70% ethanol prior to entry to the site, per ASPA guidelines. All equipment and clothing was sterilized with ethanol prior to moving between sites on Erebus (Tramway Ridge and Western Crater).

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A