

Genomic and experimental evidence from Antarctic geothermal sites indicates diverse thermophiles are globally cosmopolitan

Ian McDonald

ian.mcdonald@waikato.ac.nz

University of Waikato <https://orcid.org/0000-0002-4847-6492>

Stephen Noell

University of Waikato <https://orcid.org/0000-0001-7895-3846>

Nicholas Dragone

Monash University <https://orcid.org/0000-0002-8983-6482>

Manon Cano-Prasad

Te Whare Wānanga o Waikato - University of Waikato

Rebecca Pyne

Earth Sciences New Zealand

Laura Rea

University of Colorado Boulder

Nancy Bertler

Earth Sciences New Zealand

Philip Kyle

New Mexico Institute of Mining and Technology <https://orcid.org/0000-0001-6598-8062>

Jimmy Saw

The George Washington University <https://orcid.org/0000-0001-8353-3854>

Steven Archer

AgResearch <https://orcid.org/0000-0001-5704-2701>

Craig Herbold

<https://orcid.org/0000-0003-3479-0197>

Craig Galilee

Te Whare Wānanga O Waitaha - University of Canterbury

Matthew Stott

University of Canterbury <https://orcid.org/0000-0002-2082-9460>

Stephen Cary

University of Waikato <https://orcid.org/0000-0002-2876-2387>

Article

Keywords: Thermophiles, dispersal, Antarctica, geothermal

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11236748/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Mt. Erebus, the world's southernmost active volcano, is highly isolated, yet a previous study found its heated soils contain prokaryotes found elsewhere globally. This was unexpected, as non-sporulating thermophilic microorganisms are thought to be dispersal limited. We searched for the presence of metagenome assembled genomes (MAGs) from Antarctic geothermal sites outside Antarctica, finding that 62% of MAGs had matching species globally. Three of these cosmopolitan thermophiles isolated from Mt. Erebus could survive atmospheric-like conditions, with UV being the strongest selective pressure. Moreover, hot soils from Hawai'i sampled four years after a lava flow hosted many thermophilic MAGs with identical strains elsewhere globally, indicating that dispersal/recolonisation by thermophiles can be rapid. Finally, ice core sections from Fang Glacier (Mt. Erebus) associated with volcanic events were enriched in thermophiles, implying that thermophiles could disperse via volcanic events. Our findings dispel the common conception of thermophilic microorganisms as being inherently dispersal limited.

Introduction

The phenomenon of microorganisms being transported globally via aeolian dispersal is well-studied ¹⁻³, with modelling showing that intercontinental transport has the potential to occur for microbial organisms ⁴. There are multiple nuances to this global aeolian dispersal, however, with transport between hemispheres having a lower likelihood of occurring ⁴. Moreover, cell size, seasonality (windspeed variation and the nature of particulates in the air between seasons), and source ecosystem type strongly influence atmospheric residence times for prokaryotic cells, which tend to be on the order of days to weeks ^{4,5}. The most important nuance, however, is that not all microbial organisms can survive aeolian transport ⁶. Some microbial species are at the extreme ends of distribution, being extremely broadly distributed globally (e.g., ⁷) or endemic to a given region (e.g., the thermophilic genus *Venenivibrio*, phylum *Aquificota*, endemic to Aotearoa-New Zealand ⁸). However, most prokaryotic species fall into the middle range of being neither extremely narrowly nor globally distributed ³. Indeed, geographic distance can often be a primary determinant of microbial population structure ⁹, especially when taxonomic resolution is considered ¹⁰. It is often thought that terrestrial microorganisms are more capable of long-distance dispersal than marine microorganisms, due to their ability to "hitch a ride" on dust particles, providing protection from UV damage ^{11,12}. Recently, an important, but not mutually exclusive, distinction has been made between microorganisms with "disperser" and/or "colonizer" traits ¹³. Dispersal-associated traits allow greater chances of survival during aeolian dispersal conditions ⁶ (high UV exposure, freezing conditions, starvation, desiccation ^{12,14}), while colonizer-associated traits allow for the quick establishment of a species upon arrival at a site ¹³.

In contrast, thermophilic prokaryotes (optimal growth temperature $\geq 50^{\circ}\text{C}$) are generally thought to be much more severely geographically isolated than mesophilic prokaryotes. For instance, hot spring

cyanobacteria have been found to be geographically isolated^{15,16}, despite many mesophilic cyanobacteria being capable of long-range dispersal¹⁷. Thermophilic microorganisms from Yellowstone National Park were also found to be geographically isolated from each other, with volcanic caldera history being the most important factor for distribution and no evidence for contemporary dispersal¹⁸. Thermophilic microorganisms from a variety of other hot spring environments have similarly indicated a lack of evidence for dispersal between regions, with few widespread lineages^{19–21}. A survey of the global dispersion of microorganisms from different habitats inferred the lowest rates of dispersal for microorganisms from hot springs of any environment studied except the subsurface²². This dispersal limitation is generally attributed to the discontinuous and often distant nature of potential habitats for thermophiles (and extremophiles more broadly), given that these microorganisms often have specific growth requirements for intemperate habitats^{21,23}. The strong environmental filtering of many geothermal environments has also been cited as a significant factor restricting the dispersal of thermophiles^{24,25}. Endospore-forming *Geobacillus* (and related genera) are a notable exception to this trend, with their spores being known to be extremely widespread in distribution^{26,27}.

Antarctic geothermal soils provide an excellent study site to explore the dispersal capabilities of thermophilic microorganisms. These habitats are geographically and environmentally isolated, being surrounded by snow- and ice-covered soils year-round. Antarctica is also relatively isolated with regards to off-continent atmospheric inputs, with most aerial dispersal of microbes in Antarctica occurring from other sites in Antarctica^{6,28}. However, geothermally heated soils from Mt. Erebus, the southernmost active volcano in the world and largest geothermally active region in Antarctica, were found to harbour a large number of prokaryotes highly similar (> 97% ID based on the V4-V5 region of 16S rRNA gene sequence) to species found elsewhere globally²⁹.

In this study, we sought to build on these previous findings, starting by exploring the distribution of recently published metagenome assembled genomes (MAGs) from four Antarctic geothermal sites: Tramway Ridge, Western Crater (both from Mt. Erebus), Mt. Melbourne, and Mt. Rittmann³⁰. Using MAGs provides clearer taxonomic resolution, more confidence in matches, and information about genomic content that could facilitate the exploration of disperser and colonizer traits. We sought to address the following questions: (1) Can we find support, using MAGs and cultures from Antarctic geothermal sites, for the presence of a population of globally distributed, thermophilic microorganisms? If so, (2) how frequent are these dispersal events (extremely rare vs. common)? (3) What traits might allow long-range aeolian dispersal? (4) What are the potential mechanisms for thermophiles to get into the atmosphere? Our null hypothesis was that thermophiles exhibit strong geographic localization and little to no global dispersal. Our alternate hypothesis was that there is a globally distributed population of thermophiles, with dispersal events happening regularly (on the scale of months to years) and similar traits to mesophiles enabling this dispersal. Finally, we explored the hypothesis that volcanic events play a role in dispersing thermophiles.

Results

Global distribution of Antarctic geothermal thermophilic microorganisms

Given that our first question focuses on whether thermophiles are capable of global dispersal, we chose to focus on the most isolated geothermally active sites in the world: geothermally heated soils in Antarctica (Tramway Ridge and Western Crater from Mt. Erebus, as well as Mts. Melbourne and Rittmann; Table 1). These sites are not only very geographically isolated from other geothermal sites but also are surrounded by an environment hostile to thermophiles (extreme cold, snow/ice), are extremely isolated from human contact (most are within Antarctic Specially Protected Areas), and are ecologically simple (the only visible life present consist of scattered mosses and cyanobacterial mats). The largest of these sites, Mt. Erebus, is the southernmost active volcano on Earth and harbours extensive, geothermally heated soils up to 65°C. We reasoned that, if the null hypothesis is correct, these highly geographically and environmentally isolated sites should be almost exclusively made up of non-cosmopolitan, potentially endemic species. We thus examined the global distribution of recently published³⁰, de-duplicated metagenome assembled genomes (MAGs) binned from these sites, filtered thermophiles based on predicted optimal growth temperature $\geq 50^{\circ}\text{C}$.

Table 1

Sampling locations and sample information for samples used in this paper. n.m.: not measured. n.a.: not applicable.

Sample	Site	Latitude	Longitude	Sampling year	Soil pH	Temp. (°C)
Mel1-3	Mt. Melbourne, Antarctica	-74.350016	164.700139	2011	6.43	52
Ritt1-3	Mt. Rittmann, Antarctica	-73.448445	165.497444	2011	6.73	69
WC-50	Mt. Erebus - Western Crater, Antarctica	-77.531962	167.117093	2019	9.2	50
TR	Mt. Erebus - Tramway Ridge, Antarctica	-77.518333	167.111389	2009	6.3	65
F8-8-22	Fissure 8, Hawaii	19.4615	-154.90906	2022	n.m.	55
F8-9-22	Fissure 8, Hawaii	19.4615	-154.90906	2022	n.m.	56.7
F8-11-22	Fissure 8, Hawaii	19.4615	-154.90906	2022	n.m.	55
F8-14-22	Fissure 8, Hawaii	19.4615	-154.90906	2022	n.m.	57.6
F8-15-22	Fissure 8, Hawaii	19.4615	-154.90906	2022	n.m.	55
Fang1 - Fang25 ice cores	Mt. Erebus – Fang Glacier, Antarctica	-77.4954	167.2311	2007	n.a.	n.a.
TR air	Mt. Erebus - Tramway Ridge, Antarctica	-77.518333	167.111389	2023	n.a.	n.a.
WC air	Mt. Erebus - Western Crater, Antarctica	-77.531962	167.117093	2023	n.a.	n.a.

We used branchwater to search the Sequence Read Archive (SRA) for metagenomes containing any of these MAGs, with YACHT used to generate MAG-specific, dataset-specific containment-based ANI thresholds at which each MAG could be classified as present in each metagenome (at an ANI of 96% to reduce false positives) (see Table S3 for cANI thresholds used for each MAG). We defined “cosmopolitan” MAGs as those that had a matching species (ANI $\geq$ 96%) present in at least two metagenomes (to avoid spurious hits on singleton metagenomes) outside of Antarctica. We classified other MAGs as “unknown,” given that the absence of a MAG in the SRA does not necessarily indicate that that MAG is endemic due to sampling bias. See Methods for a further explanation.

Based on these criteria, we identified 18 cosmopolitan MAGs and 11 MAGs with unknown global distribution MAGs (62% cosmopolitan). These thermophilic, cosmopolitan MAGs were from a variety of

phyla, including *Acidobacteriota*, *Armatimonadota*, *Chloroflexota*, CSP1-3 (also known as GAL-15, now *Ca. Syusimicrobiota*³¹), *Deinococcota*, *Dormibacterota*, *Gemmatimonadota*, and *Thermoproteota* (Figure S2, Table S4). These MAGs were globally distributed, with the average number of unique geographic locations per MAG being eight (median: 5 locations) (Fig. 1A, Table S4). The containing metagenomes came from a variety of habitat types, with non-hydrothermal soil and hydrothermal environments being the most common types (Fig. 1B).

If thermophiles are inherently less dispersed than mesophiles (null hypothesis), we would predict that a greater proportion of mesophiles from these Antarctic geothermal sites would be globally cosmopolitan. However, the opposite was true: 32 out of 82 mesophilic MAGs were globally cosmopolitan (39% of MAGs) (Table S4). However, the cosmopolitan mesophiles tended to have a greater distribution (mean global locations: 25, median: 6; Table S4) than thermophiles. Across both thermophiles and mesophiles, we did not see a particularly strong phylogenetic signature, with cosmopolitan and non-cosmopolitan representatives found in all phyla with more than one MAG (Figure S2). Three exceptions stood out: the phylum *Proteobacteria* (*Pseudomonadota*) had a large proportion of cosmopolitan MAGs (9/11 MAGs), the phylum *Bacteroidota* did not have any cosmopolitan MAGs, and the class *Ktedonobacteria* (phylum *Chloroflexota*) did not have any cosmopolitans.

Cultures validate the ability of certain thermophiles to survive dispersal conditions

We were able to culture three cosmopolitan thermophiles from Mt. Erebus: two from surface Tramway Ridge soils from the *Deinococcota* phylum, family *Thermaceae*, *Calidithermus* sp. Yb (99.6% ANI to Erebus-TR_bin.4) and *Thermus* sp. SS11024-3 (99.4% ANI to Erebus-TR_bin.12); and a strain of *Geobacillus thermoleovorans* (st. GW348) from fumarolic steam at Tramway Ridge (99% ANI to *Geobacillus thermoleovorans*, GCF_001610955.1) (Table 2). The *Thermaceae* isolates had matching MAGs from our metagenome data, while the *Geobacillus* species was not identified as present in either Tramway Ridge or Western Crater surface samples (using YACHT, 95% ANI). These species were found to have a spectrum of global distribution and traits (Fig. 1, Table 2). We used these isolates to test their capability of surviving aeolian dispersal conditions (UV exposure, desiccation, starvation, and freezing), with the hypothesis that the species with the broadest distribution (*Calidithermus* sp. Yb and *G. thermoleovorans* st. GW348) would have the greatest tolerance and the species with the narrowest distribution (*Thermus* sp. SS11024-3) would have the least tolerance. Given the known sporulation ability of *G. thermoleovorans*, we tested whether it would show increased tolerance to dispersal conditions when increased spores were present in the culture. Our use of *G. thermoleovorans* also served as a positive control, given that this genus is known to be extremely widespread globally³².

Table 2

Genome information, taxonomy, and global distribution (based on branchwater results) of the three thermophilic species used for survival tests. *: number of DNA repair genes associated with UV resistance¹⁰².

Genus, species	Y1b	SS11024-3	GW348
	<i>Calidithermus sp. nov</i>	<i>Thermus sp. nov</i>	<i>Geobacillus thermoleovorans</i> st. GW348
Genome size, Mbp	3.47	2.11	3.32
Genome completeness, %	100	99.58	100
Genome contamination, %	0.86	1.29	0.04
GC content, %	68.43	64.87	53
Coding sequences	3332	2108	3371
DNA repair genes*	78	52	48
Pigmentation?	Yellow	Yellow	None
Matching MAG	Erebus-TR_bin.4 (99.6% ANI)	Erebus-TR_bin.12 (99.4% ANI)	NA
Global distribution	3 continents; 7 locations	1 continent; 5 locations	3 continents; 20 locations

We found that all three species (and both vegetative and spore-enriched *G. thermoleovorans*) were highly resilient to freezing, starvation, and desiccation up to ~ 30 days (Fig. 2A). Of the four tested species/physiological states, vegetative *G. thermoleovorans* cells showed the most vulnerability to the treatment, with percentage survival falling to 10% within 1 day and staying around 5–10% for the remainder of the experiment. All others exhibited similar survival rates (> 50% surviving) for the entire course of the experiment.

For UV exposure, we chose to use UV-C (200-280nm), which is the most lethal of the UV wavelength ranges. We exposed liquid cultures to a maximum of 1800 J/m² of UV-C (18 mins exposure under a germicidal UV-C lamp), which is roughly equivalent to 5 days of exposure in the stratosphere at 20 km altitude at a mid-latitude location in boreal springtime¹⁵. We observed much less resilience of our cultures to UV-C exposure than freezing/desiccation/starvation (Fig. 2B). *Thermus* sp. SS11024-3 was the most susceptible, with only 2.4×10^{-5} % of cells surviving at 1800 J/m² (18 min exposure). Spore-enriched *G. thermoleovorans* st. GW348 exhibited the most UV resistance, with 0.01% of cells surviving at the maximum UV-C dose. *Calidithermus* sp. Yb initially showed the most UV resistance at 500 J/m² UV-C (5 min exposure; 0.8% of cells surviving) but then fell to similar survival rates as vegetative *G. thermoleovorans*.

We also explored the temperature tolerance of the two *Thermaceae* species to determine if they would be capable of growth at mesophilic temperatures encountered in non-geothermally heated soils. We found that *Calidithermus* sp. Yb was capable of appreciable growth at 35 and 45°C, while the lowest temperature at which *Thermus* sp. SS11024-3 showed appreciable growth was 45°C (Table S6).

Exploring traits that may enable dispersal and colonization in thermophiles

We next sought to explore what traits may be enabling the survival of thermophiles under aeolian dispersal conditions (high UV, low temperatures, starvation conditions¹²). We used the framework recently proposed¹³ for distinguishing between dispersal traits (GC content^{33,34} [although this is a complex trait^{35,36}], DNA repair^{37,38}, dormancy-related genes¹³, and the metabolic potential for trace gas utilization^{39,40}) and colonization traits (metabolic generalism^{7,41,42} (genome size⁷, metabolic flexibility^{7,42}), wider tolerance to different growth conditions⁷, and faster maximal growth rates¹³). We acknowledge that this analysis is challenging to conduct, given that we are categorizing MAGs into binary bins of “cosmopolitan” and “MAGs with unknown global distribution,” which oversimplifies the situation. However, our goal is to provide preliminary evidence of the types of adaptations required for global dispersal and compare between thermophiles and mesophiles in the same sample set to determine if similar adaptations are required for both.

We did not find any significant differences (two-sided Wilcoxon test, Benjamini-Hochberg adjustment for multiple tests, $p > 0.05$) between Antarctic geothermal cosmopolitan and unknown MAGs in any of the traits studied, including other traits not known to be linked to dispersal or colonisation, except for temperature range (Fig. 3A-C, Figure S3). For temperature range, cosmopolitan MAGs were predicted to have a median range around 24°C, compared to around 20°C for unknowns. Cosmopolitans did tend to have a greater GC content than unknowns (Fig. 3A), and cosmopolitans also tended to be more likely to have greater amounts of dormancy-related genes (Fig. 3B). Mesophilic, cosmopolitan MAGs from Antarctic geothermal sites had significantly (Wilcoxon $p < 0.05$, Benjamini-Hochberg adjusted for multiple tests) greater GC content than unknown MAGs, but the two categories were not significantly different in any other traits, including dormancy-related genes (Figure S4A-D).

We then looked for the presence of a correlation between either of these traits (GC content and temperature range) and the maximum distance between Mt. Erebus and SRA metagenome samples with a matching species. Longer distances indicate either that that species has arrived at that more distant point via a single, longer trip in the atmosphere, or that it has successfully survived multiple, shorter trips, with multiple colonization events having taken place. We found that, for both thermophiles and mesophiles, there was a significant ($p = 0.009$ or 2.21×10^{-5} for thermophiles or mesophiles, Benjamini-Hochberg adjusted), strong ($R^2 = 0.42$ or 0.46 for thermophiles or mesophiles), positive linear correlation between GC content of cosmopolitan MAGs and maximum distance (Fig. 3D, Figure S4E). Both mesophiles and thermophiles had similar linear regression equations, with MAGs having maximum

distances > 30,000km having GC > 0.66 or 0.62 (thermophiles or mesophiles). There was no significant correlation between growth temperature range and maximum distance for thermophiles (adjusted $p = 0.31$).

We also found that cosmopolitan thermophilic MAGs were more likely (6/18 vs. 2/11 for cosmopolitan vs. unknown, respectively) to have predicted hydrogenase genes than the other categories of MAGs, as well as having a greater diversity of hydrogenase gene types (Fig. 3E). Four of the six cosmopolitan MAGs with hydrogenases had high affinity ⁴³ (Group 1h or Group 2a) hydrogenases, as opposed to only one MAG from the unknown group. Additionally, cosmopolitan MAGs were more likely to be phototrophic. This was not the case for mesophiles, with similar proportions of cosmopolitans or unknowns having various dispersal-related predicted metabolisms (Figure S4F).

Evidence indicates dispersal of thermophilic microorganisms can be frequent

Our above results do not indicate the frequency with which dispersal events might occur for thermophiles. To begin to answer this question, we collected soil samples from geothermally heated soils (> 50°C) in Fissure 8 (Hawaiʻi Island, USA) within 4 years of the formation of these geothermal hot spots due to a lava flow (Table 1). We binned 19 predicted thermophilic MAGs from metagenomes generated from these samples (Table S4) and performed the same branchwater analysis as above. We predicted that, if the null hypothesis is correct, all these MAGs would have arrived either locally from nearby geothermal sites or globally from known sporulating thermophiles (e.g., *Geobacillus*). If our alternative hypothesis is true, we would still expect many Fissure 8 MAGs to have arrived from local sources, given their proximity, but we would also expect to see a non-zero proportion of the MAGs to be globally cosmopolitan, non-*Geobacillus*-type species with matching strains globally (given their recent arrival).

Given that we were looking for evidence of recent dispersal events (< 4 years), we used a cANI threshold of 100% to identify identical strains in metagenomes globally using branchwater, reasoning that recent arrivals to the site should have little to no divergence from the site of origin. We identified 8 MAGs that had identical strains present globally, all but one of which were from metagenomes outside of Hawaiʻi (Figure S5; Table 3). These MAGs were almost exclusively found in samples from the northern hemisphere (Figure S5). Of these 8 MAGs, several came from identical phyla (and, in the case of *Chthonomonas* and *Pyrinomonas*, identical genera) as Antarctic geothermal cosmopolitan thermophiles (Table S4).

Table 3

Searching for identical strains (cANI = 100%) of Fissure 8 MAGs in either recently published metagenomes from Hawaiian hot soils or the Sequence Read Archive globally. Global loc.: number of locations the MAG was identified in at 100% cANI globally in the SRA.

MAG	Hawai'i	Global	Global loc.	Phylum	Class
F8-12-22_bin.26	FALSE	FALSE	0	Acidobacteriota	Terriglobia
F8-14-22_bin.11	FALSE	FALSE	0	Thermoproteota	Thermoprotei_A
F8-14-22_bin.23	FALSE	FALSE	0	Deinococcota	Deinococci
F8-14-22_bin.9	FALSE	FALSE	0	Chloroflexota	Chloroflexia
F8-8-22_bin.5	FALSE	FALSE	0	Chloroflexota	Dehalococcoidia
F8-9-22_bin.4	FALSE	FALSE	0	Chloroflexota	UBA2235
F8-12-22_bin.19	FALSE	TRUE	1	Desulfobacterota_D	UBA1144
F8-12-22_bin.22	FALSE	TRUE	1	Chloroflexota	Dehalococcoidia
F8-14-22_bin.3	FALSE	TRUE	1	Chloroflexota	Dehalococcoidia
F8-14-22_bin.8	FALSE	TRUE	2	Armatimonadota	HRBIN17
F8-11-22_bin.13	TRUE	FALSE	0	Armatimonadota	Chthonomonadetes
F8-11-22_bin.30	TRUE	FALSE	0	Chloroflexota	Chloroflexia
F8-13-22_bin.10	TRUE	FALSE	0	Chloroflexota	Dehalococcoidia
F8-9-22_bin.1	TRUE	FALSE	0	Chloroflexota	Ktedonobacteria
F8-9-22_bin.3	TRUE	FALSE	0	Deinococcota	Deinococci
F8-12-22_bin.20	TRUE	TRUE	3	Acidobacteriota	Blastocatellia
F8-12-22_bin.5	TRUE	TRUE	1	Chloroflexota	Ktedonobacteria
F8-14-22_bin.18	TRUE	TRUE	4	Chloroflexota	Dehalococcoidia
F8-8-22_bin.9	TRUE	TRUE	8	Acidobacteriota	Terriglobia

We compared these results to those obtained using sourmash gather (the same algorithm used by branchwater) with Fissure 8 MAGs against a recently published dataset of 51 metagenomes from geothermal sites in Hawai'i spanning a range of soil temperatures, 30–69°C (closest site to Fissure 8: 5 km)⁴⁴, using the same 100% cANI threshold (Table 3). We found 6 MAGs that did not have matching strains present in either Hawaiian or global datasets, 5 that were present in the Hawaiian dataset but not global, 4 that were present in the global dataset but not Hawaiian, and 4 that were present in both Hawaiian and global datasets. MAGs present in both datasets tended to have greater numbers of global locations compared to those found only in the global dataset. None of these MAGs were from *Geobacillus* or related groups known to be sporulating, globally distributed thermophiles. While we

cannot rule out the possibility of these strains having arrived in Fissure 8 from Hawaiian geothermal sites that have yet to be sampled, our results provide at least preliminary evidence of these thermophiles having recently (within 4 years) arrived due to long-range aeolian dispersal.

Exploring potential mechanisms for thermophiles to enter the atmosphere for dispersal

Finally, we examined how these cosmopolitan thermophiles might be achieving the atmospheric height necessary for long-range aeolian dispersal. Given that thermophiles inhabit hot soils surrounding volcanic areas, and diatoms are known to be dispersed long distances in volcanic eruptions⁴⁵, we hypothesized that volcanic events could provide a novel mechanism for thermophiles to enter higher atmospheric layers for dispersal, in addition to normal, wind-driven aeolian dispersal. To assess this hypothesis, we used 16S rRNA gene sequencing on tephra (associated with a volcanic event, 20 sections) and ice (5 sections) sections of a 12.5m ice core collected from Fang Glacier, Mt. Erebus, approximately 4,000m from the caldera (Fig. 4A, Table S7). Based on water isotope dating, the first meter of this core represents approximately 2.5 years of accumulation, suggesting the total 12.5m core is likely no more than 50 years old (Figure S6, Table S8).

These ice core sections were extremely low biomass, with a total of 143 ASVs (average reads per ASV: 361) (Table S9). An additional 36 ASVs were removed from the data set based either on their presence in one of the controls or taxonomic classification (Table S10). The three most abundant phyla identified were *Pseudomonadota* (35.8% of total reads), *Bacillota* (12.5%), and *Actinomycetota* (11.1%) (Fig. 4B, Figure S7). When comparing the ice and tephra portions of the core, we did not find noticeable differences between the two types of sections in the number of reads recovered or in alpha diversity (richness and Shannon diversity, Fig. 4C-D).

If volcanic events are a potential mechanism for dispersal of thermophiles, we would expect to see an enrichment of thermophiles in the tephra layers compared to the ice layers. Our findings support this hypothesis. Of the 143 ASVs recovered, only 16 ASVs (11.2% of total) were found in both tephra and ice layers. 116 taxa were found exclusively in the tephra layers (the 25 most abundant of these ASVs are shown in Fig. 5). Of these ASVs, many are associated with thermophilic or moderately thermophilic taxa – including *Vulcanibacterium*, *Rubrobacter naiadicus*, *Geobacillus thermoleovorans* (which, as noted previously, was cultured from Erebus fumarolic steam) and *Thermogemmatispora* (two MAGs from this genus were identified in the Fissure 8 samples, F8-12-22_bin.5 and F8-9-22_bin.1; Table S4). The ASV most associated with tephra layers, ASV_1, could not be identified to a genus, but is associated with the family *Hydrogenophilaceae* (phylum *Pseudomonadota*) whose described genera are all thermophiles⁴⁶. While we did find many non-thermophilic ASVs to be associated with the tephra layers, no thermophilic ASV was found to only be associated with the ice (non-tephra) layers. We also identified 10 ASVs that were exclusively found in the ice layers, all of which have previously been identified in Antarctic or high-alpine environments (e.g., *Modestobacter*, *Hymenobacter terrae*, *Asinibacterium*, etc.)^{47,48}. Interestingly,

we found that the thermophilic taxa from the tephra layers were also present in a dataset of 166 Antarctic soils from four Antarctic bioregions⁴⁷. These thermophilic ASVs were not abundant in these soil communities (average 0.86% of total reads per sample, range: 0.008–8.9% of reads) and are unlikely to be active members of the community; rather, these ASVs are likely to have been deposited due to volcanic activity either on Erebus or other Antarctic geothermal sites.

We also collected air samples from Tramway Ridge and Western Crater on Mt. Erebus, to provide corroborating evidence that thermophiles can become airborne. After removing potential contaminants, we were left with 78 ASVs, mainly from phyla such as *Actinomycetota*, *Bacillota*, *Chloroflexota*, and *Pseudomonadota* (Figure S8A-B, Table S11). These samples were also low biomass (815 and 616 reads from Tramway Ridge and Western Crater, respectively). We identified several thermophilic genera in these samples: *Thermus*, *Acidothermus*, and *Chthonomonas*, the latter of which genus has two MAGs belonging to it in our data set, one from Mt. Rittmann (Ritt_bin.13.permissive_final) and one from Fissure 8 (F8-11-22_bin.13). We also found several other taxonomic groups that contain thermophilic species (*Armatimonadaceae*, *Acidibacillus*, and *Candidatus Koribacter*). Moreover, out of the 43 unique genera in our data set, 16 were present in previously published data from various geothermally heated soils around Mt. Erebus⁹⁷, especially at higher-temperature, exposed sites (Figure S8C). These results, while preliminary, indicate that thermophilic microorganisms from geothermally heated soils on Erebus can become airborne and dispersing locally, as previous metagenomic evidence indicated³⁸.

Discussion

We present here strong evidence for the presence of globally distributed, cosmopolitan thermophilic microorganisms beyond previously known endospore-forming *Bacillota* members²⁶. Our results (1) indicate that there is nothing *a priori* that prevents non-spore-forming thermophiles from achieving global distribution via aeolian dispersal (Fig. 1, Fig. 2); (2) identify several potential routes implicated in the successful dispersal of thermophiles (greater GC content, expanded growth temperature tolerance) (Fig. 3, Figure S4); (3) provide evidence that this dispersal can happen on a regular basis (Table 3); and (4) expand the types of microbes dispersed during a volcanic event beyond diatoms to include a variety of thermophiles (Fig. 4). Below, we discuss these four primary results in turn.

The taxa we identified as cosmopolitan in distribution are highly similar to thermophiles previously identified from Tramway Ridge as potentially cosmopolitan based on 16S rRNA gene sequence²⁹; these include OTUs from the genera *Meiothermus* and the phyla CSP1-3 (*Ca. Sysuimicrobiota*, previously GAL15), *Chloroflexota*, and *Acidobacteriota*. Interestingly, we found that an archaeal MAG from Tramway Ridge previously identified as highly likely to be endemic, *Ca. Australlarchaeum erebusii*⁴⁹, has a matching species in a South American metagenome sample. This MAG had also previously shown evidence of dispersal between Antarctic geothermal sites³⁰; further study is needed to explore how divergent the South American strains are from the Erebus strains. The South American origin of this species and of several other cosmopolitan MAGs from Antarctic geothermal sites matches the well-

known South American (especially Patagonian) origin for most dust particles found in Antarctica⁵⁰. We also note that a species from one of the two genera common to both Erebus and the Fissure 8 samples, *Chthonomonas* (non-spore forming), has previously been identified as being capable of dispersal widely across Aotearoa-New Zealand⁵¹.

Our branchwater results do not rule out the possibility that geographic isolation has led to distinct local populations of thermophiles within species post-colonization (as previously found²¹), since our results only give presence information. Indeed, the strong environmental selection found at many geothermal sites could predispose successful dispersers towards diversifying quickly. Additionally, our results cannot give evidence about activity levels of present species; metatranscriptomic and/or culturing studies would be needed. We also note the concern that our results could be due to contaminants being present in metagenome sequences. However, none of the cosmopolitan thermophiles we identified are known contaminants, which tend to be from *Pseudomonadota*, *Bacillota*, and *Actinomycetota*^{52,53}. Additionally, negative controls run in parallel to the Fissure 8 samples did not return any of the same species identified in those metagenomes⁵⁴.

Previous studies have indicated that strong selective forces act on microorganisms during atmospheric dispersal, especially UV^{6,12}. Our survival tests (Fig. 2) indicate this is also the case for thermophiles. Thus, traits associated with surviving UV exposure would be the most likely to assist with dispersal. For our UV experiments, we used a maximum UV-C dose equivalent to 5 days at 20 km altitude, although in our experiments, the cultures were in liquid culture, whereas in the atmosphere they would most likely be frozen as single cells. A previous study found that *Deinococcus radiodurans* had significantly greater survival rates during gamma irradiation when the cells were frozen for the irradiation⁵⁵; thus, it is likely that our UV experiments are under-estimating the survival rate of our test species. Microbes dispersing via the atmosphere would not necessarily have to reach an altitude of 20 km for long-range travel and thus would have lower UV exposure, but 5 days could be sufficient time for a particle to travel from Patagonia to West Antarctica⁵⁶.

Our results identified several traits that appear to be useful for increasing the global dispersal of thermophiles: sporulation (or other dormancy-type resistant state), increased GC content, and/or increased temperature growth range. The first two traits are most likely useful for reducing UV damage, while the third trait is a colonization trait. Spores are known to be more resistant to harsh conditions, including UV exposure⁵⁷, due to their protective layers. Unfortunately, given that there is no common genetic basis for sporulation across phyla, we were unable to include this strategy in our MAG trait analysis. However, in addition to our *Geobacillus* spore-enriched survival tests, we did observe that the genus *Thermogemmatispora*, which is known to be sporulating^{58,59}, was identified both in the Fissure 8 samples (F8-12-22_bin.5) as well as in the Fang glacier ice core (Table 3, Fig. 5). With regards to increased GC content, this trait seems to be a common trait across cosmopolitan thermophiles and mesophiles (Fig. 3D, Figure S4E). Greater GC content has previously been postulated to be correlated with increased exposure to sunlight, with the reasoning being that pyrimidine dimers (formed from

adjacent thymidines or TC bases) are one of the major products of ultraviolet damage; thus, decreasing the amount of thymidines would decrease the potential for these harmful dimers^{33,60}. However, GC content is a complex trait, and it is likely that UV alone is not sufficient to explain the differential GC contents in these organisms. Other potential traits that could be involved in UV resistance such as pigment formation or adhesion to particles also do not have well-defined genetic bases that would allow for their inclusion in our analysis. Adhesion to particles could be helpful for increasing the availability of moisture to the cells and thus is likely to be an important trait¹¹. DNA repair mechanisms are also likely to be important for repairing lesions from UV damage; the non-significant results we observed could be due to a lack of phylogenetic coherence in this trait.

Recently, it has been proposed that trace gas metabolism may aid in dispersal of microorganisms, either through active trace gas metabolism during dispersal to aid in maintenance costs^{39,61} or to assist with colonisation upon arrival in a new habitat³⁹. Our results provide preliminary evidence in support of the hypothesis that trace gas metabolism may be an important dispersal trait for thermophiles (but not mesophiles), with cosmopolitan thermophiles enriched especially in Group 1h hydrogenases (high affinity), in addition to CO dehydrogenases (Fig. 3D). However, we cannot distinguish between the two proposed uses of trace gas metabolism (during transport or to support colonisation upon arrival).

One of the commonly given reasons for why thermophiles are dispersal limited is the infrequency of geothermal habitats, making it less likely for dispersed thermophiles to find a high-temperature habitat²⁶. Our traits-based results indicated that cosmopolitan thermophiles are more likely to be able to grow even under mesophilic conditions than non-cosmopolitan thermophiles. Our culturing results also help support this. *Calidithermus* sp. Yb, a widely distributed thermophile found in a wide variety of habitats beyond hydrothermal, was capable of significant growth even at 35°C (Table S6), a temperature likely to be experienced in many non-polar mesophilic soils. *Thermus* sp. SS11024-3, on the other hand, was almost exclusively found in hydrothermal habitats globally, and was incapable of growth at temperatures below 45°C. Another potential answer to this conundrum lies in several previously overlooked non-geothermal habitats globally, identified in our branchwater results, that experience thermophilic temperatures (Table S5). For instance, several cosmopolitan MAGs were found to be present in compost, landfill, and/or mine fire and peat fire impacted soils, metagenomes, all of which can reach thermophilic temperatures^{62,63}.

Our results from Fissure 8 provide preliminary evidence that some thermophilic species are capable of frequently dispersing between distant sites via aeolian dispersal. It is possible that the cosmopolitan MAGs in the Fissure 8 samples that have matching strains globally but not in Hawai'i may have come from Hawaiian sites that have yet to be sampled. Still, the presence of identical strains elsewhere, from samples collected prior to the eruptive event (all global matching metagenomes were sampled prior to 2021; Table S5), indicates either recent arrival of these organisms or regular dispersal between Hawaiian sites and globally dispersed sites. The fact that these identical strains were found in metagenomes almost exclusively from the Northern Hemisphere (Figure S5) is in line with expectations about the origin

of species, given that the hemispheres are relatively isolated from each other. Further studies are needed on the evolutionary history of these MAGs at Fissure 8 and identical strains globally to provide further evidence of their dispersal history.

The Fang Glacier ice core data provide strong evidence that thermophiles are dispersed during volcanic events at least locally (~ 4km from the Main Crater). Given the < 50-year age of this core, the tephra layers present can conclusively be attributed to eruptions from Mt. Erebus and likely correspond to ~ 1-year-long periods of higher eruptivity from the volcano ⁶⁴. Volcanic activity beyond eruptions could also be a vector for thermophilic dispersal, such as the continual Erebus plume which can shift directions depending on wind direction (see Figure S9). Further sampling of Mt. Erebus hot soils, especially within the Main Crater itself, would help identify whether any of the thermophiles identified in either of these sampling sets are growing in an area where they would easily be picked up by volcanic activity.

Interestingly, the ASV most associated with the tephra layers came from the same family (*Hydrogenophilaceae*) as DNA signatures from an ice core collected from Lake Vostok ⁶⁵, Antarctica, ~ 1400 km from Mt. Erebus. The presence of this thermophile in Lake Vostok was attributed to subglacial hydrothermal activity below Lake Vostok, but given our results, an alternate explanation could be atmospheric deposition from volcanic activity. Our results are also supported by a previous study that found that the air above the McMurdo Dry Valleys, 100 km from Mt. Erebus, contained many presumed thermophiles, attributed to Mt. Erebus and other continental Antarctic geothermal sites ⁶⁶. Our results cannot indicate how far these thermophiles may be dispersed during a volcanic event, but volcanic eruptions tend to eject particulate matter 5–20 km into the atmosphere, which could easily lead to long-range aeolian dispersal. Our findings also give no indication of whether these thermophiles are still viable after the eruption.

Overall, our results provide evidence to reject the common idea that most thermophiles are dispersal limited and unable to be globally distributed. We argue that past studies on the dispersal capabilities of thermophiles have been incapable of fully answering this question because of a narrowness in taxonomic or environmental scope. On taxonomy scope, many studies have been limited to a few taxonomic groups, such as thermophilic cyanobacteria ¹⁵, archaea ²¹, or fermentative bacterial H₂ producers ²⁰. Our analyses examined dispersal capabilities of thermophiles without filtering for specific taxonomy or metabolic capability. Additionally, all studies we could find on this topic focused on hot spring environments. Our analyses expand on this by exploring thermophiles from geothermally heated soils and examining potential dispersal agnostic of the habitat type successfully colonized. Although our results cannot address whether hot soil or hot spring thermophiles are more widely dispersed, it is reasonable to suggest that thermophiles from hot soils would be more likely to disperse due to attachment to soil particles ¹¹.

Further research is needed to continue to expand on several of our findings. For instance, studies are needed to examine whether thermophiles can survive volcanic events and thus whether volcanic events are truly a vector for thermophiles to disperse globally. Studies to determine the viability and activity of

thermophiles in mesophilic samples are also needed. The sites used in this study did not have particularly strong environmental selective forces such as extremes in pH or extremely high temperatures; further studies are needed to examine whether microbes from other extreme types of environments (acidophiles, halophiles, etc.) are globally dispersed. The use of tools such as branchwater that allow for the rapid searching of the SRA will make these types of studies more feasible in the future.

Methods

Sample locations

Data used in this research has been compiled from several complementary, but independent, projects. These include geothermally heated soils from Antarctic volcanic features: Tramway Ridge and Western Crater on Mt. Erebus; Cryptogram Ridge on Mt. Melbourne, and Mt. Rittmann (these sites are described in full in ³⁰); biofilms near steam vents at a site (Fissure 8), sterilized by a lava flow 4 years prior to sampling, from the Big Island, Hawai'i (described in ⁵⁴); and an ice drill core collected from the cirque of Fang Glacier on the northern flank of Mt. Erebus. Physicochemical parameters, sampling year, latitude and longitude, and associated metadata for these features are presented in Table 1.

Branchwater analysis

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^{67,68}; this tool is a k-mer based approach for searching metagenomes (SRA was accessed 09-04-2026). The methods for generating these MAG sequences are presented in ³⁰. Putative hits of MAGs in metagenomes by branchwater are defined by a given containment level and containment-based ANI (cANI). Containment generally refers to the amount of genetic material from a MAG present in a metagenome, while cANI refers to the percentage similarity of that material from the hit to the MAG. A containment of 0.3 (30%) using a k-mer size of 21 (branchwater default) is generally regarded as sufficient genetic material to determine whether a species is present ^{68,69}. However, cANI is not necessarily equivalent to ANI, and the relationship between cANI and ANI can vary by genome, although a cANI threshold of 0.97 is generally given as indicative of a species-level match (see the branchwater webserver, <https://branchwater.sourmash.bio>). Past studies using branchwater have used a wide range of containment and cANI thresholds, but they were less concerned about the possibility of false positives ^{68,70}.

Thus, we chose to first filter branchwater hits to those with ≥ 0.3 containment. Then, for each MAG, we sought to identify what cANI level was indicative of a matching species (using a stricter ANI of 96% to reduce false positives) using a separate program, YACHT v.1.3.1 ⁷¹, which uses hypothesis testing and containment to identify the presence of a genome in a metagenome at a given ANI threshold. In YACHT, we used a k-mer size of 21 to match branchwater and minimum coverage of 1.0 to keep false positives low. For each MAG, based on the branchwater results, we selected the metagenome with the lowest

containment value ≥ 0.3 and downloaded it from the SRA using the `fasterq-dump` command from the SRA Toolkit v. 3.0.2. We then ran YACHT to identify whether that MAG was present ($\text{ANI} \geq 96\%$) in that metagenome. If present, we used the corresponding cANI value for that branchwater hit as the cANI threshold indicative of a matching species. If not present, we moved to the next highest cANI value in the branchwater results and repeated the downloading of the metagenome with the lowest containment and running YACHT. This was repeated until either the MAG was identified as present in a metagenome, or a cANI value of 0.97 was reached. We used 0.97 as the default threshold as this is generally used as a conservative threshold for presence of a matching species in a metagenome at $k = 21$ ⁶⁹. This generated MAG-specific, dataset-specific cANI thresholds for presence of a near-identical species.

As in⁷⁰, metagenomes from synthetic data based on the “organism” classification in NCBI were removed. Metagenomes were classified as air, brackish, freshwater, host-associated, hydrothermal, ice, marine, plant, soil, or wetland based on the “organism” classification in NCBI, with classification refined, if necessary, using additional metadata available in NCBI or removed if no classification was possible. Metagenome hits within 10 m were consolidated into a single “location” prior to mapping, with spatial operations performed using the `sf` package v. 1.1-0 in R⁷². Correlation between proportion of hydrothermal habitats per MAG and number of locations globally was calculated using the nonlinear least squares (`nls`) function in the base R Stats Package, v. 4.5.2.

A phylogenomic tree for all Antarctic geothermal MAGs was produced with GToTree v. 1.8.16⁷³, using the prepackaged single-copy gene-set for Bacteria and Archaea (25 target genes). Briefly, Prodigal v. 2.6.3⁷⁴ was used to predict genes on input genomes provided as fasta files. Target genes were identified with HMMER3 v. 3.4⁷⁵, individually aligned with Muscle v. 5.1⁷⁶, trimmed with trimAl v. v1.5.rev1⁷⁷, and concatenated prior to phylogenetic estimation with FastTree2 v. 2.1.11⁷⁸. Four MAGs were removed from the tree due to having too few hits to the targeted single copy genes: Erebus-WC_bin.38.permissive_final, Erebus-WC_bin.59.orig_final, Mel_bin.5.orig_final, and Ritt_bin.10.strict_final. The resulting tree was visualized and annotations added using iTOL⁷⁹.

Isolation of Erebus bacteria and survival experiments

In this work, we isolated three thermophilic taxa: strains from the genera *Geobacillus*, *Thermus*, and *Caldithermus* strain. The two *Thermaceae* strains were isolated from geothermally heated soils from Tramway Ridge on Mt Erebus, collected in either November 2023 or 2024 from 10 cm depth. The top 10 cm were removed using a sterile scoop and soil was collected into a sterile 50 mL Falcon tube using a sterile spatula. The soil temperature was checked prior to sampling using a Checktemp 1C electronic temperature sensor (Hanna Instruments, Rhode Island, United States) and was found to be $\sim 65^\circ\text{C}$. The soil was kept warm until return to New Zealand for culturing. For *Thermus* isolation, soil crumbs were plated onto 1% v/v R2A plates, solidified using 3% w/v Gelrite (Duchefa Biochemie, The Netherlands) using 1 g/L CaCl_2 hexahydrate (Sigma-Aldrich) as the cation. Plates were incubated in a sealed container at 65°C and colonies picked and re-streaked until pure. The *Caldithermus* strain was isolated in a similar manner except using a 10% v/v R2A plates (pH 8.0) and incubations were undertaken at 55°C .

For *Geobacillus*, sterile glass wool incubated for one year (November 2023 – November 2024) in a borehole (temperature in borehole during the incubation: 62°C) in Tramway Ridge soils with fumarolic steam as the sole inoculant was diluted and used to inoculate enrichment cultures in pH 7 FS1 media⁸⁰ with a headspace of (v/v) 80% nitrogen, 10% CO₂, 10% H₂, and 0.06% CO, incubated at 65°C. One of the cultures with growth was plated on 50% R2A solidified using 1.5% w/v Gelrite with 1 g/L MgCl₂ hexahydrate (Sigma-Aldrich) as the cation under aerobic conditions to check for the presence of aerobic microorganisms, resulting in the isolation of a *Geobacillus* strain.

DNA was extracted from *Thermus* or *Geobacillus* cell pellets using a CTAB extraction method, as previously described²⁹, whereas DNA was extracted from the *Caldithermus* strain using a NucleoSpin Microbial DNA kit as per the manufacturer's instructions. The resultant DNA was then cleaned using the Zymo Clean and Concentrate 25 kit. Genomic DNA was quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Massachusetts, United States). For whole genome sequencing for *Thermus* and *Geobacillus*, genomic DNA was sequenced on an Illumina HiSeq sequencer using 150bp PE sequencing. Next generation sequencing library preparations were constructed following the manufacturer's protocol (NEBNext® Ultra™ DNA Library Prep Kit for Illumina®). Image analysis and base calling were conducted by the HiSeq Control Software (HCS) + RTA 2.7 (Illumina) on the HiSeq instrument. For the *Caldithermus* strain, the genome was sequenced using the Illumina NovaSeq X Plus platform. Libraries were prepared using the tagmentation-based and PCR-based Illumina DNA Prep kit and custom IDT 10 bp unique dual indices (UDI) and a target insert size of 280 bp. The libraries were quality-checked by Agilent Tape station (Agilent, Santa Clara, CA). For all genomes, the raw reads were quality trimmed and sequencing adapters removed using trimmomatic (v. 0.39)⁸¹ (parameters: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:2:True MINLEN:70 SLIDINGWINDOW:4:15) and assembled using SPAdes v. 4.0.0⁸² with default parameters, with quality checking using QUAST v. 5.2.0⁸³. To match isolate genome sequences to MAGs, we used the EZBioCloud ANI calculator⁸⁴.

To determine the ability of these isolates to survive conditions similar to those they might encounter during aeolian dispersal, we conducted two types of experiments. To assess the impact of UV exposure on survival, we loosely followed the methods presented in^{14,15}. First, we centrifuged cultures in the late exponential growth phase (determined via OD readings) for 15 minutes at 2500 × *g* and resuspended the pellet in either 30 mM TRIS, pH 8 (for *Thermaceae* species) or 20 mM phosphate buffer, pH 6.5 (for *Geobacillus*). This cell slurry was split into triplicate 8 mL volumes in sterile 55 × 15 mL polystyrene petri dishes. These were placed onto stir plates with a sterilized stir bar and the plate turned on at a low rate inside a biosafety cabinet. The UV light in the biosafety cabinet was then turned on (~ 40 μW/cm² output) and subsamples taken at time points of 0, 2, 5, 8, 12, and 18 minutes. CFUs present in these subsamples were determined via plate counts on 50% R2A (either pH 8 or 6.5 for *Thermaceae* or *Geobacillus* species, respectively) and survival calculated as a percentage of the CFUs in the time 0 timepoint.

For freezing/starvation/desiccation experiments, we loosely followed methods presented in ^{85,86}. Late exponential cultures were centrifuged and the cell pellet was washed twice as before, with re-suspension in TRIS or phosphate buffer at the appropriate pH. 100 μ L of this cell slurry was pipetted onto the lid of sterile 55 $\times$ 15 mL polystyrene petri dishes, with triplicates per time point. This slurry was vigorously spread onto the lid and allowed to dry at room temperature in a laminar flow hood; once completely dried, closed dishes were placed into a sealed box with 100g of anhydrous magnesium chloride (Sigma-Aldrich) and placed at -20°C (actual temperature during the experiment varied between -16 and -26°C). At various time points, triplicate dishes were removed from the freezer and 200 μ L of 30 mM TRIS (pH 8 or 6 depending on the species) was pipetted onto the lids. This liquid was vigorously spread around the lid to resuspend as much biomass as possible, then diluted and CFUs determined via plate counts (as above), with survival calculated as a percentage of CFUs present in the time 0 timepoint (which was re-suspended after drying but prior to freezing). For *Geobacillus*, we tested whether there would be increased tolerance to dispersal conditions when increased spores were present. Thus, we conducted the same experiments on overnight *Geobacillus* cultures and on those left at room temperature for 7 days (which is known to increase spore numbers in *Geobacillus* ⁸⁷).

Growth temperature tolerance of the two *Thermaceae* species was determined by monitoring the optical density (OD₆₀₅) of triplicate cultures in TYE Castenholz media pH 8 ^{88,89}, inoculated with 20% v/v of culture grown until the late/mid-exponential stage of growth, at various temperatures (from 25–65 or 70°C) over the course of 7 days. OD readings of each culture were taken regularly by removing 100 μ L into a 96-well polystyrene flat-bottom plate (Biofil, Guangzhou, China) and measuring the OD₆₀₅ on an Absorbance 96 microplate reader (Byonoy, Hamburg, Germany).

Fissure 8, Hawai'i, MAGs and metagenomes

MAGs from Fissure 8 samples with soil temperatures > 50°C were used from a previously published dataset (BioProject PRJNA1459608) ⁵⁴. These were filtered to only include predicted thermophilic MAGs based on predicted optimal growth temperature from GenomeSPOT v. 1.0.1 ⁷¹ (19 MAGs, MAG BioSamples SAMN57599648 through SAMN57599666). To identify the presence of these MAGs in recently published metagenomes from other geothermal sites on the Hawaiian Archipelago ⁴⁴, we downloaded all metagenomes associated with the BioProject PRJEB52128 and used the sourmash gather function (using k-mer size of 21 to match branchwater) to search for Fissure 8 MAGs in these metagenomes ⁹⁰.

MAG characterization

To investigate genomic traits or predicted phenotypic traits that might enable aeolian dispersal of cosmopolitan MAGs, we used several bioinformatics programs. Basic genomic properties (coding density, GC content, genome size, and MAG completeness and contamination) were obtained using checkM2 v. 1.01 ⁶⁸. Optimal growth temperature was estimated using Tome v. 1.0.0 ⁶⁹ and maximal growth rate via gRodon2 v. 2.7.2 ⁷⁰. Temperature, pH, and salinity tolerance were predicted using

GenomeSPOT; we did not include analysis of aerobic/anaerobic predictions from this program given its known inadequacies for thermophiles. CAZymes were predicted from MAG protein sequences using run_dbcan v. 4.1.4 ⁷². Protein sequences from MAGs were classified into COG functional categories using COGclassifier v. 1.0.5 ⁷³. To identify potential UV resistance genes, we searched for predicted genes with matching COG IDs as those found to be important in *Deinococcota* for UV and radiation resistance ⁷⁴ (Table S1). To identify potential dormancy-related genes, we searched for predicted genes with the following COG IDs, based on ⁷⁵: COG4575 (ElaB / YqjD / YgaM), COG3448 (Dri), COG3130 (Rmf), and COG1544 (RaiA / HPF). Note: we could not identify COG categories for other dormancy-related genes such as Balon and Sra. All MAG genome sizes were corrected for MAG completeness.

We estimated the metabolic potential of each MAG using a method described by Greening and colleagues ⁷⁶. In short, we used DIAMOND v. 2.1.6 (query cover > 80%) to search quality-filtered assembled ORFs of the MAGs against custom protein databases (2022 version for hydrogenases, accessible at <https://github.com/GreeningLab/HydDB>; v.3 for all other marker genes, ⁷⁷) of representative metabolic marker genes designed by the authors. The metabolic marker genes we focused on were those likely to be related to maintenance metabolism during aeolian dispersal: rhodopsin (RHO), phototrophy (PsaA, PsbA, energy-converting microbial rhodopsin), methane cycling (McrA, MmoA, PmoA), hydrogen cycling (using HydDB ⁷⁸), carbon monoxide oxidation (CoxL, CoxS), and carbon fixation (RbcL, AcsB, AclB, Mcr, HbsT, HbsC). We filtered results based on different identity thresholds, as outlined in the respective papers ^{76,78,79}.

Fang Glacier ice core analysis

Two ice cores (15 m and 12.5 m depth) were drilled on the Fang Glacier cirque for the purposes of seismic studies on an unrelated project in November 2007. The ice cores were retrieved and immediately cut into 1 m lengths, wrapped in polypropylene layflat and packed in polyurethane ice core storage boxes. These were maintained at -40°C while being transported to the National Ice Core Facility in Wellington, Earth Sciences New Zealand. All ice cores were maintained at -36°C until use. Only the 12.5m core was used for this project, as the other was visibly much more dirty and not suitable for DNA extractions.

All tools (i.e. knives, cutting boards) and work surfaces were prepared using strict contamination-control procedures. Briefly, all cutting tools and handling equipment were cleaned over three days using extended Milli-Q water soaking, repeated rinsing, and isopropanol sterilization, and were dried to prevent contact contamination before use. Prior to commencing core examination and cutting, the ice core facilities air system was turned off, the cold laboratory floor was vacuumed to remove particulate matter, benches were wiped with clean paper towels to remove residual dust, and all work surfaces were disinfected with 100% isopropyl alcohol (IPA). After each use, all materials were re-sterilized by rinsing with milliQ water and then disinfected through soaking in IPA (for knives and small tools) or through wiping with IPA 3–4 times. All other material, including Whirl-Pak bags (Filtration Group, Pleasant Prairie, WI, USA), 0.2 µm Sterivex filters (Merck Millipore, Darmstadt, Germany), and 60 mL plastic syringes

(Thermo Fisher Scientific, Auckland, NZ) were purchased new and sterilized. The packing of these materials was disinfected with IPA prior to being introduced into the ice core facility sampling environment.

Ice core processing was performed at the Earth Science New Zealand National Ice Core Research facility in Lower Hutt, NZ. All core cutting took place in the -20°C cold laboratory which had been decontaminated as described previously. All personnel involved in the processing wore insulated layers underneath cleanroom clothing (including one-piece non-shedding coveralls, boots and shoe covers, nitrile gloves, laboratory goggles, and masks). Gloves were replaced anytime a different core section was handled. One meter core sections were removed from the crates in their storage bag and were placed onto a light box. Sections of core that contained tephra were identified visually. After identification, samples were removed from their bags and sectioned following the cut diagram in Figure S1 using a vertical bandsaw for lengthwise cuts down the core and a circular saw for horizontal cuts across the core diameter to isolate the tephra layers. Horizontal cuts were performed at least 2 cm away from each side of the tephra layer to allow the outside material to be removed. Approximately 1/3 of each core section was removed for microbial analysis, the remaining core sections were re-archived. After cutting, each removed section of the core was placed into sealed sterilized glass containers.

To gain a first approximation of the age of this ice core, we used the top section of the core (1 meter length, re-archived core section) cut into $\sim 2\text{cm}$ sections and melted for water isotopes. Water samples were analysed on a Los Gatos Liquid Water Isotope Analyser by high resolution laser absorption at Earth Sciences New Zealand, Wellington, NZ. All results are reported with respect to VSMOW2, normalized to our internal standards: CM2 with reported values of -19.10‰ for $\delta^{18}\text{O}$, -134.9‰ for $\delta^2\text{H}$, and USGS 49 with reported values of -50.55‰ for $\delta^{18}\text{O}$, -394.7‰ for $\delta^2\text{H}$. The analytical precision for this instrument is 0.2‰ for $\delta^{18}\text{O}$ and 1.5‰ for $\delta^2\text{H}$.

For the sections used for microbial analysis, to remove potential contaminants, we used sterilized knives to scrape away the outside 0.5–1 cm of ice from each cut core section. Once removed, the samples were placed in new sealed sterilized glass containers and were brought into the ice core facilities room temperature clean room environment and were left to thaw in a decontaminated biosafety cabinet. After the sections had melted, samples were agitated to homogenize and filtered through a $0.2\text{ }\mu\text{m}$ Sterivex filter. Filters were placed in sterile Whirl-Pak bags and frozen at -20°C .

DNA extractions were performed in a decontaminated biological safety cabinet at the University of Canterbury following the suggestions on laboratory practices described in ⁵³ for low biomass samples. The Sterivex filter chambers were broken open with flame sterilized pliers, and the filters were removed from the cartridges with sterilized surgical blades and flame sterilized forceps. DNA was extracted from the filters using the Qiagen DNeasy Powersoil Pro Kit (Qiagen, Germantown, MD, USA) following the manufacturer's recommended protocol until the MB spin column loading (step 8, Kit Handbook – "Protocol: Experienced User"). More specifically, each filter was split in half, and each half was extracted in two replicate reactions. Lysate from both reactions were then combined into a single MB spin column.

After this, the manufacture's recommended protocol was once again followed until elution where DNA was eluted in 30 µL solution C6. A total of five extraction blanks were included to control for possible contamination introduced during DNA extraction and downstream processing and were also extracted with these modifications (including the combined lysate and 30 µL elution volume).

The extracted DNA from the 25 samples, the five extraction blanks, and three no-template controls were used for replicate PCR amplifications to assess the prokaryotic (bacterial/archaeal) communities following methods described in ^{81,82}. We used PCR primers targeting the hypervariable V4 region of the 16S rRNA gene (515F: 5'- GTGCCAGCMGCCGCGGTAA-3' and 806-R: 5'-GGACTACHVGGGTWTCTAAT-3') ⁸³ which included Illumina adapters and unique 12-bp barcode sequences ⁸⁴. Duplicate PCRs were performed in 25 µL reaction volumes with 5 µL template with PlatinumTM II Hot-Start PCR Master Mix (2X) (Invitrogen, Carlsbad, CA, USA). Cycling parameters for the 515F/806R primer set consisted of an initial denaturation step at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C (15 s), annealing at 60°C (15 s), extension at 68°C (60 s), and a final extension step at 72°C for 10 min. After confirming amplification via gel electrophoresis, duplicated amplified products were combined, cleaned, and normalized to equimolar concentrations using Sera-Mag SpeedBeads and were sequenced on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) using the V2 2 × 250 bp paired-end Illumina sequencing kit at the University of Colorado Boulder. The sequencing data was then demultiplexed using idemp (<https://github.com/yhwu/idemp>). Raw demultiplexed sequence data can be accessed in the NCBI Sequence Read Archive, BioProject ID PRJNA1460789.

The 16S rRNA gene sequences were processed with the DADA2 pipeline v 1.26⁹¹ paired with Cutadapt v 2.1⁹² to remove sequencing primers. Sequences were trimmed (truncLen = c(240,160)) and quality filtered with standard filtering parameters (maxEE = c(2,2), truncQ = 2, maxN = 0, rm.phix=TRUE), paired-end reads were merged with default settings into contigs, and clustered into amplicon sequence variants (ASVs) with samples unpooled at the inference step to preserve rare taxa. Taxonomic assignments were generated using the DADA2 package, which implements a naive Bayesian classifier⁹³ via the assignTaxonomy function, trained against the SILVA reference database (v138.2) ⁹⁴. A minimum bootstrapping threshold required to return a taxonomic classification of 50% was used for analysis. More details on the parameters used can be found at https://github.com/fiererlab/dada2_fiererlab.

Downstream community analyses were performed in R v.4.3.3 ⁹⁵. Prior to downstream analysis, ASVs associated with chloroplasts (10 ASVs) and mitochondria (11 ASVs) were removed. No ASVs were associated with eukaryotes or were unassigned at the phylum level. Filtering of potential contaminants was performed following recommendations described in ⁵³, through taxonomic comparison with well-known laboratory contaminants ^{96,97}, and by comparison against the extraction blanks and no template controls with the prevalence method of Decontam v1.32.0 (Bioconductor) ⁹⁸.

ASV-based analyses on the filtered ASV table were performed using the R packages mctoolsr (<https://github.com/leffj/mctoolsr/>), vegan ⁹⁹, and phyloseq ¹⁰⁰. Richness was calculated using the

function “specnumber” within vegan. Indicator taxa analysis (MultiPATT) was performed using the “indicspecies” package with the “r.g” association function and a significance level of 0.05¹⁰¹. Additional statistical tests were performed using the base R functions “wilcox.test” and the packages “rstatix” (<https://github.com/kassambara/rstatix>). Plotting was performed using the R packages ggplot2 and cowplot.

To shed light on the potential spread of volcanic-associated thermophiles to other Antarctic environments, we reprocessed the dataset described in⁴⁷ with the DADA2 pipeline following the methods and using the reference datasets described previously. This dataset includes 200 soils from across Antarctica downloaded from NCBI (BioProject accession: PRJNA1214840). For the 16S rRNA gene sequences, ASVs associated with chloroplasts, mitochondria, eukaryotes, and those unassigned at the phylum level (1064 ASVs) were removed. Following filtering we used a cutoff of 2000 reads per sample as a threshold for inclusion in our analysis. This left 166 samples that had sufficient prokaryotic 16S rRNA gene reads for downstream analyses

Erebus air sampling

Preliminary air data were collected to investigate the presence of thermophilic microorganisms in the air around geothermal sites on Mt. Erebus, Antarctica. One sample was collected from each of Tramway Ridge and Western Crater, as well as a control sample on the sea ice in front of Scott Base in November 2023. We used a SASS3100 air sampler (Research International, USA) equipped with a biological removable filter (P/N26 7100-134-242-01, Research International, Inc.) cleaned with 6% v/v hydrogen peroxide immediately prior to the filter being installed. The sampler was attached to a tripod (1.5 m high) and turned on at a rate of 300 L/min for upward of 4 hours and was positioned at the edge of the geothermally heated area (downwind). The total active sampling time at Western Crater was unknown due to battery failure.

Aerosol filters were collected onsite in sterile plastic packaging and transported frozen until processed. These were processed in a Class II biosafety cabinet that was decontaminated prior to use with hydrogen peroxide to minimise contamination, with all materials similarly treated before introduction into the cabinet. DNA was extracted using the DNeasy PowerSoil Pro Kit (Qiagen; Qiagen Inc., Germantown, MD, USA) with the following modifications. Filters were placed into sterile 15 mL Falcon tubes, and 1600 µL of CD1 lysis buffer was added along with 10 µL of 20% sodium dodecyl sulfate (SDS). Tubes were gently tilted to ensure complete saturation of the filter and incubated for 5 min at room temperature. Samples were then subjected to initial mechanical disruption using a FastPrep instrument (MP Biomedicals, USA) at 4.0 m s⁻¹ for 20 s using a 15 mL tube adapter. Following this, filters were aseptically transferred using sterile tweezers into 10 mL sterile syringes, and the lysate was expressed into PowerBead tubes supplied with the kit. Samples were then processed according to the manufacturer’s protocol from step 2 onwards, including bead beating at 5.0 m s⁻¹ for 45 s. Where necessary, split lysates were recombined onto a single silica spin column prior to elution in 50 µL of

nuclease-free water. Sample controls were included by opening sterile filters inside the biosafety cabinet and processing them identically to experimental samples.

Airborne microbial DNA samples were subjected to 16S rRNA gene amplicon sequencing without prior dilution due to low biomass. Bacterial 16S rRNA genes were amplified using Illumina-adapted primers targeting the V1–V3 region, Illumina_Ba9F (5'-GAGTTTGATCMTGGCTCAG-3') and Illumina_Ba515Rmod1 (5'-CCGCGGCKGCTGGCAC-3'). PCR reactions were performed in 20 µL volumes containing 10 µL of 2× Taq PCR Master Mix (Qiagen), 4 µL of each primer (2 µM), 5 µL of DNA template, and 1 µL nuclease-free water. Each sample was amplified in duplicate to ensure sufficient DNA yield. PCR plates were sealed with adhesive PCR seals (Thermo Fisher Scientific), briefly centrifuged, and amplified under the following thermocycling conditions: initial denaturation at 95°C for 3 min, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 52°C for 30 s, and extension at 72°C for 45 s, with a final extension at 72°C for 5 min and a hold at 4°C. Amplicons were quantified using the Quant-iT™ dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions, and fluorescence was measured using a Varioskan microplate reader with excitation at 485 nm and emission at 530 nm. PCR products were pooled and concentrated at 35°C using a centrifugal evaporator until approximately 100 µL remained. Concentrated pools were quantified using the Qubit dsDNA High Sensitivity Assay (Thermo Fisher Scientific) following a 1:10 dilution. Fragment size selection was performed using NucleoMag NGS Clean-up and Size Selection beads (Macherey-Nagel) at a 0.5× bead-to-sample ratio, and the resulting supernatant was purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol, with DNA eluted in 50 µL nuclease-free water. Final libraries were quantified in triplicate using the Qubit HS assay, normalised to a minimum concentration of ~ 10 ng µL⁻¹ in a final volume of 100 µL, and sequenced on an Illumina MiSeq platform at Massey University (Palmerston North, New Zealand).

Amplicon sequence variants (ASVs) were inferred from the DNA sequences and taxonomy assigned as above for the ice core data. Read tracking through the pipeline is presented in Table S2. Potential contaminating ASVs were removed using the decontam R package, v. 1.26.0, with the “prevalence” method, threshold of 0.7⁹². In the remaining set of ASVs, we identified several that were unique to the Scott Base samples; these were removed as well (567 ASVs removed as contaminants overall). Low abundance (1 read across the two samples; 41 ASVs) and erroneous assignment ASVs (assigned to *Eukaryota*, chloroplast, or mitochondria; 5 ASVs) were also removed, resulting in a total of 78 ASVs. ASV abundance barplots and heatmaps were plotted using phyloseq (v. 1.50.0) and ggplot2. Abundances of matching taxa from Erebus soils were plotted using ASV data from⁹⁷.

Declarations

Acknowledgements

This work was supported by the Royal Society of New Zealand (Marsden Grant 18-UOW-028). The authors wish to acknowledge the contribution of REANNZ to the results of this research. New Zealand's

national compute and analytics services and team are supported by the Research Education Advanced Network New Zealand (REANNZ) and funded jointly by REANNZ's collaborator institutions and through the Ministry of Business, Innovation and Employment. URL: www.reannz.co.nz. The authors also wish to acknowledge the logistical and fieldwork support of Antarctica NZ through the event K023. The authors especially acknowledge the expert assistance of field trainers Jon Tyler and Pepper Cook. The ice cores from Fang Glacier on Mount Erebus were drilled to provide shot holes for a passive seismic calibration experiment for the Mount Erebus Volcano Observatory under the direction of Philip Kyle as part of the United States Antarctic Program. The authors also wish to thank Jay Kyne and Terry Gacke and members of the G081 team for their support in drilling the Fang Glacier cores.

Data Availability

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>.

Author Contributions

S.E.N., S.C.C., C.W.H., and I.R.M. conceived the study. S.E.N., N.B.D., and L.T.R. performed bioinformatic analyses. S.E.N., M.C., C.G., and M.B.S. isolated the bacteria and performed laboratory experiments. N.B.D., R.P., N.B., and M.B.S. conceived and executed the ice core cutting and processing. S.E.N., M.B.S., S.C.C., I.R.M., and P.K. carried out fieldwork on Mt. Erebus at various times. S.D.J.A., C.W.H., and J.S. provided assistance with formal analysis. J.S. also assisted with data curation. M.B.S., S.C.C., I.R.M., and N.B. provided supervision, resources, and project administration. S.E.N. and N.B.D. wrote the initial manuscript with input from all authors. All authors read and approved the final manuscript.

Ethics declaration

Ethics declaration: not applicable.

Competing Interests

The authors declare that they have no competing interests.

References

1. van der Gast CJ (2015) Microbial biogeography: the end of the ubiquitous dispersal hypothesis? *Environ Microbiol* 17:544–546
2. Whitfield J (2005) Is Everything Everywhere? *Science* (1979). 310, 960–961
3. Martiny JBH et al (2006) Microbial biogeography: Putting microorganisms on the map. *Nat Rev Microbiol* 4:102–112
4. Burrows SM et al (2009) Bacteria in the global atmosphere – Part 2: Modeling of emissions and transport between different ecosystems. *Atmos Chem Phys* 9:9281–9297
5. Wilkinson DM, Koumoutsaris S, Mitchell EAD, Bey I (2012) Modelling the effect of size on the aerial dispersal of microorganisms. *J Biogeogr* 39:89–97
6. Archer SDJ et al (2019) Airborne microbial transport limitation to isolated Antarctic soil habitats. *Nat Microbiol* 4:925–932
7. Zhang D et al (2021) Comparative Genomics of *Exiguobacterium* Reveals What Makes a Cosmopolitan Bacterium. *mSystems* 6
8. Power JF et al (2024) A genus in the bacterial phylum Aquificota appears to be endemic to Aotearoa-New Zealand. *Nat Commun* 15:179
9. Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Sci* (1979) 301:976–978
10. Finlay BJ (2002) Global Dispersal of Free-Living Microbial Eukaryote Species. *Sci* (1979) 296:1061–1063
11. Schiro G, Chen Y, Blankinship JC, Barberán A (2022) Ride the dust: linking dust dispersal and spatial distribution of microorganisms across an arid landscape. *Environ Microbiol* 24:4094–4107
12. Heitkämper T et al (2024) Flying microbes—survival in the extreme conditions of the stratosphere during a stratospheric balloon flight experiment. *Microbiol Spectr* 12
13. Echenique-Subiabre I et al (2025) Traits determine dispersal and colonization abilities of microbes. *Appl Environ Microbiol* 91
14. Yang Y, Itahashi S, Yokobori S, Yamagishi A (2008) UV-resistant bacteria isolated from upper troposphere and lower stratosphere. *Biol Sci Space* 22:18–25
15. Papke RT, Ramsing NB, Bateson MM, Ward DM (2003) Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* 5:650–659
16. Podar PT, Yang Z, Björnsdóttir SH, Podar M (2020) Comparative Analysis of Microbial Diversity Across Temperature Gradients in Hot Springs From Yellowstone and Iceland. *Front Microbiol* 11
17. Marshall WA, Chalmers MO (1997) Airborne dispersal of antarctic terrestrial algae and cyanobacteria. *Ecography* 20:585–594
18. Takacs-Vesbach C, Mitchell K, Jackson-Weaver O, Reysenbach AL (2008) Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ Microbiol* 10:1681–1689

19. Valverde A, Tuffin M, Cowan DA (2012) Biogeography of bacterial communities in hot springs: A focus on the actinobacteria. *Extremophiles* 16:669–679
20. Boyd ES, Hamilton TL, Spear JR, Lavin M, Peters JW (2010) [FeFe]-hydrogenase in Yellowstone National Park: evidence for dispersal limitation and phylogenetic niche conservatism. *ISME J* 4:1485–1495
21. Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Sci* (1979) 301:976–978
22. Louca S (2022) The rates of global bacterial and archaeal dispersal. *ISME J* 16:159–167
23. Staley JT, Gosink JJ (1999) Poles Apart: Biodiversity and Biogeography of Sea Ice Bacteria. *Annu Rev Microbiol* 53:189–215
24. Menzel P et al (2015) Comparative Metagenomics of Eight Geographically Remote Terrestrial Hot Springs. *Microb Ecol* 70:411–424
25. DePoy AN, King GM (2023) Distribution and diversity of anaerobic thermophiles and putative anaerobic nickel-dependent carbon monoxide-oxidizing thermophiles in mesothermal soils and sediments. *Front Microbiol* 13
26. Marchant R, Banat IM, Franzetti A (2011) Thermophilic bacteria in cool soils: metabolic activity and mechanisms of dispersal. *Biogeogr Microscopic Organisms Is Everything Small Everywhere* 1:43–57
27. Rahman TJ, Marchant R, Banat IM (2004) Distribution and molecular investigation of highly thermophilic bacteria associated with cool soil environments. *Biochem Soc Trans* 32:209–213
28. King-Miaow K, Lee K, Maki T, LaCap-Bugler D, Archer SDJ (2019) Airborne Microorganisms in Antarctica: Transport, Survival and Establishment. 163–196. 10.1007/978-3-030-02786-5_8
29. Herbold CW, Lee CK, McDonald IR, Cary SC (2014) Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. *Nat Commun* 5:3875
30. Noell SE et al (2025) Antarctic Geothermal Soils Exhibit an Absence of Regional Habitat Generalist Microorganisms. *Environ Microbiol* 27
31. Liu L et al (2024) Insights into chemoautotrophic traits of a prevalent bacterial phylum CSP1-3, herein Syuimicrobiota. *Natl Sci Rev* 11
32. Marchant R et al (2008) Thermophilic bacteria in cool temperate soils: are they metabolically active or continually added by global atmospheric transport? *Appl Microbiol Biotechnol* 78:841–852
33. Singer CE, Ames BN (1970) Sunlight Ultraviolet and Bacterial DNA Base Ratios. *Sci* (1979) 170:822–826
34. Pullerits K et al (2020) Impact of UV irradiation at full scale on bacterial communities in drinking water. *NPJ Clean Water* 3:11
35. Bak AL, Atkins JF, Meyer SA (1972) Evolution of DNA Base Compositions in Microorganisms. *Sci* (1979) 175:1391–1393

36. Hu E-Z, Lan X-R, Liu Z-L, Gao J, Niu (2022) D.-K. A positive correlation between GC content and growth temperature in prokaryotes. *BMC Genomics* 23:110
37. Gascon J, Oubina A, Perez-Lezaun A, Urmeneta J (1995) Sensitivity of selected bacterial species to UV radiation. *Curr Microbiol* 30:177–182
38. Fernández Zenoff V, Siñeriz F, Farías ME (2006) Diverse Responses to UV-B Radiation and Repair Mechanisms of Bacteria Isolated from High-Altitude Aquatic Environments. *Appl Environ Microbiol* 72:7857–7863
39. Barbieri Oliveri L, Leung PM (2025) Trace gas oxidation as a novel microbial dispersal trait. *Current Opinion in Microbiology* vol. 88 Preprint at <https://doi.org/10.1016/j.mib.2025.102666>
40. Fantom N et al (2025) Metabolism of CO and H₂ by pioneer bacteria in volcanic soils and the phyllosphere. *ISME J* 19
41. Sriswasdi S, Yang CC, Iwasaki W (2017) Generalist species drive microbial dispersion and evolution. *Nat Commun* 8:1162
42. Chen YJ et al (2021) Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. *ISME J* 15:2986–3004
43. Søndergaard D, Pedersen CNS, Greening C (2016) HydDB: A web tool for hydrogenase classification and analysis. *Sci Rep* 6:34212
44. Saw JH et al (2026) Amplicon and metagenomic data from fumarole-associated geothermal features of Hawai'i. *bioRxiv* <https://doi.org/10.64898/2026.02.26.708227>
doi:10.64898/2026.02.26.708227
45. Van Eaton AR, Harper MA, Wilson CJ (2013) N. High-flying diatoms: Widespread dispersal of microorganisms in an explosive volcanic eruption. *Geology* 41:1187–1190
46. Boden R, Hutt LP, Rae AW (2017) Reclassification of *Thiobacillus aquaesulis* (Wood & Kelly, 1995) as *Annwoodia aquaesulis* gen. nov., comb. nov., transfer of *Thiobacillus* (Beijerinck, 1904) from the Hydrogenophilales to the Nitrosomonadales, proposal of Hydrogenophilalia class. nov. within the 'Proteobacteria', and four new families within the orders Nitrosomonadales and Rhodocyclales. *Int J Syst Evol Microbiol* 67:1191–1205
47. Dragone NB et al (2025) A comprehensive survey of soil microbial diversity across the Antarctic continent. *Polar Biol* 48:50
48. Dragone NB et al (2023) Genetic analysis of the frozen microbiome at 7900 m a.s.l., on the South Col of Sagarmatha (Mount Everest). *Arct Antarct Alp Res* 55
49. Herbold CW et al (2024) Nutritional niches of potentially endemic, facultatively anaerobic heterotrophs from an isolated Antarctic terrestrial hydrothermal refugium elucidated through metagenomics. *Environ Microbiome* 19:104
50. Gili S et al (2016) Provenance of dust to Antarctica: A lead isotopic perspective. *Geophys Res Lett* 43:2291–2298

51. Lee KC et al (2016) The *Chthonomonas calidirosea* Genome Is Highly Conserved across Geographic Locations and Distinct Chemical and Microbial Environments in New Zealand's Taupō Volcanic Zone. *Appl Environ Microbiol* 82:3572–3581
52. Laurence M, Hatzis C, Brash DE (2014) Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS ONE* 9:1–8
53. Fierer N et al (2025) Guidelines for preventing and reporting contamination in low-biomass microbiome studies. *Nat Microbiol* 10:1570–1580
54. Espinosa A et al (2026) Dataset of amplicons, metagenomes, and metagenome-assembled genomes from geothermal biofilms four years after a volcanic eruption in Hawai'i. *SSRN Preprint* at <https://doi.org/10.2139/ssrn.7389287>
55. Dartnell LR, Hunter SJ, Lovell KV, Coates AJ, Ward JM (2010) Low-Temperature Ionizing Radiation Resistance of *Deinococcus radiodurans* and Antarctic Dry Valley Bacteria. *Astrobiology* 10, 717–732
56. Li F, Ginoux P, Ramaswamy V (2010) Transport of Patagonian dust to Antarctica. *J Geophys Research: Atmos* 115
57. Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P (2000) Resistance of *Bacillus* Endospores to Extreme Terrestrial and Extraterrestrial Environments. *Microbiol Mol Biol Rev* 64:548–572
58. King CE, King GM (2014) Description of *Thermogemmatispora carboxidivorans* sp. nov., a carbon-monoxide-oxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *Int J Syst Evol Microbiol* 64:1244–1251
59. Yabe S, Yokota A (2020) *Thermogemmatispora*. in *Bergey's Manual of Systematics of Archaea and Bacteria* (ed. Oren, A.) 1–7Wiley. 10.1002/9781118960608.gbm01562
60. Jones D, Baxter B (2016) Bipyrimidine Signatures as a Photoprotective Genome Strategy in G + C-rich. *Halophilic Archaea* 6:37
61. Lappan R et al (2024) The atmosphere: a transport medium or an active microbial ecosystem? *ISME J* 18
62. Finore I et al (2023) Thermophilic bacteria and their thermozymes in composting processes: a review. *Chem Biol Technol Agric* 10:7
63. Hao Z et al (2017) Heat Generation and Accumulation in Municipal Solid Waste Landfills. *Environ Sci Technol* 51:12434–12442
64. Jones KR, Johnson JB, Aster R, Kyle PR, McIntosh WC (2008) Infrasonic tracking of large bubble bursts and ash venting at Erebus Volcano, Antarctica. *J Volcanol Geoth Res* 177:661–672
65. Bulat SA et al (2004) DNA signature of thermophilic bacteria from the aged accretion ice of Lake Vostok, Antarctica: implications for searching for life in extreme icy environments. *Int J Astrobiol* 3:1–12

66. Bottos EM, Woo AC, Zawar-Reza P, Pointing SB, Cary SC (2014) Airborne Bacterial Populations Above Desert Soils of the McMurdo Dry Valleys, Antarctica. *Microb Ecol* 67:120–128
67. Irber L, Pierce-Ward NT, Brown CT (2022) Sourmash Branchwater Enables Lightweight Petabyte-Scale Sequence Search. *bioRxiv* <https://doi.org/10.1101/2022.11.02.514947>
doi:10.1101/2022.11.02.514947
68. Lumian J et al (2024) Biogeographic distribution of five Antarctic cyanobacteria using large-scale k-mer searching with sourmash branchwater. *Front Microbiol* 15
69. Hera MR, Pierce-Ward NT, Koslicki D (2023) Deriving confidence intervals for mutation rates across a wide range of evolutionary distances using FracMinHash. *Genome Res.*
<https://doi.org/10.1101/gr.277651.123>
70. Grettenberger C, Gold DA, Brown CT (2025) Distribution of early-branching Cyanobacteria and the potential habitats that gave rise to the earliest oxygenic phototrophs. *mSphere* 10
71. Koslicki D, White S, Ma C, Novikov A (2024) YACHT: an ANI-based statistical test to detect microbial presence/absence in a metagenomic sample. *Bioinformatics* 40
72. Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. *R J* 10:439–446
73. Lee MD (2019) GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35:4162–4164
74. Hyatt D et al (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119
75. Eddy SR, Accelerated Profile HMM, Searches (2011) *PLoS Comput Biol* 7:e1002195
76. Edgar RC (2022) Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun* 13:6968
77. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973
78. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* 5:e9490
79. Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 52:W78–W82
80. Stott MB et al (2008) Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environ Microbiol* 10:2030–2041
81. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
82. Bankevich A et al (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19:455–477
83. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075

84. Chalita M et al (2024) EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. *Int J Syst Evol Microbiol* 74
85. Musilova M, Wright G, Ward JM, Dartnell LR (2015) Isolation of Radiation-Resistant Bacteria from Mars Analog Antarctic Dry Valleys by Preselection, and the Correlation between Radiation and Desiccation Resistance. *Astrobiology* 15:1076–1090
86. Walsh RL, Camilli A (2011) *Streptococcus pneumoniae* is desiccation tolerant and infectious upon rehydration. *mBio* 2
87. DeFlaun MF et al (2007) Isolation and characterization of a *Geobacillus thermoleovorans* strain from an ultra-deep South African gold mine. *Syst Appl Microbiol* 30:152–164
88. Raposo P Transfer of *Meiothermus chliarophilus* et al (1995) Tenreiro Nobre 1996, *Meiothermus roseus* Ming. 2016, *Meiothermus terrae* Yu 2014 and *Meiothermus timidus* Pires 2005, to *Calidithermus* gen. nov., as *Calidithermus chliarophilus* comb. nov., *Calidithermus roseus* comb. nov., *Calidithermus terrae* comb. nov. and *Calidithermus timidus* comb. nov., respectively, and emended description of the genus *Meiothermus*. *Int. J. Syst. Evol. Microbiol.* 69, 1060–1069 (2019)
89. Williams RAD (2021) The Genus *Thermus*. *Thermophilic Bacteria*. CRC, pp 51–62. doi:10.1201/9781003068334-3.
90. Irber L et al (2024) sourmash v4: A multitool to quickly search, compare, and analyze genomic and metagenomic data sets. *J Open Source Softw* 9:6830
91. Callahan BJ et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583
92. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10
93. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* 73:5261–5267
94. McLaren MR, Silva (2024) SSU taxonomic training data formatted for DADA2 (Silva version 138.2) (Version 2). *Zenodo* 10.5281/zenodo.14169026 doi:https://doi.org/10.5281/zenodo.3986799
95. R Core Team (2020) R: A language and environment for statistical computing. <https://www.r-project.org/>
96. Salter SJ et al (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:1–12
97. Eisenhofer R et al (2019) Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol* 27:105–117
98. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ (2018) Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6:226
99. Oksanen J et al (2020) vegan: Community Ecology Package. Preprint at <https://cran.r-project.org/package=vegan>

100. McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8:e61217

101. De Cáceres M, Legendre P (2009) Associations between species and groups of sites: indices and statistical inference. Ecology 90:3566–3574

102. Abbaszadeh J (2024) The ecology and evolutionary history of the Deinococcaceae family. (University of Waikato)

Figures

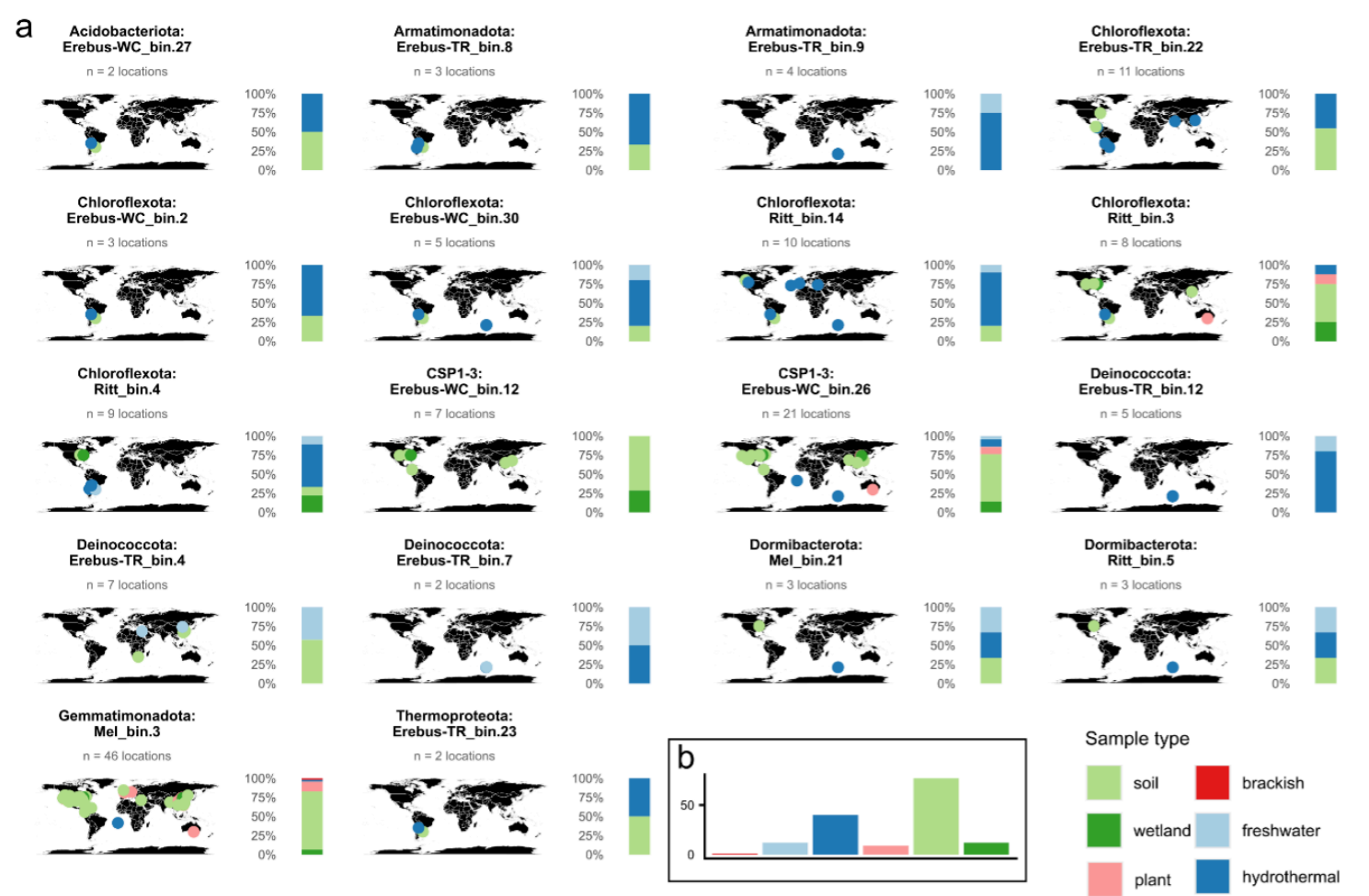


Figure 1

(A) Maps for all globally cosmopolitan thermophiles from Antarctic geothermal sites, identified by the phylum of origin and MAG name. Each dot represents a metagenome from the Sequence Read Archive (SRA) (with metagenomes within 10 m consolidated into a single location) identified as containing a near-identical species (containment $\geq 30\%$, ANI $\geq 96\%$). Habitats from the SRA metadata for each metagenome were condensed into 10 categories for ease of visualization. The stacked bar plot next to each MAG indicates the proportion of each sample type in which that MAG was found. (B) Across all MAGs, the frequency with which MAGs were identified in each type of habitat.

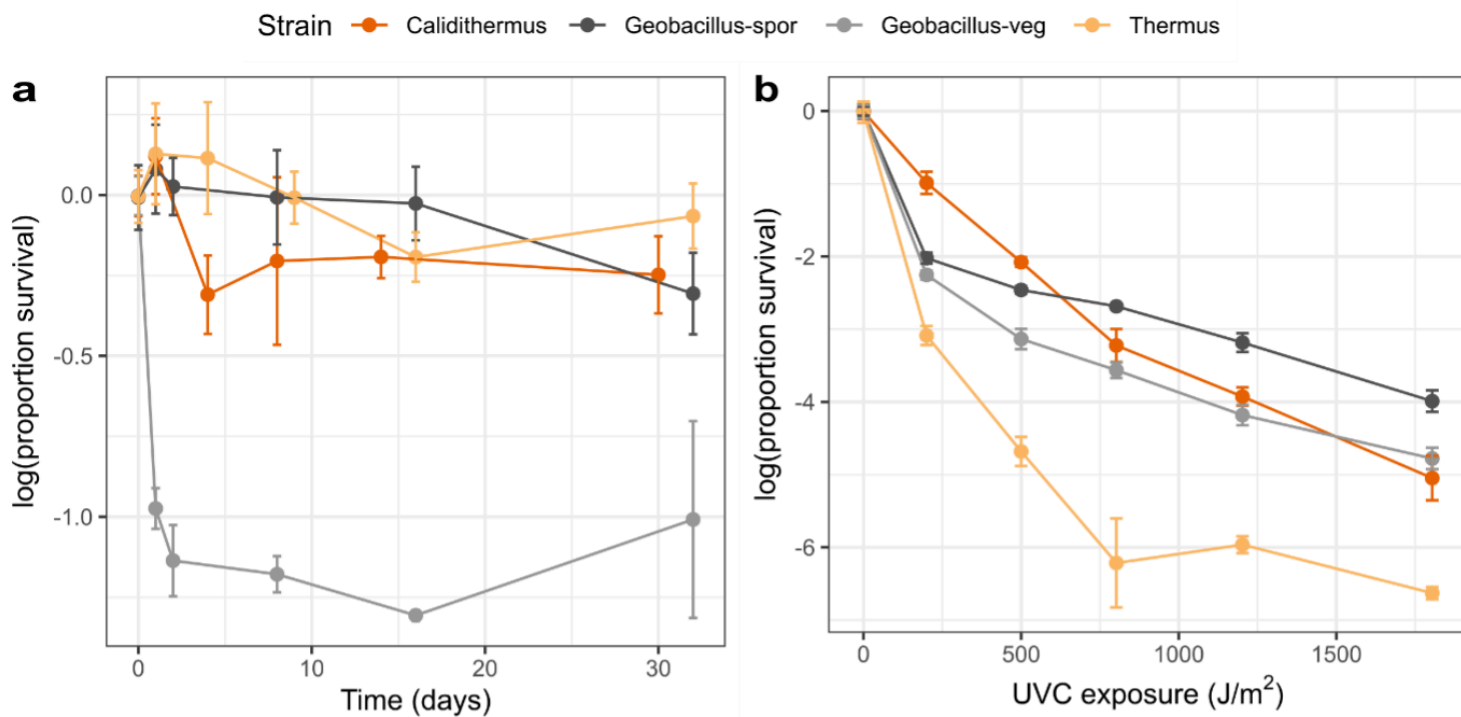


Figure 2

Results of experiments testing the ability of different species and, in the case of *Geobacillus*, physiological states (vegetative cells, “veg”, vs. culture enriched in spores, “spor”) to survive either (A) being frozen, desiccated, and starved for multiple days or (B) UV-C exposure. The results were calculated as a proportion of cells surviving compared to the time zero concentration. Error bars represent standard deviation around the mean for triplicate cultures. Where error bars are not visible, the standard deviation is too small to be visible.

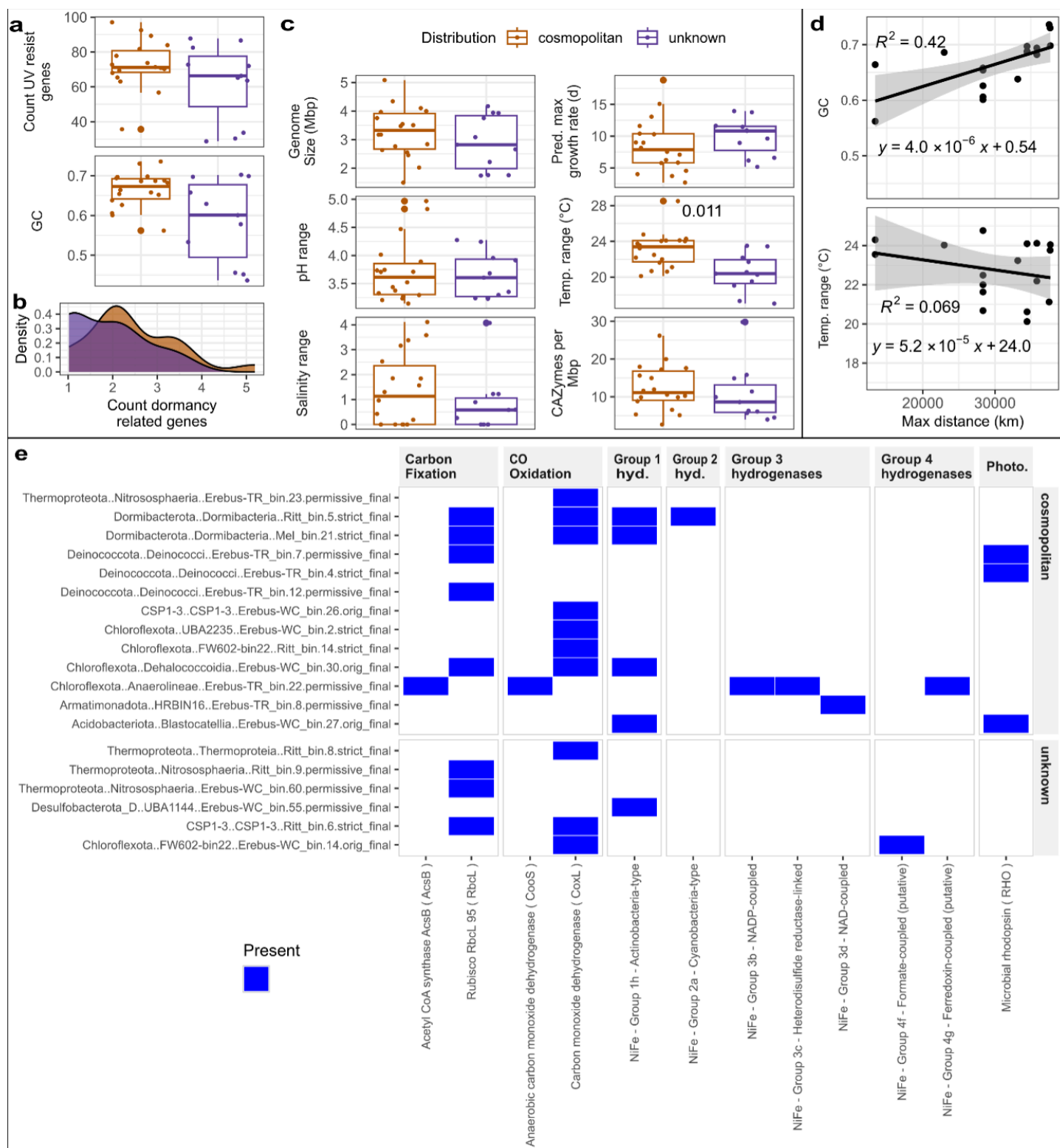


Figure 3

(A-C) Predicted traits based on MAG sequences, compared between Antarctic geothermal cosmopolitan or unknown MAGs. (A-B) Traits predicted to be associated with dispersal success. (C) Traits predicted to be associated with colonization success. Presented p -values are the result of Wilcoxon tests (Benjamini-Hochberg adjusted for multiple tests; $n = 18$ for cosmopolitan, 11 for unknown for each boxplot). Where no p -values are shown, the difference was not significant ($p > 0.05$). (D) Linear regression analysis for

the two variables identified from A-D as being significantly or very different in medians between the two groups, correlated with the maximum distance between Mt. Erebus and the furthest distant metagenome from the SRA with a matching species. For temperature range, a single outlier with a temperature range of 28°C was removed (n = 18 for GC, 17 for temperature range). (E) Predicted metabolic functions associated with dispersal success; only MAGs containing at least one of these predicted metabolisms are shown. Hyd: hydrogenases. Photo: Phototrophy.

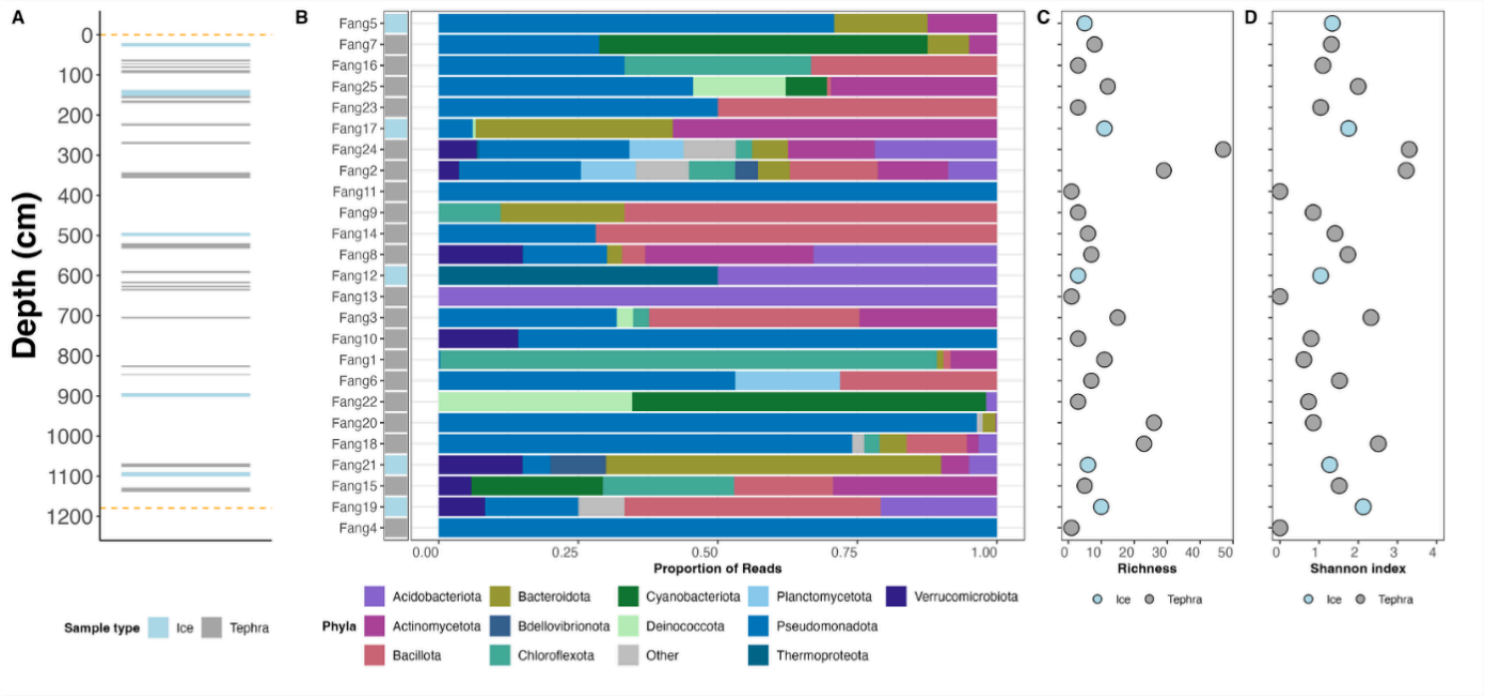


Figure 4

Microbial communities in the Fang Glacier ice core layers. (A) Diagram of the 25 layers extracted for our study. (B) Community composition of the most 12 most abundant bacterial families identified in our layers. (C) Richness of the microbial communities in each layer. (D) Shannon diversity of each layer. In (B) – (D), the samples are ordered (top to bottom) based on the layers displayed in (A).

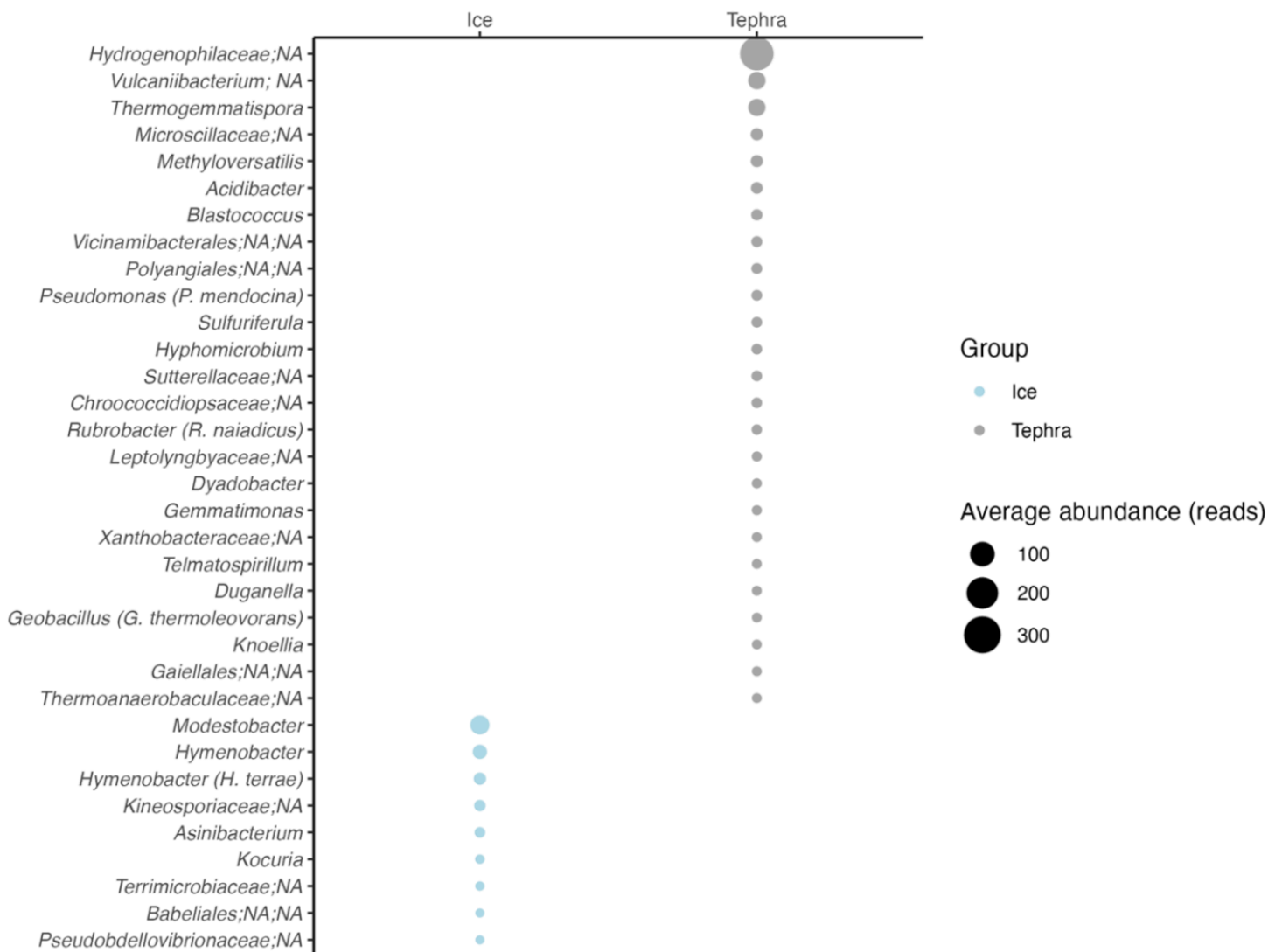


Figure 5

The most abundant organisms found to be associated with the ice layers (10 genera) and the tephra layers (25 genera) from the Fang Glacier ice core. Genera are organized from most abundant to least abundant. If an organism could not be assigned to a genus, the next most specific taxonomic identification is listed.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS11airasvsinfo.xlsx](#)
- [TableS9fangcoreasvinfo.xlsx](#)
- [TableS3cANIthresholds.csv](#)
- [TableS4maginfovf.xlsx](#)
- [TableS10fangasvsremoved.xlsx](#)

- [TableS1DNARepairGenes.csv](#)
- [Globalcosmopolitanthermophilesmanuscriptsupplementaryv1.docx](#)
- [TableS8waterisotopes.xlsx](#)
- [TableS5rawbranchwater.csv](#)
- [nrsoftwarepolicyfilled.pdf](#)
- [nrreportingsummaryfilled.pdf](#)