Metagenomics untangles potential adaptations of Antarctic endolithic bacteria at the fringe of habitability

Claudia Coleine, Davide Albanese, Angelique E. Ray, Manuel Delgado-Baquerizo, Jason E. Stajich, Timothy J. Williams, Stefano Larsen, Susannah Tringe, Christa Pennacchio, Belinda C. Ferrari, Claudio Donati, Laura Selbmann

PII: S0048-9697(24)00425-X

DOI: https://doi.org/10.1016/j.scitotenv.2024.170290

Reference: STOTEN 170290

To appear in: *Science of the Total Environment*

Received date: 21 November 2023

Revised date: 17 January 2024

Accepted date: 17 January 2024

Please cite this article as: C. Coleine, D. Albanese, A.E. Ray, et al., Metagenomics untangles potential adaptations of Antarctic endolithic bacteria at the fringe of habitability, *Science of the Total Environment* (2023), https://doi.org/10.1016/j.scitotenv.2024.170290

Metagenomics untangles potential adaptations of Antarctic endolithic bacteria at the fringe of habitability

Claudia Coleine[1,*,†], Davide Albanese[2,†], Angelique E. Ray[3,†], Manuel Delgado-Baquerizo[4,5], Jason E. Stajich[6,7], Timothy J. Williams[3], Stefano Larsen[2], Susannah Tringe[8], Christa Pennacchio[8], Belinda C. Ferrari[3], Claudio Donati[2,*], Laura Selbmann[1,9]

[1]*Department of Ecological and Biological Sciences, University of Tuscia, Largo dell'Università, 01100, Viterbo, Italy*

[2]*Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38098, San Michele all'Adige, Italy*

[3]*School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Randwick, New South Wales, 2052, Australia*

[4]*Laboratorio de Biodiversidad y Funcionamiento Ecosistémico. Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Av. Reina Mercedes 10, E-41012, Sevilla, Spain*

[5]*Unidad Asociada CSIC-UPO (BioFun). Universidad Pablo de Olavide, 41013 Sevilla, Spain*

[6]*Department of Microbiology and Plant Pathology and Institute of Integrative Genome Biology, University of California, Riverside, Riverside, CA, 92507, USA*

[7]*Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA, USA*

[8]*Department of Energy Joint Genome Institute, One Cyclotron Road, Berkeley, CA, 94720, USA*

[9]*Mycological Section, Italian Antarctic National Museum (MNA), Via al Porto Antico, 16128, Genoa, Italy*

[†]These authors contributed equally to this work.

Correspondence: coleine@unitus.it (Largo dell'Università snc, Viterbo, 01100, Italy); claudio.donati@fmach.it (Via E. Mach, San Michele all'Adige, 38098, Italy)

Running title: Catalog of Antarctic bacterial genomes

Abstract

Survival and growth strategies of Antarctic endolithic microbes residing in Earth's driest and coldest desert remain virtually unknown. From 109 endolithic microbiomes, 4,539 metagenome-assembled genomes were generated, 49.3% of which were novel candidate bacterial species. We present evidence that trace gas oxidation and atmospheric chemosynthesis may be the prevalent strategies supporting metabolic activity and persistence of these ecosystems at the fringe of life

and the limits of habitability.

Keywords

Antarctica, Extremophiles, Habitability, Adaptation, Metagenomics, MAGs

1 Introduction

Permanently ice-free areas cover less than 1% of the Antarctic continent (Lee et al. 2017) and include the coldest, driest and the most oligotrophic environments of Earth. Even so, Antarctic rocks are unexplored and isolated ecosystems that support highly diverse microbial communities; in such regions, highly adapted life forms subjected to a combination of poly-stresses still perpetuate (Dragone et al. 2021; Montgomery et al. 2021). Endolithism lifestyle represents adaptation at the edge inhabitable conditions; it is a specialized colonization of microorganisms dwelling inside airspaces of rocks. Airspaces within rocks offer microbiota a protected and buffered microenvironment, allowing life to expand into different extreme conditions (Friedmann, 1982; Archer et al. 2017). Endolithic communities constitute simple food webs of varying complexity. Lichen-associated or free-living chlorophycean algae and *Cyanobacteria* function as primary producers, whilst fungi and more heterotrophic bacteria support key ecosystem services such as nutrient cycling, rock weathering, and proto-soil formation (de La Torre et al. 2003; Archer et al. 2017). Recent scientific studies considerably advanced our understanding of endolithic microbial biodiversity, environmental preferences, and extraordinary resistance to multiple stresses (Archer et al. 2017; Coleine et al. 2020; Gevi et al. 2022). For instance, it was recently found that the majority of new bacterial species belong to monophyletic bacterial clades that diverged from related taxa in a range from 1.2 billion to 410 million of years and are functionally distinct from known related taxa (Albanese et al. 2021). More recently, is has been presented the first predicted viral catalog comprising > 75,000 viral operational taxonomic units (vOTUS), with potential functions that indicate that they might influence other rock's microbial components (Ettinger et al. 2023).

However, despite a number of studies being conducted at the community level, we still lack the most basic knowledge of how Antarctic endoliths survive the challenging conditions. A comprehensive genome catalog is the necessary first step to clarifying the metabolic features and capabilities of these microorganisms and to elucidate how they survive such harsh conditions. Learning more about life under the extreme conditions is critical towards defining the fringe of habitability on Earth (Merino et al. 2019).

To address this knowledge gap, we conducted a field survey including 109 endolithically colonized rocks, covering a plethora of regions and environments found in ice-free Antarctica, which includes a broad range of geo-environmental (e.g. altitudinal gradient, different rock typologies) and geographical distributions (i.e. Antarctic Peninsula, Northern Victoria Land, and McMurdo Dry Valleys; Figure 2a-c; Supplementary Table S1). We herein present the first Antarctic Rock Genomes Catalog (ARGC), which is the most comprehensive resource of bacterial metagenome-assembled genomes (MAGs) from terrestrial Antarctica to date.

2 Material and Methods

2.1 Study area

Rocks colonized by endolithic communities were collected in thirty-eight sites in Antarctica including Antarctic Peninsula (*n* = 3), McMurdo Dry Valleys, Southern Victoria Land (*n* = 27), and Northern Victoria Land (*n* = 79) during more than 20 years of Italian Antarctic Expeditions. Different rock typologies (sandstone *n* = 50, granite *n* = 43, quartz *n* = 5 , and basalt/dolerite *n* = 2) were collected along a latitudinal transect (ranging from -62.10008 -58.51664 to -77.874 160.739) and selecting different environmental conditions: sun exposure (northern sun exposed and southern shady rocks), altitude, distance from sea (up to 3,100 m above sea level (a.s.l.). This selection has been made to provide a comprehensive overview of Antarctic endolithic diversity (Figures 1, 2, Supplementary Table 1). The presence of endolithic colonization was assessed by direct observation *in situ* by using magnifying lens. Rocks were excised aseptically using a geologic hammer and sterile chisel, and rock samples, preserved in sterile plastic bags, and immediately preserved at −20 °C upon collection to avoid contamination. Rocks were then transported to University of Tuscia and stored at −20 °C in the Culture Collection of Antarctic fungi of the Mycological Section of the Italian Antarctic National Museum (MNA-FCC), until downstream analysis.

Figure 1. Examples of samples collected in the Victoria Land, Continental Antarctica. **a)** Finger Mt., **b)** Linnaeus Terrace, **c)** Battleship Promontory.

2.2 DNA extraction, library preparation, and sequencing

Metagenomic DNA was extracted from 1 g of crushed rocks using DNeasy PowerSoil Pro Kit (Qiagen, German), quality checked by electrophoresis using a 1.5% agarose gel and Nanodrop spectrophotometer (ThermoFisher, USA) and quantified using the Qubit dsDNA HS Assay Kit (Life Technologies, USA) according to Coleine et al. (2021). Shotgun metagenomic sequencing paired-end libraries were constructed and sequenced as 2×150 bp using the Illumina NovaSeq platform (Illumina Inc, San Diego, CA) at the Edmund Mach Foundation (San Michele all'Adige, Italy) and at the DOE Joint Genome Institute (JGI).

2.3 Sequencing reads preparation, assembly and binning

The metashot/mag-illumina v2.0.0 Nextflow-based (Di Tommaso et al. 2017) workflow (https://github.com/metashot/mag-illumina, parameters: --metaspades_k 21,33,55,77,99) was used to perform raw reads quality trimming and filtering, assembly and contings binning on the 91 metagenomic samples. In brief, adapter trimming, contaminant (artifacts and and spike-ins)

and quality filtering were performed using BBDuk (BBMap/BBTools v38.79, https://sourceforge.net/projects/bbmap/). During the quality filtering procedure i) raw reads were quality-trimmed to Q6 using the Phred algorithm; ii) reads that contained 4 or more "N" bases, had an average quality below 10, shorter than 50 bp or under 50% of the original length were removed. Samples were then assembled individually with SPAdes (Nurk et al. 2017) v3.15.1 (parameters --meta -k 21,33,55,77,99).

Metagenomic contigs were binned into candidate metagenome-assembled genomes (MAGs) using MetaBAT 2 (Metagenome Binning based on Abundance and Tetranucleotide frequency) (Kang et al. 2019) v2.12.1. Briefly, high-quality reads were mapped on assembled contigs using Bowtie2 (Langmead and Salzberg 2012) v2.3.4.3. Samtools (Li et al. 2009) (htslib v1.9) was used to create and sort the BAM files. The depth of coverage was estimated by applying the MetaBAT2 script "jgi_summarize_bam_contig_depths". Finally, contigs sequences and the depth of coverage estimates were used by MetaBAT2 to recover the 10,677 bins.

2.4 Quality assessment, filtering and dereplication

The resulting bins were combined with the 1650 metagenomic bins from (Albanese et al. 2021) and analyzed using themetashot/prok-quality (Albanese and Donati 2021) v1.2.3 (parameters --gunc_filter --gunc_db gunc_db_2.0.4.dmnd) workflow. Briefly, completeness, redundant and non-redundant contamination (Orakov et al., n.d.) estimates were obtained by CheckM (Parks et al. 2015) v1.1.2 and GUNC (Orakov et al., n.d.). Bins with completeness estimates of <50%, more than 10% contamination and that did not pass the GUNC filter were discarded, resulting in a total of 4,540 filtered prokaryotic MAGs. MAGs were classified into "high-quality draft" (HQ) with >90% completeness and <5% contamination and "medium-quality draft" (MQ) with completeness estimates of ≥50% and less than 10% contamination. Species-level operational taxonomic units (OTUs) were identified by clustering HQ and MQ MAGs at 95% average nucleotide identity (ANI) using dRep (Olm et al. 2017) v2.6.2, resulting in a total of 2,279 OTUs. For each species-level OTUs, the MAG with the highest quality score was chosen as representative. The score was computed using the formula: score = completeness - 5 x contamination + 0.5 x log(N50) (Albanese and Donati 2021).

2.5 Taxonomic classification of prokaryotic MAGs

Species-level OTUs representative MAGs were taxonomically classified using the metashot/prok-classify v1.2.1 workflow (https://github.com/metashot/prok-classify, parameters:

--gtdbtk_db release202). The workflow includes the genome taxonomy database toolkit (GTDB-Tk) (Chaumeil et al. 2019) v1.5.0 and the GTDB release 202, following the recently proposed nomenclature of prokaryotes (Parks et al. 2022). A single OTU was classified as archaea and was removed from subsequent analyses. Approximately-maximum-likelihood phylogenetic tree from the GTDB protein alignments of the 2,278 bacterial OTU representatives was inferred using FastTree (Price, Dehal, and Arkin 2010) v2.1.11 (default parameters).

2.6 Bacterial OTU coverage estimates in metagenomes

The metashot/containment v1.0.0 workflow (https://github.com/metashot/containment, parameters: --min_identity 0.95 --winner_takes_all --sketch_size 10000) was used to determine the presence of the reconstructed bacterial OTU in the 109 Antarctic samples. Briefly, for each metagenome we applied the Mash Screen algorithm (Ondov et al. 2019) (Mash v2.1) in order to calculate the containment score for each OTU (i.e., the estimate of the similarity of an OTU representative to a sequence contained within the metagenome), its p value and the OTU median-multiplicity, as a proxy for the OTU coverage. The Mash Screen algorithm demonstrated to be in good agreement with the mapping-and-consensus procedure described in (Albanese et al. 2021).

2.7 Functional annotation of bacterial MAGs

Functional annotation was performed using the workflow metashot/prok-annotate (https://github.com/metashot/prok-annotate, commit da2d0bb, parameters: --run_eggnog --eggnog_db emapperdb-5.0.2). Input MQ and HQ bacterial MAGs (n=4,539) were processed as follows: (i) 16,830,059 translated coding DNA sequences (CDSs) were predicted using Prokka (Seemann 2014) v1.14.5 which in turn wraps the gene predictor Prodigal (Hyatt et al. 2010) and (ii) functionally annotated using EggNOG-mapper (Cantalapiedra et al. 2021) (v2.1.4, parameters -m diamond --itype protein) against the eggNOG Orthologous Groups (OGs) database (Huerta-Cepas et al. 2019) v5.0.2. The eggNOG database integrates functional annotations collected from several sources, including Gene Ontology (GO) terms, KEGG functional orthologs (Kanehisa et al. 2014) and COG categories (Tatusov et al. 2000). For each species-level OTU, a target gene/ortholog was marked as "present" if more or equal than 80% of the HQ genomes which belong to the OTU encoded that gene/ortholog.

Translated CDSs were de-replicated at 95%, 80% and 50% identity and an alignment fraction threshold of 80% using MMseqs2 (Mirdita, Steinegger, and Söding 2019) v13-1 with the parameters "easy-linclust -e 0.001 --min-seq-id [IDENTITY] -c 0.80". 50% protein cluster

representatives were searched against the UniProt Reference Clusters (UniRef50, release 2022_01, 23-Feb-2022, http://www.uniprot.org) with an identity threshold of 50% using the MMseqs2's easy-search protocol (parameters -e 0.001 --min-seq-id 0.5 --cov-mode 2 -c 0.8).

Moreover, translated CDSs (n=16,830,059) were searched against the "Greening lab metabolic marker gene databases" (Greening 2021) using an identity threshold of 50% (parameters: easy-search --min-seq-id 0.5 --cov-mode 2 -c 0.8). Best hits were further filtered for some marker gene according to (Chen et al. 2021): [NiFe]-hydrogenases, [FeFe]-hydrogenases, CoxL, AmoA, NxrA and NuoF were filtered at 60% identity threshold, AtpA, YgfK, HbsT, ARO, and PsbA at 70%, and PsaA at 80%. For each species-level OTU, a target gene was marked as "present" if more or equal than 80% of the HQ genomes that belong to the OTU encoded that gene.

2.8 Phylogenetic Analysis of RuBisCO and [NiFe]-hydrogenase

A total 978 putative RuBisCO sequences and 2433 putative [NiFe]-hydrogenase sequences were yielded from the recovered MAGs. All sequences obtained were further classified into subforms using previously published databases and BLAST+ (ver. 2.12.0) (Camacho et al. 2009). In addition, separate phylogenetic analyses were conducted to visualize the forms of RuBisCO and [NiFe]-hydrogenase present. The extracted RuBisCO sequences were analyzed against 3129 reference sequences obtained through previous phylogenetic analysis of the Genome Taxonomy Database (Ray et al., 2022). [NiFe]-hydrogenase sequences were analyzed against 2019 reference sequences obtained from the HydDB (Søndergaard, Pedersen, and Greening 2016)(Søndergaard et al., 2016) and previous phylogenetic analysis (Ray et al., 2022).

Multiple sequence alignment was conducted using MAFFT (ver.7.407), applying the L-INS-i iterative refinement method (Katoh et al. 2002; Katoh and Standley 2013). To remove poorly aligned regions, the resulting alignments were trimmed using trimAl (ver. 1.4.1), with a gap threshold of 0.5 (Capella-Gutiérrez et al., 2009). Sequences with more than 50% gaps after alignment were removed. Maximum likelihood phylogenetic trees were constructed using IQ-Tree (ver. 1.6.10) (Nguyen et al. 2015)(Nguyen et al., 2015), applying 1000 ultrafast bootstrap iterations, hill-climbing nearest neighbor interchange (NNI) search and incorporating additional SH-like approximate likelihood ratio tests (SH-alrt) (Guindon et al., 2010). ModelFinder was used to determine the best evolutionary model, which was LG+F+R10 for the RuBisCO tree and LG+R10 for the [NiFe]-hydrogenase tree. Sequences that failed the chi2 test during tree building were removed.

The final consensus trees, comprising 1,347 RuBisCO sequences and 2,706 [NiFe]-hydrogenase sequences, were uploaded to Interactive Tree Of Life (iTOL) v6 (Letunic and Bork 2016) for visualization. Branches were color-coded according to the form of [NiFe]-hydrogenase or RuBisCO, and bootstrap values 90-100 were indicated by circles on the corresponding branches, with size corresponding to values. The phyla of the MAGs from which each sequence originated are displayed as a color-coded outer ring. In the [NiFe]-hydrogenase tree, if the RuBisCO large subunit co-occurred within the originating genome, then these sequences are marked by an outer pie chart, which depicts the proportion of RuBisCO forms detected. Complete trees showing all RuBisCO and [NiFe]-hydrogenase sequences are provided (Supplementary Information). Collapsed versions are also provided, focusing upon clades where sequences from the MAGs studied here were identified, namely RuBisCO form I and forms 1h, 1l, 1m, 1e, 2a, 3b, 3d [NiFe]-hydrogenase.

2.9 Downstream analysis

Downstream analysis was performed using the R environment (https://www.R-project.org/) v4.0.3 and the packages tidyverse v1.3.0, ggtree v2.4.1 and phytools v0.7-70.

3 Results and Discussion

Following quality filtering (see Online Methods), 2,636 high-quality (HQ with ≥90% completeness and <5% contamination) and 1,903 medium-quality (MQ with ≥50% completeness and <10% contamination) bacterial MAGs were classified (Figure 2d; Supplementary Table S2, Supplementary Figures S1-5). The ARGC provides a complete picture of sandstone microbiomes across Antarctica, as revealed by the accumulation curves, which indicate that most species were retrieved; whilst, diversity in granite requires further elucidation (Supplementary Figure S5). MAGs were then grouped at 95% average nucleotide identity (ANI) into 2,278 species-level bacterial operational taxonomic units (OTUs) (Figure 2e, f), 8.6 times more than previously reported (Albanese et al. 2021). All the OTUs can be assigned to known phyla, while 2,277, 2,262, 2,164 (95%), and 1,433 (63%) to known classes, orders, families and genera, respectively. Notably, 98.3% of species-level OTUs were distinct from the Genome Taxonomy Database (GTDB) reference genomes, representing 2,239 new candidate species (Figure 2e; Supplementary Table S3). On a phyla level, *Actinobacteriota* and *Proteobacteria* were dominant,

with many new genomes of *Acidobacteriota*, *Chloroflexota*, and *Bacteroidota* also uncovered. *Actinomycetia* and *Thermoleophilia*, *Alphaproteobacteria*, and *Chloroflexia* classes were the most abundant and recurrent in the dataset (Figure 2g, Supplementary Figure S6; Supplementary Tables S4, S5). The dominant orders were *Mycobacteriales* (38%), *Actinomycetales* (15%), *Solirubrobacterales* (14%), *Acetobacterales* (12%), and *Thermomicrobiales* (7%) (Supplementary Table S6, S7).

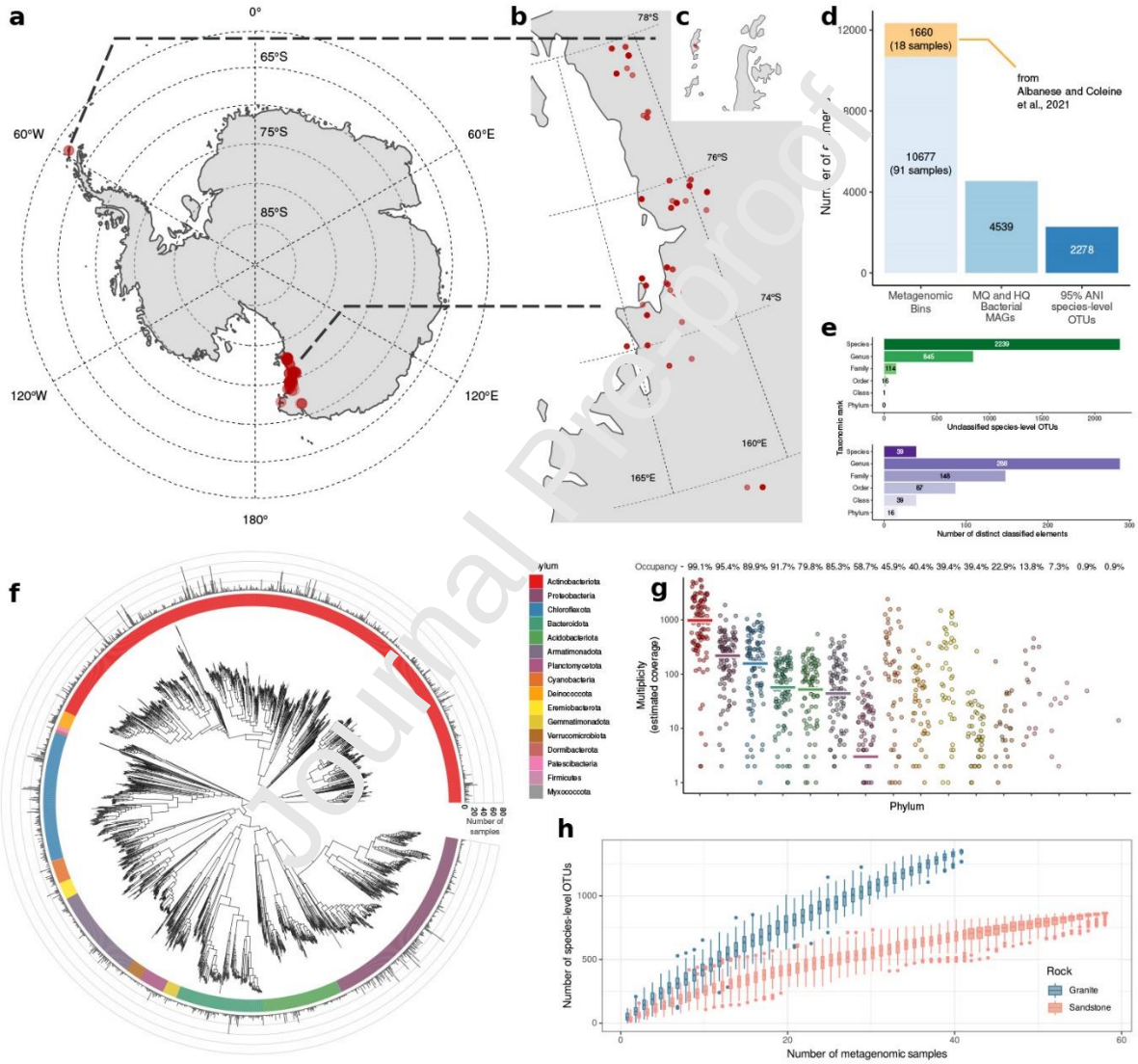


Figure 2. Study area and MAGs characteristics. a-c, Map of Antarctica (a) and sampling sites (Victoria Land, b; Peninsula, c) (red dots). **d,** Number of MAGs and their quality-based classification. **e,** Upper bar plot: number of unclassified OTUs. Bottom bar plot: number of species, genera, families, orders, classes and phyla. **f,** Phylogenetic tree of the 2,278 OTUs built from the multiple sequence alignment of 120 GTDB marker genes. Barplot in the outer circle indicates the number of samples in

which each OTUs was found. **g,** Phylum-level Mash Screen multiplicity for each sample, indicating sequence coverage. Horizontal lines represent the median values. The occupancy value indicates the percentage of samples that contains the underlying phylum. **h,** Number of OTUs as a function of the number of rock samples.

To predict metabolic competencies, we retrieved 16,830,059 protein coding sequences (CDS) based on Prodigal analysis (see Methods). These CDS were dereplicated into 9,632,227, 6,997,885, 4,538,534 protein groups using MMseqs2 with identity thresholds of 95%, 80% and 50% respectively. Moreover, 50% protein representatives were searched against the UniProt Reference Clusters (Suzek et al. 2007) (UniRef, see Methods); since only 52.4% of the proteins displayed at least one match within the database, this resource should lay the foundation for future Antarctic terrestrial catalog.

During functional analysis, we focused on two widespread survival and growth strategies that allow microbiomes to persevere in extreme, oligotrophic environments: autotrophic metabolism, particularly trace gas based chemosynthesis, and cold resistance adaptations. In cold edaphic deserts, energy generation through trace gas oxidation supports both microbial persistence and growth, with increased carbon fixation activity observed with aridity (Chen et al. 2021; Ortiz et al. 2021; Ray et al. 2022). However, the significance of this strategy to endolithic microbiomes where photosynthetic microorganisms are more prevalent is questionable (Wierzchos et al., 2012).

High-affinity [NiFe]-hydrogenase genes, including forms 1h, 1l, 1m and 2a, are widely represented in our dataset, occurring in 41.1% of all dereplicated MAGs, including *Ca.* Dormibacterota (88.9%), *Eremiobacterota* (80.2%), *Actinobacteriota* (59.1%), *Gemmatimonadota* (57.1%), *Chloroflexota* (53.0%), *Acidobacteriota* (43.9%), *Verrucomicrobiota* (25.8%), *Planctomycetota* (13.4%), *Cyanobacteria* (7.5%), *Bacteroidota* (7.3%), *Proteobacteria* (6.1%), and *Armatimonadota* (4.8%) (Figure 3). The oxidation of trace levels of hydrogen gas plays a key role for persistence in dormant state and is a wide- spread ability in both Bacteria and Archaea in terrestrial and marine ecosystems (Greening, C. & Grinter, 2022; Lappan et al. 2023). The same strategy may be therefore crucial to support endolithic microbiomes whose active metabolism is, as average, limited to 1,000 h per year only (Friedmann et al. 1987).

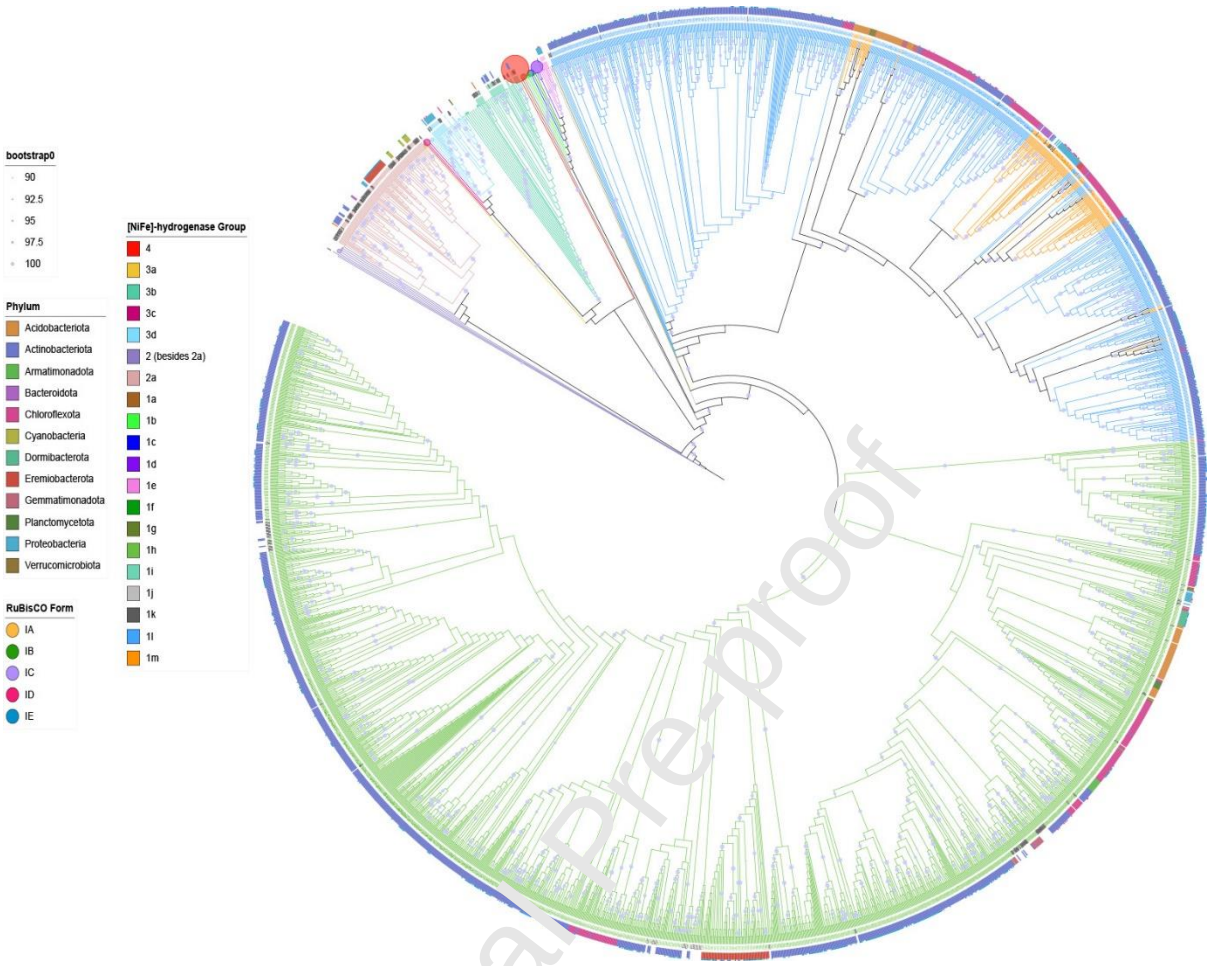


Figure 3. **Phylogenetic tree of [NiFe]-hydrogenase**. Maximum likelihood phylogenetic tree of [NiFe]-hydrogenase gene sequences obtained from our MAGs (n = 2433), with reference sequences obtained from the HydB and previous phylogenetic analysis. Branches and reference gene labels are colored according to the group of [NiFe]-hydrogenase. Bootstrap values >90% are depicted as filled circles on branches, with size reflecting value, and 1000 ultrafast bootstrap iterations applied. The phyla of the originating MAGs assembled in this study are displayed in a color-coded outer ring. In cases where RuBisCO large subunit gene/s co-occurred within these genomes, the proportion of forms present is indicated by external pie charts.

Autotrophic metabolisms are critical under such strict oligotrophic conditions and were indeed pervasive amongst the bacterial MAGs uncovered. Specifically, representatives from 7 of the 15 phyla presented signatures for carbon fixation. Phototrophic metabolism, mostly largely present in *Cyanobacteria*, is based on photolysis and requires water to take place. Data presented here suggests that trace gas oxidation may produce enough energy to not only support persistence but

also to fuel the Calvin-Benson-Bassham (CBB) cycle in a subset of the residing bacterial taxa, through the process of atmospheric chemosynthesis. This process is limited to cold soil deserts, while scarce to no carbon fixation activity has been observed yet in other environments (Ray et al. 2022; Ji et al. 2017). Here we provide clear evidence that atmospheric chemosynthesis could be extended to endolithic populations and may be a key adaptation for Carbon accumulation under highly dry conditions, with this process also proposed to be water-producing (Cowan et al. 2022). High-affinity [NiFe]-hydrogenases co-occurred alongside light-independent RuBisCO (1E/D) in 72.2% of *Ca*. Dormibacterota, 62.3% of *Eremiobacterota*, 20.6% of *Actinobacteriota*, 8.8% of *Chloroflexota*, 2.9% of *Gemmatimonadota* and 2.5% of *Proteobacteria* MAGs (Supplementary Figure S7), with RuBisCO form IE dominant accounting for 92.7% of those detected. These genetic indicators suggest that atmospheric chemosynthesis, as a fundamental process for primary production in hyper-arid cold environments, may be extended beyond soils to endolithic niches. RuBisCO form ID, showing a CO$_2$ high affinity, is better adapted to a higher O_2/CO_2 ratio and requires less energetic or nutrient investment to attain high carboxylation rates; this finding suggests that although uncommon, other RuBisCO forms may play a role in this chemoautotrophic process (Rickaby et al. 2019). We propose that the plethora of RuBisCO forms found, displaying various efficiency, specificities, and affinities, enables the community to modulate its activity shifting from dormant to active state; this is paramount to adapt and exploit extreme and fluctuating microenvironments.

Aerobic respiration was predominant among endolithic MAGs (Supplementary Table S8; Figure 4); yet, the ability to use alternative e- acceptors via formate dehydrogenase, were limited to rare phyla, particularly in *Thermoanaerobaculia*, which was represented by one single family of anaerobic bacteria. The presence of additional chemosynthetic pathways, alternative to atmospheric chemosynthesis, using e- donors via Arsenate reductase were also found in a few (7) phyla, particularly abundant in *Bacilli*. This plethora of abilities to exploit various e- donors or acceptors increase the possibility of adaptability and survival of the whole community.

Lastly, below-freezing temperatures are a main challenge to life that can influence metabolic activity; reaching temperatures as low as -89°C, Antarctica is the coldest continent on the planet. We found that Antarctic endolithic bacteria encompass an innate adaptive capacity to cope with

life in the persistent cold and the associated stresses. In fact, well-established genes involved in cold adaptation such as anti-freezing proteins (AFPs; e.g. 05934, K03522, K02959, K02386, K01993, K01934, K00658, K00627, K00324) were ubiquitous in all rock typologies and across all sampled areas (Supplementary Figure S8). This suggests the pivotal role of cold adaptation for survival at temperatures below 0°C (Wong et al. 2019; Liao et al. 2021).

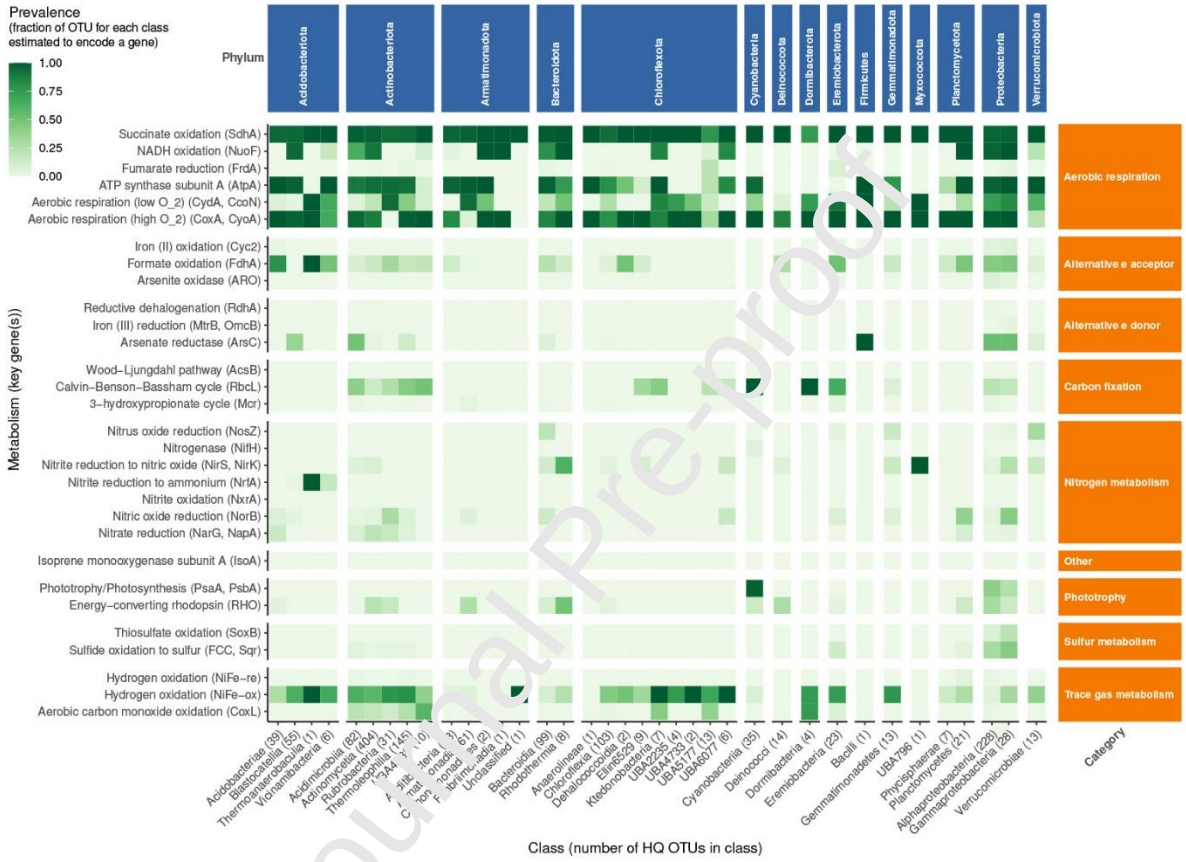


Figure 4. Metabolic potential of the species-level OTUs in Antarctic endolithic communities. The squared green cells represent the proportion of HQ OTUs in each class estimated to encode a particular metabolism. The analysis includes 1503 HQ OTUs partitioned in 37 classes and 15 phyla (blue rectangles), encompassing 30 key metabolisms partitioned in 9 categories (orange rectangles). NiFe-re and NiFe-ox indicates NiFe hydrogenases involved in H_2 production (groups 3 and 4) H_2 oxidation (groups 1 and 2a) respectively.

Conclusions

Our study provides insights on the diversity of endolithic bacterial taxa thriving in the prohibitive

conditions of Antarctica, and further identified survival strategies supporting their endurance at the limit of habitability. This resource represents the largest effort to date to capture the breadth of bacterial genomic diversity from Antarctic rocks. We also unearthed the key and targeted adaptation strategies that allow microbes to spread and perpetuate in the harshest ecosystems. These results represent the foundation to untangle adaptability at the edge of sustainability on Earth and on other dry Earth-like planetary bodies.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Metagenomes raw data are available under the NCBI accession numbers listed in Supplementary Table 9. MAGs and annotations for high-quality MAGs are available at the zenodo repository (DOI: 10.5281/zenodo.7313591).

Acknowledgements

C.C. is supported by the European Commission under the Marie Sklodowska-Curie Grant Agreement No. 702057 (DRYLIFE). C.C. and L.S. wish to thank the Italian National Program for Antarctic Research for funding sampling campaigns and research activities in Italy in the frame of PNRA projects. The Italian Antarctic National Museum (MNA) is kindly acknowledged for financial support to the Mycological Section of the MNA and for providing rock samples used in this study stored in the Culture Collection of Antarctic fungi (MNA-CCFEE), University of Tuscia, Italy. M.D-B. is supported by a project from the Spanish Ministry of Science and Innovation (PID2020-115813RA-I00), and a project of the Fondo Europeo de Desarrollo Regional (FEDER) and the Consejería de Transformación Económica, Industria, Conocimiento y Universidades of the Junta de Andalucía (FEDER Andalucía 2014-2020 Objetivo temático '01 – Refuerzo de la investigación, el desarrollo tecnológico y la innovación') associated with the research project P20_00879 (ANDABIOMA). J.E.S. is a CIFAR fellow in the Fungal Kingdom: Threats and Opportunities program. B.C.F. acknowledges support from the Australian Research Council Discovery Project (DP220103430). Part of this work (proposal

10.46936/10.25585/60000791) was conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231.

Contributions

C.C., D.A., A.E.R., B.C.F., C.D., and L.S., designed the study; C.C. performed DNA extraction and quality check control; D.A., C.C., and A.R. analyzed the data; C.C., D.A., A.R., B.C.F., C.D., and L.S., interpreted the results and wrote the paper with input from all authors. The authors read and approved the final manuscript.

Conflict of interests

The authors declare that they have no competing interests.

References

Albanese, D., Donati, C., 2021. Large-Scale Quality Assessment of Prokaryotic Genomes with Metashot/prok-Quality. *F1000Research*, 10.12688/f1000research.54418.1.

Albanese, D., Coleine, C., Rota-Stabelli, O., Onofri, S., Tringe, S. G., Stajich, J. E., ... & Donati, C. (2021). Pre-Cambrian roots of novel Antarctic cryptoendolithic bacterial lineages. Microbiome, 9(1), 1-15.

Archer, S. D., De los Ríos, A., Lee, K. C., Niederberger, T. S., Cary, S. C., Coyne, K. J., ... & Pointing, S. B. (2017). Endolithic microbial diversity in sandstone and granite from the McMurdo Dry Valleys, Antarctica. Polar Biology, 40, 997-1006.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. BMC bioinformatics, 10, 1-9.

Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Molecular biology and evolution, 38(12), 5825-5829.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 25(15), 1972-1973.

Chen, Y. J., Leung, P. M., Wood, J. L., Bay, S. K., Hugenholtz, P., Kessler, A. J., ... & Greening, C. (2021). Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. The ISME journal, 15(10), 2986-3004.

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to

classify genomes with the Genome Taxonomy Database..

Coleine, C., Gevi, F., Fanelli, G., Onofri, S., Timperio, A. M., & Selbmann, L. (2020). Specific adaptations are selected in opposite sun exposed Antarctic cryptoendolithic communities as revealed by untargeted metabolomics. PLoS One, 15(5), e0233805.

Cowan, D. A., Ferrari, B. C., & McKay, C. P. (2022). Out of thin air? Astrobiology and atmospheric chemotrophy. Astrobiology, 22(2), 225-232.

de la Torre, J. R., Goebel, B. M., Friedmann, E. I., & Pace, N. R. (2003). Microbial diversity of cryptoendolithic communities from the McMurdo Dry Valleys, Antarctica. Applied and environmental microbiology, 69(7), 3858-3867.

Dragone, N. B., Diaz, M. A., Hogg, I. D., Lyons, W. B., Jackson, W. A., Wall, D. H., ... & Fierer, N. (2021). Exploring the boundaries of microbial habitability in soil. Journal of Geophysical Research: Biogeosciences, 126(6), e2020JG006052.

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nature biotechnology, 35(4), 316-319.

Ettinger, C. L., Saunders, M., Selbmann, L., Delgado-Baquerizo, M., Donati, C., Albanese, D., ... & Coleine, C. (2023). Highly diverse and unknown viruses may enhance Antarctic endoliths' adaptability. Microbiome, 11(1), 1-8.

Friedmann, E. I. (1982). Endolithic microorganisms in the Antarctic cold desert. Science, 215(4536), 1045-1053.

Friedmann, E. I., McKay, C. P., & Nienow, J. A. (1987). The cryptoendolithic microbial environment in the Ross Desert of Antarctica: satellite-transmitted continuous nanoclimate data, 1984 to 1986. Polar Biology, 7, 273-287.

Gevi, F., Leo, P., Cassaro, A., Pacelli, C., de Vera, J. P. P., Rabbow, E., ... & Onofri, S. (2022). Metabolomic profile of the fungus Cryomyces antarcticus under simulated Martian and space conditions as support for life-detection missions on Mars. Frontiers in Microbiology, 13, 749396.

Greening, C., & Grinter, R. (2022). Microbial oxidation of atmospheric trace gases. Nature Reviews Microbiology, 20(9), 513-528.

Greening, Chris. 2021. "Greening Lab Metabolic Marker Gene Databases," April. https://doi.org/10.26180/c.5230745.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology, 59(3), 307-321.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., ... & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic acids research, 47(D1), D309-D314.

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics, 11, 1-11.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), 357-359.

Lappan, R., Shelley, G., Islam, Z. F., Leung, P. M., Lockwood, S., Nauer, P. A., ... & Greening, C. (2023). Molecular hydrogen in seawater supports growth of diverse marine bacteria. Nature Microbiology, 8(4), 581-595.

Lee, J. R., Raymond, B., Bracegirdle, T. J., Chadès, I., Fuller, R. A., Shaw, J. D., & Terauds, A. (2017). Climate change drives expansion of Antarctic ice-free habitat. Nature, 547(7661), 49-54.

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research, 44(W1), W242-W245.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. bioinformatics, 25(16), 2078-2079.

Liao, L., Gao, S., Xu, Y., Su, S., Wen, J., Yu, Y., & Chen, B. (2021). Complete genome sequence of Marinomonas arctica BSI20414, a giant antifreeze protein-producing bacterium isolated from Arctic sea ice. Marine Genomics, 57, 100829.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic acids research, 42(D1), D199-D205.

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2:

an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ, 7, e7359.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Ji, M., Greening, C., Vanwonterghem, I., Carere, C. R., Bay, S. K., Steen, J. A., ... & Ferrari, B. C. (2017). Atmospheric trace gases support primary production in Antarctic desert surface soil. Nature, 552(7685), 400-403.

Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., & Giovannelli, D. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. Frontiers in microbiology, 10, 780.

Mirdita, M., Steinegger, M., & Söding, J. (2019). MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics, 35(16), 2856-2858.

Montgomery, K., Williams, T. J., Brettle, M., Bergquist, J. F., Ray, A. E., Zhang, E., ... & Ferrari, B. C. (2021). Persistence and resistance: survival mechanisms of Candidatus Dormibacterota from nutrient-poor Antarctic soils. Environmental Microbiology, 23(8), 4276-4294.

von Haeseler, A., Schmidt, H. A., Bui, M. Q., & Nguyen, L. T. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies..

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome research, 27(5), 824-834.

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. The ISME journal, 11(12), 2864-2868.

Ondov, B. D., Starrett, G. J., Sappington, A., Kostic, A., Koren, S., Buck, C. B., & Phillippy, A. M. (2019). Mash Screen: high-throughput sequence containment estimation for genome discovery. Genome biology, 20, 1-13.

Orakov, A., Fullam, A., Coelho, L. P., Khedkar, S., Szklarczyk, D., Mende, D. R., ... & Bork, P. (2021). GUNC: detection of chimerism and contamination in prokaryotic genomes. Genome biology, 22, 1-19.

Ortiz, M. et al. Multiple energy sources and metabolic strategies sustain microbial diversity in

Antarctic desert soils. Proc. Natl. Acad. Sci. U. S. A. 118, (2021).

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P. A., & Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic acids research, 50(D1), D785-D794.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. "CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes." Genome Research 25 (7): 1043–55.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one, 5(3), e9490.

Ray, A. E., Zaugg, J., Benaud, N., Chelliah, D. S., Bay, S., Wong, H. L., ... & Ferrari, B. C. (2022). Atmospheric chemosynthesis is phylogenetically and geographically widespread and contributes significantly to carbon fixation throughout cold deserts. The ISME Journal, 16(11), 2547-2560.

Rickaby, R. E., & Hubbard, M. E. (2019). Upper ocean oxygenation, evolution of RuBisCO and the Phanerozoic succession of phytoplankton. Free Radical Biology and Medicine, 140, 295-304.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068-2069.

Søndergaard, D., Pedersen, C. N., & Greening, C. (2016). HydDB: a web tool for hydrogenase classification and analysis. Scientific reports, 6(1), 34212.

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., & Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics, 23(10), 1282-1288.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research, 28(1), 33-36.

Wierzchos, J., Ríos, A. D. L., & Ascaso, C. (2012). Microorganisms in desert rocks: the edge of life on Earth.

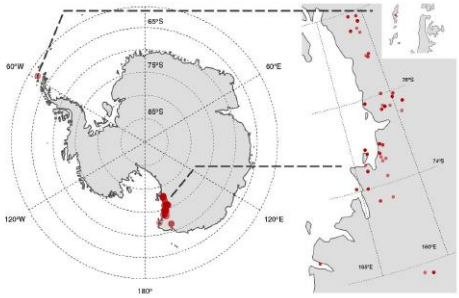
Wong, C. M. V. L., Boo, S. Y., Voo, C. L. Y., Zainuddin, N., & Najimudin, N. (2019). A comparative transcriptomic analysis provides insights into the cold-adaptation mechanisms

of a psychrophilic yeast, Glaciozyma antarctica PI12. Polar Biology, 42, 541-553.

Graphical abstract

Antarctic Rock Genomes Catalog (ARGC)

- 109 endolithic microbiomes
- 4,539 metagenome-assembled genomes (MAGs)
- 49.3% of MAGs novel candidate bacterial species



Trace gas oxidation and **atmospheric chemosynthesis** may be the prevalent strategies supporting persistence of these ecosystems at the fringe of life.

- Untangling adaptability on Earth and on other dry Earth-like planetary bodies.

- Informing us on the fate of microbial life in a warming and drying world.

Highlights

- We unearthed survival strategies of Antarctic endolithic microbes.
- We generated 4,539 metagenome-assembled genomes (MAGs).
- 49.3% of MAGs were novel candidate species.
- Trace gas oxidation and atmospheric chemosynthesis support survival.
- Cold adaptation is pivotal for surviving in the coldest and driest desert on Earth.