Sampling strategies to assess microbial diversity of Antarctic cryptoendolithic communities

Claudia Coleine¹, Jason E. Stajich^{2*}, Nuttapon Pombubpa², Laura Zucconi¹, Silvano Onofri¹, Laura Selbmann^{1,3*}

¹*Department of Ecological and Biological Sciences, University of Tuscia, Viterbo 01100 Italy*

²*Department of Microbiology and Plant Pathology, University of California, Riverside, United States*

³*Italian National Antarctic Museum (MNA), Mycological Section, Genoa, Italy*

*Correspondence: Laura Selbmann, selbmann@unitus.it; Jason E. Stajich, jason.stajich@ucr.edu

Abstract

Describing the total biodiversity of an environmental metacommunity is challenging due to the presence of cryptic and rare species and incompletely described taxonomy. How many samples to collect is a common issue faces ecologists when designing fieldwork sampling: collecting many samples may indeed capture the whole metacommunity structure, but can be prohibitively costly and lead to an enormous amount of data to analyse. Conversely, too few samples may yield inadequate and incomplete data which can prohibit complete assessment of community diversity. High-throughput sequencing allows examination of large numbers of samples enabling comprehensive biodiversity assessments. In this study, we sought to estimate how the scale of sampling affects accuracy of community diversity description in order to develop strategies to exhaustively describe the microbial diversity of cryptoendolithic communities in the McMurdo Dry Valleys in Antarctica accounted as the closest Martian analogue on Earth, exhibiting extreme conditions such as low temperatures, wide thermal fluctuations, low nutrient availability and high UV radiation. We found that sampling effort, based on accumulation curves analysis, had a considerable impact on assessing species richness and composition in these ecosystems, confirming that a sampling as large as nine rock specimens was necessary to detect almost all fungal species present, but was not sufficient to capture whole bacterial assemblage.

Keywords: Antarctica, Cryptoendolithic communities, Bacteria, Fungi, Metabarcoding, Sampling effort.

Introduction

How much sampling is appropriate to assess the microbial diversity? The number of samples to collect is an important consideration for experimental design of microbial community studies. Large numbers of samples may be required to test ecological hypotheses that depend on accurately capturing the entire metacommunity (Bonar *et al.*, 2011), but these requirements may be prohibitively expensive, but too little sampling may yield inadequate data. Therefore, a cost-benefit analysis to maximize data quality and completeness (Gotelli and Colwell, 2001) must be incorporated when designing a field campaign while considering minimal required sampling efforts to achieve the goals (Kasprzak, 1993).

Environmental samples typically contain many rare (low-abundant) taxa or species potentially difficult to identify (Grey *et al.*, 2018). High-throughput sequencing (HTS), particularly environmental metabarcoding (eDNA), have provided new opportunities to analyse large numbers of sample for low costs (Caporaso *et al.*, 2012), leading to a more accurate exploration of biodiversity and community's composition (see Buée *et al.*, 2009; Jumpponen and Jones, 2009; Blaaid *et al.*, 2012).

Accumulation curves are used to assess if representative microbial diversity is fully sampled with an HTS approach. These assessments may be used to test if sufficient sampling efforts have been performed to accurately estimate the total species observed (Gotelli and Colwell, 2001). Theoretically, finding out how many species characterize a microbial assemblage means sampling more and more individuals until no additional species are found, but often the number of individuals that must be sampled to reach this result may often be prohibitively large (Chao *et al.*, 2009). Exhaustive description of biodiversity meaning detection of all organisms in a community is challenging (Ariño *et al.*, 2008): a sampling plan must be feasible and cost effective, answering a series of biological questions and leading to the best results and proper interpretation of the analysis. Development of a clear community research questions is the key to designing a good sampling plan and more necessary in the extremes and in difficult to access ecosystems such as Polar regions.

The McMurdo Dry Valleys in Southern Victoria Land (Continental Antarctica) are considered one of the most extreme environment on Earth; for the different stressors characterizing this area, such as extreme temperature fluctuations (Doran *et al.*, 2002), dryness, oligotrophy and high levels of solar and UV radiation (Wynn-Williams, 1990; Smith *et al.*, 1992), it is accounted as the closest Martian analogue on Earth (Horowitz *et al.*, 1972; McKay, 1993; Wynn-Williams and Edwards, 2000; Onofri *et al.*, 2004; Baqué *et al.*, 2013). In these ice-free areas, the prohibitive factors limit most life forms and microorganisms can only find refuge in cryptic niches, developing within airspaces of porous rocks and forming cryptoendolithic communities (Friedmann and Ocampo, 1976; Friedmann, 1982; 1988); the endolithic lifestyle is considered a stress avoidance strategy (Pointing and Belnap, 2012; Pointing *et al.*, 2014), providing microorganisms a means to survive in extreme aridity.

Despite recent molecular studies have contributed provided new insight into the biodiversity and distribution of Antarctic cryptoendolithic communities and their response to environmental pressure (de La Torre *et al.*, 2003; Wei *et al.*, 2016; Archer *et al.*, 2017; Selbmann *et al.*, 2017; Coleine *et al.*, 2018a, b), the sampling effort necessary for a proper whole metacommunity description remain still unexplored. In this survey, we aimed to determine the minimum sampling effort required to sufficiently estimate the scale of microbial biodiversity and composition of the Antarctic cryptoendolithic communities in McMurdo Dry Valleys, based on fungal and bacterial metabarcoding profiles. Our findings provide recommendations for optimizing future experimental sampling design for surveys of these communities. These sampling suggestions are critical to reduce possible impact in these strictly protected and sensitive environmental areas (ASMA, Antarctic Specially Managed Areas, and ASPA, Antarctic Specially Protected Areas), with very limited access.

The results include an assessment of the variability of community structure in a sampling of as many as nine rock samples collected from a single locality of the McMurdo Dry Valleys and an estimate of minimum sampling efforts required to adequately characterize fungal and bacterial diversity.

Materials and Methods

Sampling area

Nine homogenous colonized by lichen-dominated cryptoendolithic communities sandstones were collected by L. Selbmann and L. Zucconi during the XXVI Italian Antarctic Expedition (December 2010-January 2011) at Battleship Promontory 76°54'36''S 160°56'05''E (McMurdo Dry Valleys, Southern Victoria Land, Antarctica) (Fig.1) at 1000 m elevation and 33.5 km distance from sea within an area of about 100 m². The presence of endolithic colonization was assessed by direct observation *in situ*. Rocks were excised using a geological hammer placed in sterile bags and stored at -20°C until processing to extract DNA and perform molecular analysis at University of Tuscia, Italy.

DNA extraction and metabarcoding sequencing

Sandstones were crushed under sterile conditions to collect 1 g of powdered rock used for DNA extraction with MOBIO Power Soil DNA Extraction kit (MOBIO Laboratories, Carlsbad, CA, USA) following the manufacturer's instructions. The Ribosomal Internal Transcribed Sequence 1 region (ITS1) and V4 region of the 16S rRNA gene were targeted to assess the fungal and bacterial community membership, respectively. We amplified the ITS1 region using barcoded primers ITS1F/ITS2, developed for shorter read length (Smith and Peay, 2014) and V4 region using the new developed barcoded F515/R806 primer set as described by Caporaso and colleagues (2012) (Table 1). PCR was carried out with a total volume of 25 µl, containing 1 µl of each primer, 12.5 µl of Taq DNA Polymerase (Thermo Fischer Scientific Inc., Waltham, MA, USA), 9.5 µl of nuclease-free water (Sigma-Aldrich, UK) and 5 ng of DNA template using an automated thermal cycler (BioRad, Hercules, CA). The ITS1 locus was amplified following initial denaturation at 94°C for 1 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s, extension at 68°C for 90 s, followed by a final extension at 68°C for 7 min. The PCR for the V4 region followed a protocol of an initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 45 s, annealing at 50°C for 1 min, extension at 72°C for 90 s, followed by a final extension at 72°C for 10 min. Amplicons, quantified by Qubit dsDNA HS Assay Kit (Life Technologies, USA), were pooled and purified with Qiagen PCR CleanUp kit (Macherey-Nagel, Hoerd, France). Paired-end sequencing (2×300 bp) was carried out on Illumina MiSeq platform at the Institute for Integrative Genome Biology, University of California, Riverside, California.

DNA extraction was performed in duplicate and replicates were sequenced separately.

Processing of metabarcoding data

The ITS1 and 16S datasets were processed separately with the AMPtk (Palmer et al., 2018; <https://github.com/nextgenusfs/ampatk>) tool. Briefly, overlapping 2×300 base pair MiSeq reads were merged using USEARCH v8 (Edgar and Flyvbjerg, 2015). Barcodes and primers were removed from the amplicons sequencing data and reads were demultiplexed using the script `split_libraries.py` in QIIME v1.9.1 (Caporaso et al., 2010). Reads were then subjected to quality trimming based on accumulation of expected errors less than 1.0 (Edgar and Flyvbjerg, 2015) and chimeric reads were removed in AMPtk as implemented by USEARCH (Edgar, 2010). A 97% sequence similarity clustering threshold was utilized to create Operational Taxonomic Units (OTUs) using VSEARCH (v 2.3.2) (Rognes et al., 2016) tool. Taxonomy assignment was performed with SINTAX/UTAX (Edgar, 2010) (Supplementary Materials Tables S2, S3)

Singletons (OTUs represented by 1 reads in whole dataset) and rare taxa (<5 reads) were discarded as likely false positives due to sequencing errors, as recommended by Lindahl et al. (2013).

Reads obtained from duplicates were merged. Fungal and bacterial rarefaction curves were calculated using the 'rarecurve' function in the package 'vegan' (v. 2.3-4; Oksanen et al., 2013) in R 3.2.0 (R-Development-Core-Team 2015).

Raw sequence data were submitted to the GenBank database under BioProject accession number PRJNA379160.

Biodiversity and NMDS ordination plots

Alpha diversity from the OTU tables was estimated using Primer-E v7 (Auckland, New Zealand) software, calculating biodiversity indices as follows: (i) the species richness (S) as a count of the total number of OTUs found in each sample; (ii) the Shannon index (H'), a phylotype-based approach constructed using OTU groupings (Shannon and Weaver, 1949; Ludwig, 1988); (iii) the Simpson's Index of Dominance ($1-D$), measuring the probability that two individuals randomly selected from a sample will belong to the same species (Simpson, 1949). Statistical analysis was performed by one-way analysis of variance (one-way ANOVA) and pairwise multiple comparison procedure (Tukey test) carried out using the statistical software SigmaStat 2.0 (Jandel, USA).

Changes in fungal and bacterial assemblages' composition were evaluated with non-metric multidimensional scaling (NMDS) analysis, performed with PAST (v.2.17) (PAleontological Statistics) (Hammer et al., 2001) and their confidence intervals were investigated through permutational PERMANOVA (Anderson et al., 2013). We used Jaccard (presence-absence) and Bray-Curtis (relative abundance) metrics to calculate pairwise community distance matrices and examine differences in beta diversity. Abundance data were square-root transformed to limit the influence of OTUs with high sequence counts and analyses were carried out with 999 permutations.

A small probability p-value (<0.05) indicated a significant difference of community composition among all samples.

Accumulation curves and extrapolation analysis

Species accumulation curves analysis was implemented to compare microbial assemblages and to estimate the expected number of new species given additional sampling effort, which can give insights to proper sampling protocols (Colwell and Coddington, 1994; Moreno and Halpern, 2000).

Curves were calculated using PRIMER-E v7 implementing Plots>Species Accum Plot function. The increasing total number of different species observed (S) was plotted against samples successively pooled (referred to as the Sobs curve; Sobs: total number of species observed in all samples pooled). We used the 'permute' option to enter samples utilizing randomization procedures (Colwell and Coddington, 1994; Gotelli and Entsminger, 2001). When the curve reaches an asymptote, and the number of OTUs does not increase by adding further samples, it indicates sufficient samples have been collected to accurately characterize the community (Pielou, 1977).

We also computed an extrapolation model (sample-based extrapolation curve) with 95% unconditional confidence intervals, using the analytical formulas of Colwell et al. (2012). The model attempts to predict the asymptotic number of species that would be found for an increasing number of samples, estimating the number of samples needed to reach a plateau, by using EstimateS v9 (Statistical Estimation of Species Richness and Shared Species from Samples) (Colwell et al., 2012).

For all tests, sample-based incidence data were used and analysis carried out with 999 permutations.

Quantitative PCR

Quantification of fungal and bacterial rRNA gene copies was estimated by using quantitative PCR assay. To estimate fungal and bacterial gene abundances, standard curves were generated using a 10-fold serial dilution of a plasmid containing a copy of *Cryomyces antarcticus* ITS rRNA gene or *Escherichia coli* 16S rRNA gene. Amplicons were generated in 100 μ l reactions, containing 50 μ l of 2x PCR BioMix™ (Bioline,

London, UK), 1 μ l (5 pmol) of both forward and reverse primers, 43 μ l of DEPC water and 5 μ l template of genomic DNA. For fungi, the primers used were NS91F (5'-GTCCCTGCCCTTTGTACACAC-3') and ITS51R (5'-ACCTTGTTACGACTTTTACTTCCTC-3'), while Eub338 (5'-ACTCCTACGGGAGGCAGCAG-3') Eub518 (5'-ATTACCGCGGCTGCTGG-3') primer set was utilized for bacterial component (Fierer et al., 2005). Amplification of ITS region was carried out as follows: 95 °C for 3 min and the 45 cycles of 95 °C 40 s, 55 °C 30 s and final extension at 72 °C 40 s; while 16S region was amplified as follows: 95 °C for 3 min and the 45 cycles of 95 °C 40 s, 53 °C 30 s and final extension at 72 °C 40 s. PCR products were purified using the NucleoSpin® PCR Clean-up kit (MACHEREY-NAGEL, GmbH & Co. KG), quantified with Qubit dsDNA HS Assay kit and then cloned using the pGEM®-T Easy Vector Systems (Promega, Madison, Wisconsin, US). Plasmids were isolated using the NucleoSpin Plasmid kit (Macherey-Nagel, GmbH & Co. KG). Five standards were utilized for qPCR in series from 10^3 to 10^7 copies.

The 25 μ l qPCR reactions contained 12.5 μ l iQ™ SYBR® Green Supermix (Bio-Rad, Hercules, California, US), 1 μ l of each pM forward and reverse primers, 0.3 ng of environmental DNA or standard and 9.5 μ l nuclease-free water. The reactions were carried out on a Quantitative real-time BioRad CFX96™ PCR detection system (Bio-Rad, Hercules, California, US). Primers were the same used to generate standard curves and the PCR protocols were those mentioned above.

Melting curve and gel electrophoresis analyses confirmed that the amplified products were of the appropriate size. Fungal and bacterial gene copy numbers were generated using a regression equation for each assay relating the cycle threshold (C_t) value to the known number of copies in the standards.

All qPCR reactions were run in triplicate.

Fungal-to bacterial (F/B) ratio was also calculated based on log-transformed abundance values. Means and standard deviations were calculated and statistical analysis were performed using one-way analysis of variance (ANOVA) and pairwise multiple comparison procedures, carried out using the statistical software SigmaStat v2.0 (Jandel, USA). Significant differences were calculated by Tukey test with p value < 0.05.

Results

Metabarcoding analysis

Fungal and bacterial reads were retrieved from all samples. For ITS 558,579 reads passed quality filtering and the per-sample number of rRNA gene reads averaged 62,064 with a minimum of 36,246 to a maximum of 86,668 reads. A total of 624,195 16S sequences reads were generated across the samples with an average of 69,355 reads, minimum of 49,597 and maximum 111,307 reads per sample. After removal of global singletons and rare taxa (<5 reads) (56 out of 245 fungal OTUs and 198 out of 715 bacterial OTUs).

Sequence reads were grouped, using a 97% identity threshold, into 189 fungal and 517 bacterial OTUs, respectively; the number of reads for each sample is reported in Table 2. Rarefaction curves were constructed and results are reported in Figs. 1S, 2S.

Biodiversity and community composition

Species richness (S), Shannon (H') and Simpson (1-D) indices were calculated for each rock sample. The observed fungal richness recorded ranged from 153 to 101 OTUs. In contrast, bacteria exhibited a higher diversity richness, ranging from 160 to 470 OTUs. Globally, the biodiversity, estimates from the 16S and ITS sequences, indicated that H' ranged from 1.58 to 2.63 in fungi and 1.74 to 3.70 bacteria; 1-D estimate of diversity ranged from 0.72 to 0.98 in ITS sequences and from 0.68 to 0.87 in 16S sequences (Table 3). Pairwise comparisons of biodiversity measures using Tukey's test reported that, in generally, biodiversity varied significantly amongst samples and even, between fungi and bacteria (p < 0.05, data not shown).

A non-metric multidimensional scaling (NMDS) ordination was computed to investigate changes in communities composition using both incidence data (Jaccard dissimilarity index) and the read-abundance

data (Bray–Curtis metrics), separately, to avoid the uncertainty whether read abundance was a good indicator of OTU abundance in the samples (Amend *et al.*, 2010). Since both approaches produced similar outputs, results based on abundance data are reported only. NMDS plots resulted in two-dimensional solutions with stress values of 0.095 (fungi) and 0.076 (bacteria). Fungal assemblages had more similarity ($p < 0.05$) among all samples than did the bacteria, where only few samples were similar ($p > 0.05$) (Figure 2).

Estimating and determining sampling effort

The species accumulation curve is a graph of the expected number of detected species as a function of sampling effort, representing the sampling process (Palmer, 1990). Fungal randomized accumulation curve reached saturation with our sampling effort and the asymptote indicated that 100% of the species present were detected when nine samples are included (Figure 3A). In contrast, the total bacterial diversity was much higher and accumulation curves showed a continued increase in species per sample, indicating that biodiversity of the entire resident community was not captured and sampling was not exhaustive (Figure 3B). An extrapolation analysis of empirical sample-based rarefaction curves estimated the number of bacterial species expected to be found in a larger number of samples from the same assemblage. Figure 4 showed that, increasing the number of samples, the sampling would require approximately four times more effort (≈ 40 samples) for a comprehensive description of bacterial diversity.

Fungal and bacterial abundance

The abundance of fungi and bacteria, as determined using qPCR, varied significantly ($p < 0.05$) along whole sampling ranging from 5.3×10^5 to 2.8×10^6 fungal and from 5.8×10^4 to 1.3×10^5 bacterial gene copies, respectively (see Fig. 5 and Table 1S). Particularly, the lowest fungal abundance was estimated in sample 3 and 5 (527,384 and 822,113 copies, respectively), while the highest in sample 6 (2,784,953 copies); in bacterial compartment, 8 and 9 showed the lowest abundance (55,679 and 57,923 copies, respectively), while the highest value (128,796) was retrieved in sample 3 (Table 1S). Additionally, fungal and bacterial abundances significantly varied ($p < 0.05$) along whole sampling; this resulted more remarkable in fungi, where almost all samples are different, while bacterial abundances were more homogenous. Fungal-to bacterial ratio (F/B) (based on log copy numbers) was also calculated from these qPCR results. These findings showed a slightly higher dominance of fungi in these ecosystems varying between 1.25 ± 0.06 (mean $\pm$ SD).

Discussion

The McMurdo Dry Valleys in Southern Victoria Land, is one of most extreme ice-free areas known on Earth. The extreme aridity, temperature swings, and UV radiation exposure have been used to classify it as the closest terrestrial analogue of Martian environment on Earth. In this region, where climatic conditions are too harsh to allow colonization on the surface of rocks, endolithic communities are almost the only life-form possible. Rock niches provides buffered conditions for microorganisms to find a refuge for survival (Friedmann and Ocampo-Friedmann, 1976; Nienow and Friedmann, 1993).

Recent molecular advances have provided new insights into microbial diversity and community composition in Antarctic cryptoendolithic communities. These ecosystems harbour a very low diversity of both eukaryotic and prokaryotic assemblages, dominated by a restricted number of highly specialized species (Archer *et al.*, 2017; Selbmann *et al.* 2017; Coleine *et al.*, 2018 a, b). To better define the minimum sampling effort needed to capture the community diversity and to inform and optimize future sampling protocols we evaluated how well our sampling recovered whole metacommunity composition from Antarctic cryptoendolithic communities.

The high degree of specialization shaped by the presence of very few predominant species and rare taxa as determined by the Shannon's and Simpson's dominance indices (0.83 ± 0.08 and 0.77 ± 0.07 in fungi

and bacteria, respectively), suggested that these microbial communities are highly specialized, adapted and resistant, but they may be dramatically influenced by any disturbance events which may irreparably affect their fragile equilibrium and future in the face of global change (Selbmann et al., 2017; Coleine et al., 2018 a, b). Analysis of these communities and correlations to abiotic and biotic parameters will help determine and predict the responses of ecosystem processes to climate change.

Overall, a considerable amount of heterogeneity has been observed between rock samples, both for fungal and bacterial assemblages; the high degree of variability may be related to the breadth of the sampling area (100 m²). A more homogeneous dataset may be observed from a selection of rocks collected in strict proximity to each other. A remarkable variability in biodiversity and structure in endolithic communities collected from the same location was reported by Selbmann and colleagues (2017) or in sandstones collected in localities with similar environmental parameters (altitude and sea distance) as recently reported by Coleine et al. (2018a). Specifically, in our survey, richness and Shannon's biodiversity index values significantly varied among all samples: number of observed species ranged from 93 to 153 and 160 to 470 in fungi and bacteria, respectively, while biodiversity values from 1.58 to 2.63 and 1.74 to 3.70 in fungi and bacteria, respectively. The higher values of bacteria are consistent with a reconstruction based on a tree of life that found bacteria generally more diverse than fungi (Hinchliff et al., 2015); richness in species in individual communities varied according to the habitat, but it seems to be definitively higher for bacteria than fungi. For example, in soil communities, bacterial species are generally 2–3 times more numerous than fungi (Urbanová et al., 2015).

Changes in community's composition was reported both in fungi and bacteria, as suggested by NMDS ordinations, even in samples with comparable species richness. Moreover, community composition was considerably more different among observations of bacteria in the samples. A PERMANOVA analysis showed that community structure varied in almost all samples ($p > 0.05$) and these data were consistent when analyzed as presence-absence or abundance matrices.

Substantial heterogeneity was even observed considering fungal and bacterial relative abundances estimated with qPCR assay. Abundance data was significantly variable along the sampled area (values ranged 5.3×10^5 to 2.8×10^6 for fungi and from 5.8×10^4 to 1.3×10^5 for bacteria). The fungal/bacterial (F/B) ratio, a common metric to assess environmental impacts and the functional implications of microbial communities (Raeymaekers, 2000; Fierer et al., 2005), was also assayed in order to compare community makeup. We found a weak fungal dominance (F/B: 1.25); this may be explained by the ability of fungi to grow at lower water activity than prokaryotes, the accuracy of F/B ratio from Antarctic endolithic communities remains unexplored; future research into the microbial members of these environments is needed, especially gathering a basic understanding of the ecology of bacteria and fungi in these ecosystems in order to better assess the mechanism by which a change in F/B dominance might occur. For example, in acidic soils the fungi are predominant biomass linked to the observation that fungi are weakly influenced by pH values (Rousk et al., 2010), while bacterial biomass is highly affected by changes in pH (Griffiths et al., 2011). Soil nutrient availability and moisture may also affect the F/B ratio (Strickland and Rousk, 2010).

Based on these observations, properly describing the complete microbial biodiversity in Antarctic cryptoendolithic communities remains a challenge. The high variability in microbial assemblages in the same sampled area suggests that there are highly localized patterns in individual rocks. Species accumulation curves indicated that fungal diversity in the sampled area may be sufficiently captured with the current sampling depth and design (nine sandstone samples). The number of species found exhibited a cumulative pattern of diminishing increase with increasing the samples number. Conversely, we captured only a subset of the total estimated bacterial diversity based on the lack asymptote of the accumulation curves. We estimate a total of 40 samples from nine rocks is needed to reach a plateau in the observations of bacteria community diversity.

This work emphasizes the importance of sampling efforts on precision and provides information to determine sampling efficiency in extreme areas like the Antarctic where extensive sampling can be prohibitive. Based on these results, we determined that the prior sampling effort had been able to robustly

assess species richness and community composition in Antarctic cryptoendolithic ecosystems, especially for the fungal component. Recommendations for future eDNA metabarcoding surveys and research would need to take into account these expected scale of local biodiversity within rocks in order to fully assess cryptoendolithic community diversity. Since a large number of additional undetected bacterial species are likely to be present in the sampled area, for future studies we suggest a pooling strategy, prior to DNA extraction, to minimize local variability. The pooling method in environmental metabarcoding studies may be particularly useful for among-site comparison of soil and rock communities and it is generally utilized to reduce sample numbers for evaluations of microbial community structure (Baker et al., 2009; Osborne et al., 2011). Typically pooling of eDNA is not recommended since DNA of rare species could be diluted and important portion of microbial composition may be under sampled, especially if communities are extremely complex (Manter et al., 2010). These approaches can be considered due to the very low biodiversity and complexity of endolithic ecosystems inhabiting ice-free area of Victoria Land in Antarctica (Selbmann et al., 2017; Coleine et al., 2018a).

This pilot study sought to define optimal sampling efforts for future studies and further investigations will refine hypotheses regarding local distribution patterns of microbes. The current sampling of areas of Southern Victoria Land can be contrasted to environmentally different Northern Victoria Land. This examination of diversity using eDNA methods gives insights on optimizing sampling protocols to increase the sampling efficiency and decrease an unnecessary number of samples which can reduce the impact on delicate ecosystems and strictly protected areas.

Compliance with Ethical Standards

This article does not contain any studies with human participants or animals performed by any of the authors. Conflict of Interest: The authors declare that they have no conflict of interest. The datasets generated during the current study are available in the GenBank repository under BioProject accession number PRJNA379160.

ACKNOWLEDGMENTS

L.S., C.C. and L.Z. wish to thank the Italian National Program for Antarctic Researches (PNRA) for funding sampling campaigns and researches activities in Italy in the frame of 2009/A1.11, 2013/AZ-17, 2015/AZ1.02 and AMunDsEN PNRA_00006 projects. The Italian Antarctic National Museum (MNA) is acknowledged for financial support to the Mycological Section of the MNA for preserving rock Antarctic samples analysed in this study and stored in the Culture Collection of Fungi from Extreme Environments (CCFEE), University of Tuscia, Italy. Fungal ITS and bacterial 16S primers were made available through the Alfred P. Sloan Foundation Built Environment Program and sequencing supported by funds through United States Department of Agriculture - National Institute of Food and Agriculture Hatch project CA-R-PPA-5062-H to JES. N.P. was supported by a Royal Thai government fellowship. Data analyses were performed on the High-Performance Computing Cluster at the University of California-Riverside in the Institute of Integrative Genome Biology supported by NSF DBI-1429826 and NIH S10-OD016290.

References

- Amend AS, Seifert KA, Bruns TD (2009) Quantifying microbial communities with 454 pyrosequencing: Does read abundance count? *Mol Ecol* 19:5555-5565. [10.1111/j.1365-294X.2010.04898.x](https://doi.org/10.1111/j.1365-294X.2010.04898.x)
- Anderson MJ, Walsh DCI (2013) PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol Monogr* 83:557-574. [10.1890/12-2010.1](https://doi.org/10.1890/12-2010.1)
- Archer SD, de los Ríos A, Lee KC, Niederberger TS, Cary SC, Coyne KJ et al. (2017) Endolithic microbial diversity in sandstone and granite from the McMurdo Dry Valleys, Antarctica. *Polar Biol* 40(5):997-1006. [10.1007/s00300-016-2024-9](https://doi.org/10.1007/s00300-016-2024-9)
- Ariño AH, Belascoáin C, Jordana R (2008) Optimal sampling for complexity in soil ecosystems. In *Unifying themes in complex systems IV* (pp. 222-230). Springer, Berlin, Heidelberg. [10.1007/978-3-540-73849-7_25](https://doi.org/10.1007/978-3-540-73849-7_25)
- Baker KL, Langenheder S, Nicol GW, Ricketts D, Killham K, Campbell CD, Prosser J I (2009) Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies. *Soil Biol Biochem* 41:2292-2298. [10.1016/j.soilbio.2009.08.010](https://doi.org/10.1016/j.soilbio.2009.08.010)
- Baqué M, de Vera JP, Rettberg P, Billi D (2013) The BOSS and BIOMEX space experiments on the EXPOSE-R2 mission: Endurance of the desert cyanobacterium *Chroococcidiopsis* under simulated space vacuum, Martian atmosphere, UVC radiation and temperature extremes. *Acta Astronaut* 91:180-186. [10.1016/j.actaastro.2013.05.015](https://doi.org/10.1016/j.actaastro.2013.05.015)
- Blaalid R, Carlsen TOR, Kumar S, Halvorsen R, Ugland KI, Fontana G, Kauserud H (2012) Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. *Mol Ecol* 21(8):1897-1908. [10.1111/j.1365-294X.2011.05214.x](https://doi.org/10.1111/j.1365-294X.2011.05214.x)
- Bonar SA, Fehmi JS, Mercado-Silva N (2011) An overview of sampling issues in species diversity and abundance surveys. *Biological diversity: frontiers in measurement and assessment*: Oxford University Press Oxford 11-24.
- Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S, Martin F (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist* 184(2):449-456. [10.1111/j.1469-8137.2009.03003.x](https://doi.org/10.1111/j.1469-8137.2009.03003.x)
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335-336. [10.1038/nmeth.f.303](https://doi.org/10.1038/nmeth.f.303)
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6:1621-1624. [10.1038/ismej.2012.8](https://doi.org/10.1038/ismej.2012.8)
- Chao A, Colwell RK, Lin CW, Gotelli NJ (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90(4):1125-1133.
- Coleine C, Stajich JE, Zucconi L, Onofri S, Pombubpa N, Egidi E, Franks A, Buzzini P, Selbmann L (2018a) Antarctic cryptoendolithic fungal communities are highly adapted and dominated by Lecanoromycetes and Dothideomycetes. *Front Microbiol* 9(1392). [10.3389/fmicb.2018.01392](https://doi.org/10.3389/fmicb.2018.01392)
- Coleine C, Zucconi L, Onofri S, Pombubpa N, Stajich JE, Selbmann L (2018b) Sun exposure shapes functional grouping of fungi in cryptoendolithic Antarctic communities. *Life* 8(2):19. [10.3390/life8020019](https://doi.org/10.3390/life8020019)
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Phil Trans R Soc Lond B* 345(1311):101-118. [10.1098/rstb.1994.0091](https://doi.org/10.1098/rstb.1994.0091)
- Colwell RK, Chao A, Gotelli NJ, Lin SY, Mao CX, Chazdon RL, Longino JT (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation, and comparison of assemblages. *J Plant Ecol* 5:3-21. [10.1093/jpe/rtt044](https://doi.org/10.1093/jpe/rtt044)

- 402 de la Torre JR, Goebel BM, Friedmann E, Pace NR (2003) Microbial diversity of cryptoendolithic communities from
403 the McMurdo Dry Valleys. Antarctica. Appl Environ Microbiol 69:3858-3867. [10.1128/AEM.69.7.3858-3867.2003](https://doi.org/10.1128/AEM.69.7.3858-3867.2003)
- 404 Doran, PT, McKay CP, Clow GD, Dana GL, Fountain AG, Nylen T, Lyons WB (2002) Valley floor climate
405 observations from the McMurdo Dry Valleys, Antarctica, 1986–2000. Journal of Geophysical Research: Atmospheres
406 107(D24) ACL-13. [10.1029/2001JD002045](https://doi.org/10.1029/2001JD002045)
- 407
- 408 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.
409 [10.1093/bioinformatics/btq461](https://doi.org/10.1093/bioinformatics/btq461)
- 410 Edgar RC, Flyvbjerg H (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads.
411 Bioinformatics 31(21):3476-3482. [10.1093/bioinformatics/btv401](https://doi.org/10.1093/bioinformatics/btv401)
- 412 Fierer N, Jackson JA, Vilgalys R, Jackson RB (2005) Assessment of soil microbial community structure by use of
413 taxon-specific quantitative PCR assays. Appl Environ Microbiol 71(7):4117-4120. [10.1128/AEM.71.7.4117-4120.2005](https://doi.org/10.1128/AEM.71.7.4117-4120.2005)
- 414 Friedmann EI, Ocampo R (1976) Endolithic blue-green algae in dry valleys-primary producers in Antarctic desert
415 ecosystem. Science 193:1247-1249. [10.1126/science.193.4259.1247](https://doi.org/10.1126/science.193.4259.1247)
- 416
- 417 Friedmann EI (1982) Endolithic microorganisms in the Antarctic cold desert. Science 215:1045-1053.
418 [10.1126/science.215.4536.1045](https://doi.org/10.1126/science.215.4536.1045)
- 419 Friedmann EI, Hua M, Ocampo-Friedmann R (1988) Cryptoendolithic lichen and cyanobacterial communities of the
420 Ross Desert, Antarctica. Polarforschung 58:251-259.
- 421 Gotelli NJ, Entsminger GL (2001) Swap and fill algorithms in null model analysis: rethinking the knight's tour.
422 Oecologia 129(2):281-291. [10.1007/s004420100717](https://doi.org/10.1007/s004420100717)
- 423 Grey EK, Bernatchez L, Cassey P, Deiner K, Deveney M, Howland KL et al. (2018) Effects of sampling effort on
424 biodiversity patterns estimated from environmental DNA metabarcoding surveys. Sci Rep 8(1):8843. [10.1038/s41598-018-27048-2](https://doi.org/10.1038/s41598-018-27048-2)
- 425
- 426 Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS (2011) The bacterial biogeography of British soils.
427 Environ Microbiol 13(6):1642-1654. [10.1111/j.1462-2920.2011.02480.x](https://doi.org/10.1111/j.1462-2920.2011.02480.x)
- 428 Hammer O, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data
429 analysis. Palaeontol Electron 4:9.
- 430 Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM et al. (2015) Synthesis of phylogeny and
431 taxonomy into a comprehensive tree of life. Proc Natl Acad Sci 112(41):12764-12769. [10.1073/pnas.1423041112](https://doi.org/10.1073/pnas.1423041112)
- 432 Horowitz NH, Cameron RE, Hubbard JS (1972) Microbiology of the Dry Valleys of Antarctica. Science
433 176(4032):242-245. [10.1126/science.176.4032.242](https://doi.org/10.1126/science.176.4032.242)
- 434 Jumpponen A, Jones KL (2009) Massively parallel 454 sequencing indicates hyperdiverse fungal communities in
435 temperate *Quercus macrocarpa* phyllosphere. New Phytologist 184(2):438-448. [10.1111/j.1469-8137.2009.02990.x](https://doi.org/10.1111/j.1469-8137.2009.02990.x)
- 436 Kasprzak K (1993) Selected aspects of mathematical analysis, pp. 16-69 In: Gomy M and L Griim (eds.). Methods in
437 soil zoology. Elsevier, New York, Warszawa.
- 438 Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjoller R et al. (2013) Fungal community analysis by
439 high-throughput sequencing of amplified markers—a user's guide. New Phytologist 199(1):288-299.
440 [10.1111/nph.12243](https://doi.org/10.1111/nph.12243)
- 441 Ludwig JA (1988) Statistical Ecology: A Primer in Methods and Computing. New York, NY: John Wiley & Sons.

442 Manter DK, Weir TL, Vivanco JM (2010) Negative effects of sample pooling on PCR-based estimates of soil microbial
443 richness and community structure. *Appl Environ Microbiol* 76:2086-2090. [10.1128/AEM.03017-09](https://doi.org/10.1128/AEM.03017-09)

444 McKay CP, Nienow JA, Meyer MA, Friedmann EI (1993) Continuous nanoclimate data (1985-1988) from the Ross
445 Desert (McMurdo Dry Valleys) cryptoendolithic microbial ecosystem. In *Antarctic Meteorology and Climatology:
446 Studies Based on Automatic Weather Stations*. Bromwich, D.H., and Stearns C.R. (eds). Washington D.C.: American
447 Geophysical Union 201-207. [10.1029/AR061p0201](https://doi.org/10.1029/AR061p0201)

448 Moreno CE, Halfpenny G (2000) Assessing the completeness of bat biodiversity inventories using species accumulation
449 curves. *J Appl Ecol* 37(1):149-158. [10.1046/j.1365-2664.2000.00483.x](https://doi.org/10.1046/j.1365-2664.2000.00483.x)

450 Nienow JA, Friedmann EI (1993) Terrestrial lithophytic (rock) communities. In: Friedmann EI (ed) *Antarctic
451 Microbiology*. Wiley-Liss. New York. 343-412. [10.1007/s00248-008-9483-7](https://doi.org/10.1007/s00248-008-9483-7)

452 Onofri S, Selbmann L, Zucconi L, Pagano S (2004) Antarctic microfungi as models for exobiology. *Planet Space Sci*
453 52:229-237. [10.1016/j.pss.2003.08.019](https://doi.org/10.1016/j.pss.2003.08.019)

454 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB et al. (2013) Package 'vegan'. *Community
455 ecology package v2* (9).

456 Osborne CA, Zwart AB, Broadhurst LM, Young AG, Richardson AE (2011) The influence of sampling strategies and
457 spatial variation on the detected soil bacterial communities under three different land-use types. *FEMS Microbiol Ecol*
458 78:70-79. [10.1111/j.1574-6941.2011.01105.x](https://doi.org/10.1111/j.1574-6941.2011.01105.x)

459 Palmer MW (1990) The estimation of species richness by extrapolation. *Ecology* 71(3):1195-1198.

460 Palmer JM, Jusino MA, Banik MT, Lindner DL (2018) Non-biological synthetic spike-in controls and the AMPtk
461 software pipeline improve mycobiome data. *PeerJ* 6:e4925. [10.7717/peerj.4925](https://doi.org/10.7717/peerj.4925)

462 Pielou EC (1977) *Mathematical ecology* (No. 574.50151) Wiley.

463

464 Pointing SB, Belnap J (2014) Disturbance to desert soil ecosystems contributes to dust-mediated impacts at regional
465 scales. *Biodiversity Conserv* 23(7):1659-1667. [10.1007/s10531-014-0690-x](https://doi.org/10.1007/s10531-014-0690-x)

466 Pointing SB, Bollard-Breen B, Gillman LN (2014) Diverse cryptic refuges for life during glaciation. *Proc Natl Acad Sci*
467 111(15):5452-5453. [10.1073/pnas.1403594111](https://doi.org/10.1073/pnas.1403594111)

468 Raeymaekers L (2000) Basic principles of quantitative PCR. *Mol Biotechnol* 15(2):115-122. [10.1385/MB:15:2:115](https://doi.org/10.1385/MB:15:2:115)

469 Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics.
470 *Peer J* 4:e2584. [10.7717/peerj.2584](https://doi.org/10.7717/peerj.2584)

471 Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, et al. (2010) Soil bacterial and fungal
472 communities across a pH gradient in an arable soil. *ISME J* 4(10):1340. [10.1038/ismej.2010.58](https://doi.org/10.1038/ismej.2010.58)

473 Selbmann L, Onofri S, Coleine C, Buzzini P, Canini F, Zucconi L (2017) Effect of environmental parameters on
474 biodiversity of the fungal component in the lithic Antarctic communities. *Extremophiles* 21(6):1069-1080.
475 [10.1007/s00792-017-0967-6](https://doi.org/10.1007/s00792-017-0967-6)

476 Shannon CE, Weaver W (1949) *The Mathematical Theory of Communication*. New York, NY: John Wiley and Sons.
477 [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x)

478 Simpson EH (1949) Measurement of diversity. *Nature* 163:688. [10.1038/163688a0](https://doi.org/10.1038/163688a0)

479 Smith RC, Prezelin BB, Baker KEA, Bidigare RR, Boucher NP, Coley T et al. (1992) Ozone depletion: ultraviolet
480 radiation and phytoplankton biology in Antarctic waters. *Science* 255(5047):952-959. [10.1007/BF00014589](https://doi.org/10.1007/BF00014589)

481 Smith DP, Peay KG (2014) Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next
482 Generation DNA Sequencing. PLoS ONE 9:e90234. [10.1371/journal.pone.0090234](https://doi.org/10.1371/journal.pone.0090234)

483 Strickland MS, Rousk J (2010) Considering fungal: bacterial dominance in soils—methods, controls, and ecosystem
484 implications. Soil Biol Biochem 42(9):1385-1395. [10.1016/j.soilbio.2010.05.007](https://doi.org/10.1016/j.soilbio.2010.05.007)

485 Urbanová M, Šnajdr J, Baldrian P (2015) Composition of fungal and bacterial communities in forest litter and soil is
486 largely determined by dominant trees. Soil Biol Biochem 84:53-64. [10.1016/j.soilbio.2015.02.011](https://doi.org/10.1016/j.soilbio.2015.02.011)

487 Wei ST, Lacap-Bugler DC, Lau MC, Caruso T, Rao S, de Los Rios A et al. (2016) Taxonomic and functional diversity
488 of soil and hypolithic microbial communities in Miers Valley, McMurdo Dry Valleys, Antarctica. Front Microbiol
489 7:1642. [10.3389/fmicb.2016.01642](https://doi.org/10.3389/fmicb.2016.01642)

490 Wynn-Williams, D.D. (1990) Ecological aspects of Antarctic microbiology. In: Marshall, K.C. (ed) Adv Microb Ecol
491 71-146. Springer, US.

492 Wynn-Williams DD, Edwards HGM (2000) Proximal analysis of regolith habitats and protective biomolecules in situ
493 by laser Raman spectroscopy: overview of terrestrial Antarctic habitats and Mars analogs. Icarus 144(2): 486-503.
494 [10.1006/icar.1999.6307](https://doi.org/10.1006/icar.1999.6307)
495

Tables and figure legends

Table 1. ITS and 16S primer sequences for metabarcoding sequences.

Primers	Sequence (5'->3')	References
ITS1F	CTTGGTCATTTAGAGGAAGTAA	Smith and Peay, 2009
ITS2	GCTGCGTTCTTCATCGATGC	White <i>et al.</i> , 1990
515F	GTGYCAGCMGCCGCGGTAA	Caporaso <i>et al.</i> , 2012
806R	GGACTACNVGGGTWTCTAAT	Caporaso <i>et al.</i> , 2012

Table 2. Number of Illumina Miseq reads allocated to each sample.

Sample	Fungal-reads	Bacterial-reads
1	63,429	69,398
2	72,403	65,015
3	86,668	61,393
4	76,786	69,355
5	79,501	53,093
6	36,246	77,671
7	38,637	49,597
8	45,597	67,366
9	59,312	111,307

Table 3. Diversity metrics for fungal and bacterial rRNA sequences for each sample. Species richness (S), Shannon index (H') and Simpson's Index of Dominance (1-D) are reported.

FUNGI			
Sample	S	H'	1-D
1	101	1.58	0.78
2	151	2.04	0.88
3	120	1.28	0.85
4	156	1.76	0.72
5	145	2.11	0.98
6	135	1.88	0.78
7	140	2.63	0.88
8	149	1.49	0.74
9	153	2.54	0.9
mean	139	1.9	0.83
dev. st.	18	0.46	0.08
BACTERIA			
Sample	S	H'	1-D
1	190	1.74	0.68
2	405	3.23	0.73
3	291	2.24	0.87
4	430	3.50	0.78
5	252	3.16	0.66
6	160	2.78	0.78
7	470	3.70	0.84
8	203	2.71	0.86
9	360	3.23	0.77
mean	307	2.93	0.77
dev. st.	114	0.62	0.07

Figures legends

Figure 1. Landscape of Battleship Promontory (McMurdo Dry Valleys), visited during the XXVI Italian Antarctic Expedition (2010-2011) (A); Sandstone outcrops (B,C); Cryptoendolithic colonization just below the rock surface (D).

Figure 2. Non-Metric Multidimensional Scaling (N-MDS) ordination plots for fungal (A) and bacterial (B) groups of Antarctic endolithic communities collected at Battleship Promontory, based on square-root transformed abundance data.

Figure 3. Fungal (A) and bacterial (B) species accumulation curve produced by the analytical formulae and randomizing (999 permutations) the samples via PRIMER-E software. Values were obtained using observed species (Sobs) counts and plotted against increasing number of samples.

Figure 4. Expected species richness values (solid line) were computed for augmented reference samples (sample-based extrapolation curves) using EstimateS with 95% confidence intervals. Error bars represent the deviation standards.

Figure 5. The abundance of fungi (a) and bacteria (b), as indicated by the number of ITS or 16S ribosomal DNA (rDNA) copies measured using quantitative PCR (qPCR) assay. Error bars are the standard errors. Significant differences, calculated by Tukey test ($p < 0.05$), are indicated by different letters.

Figure 1S. The frequency of observed OTUs for each sample was used to calculate fungal rarefaction curves using the 'rarecurve' function in the R library vegan v2.3-4 (Oksanen *et al.*, 2013) in R 3.2.0 (R-Development-Core-Team 2015).

Figure 2S. The frequency of observed OTUs for each sample was used to calculate bacterial rarefaction curves using the 'rarecurve' function in the R library vegan v2.3-4 (Oksanen *et al.*, 2013) in R 3.2.0 (R-Development-Core-Team 2015).

Table 1S. The abundance of fungi and bacteria, as indicated by the number of ITS and 16S ribosomal DNA (rDNA) copies measured using quantitative PCR (qPCR) assay.

Table 2S. Fungal taxonomy results at 97% of identity. Taxonomy assignment was performed with SINTAX/UTAX databases (Edgar, 2010).

Table 3S. Bacterial taxonomy results at 97% of identity. Taxonomy assignment was performed with SINTAX/UTAX databases (Edgar, 2010).

A

Stress value: 0.095

B

Stress value: 0.076

A **B**

ITS sequences 16S sequences

3.0E+06
2.5E+06
2.0E+06
1.5E+06
1.0E+06
5.0E+05
0.0E+00