Antarctic biodiversity predictions through substrate qualities and environmental DNA

# Authors:

Paul Czechowski1\*, Michel de Lange2, Aleks Terauds3 and Mark I. Stevens4

# Affiliations:

1 University of Otago, Department of Anatomy, 270 Great King Street, Dunedin, OTA 9016, New Zealand.

2 University of Otago, Department of Biostatistics, 18 Frederick Street, Dunedin, OTA 9016, New Zealand.

3 Australian Antarctic Division, Department of Agriculture, Water and the Environment, 203 Channel Highway, Kingston, TAS 7050, Australia.

4 South Australian Museum, 61-68 North Terrace, Adelaide, SA 5000, Australia.

\*Correspondence to Paul Czechowski [paul.czechowski@otago.ac.nz](mailto:paul.czechowski@otago.ac.nz)

# Abstract:

Terrestrial biota in Antarctica face ongoing threats from invasive taxa and climate change. While Antarctic conservation science is becoming increasingly important as a means of informing and enhancing national and international Antarctic policy and protecting the Antarctic continent from further impacts, current protection regimes remain inadequate. Better management of human activity, and expansion of Antarctic protected areas may slow or even reverse impacts on the Antarctic environment. With the absence of large-scale biodiversity data, the identification of key parameters that shape the distribution of common soil organisms has high potential for predicting distribution patterns throughout vast and remote areas in Antarctica. We realize this potential by linking Antarctic soil derived environmental DNA sequence data to a selected range of soil properties beyond climatic predictors by means of the Lasso, a statistical method now commonly used in machine learning algorithms.

# Introduction:

Although only 0.3% of continental Antarctica is ice-free, many organisms including bacteria, unicellular eukaryotes, fungi, lichen, cryptogamic plants and invertebrates are scattered across the continent in extremely isolated, remote, island-like habitats, for example in soils, lakes, and cryoconite holes (Convey *et al.* 2014; Chown *et al.* 2015). Threats to this Antarctic biodiversity are posed by human activity, climate change, pollution, and invasive species. Mitigation of these threats and further alterations to the Antarctic biosphere rely on well-tailored management strategies across the continent’s bioregions (eg Coetzee *et al.* 2017).

However, effective continental-scale conservation management strategies require continental-scale data (eg Wauchope *et al.* 2019), and although the tools to gather such data are readily available (Czechowski *et al.* 2017) continental-wide surveys of Antarctic biodiversity are not being realized. Comparable large-scale systematic approaches to protect soil diversity are recognized as required globally, but often are limited to charismatic groups such as those found in the Arctic (Gillespie *et al.* 2019). Since soil nutrient status is the most important attribute of biodiverse soils, and corresponding key variables can be, and are, routinely measured economically by many laboratories around the world, we here attempt to link commonly measured soil nutrients – among other variables – to eukaryotic soil biodiversity data, in this instance from an extremely remote Antarctic region, so as to provide a suite of robust predictors for common Antarctic soil phyla beyond available climatic predictors. In doing so, we hope to contribute towards focusing future efforts to obtain comprehensive biodiversity data from regions hitherto not surveyed (Geisen *et al.* 2019).

The Prince Charles Mountains (PCMs) were first sighted by US Operation Highjump in 1946­-47 and mapped in more detail Australian (1954 to 1961) and Soviet (1983 to 1991; Kamenev *et al.* 1993) expeditioners. In 2011 we surveyed the PCMs, as the most remote terrestrial areas in eastern Antarctica, to obtain environmental DNA samples and measured corresponding soil parameters. Previously, Czechowski *et al.* (2016b) predominantly focused on invertebrates as the primary soil inhabiting metazoans and discovered major changes in soil invertebrate distribution over salinity gradients, with entire taxonomic phyla being absent in highly saline areas In this current paper we expand analyses of environmental variables influencing biotic distribution patterns, coupled with a predictive approach, to the full spectrum of eukaryotes in order to explore viable approaches of inferring biodiversity data for the entirety of ice-free terrestrial Antarctica.

# Methods:

Fieldwork was conducted in the Prince Charles Mountains (PCMs; East Antarctica, Fig. 1) from 26 November 2011 to 21 January 2012 close to Mount Menzies (MM; 73°25'29.38"S, 62°0'37.61"E), Mawson Escarpment (ME; 73°19'16.91"S, 68°19'31.20"E) and Lake Terrasovoje (LT; 70°32'23.58"S, 67°57'28.05"E) as described earlier (Czechowski *et al.* 2016a, b). A total of 154 field samples were considered for this study, with 26 from MM, 70 from ME and 58 from LT.

To determine climatic conditions of the PCMs, we used publicly available raster layers for annual mean precipitation (mm), wind speed (at 10m above ground in m s-1) and temperature (2 m in °C), as distributed with Quantarctica (www.npolar.no/quantarctica/ ) and further described elsewhere (van Wessem *et al.* 2014; Van Wessem *et al.* 2014). We attempted to mitigate the coarse rasterization of 35km px-1 by disaggregating the layers to 1km px-1 through bilinear interpolation. Subsequently we extracted mean values for the three climatic variables from a 20 km buffer surrounding each sampling location (Supporting Information, Fig. 4).

Soil geochemical composition was analyzed by APAL agricultural soil testing service ([www.apal.com.au](http://www.apal.com.au)) and these soil properties were used as predictor data for eukaryote phylum presence. Variables in the final analysis included P, S, conductivity, pH of CaCl2. Pursuing solid predictors of soil biodiversity measurements of K, ﻿NH4+, NO3−, pH of H2O, and organic C could not be included due to too many missing observations. Furthermore, the substrate mineral composition was considered through integration of X-ray diffraction spectra of the minerals ﻿quartz, calcite, feldspar, titanite, pyroxene / amphibole / garnet, micas, dolomite and kaolin / chlorite and chlorite as formerly described (Czechowski *et al.* 2016b). To adequately handle the sum-to-unity constraint of our mineral compositions, quartz was excluded from further analysis as the most common mineral across all samples. As additional predictors for most locations (MM: n=26, ME: n=69, LT: n=57), we included unpublished measurements of soil-substrate ATP (compare eg Conklin and Macgregor 1972), obtained with a Clean-Trace Luminometer (3M, Maplewood, US-MN), and slope measurements. Prior to regression, all predictors were standardized to mean of 0 and unit variance . A table and graphical summary of the predictors are provided in the Supporting Information.

Biological response data were prepared in QIIME 2020-2 (Bolyen *et al.* 2019) and R 4.0.0 (R Core Development Team 2019) from raw sequence data generated as described elsewhere (Czechowski *et al.* 2016b, 2017). In summary, 125 bp eukaryotic 18S rDNA sequences had been generated using primers ‘Euk1391f’ and ‘EukBr’ from which Amplicon Sequence Variants (ASVs; *sensu* Callahan *et al.* 2017)were re-defined with Qiime: After pre-filtering (Phred score ≥ 25), read pairs were subjected an additional round of adapter trimming using Cutadapt v1.18 (Martin 2011), before denoising with the DADA2 algorithm (v1.6.0;Callahan *et al.* 2016), retaining merged reads with an expected error value less then 3, not determined chimeric.

We identified eukaryotic sequences among our reads using a recent copy (April 2020) of the entire NCBI nucleotide collection in conjunction with Blast 2.10.0+ (Camacho *et al.* 2009). Taxonomic assignments were retrieved from references sequences at least 50% identical to queries, with an assignment error probability (*e* value) of 10-10, considering only matches with at least 90% coverage, and excluding environmental sequences. Handling of taxonomic data in R (R Core Development Team 2019) was substantially aided by packages *tidyverse* and *taxonomizr* (Wickham 2017; Sherrill-Mix 2019). Putatively contaminating reads were removed by subtracting data contained in negative controls with R package *decontam* (Davis *et al.* 2018), reads and taxa remaining in controls after this step were also excluded from the data (Supporting Information). Focusing on eukaryotes, we discarded all non-eukaryote reads.

All further analyses were performed with R package *glmnet* (Friedman *et al.* 2010). For each phylum present in at least 12 samples, we fitted a Lasso logistic regression (Tibshirani 1996) to the retained predictor variables, while disregarding read abundances as meaningless due to inherent constraints of amplicon sequencing (Czechowski *et al.* 2017). For the purpose of regression, the most biodiverse of all locations (LT; Czechowski *et al.* 2016b; also Fig 2) was set as the reference, and predictor effects at MM and ME are reported in relation. We initially retrieved the active set (variables not set to 0) estimated by Lasso, then repeated the regression on 1,000 bootstrap samples, calculated the number of times each variable was estimated to be non-zero, and report variables non-zero more than 950 times as significant. Subsequently, we calculated 95% non-parametric bootstrap confidence intervals for these estimates. We did not adjust for multiple comparisons.

# Results:

Keeping in mind the coarse raster resolution and the model character of the available climate data, the annual mean climate at MM was coldest (-32 ± 0.3 °C), windiest (10.2 ± 0.05 ms-1) and with an intermediate amount of precipitation (86 ± 1 mm), when compared to the other two locations. ME exhibited the least amount of precipitation (55.3 ± 7 mm), comparatively low wind speeds (5.4 ± 0.5 ms-1 ), and slightly higher temperatures then MM (-28.4 ± 0.6 mm). Closest to the coast, and exposed, LT appeared influenced by the highest precipitation (136 ± 16 mm), variable but moderate wind speeds (5.5 ± 1.7 ms-1) and the highest temperature in the sampling area (-24.1 ± 1.6 °C) (Supporting Information, Fig. 4). When included into our modelling approach, our three climatic variables unsurprisingly strongly correlated with the sampling locations, and consequently were excluded from further statistical consideration in favor of the sampling locations to improve predictive power. At the same, all effect of location reported below should be understood as a function of annual mean climatic variables.

Retention of eukaryotes in field-derived samples after filtering yielded 2 285 773 eukaryote reads across 145 samples. Per-ample mean coverage was 9 450 reads (min: 2, median: 2 379, max: 86 804). ASV mean coverage after filtering was 2 984 reads (min: 2, median: 132, max: 207 718). Collectively after filtering, 766 ASVs were assigned to 495 species across 25 phyla. Most prevalent phyla (and among those: taxa) by coverage were: Ascomycota (*Acanthothecis fontana*), Chlorophyta (*Coccomyxa* sp.), Basidiomycota (*Mrakia frigida*), Ciliophora (*Pseudochilodonopsis quadrivacuolata*), Nematoda (*Scottnema lindsayae*), Rotifera (*Embata laticeps*), and Tardigrada (*Mesobiotus furciger*). To the best of our knowledge all of these taxa are ubiquitous, temperate, polar, or relatively recently described.

The distribution of 173 species (100 families, 59 orders, 26 classes, 5 phyla) across the entire sample range (Fig 2) significantly correlated with a range of soil predictors. Collectively those soil taxa were defined by 265 ASVs across 1 210 855 sequences, found among 23 samples from MM, 64 from ME, and 55 samples from LT. The overall per-sample mean coverage was 9 460 (min: 2, med: 3863, max: 84 892), The overall per-ASV mean coverage was 4596, (min: 2, median: 157, max: 128 358). All taxonomic assignments prior, during, and after filtering, locations, and alignments (in MEGAN format; Huson *et al.* 2016) are available as Supporting Information.

For each predictor that significantly correlated with a phylum’s presence we report the expected effect of a one standard deviation (σ) increase of one from its mean (μ), with all other variables held at μ. Key significant results include:

1. Low levels of Basidiomycota (62 species) in high pH environments (μ = 7.15, σ = 0.88, E[present μ] = 0.6 and E[present μ +1σ] = 0.4), and a strong positive relationship of this phylum with Dolomite (μ = 0.025 %, σ = 0.05 %, E[present μ +1σ] = 0.7).
2. Very low levels of Chlorophytes (47 species) at MM presumably to harsh environmental conditions (E[present LT] = 0.61 and E[present MM] = 0.32, and in more alkaline substrates (E[present μ +1σ] = 0.46).
3. Very low levels of Ciliophorans (47 Species) at MM (E[present LT] = 0.70 and E[present MM] = 0.39), in Sulphur-rich substrates (μ = 528 mg kg -1, σ = 1410 mg kg -1, E[present μ +1σ] = 0.61), and in areas relatively rich in pyroxene, amphibole or garnet (μ = 4 %, σ = 4 %, E[present μ +1σ] = 0.52),
4. Very low levels of Nematodes (8 species; highest occurrence: *S. lindsayae*) at MM (E[present LT] = 0.47 and E[present MM] = 0.28), and in highly conductive substrates (μ = 0.55 dSm-1, σ = 1.07 dSm-1, E[present μ +1σ] = 0.35)
5. Very low levels of Tardigrades (9 species) in alkaline substrates (E[present μ] = 0.22, E[present μ +1σ] = 0.14).

ASV composition of significant phyla are listed in detail, alongside species assignments, in the Supporting Information. Observed percentages of non-zero coefficients are shown in Fig. 3, (left panels) and Table 1. 95% non-parametric bootstrap confidence intervals for non-0 estimates are provided in Fig. 3 (right panels) and further detailed also in Table 1. Directions of all predictor effects on all analyzed taxa presences, including insignificant effects, are listed in the Supporting Information.

# Discussion:

Our study unites two key methodologies that provide a case study in large scale surveys aiming at conservation of soil biodiversity ­– robust predictive statistics, such as the Lasso, now often used in machine learning algorithms (Muthukrishnan and Rohini 2016), as well as highly detailed biodiversity information derived from environmental soil DNA (Czechowski *et al.* 2017). Our expanded analyses from the original raw data (Czechowski *et al.* 2016b) utilized the availability of new algorithms for processing environmentally derived DNA sequence data (eg Callahan *et al.* 2016, 2017), expanded reference databases, and the development of packages within the R computing language (R Core Development Team 2019). Our results corroborate earlier findings regarding regional eukaryote distribution patterns as shaped by climatic and soil-age related factors (Czechowski *et al.* 2016a, b), but expand on those findings with respect to five phyla. Our study is highly relevant to other regions of Antarctica where these methodologies can be replicated, and similar approaches can be undertaken in the Arctic for conservation, and worldwide for agricultural purposes.

Identification of species with likely Antarctic occurrence highlight the power of environmental DNA to retrieve species occurrence records, provided that sufficient reference data is available. While keeping in mind the exploratory nature of our study, species identifications obtained here using a relatively short and highly conserved section of the 18S rDNA include species known or likely to occur in Antarctica (Fig. 2). *Mrakia frigida* is a Basiomycot closely related to recently described Antarctic species ﻿*Mrakia psychrophila* (Xin and Zhou 2007), and *Chloroidium angustoellipsoideum* (Chlorophyta) is in the same genus as recently described *Chloroidium antarcticum* (Darienko *et al.* 2018).Furthermore, both *Scottnema lindsayae* (Nematoda) and *Mesobiotus furciger* (Tardigrade) are both known Antarctic species with good reference data coverage (Velasco-Castrillón and Stevens 2014). Solely for *Dileptus jonesi* (Ciliophora) possible Antarctic distribution could not be confirmed here, but is conceivable.

Species distribution patterns similar to related Antarctic works give us confidence in the results reported here. The rarity of Chlorophytes, Ciliophorans, and the otherwise ubiquitous nematodes at MM in relation to the two other lower altitude and more northerly locations (ME, LT) once again reveal trends of ﻿increasing eukaryotic richness and diversity with decreasing latitude and altitude (Czechowski *et al.* 2016a; Thompson *et al.* 2020; Zhang *et al.* 2020). Beyond this, our work highlights that surprisingly high eukaryotic diversity can unexpectedly occur even in the harshest environments, such as local ice-substrate boundaries at Mount Menzies (Figs 1 and 2). The absence of Ciliophorans from Sulphur-rich substrates, and of nematodes from highly conductive soil interstices matches findings of distribution patterns being shaped by age-related salt accumulation at the surface-air interface of frozen soils described with other analytical approaches (Velasco-Castrillón *et al.* 2014; Lee *et al.* 2019).

In absence of other predictors, our work highlights the importance of neutral substrate pH, low conductivity, and key minerals (dolomite, pyroxene, amphibole, or garnet) to predict high eukaryote density in high-latitude substrates, thereby in line with, and expanding findings from other areas of Antarctica. We corroborate the negative influence of substrate alkalinity on Antarctic Basidiomycota (Arenz and Blanchette 2011). We also found that distance to coast to be a suitable proxy variable negatively related to the presence of Chlorophytes and Ciliophorans (Thompson *et al.* 2020), but additionally find soil alkalinity, sulphur content and substrates pyroxene, amphibole, or garnets to constrain distribution of the former. Among nematodes, our results indicate that *Scottnema lindsayae* can indeed occur in high altitude / high latitude environments such as MM , but is likely to be influenced by the species’ general indifference (rather than affinity) to alkaline substrates, and must be highly localized (at least at MM) if encountered at high abundance (Smykla *et al.* 2018; Zawierucha *et al.* 2019). Lastly, we confirm the negative association between tardigrade occurrence and alkaline substrates observed in Victoria Land (eg Smykla *et al.* 2018).

Based on our findings, ice-free areas with high annual mean precipitation, low wind speeds and relatively high temperatures, exhibiting substrates with a neutral pH and low conductivity, which are rich in dolomite but poor in pyroxene, amphibole, or garnets, are likely to be highly biodiverse in the Antarctic and should harbor candidates for focused conservation management. Furthermore, locations with more extreme environmental conditions may harbor endemic relic fauna equally warranting protection (Convey and Stevens 2007). Our results can be transferred to other Antarctic and possibly Arctic ecosystems, as soil pH was found to be an important factor determining bacterial and fungal community structure in Antarctic and Arctic substrates (Siciliano *et al.* 2014). At the same time Antarctic soil ecosystems are relatively simple and ­are assumed to mostly lack complex biotic interactions, although such interactions may be more present in coastal terrestrial ecosystems (Lee *et al.* 2019). Consequently, the soil eukaryote distribution patterns observed especially at MM are likely predominantly shaped by abiotic factors, and would be gradually more influenced by limited biotic interactions lower latitude substrates or more costal substrates (ME, LT).

**Conclusion:**

The approach presented here provides a case study that we hope to influence the development of large scale analysis algorithms to predict presence of eukaryote taxa by means of commonly used soil predictors in unison with readily availableclimate data, thus helping further delineation of Antarctic conservation areas (Terauds and Lee 2016), and aiding other applications in higher latitude ecosystems. In an Antarctic context we estimate that at least 500 to 1 000 soil samples should be analyzed with methods as outlined here across a large range of terrain conditions to reliably identify proxy variables suitable for more accurate predictions of continent wide eukaryote distribution patterns.

**Acknowledgments:**

﻿P.C. was supported by The University of Adelaide through an International Post-Graduate Research Scholarship, through the Royal Society of South Australia, and the University of Otago. M.S. received funding from The Australian Antarctic Division, science project 2355 and A.T. was supported by AAS 4296. Alan Cooper and M.S. received funding for this project through Australian Research Council linkage grant LP0991985. M.S. and P.C. received funding for this project from the Sir Mark Mitchell Foundation. We thank Robert McPhee for the artistic contribution of Figure 2. We thank Diana Wall for providing helpful comments pertaining to the manuscript.

# References:

Arenz BE and Blanchette RA. 2011. Distribution and abundance of soil fungi in Antarctica at sites on the Peninsula, Ross Sea Region and McMurdo Dry Valleys. *Soil Biol Biochem* **43**: 308–15.

Bolyen E, Rideout JR, Dillon MR, *et al.* 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*.

Callahan BJ, McMurdie PJ, and Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* **11**: 113597.

Callahan BJ, McMurdie PJ, Rosen MJ, *et al.* 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–3.

Camacho C, Coulouris G, Avagyan V, *et al.* 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Chown SL, Clarke A, Fraser CI, *et al.* 2015. The changing form of Antarctic biodiversity. *Nature* **522**: 431–8.

Coetzee BWT, Convey P, and Chown SL. 2017. Expanding the Protected Area Network in Antarctica is Urgent and Readily Achievable. *Conserv Lett*: 1–32.

Conklin AR and Macgregor AN. 1972. Soil adenosine triphosphate: Extraction, recovery and half-life. *Bull Environ Contam Toxicol* **7**: 296–300.

Convey P, Chown SL, Clarke A, *et al.* 2014. The spatial structure of Antarctic biodiversity. *Ecol Monogr* **84**: 203–44.

Convey P and Stevens MI. 2007. Antarctic Biodiversity. *Science (80- )* **317**: 1877–8.

Czechowski P, Clarke LJ, Breen J, *et al.* 2016a. Antarctic eukaryotic soil diversity of the Prince Charles Mountains revealed by high-throughput sequencing. *Soil Biol Biochem* **95**: 112–21.

Czechowski P, Clarke LJ, Cooper A, and Stevens MI. 2017. A primer to metabarcoding surveys of Antarctic terrestrial biodiversity. *Antarct Sci* **29**: 3–15.

Czechowski P, White D, Clarke L, *et al.* 2016b. Age-related environmental gradients influence invertebrate distribution in the Prince Charles Mountains, East Antarctica. *R Soc Open Sci* **3**: 160296.

Darienko T, Lukešová A, and Pröschold T. 2018. The polyphasic approach revealed new species of *Chloroidium* (Trebouxiophyceae, Chlorophyta). *Phytotaxa* **372**: 51.

Davis NM, Proctor DiM, Holmes SP, *et al.* 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**: 226.

Friedman J, Hastie T, and Tibshirani R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**: 1–22.

Geisen S, Briones MJI, Gan H, *et al.* 2019. A methodological framework to embrace soil biodiversity. *Soil Biol Biochem* **136**: 107536.

Gillespie MAK, Alfredsson M, Barrio IC, *et al.* 2019. Circumpolar terrestrial arthropod monitoring: A review of ongoing activities, opportunities and challenges, with a focus on spiders. *Ambio*.

Huson DH, Beier S, Flade I, *et al.* 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data (T Poisot, Ed). *PLOS Comput Biol* **12**: e1004957.

Kamenev E, Andronikov a. V, Mikhalsky E V, *et al.* 1993. Soviet geological maps of the Prince Charles Mountains, East Antarctic Shield. *Aust J Earth Sci* **40**: 501–17.

Lee CK, Laughlin DC, Bottos EM, *et al.* 2019. Biotic interactions are an unexpected yet critical control on the complexity of an abiotically driven polar ecosystem. *Commun Biol* **2**: 62.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10.

Muthukrishnan R and Rohini R. 2016. LASSO: A feature selection technique in predictive modeling for machine learning. In: 2016 IEEE International Conference on Advances in Computer Applications (ICACA). IEEE.

R Core Development Team. 2019. R: A language and environment for statistical computing.

Sherrill-Mix S. 2019. taxonomizr: Functions to Work with NCBI Accessions and Taxonomy.

Siciliano SD, Palmer AS, Winsley T, *et al.* 2014. Soil fertility is associated with fungal and bacterial richness, whereas pH is associated with community composition in polar soil microbial communities. *Soil Biol Biochem* **78**: 10–20.

Smykla J, Porazinska DL, Iakovenko NS, *et al.* 2018. Geochemical and biotic factors influencing the diversity and distribution of soil microfauna across ice-free coastal habitats in Victoria Land, Antarctica. *Soil Biol Biochem* **116**: 265–76.

Terauds A and Lee JR. 2016. Antarctic biogeography revisited: updating the Antarctic Conservation Biogeographic Regions (R Heikkinen, Ed). *Divers Distrib* **22**: 836–40.

Thompson AR, Geisen S, and Adams BJ. 2020. Shotgun metagenomics reveal a diverse assemblage of protists in a model Antarctic soil ecosystem. *Environ Microbiol* **22**: 4620–32.

Tibshirani R. 1996. Regression Shrinkage and Selection via the Lasso. *Society* **58**: 267–88.

Velasco-Castrillón A, Schultz MB, Colombo F, *et al.* 2014. Distribution and diversity of soil microfauna from East Antarctica: Assessing the link between biotic and abiotic factors (X Wang, Ed). *PLoS One* **9**: e87529.

Velasco-Castrillón A and Stevens MI. 2014. Morphological and molecular diversity at a regional scale: A step closer to understanding Antarctic nematode biogeography. *Soil Biol Biochem* **70**: 272–84.

Wauchope HS, Shaw JD, and Terauds A. 2019. A snapshot of biodiversity protection in Antarctica. *Nat Commun* **10**: 946.

Wessem JM van, Reijmer CH, Lenaerts JTM, *et al.* 2014. Updated cloud physics in a regional atmospheric climate model improves the modelled surface energy balance of Antarctica. *Cryosph* **8**: 125–35.

Wessem JM Van, Reijmer CH, Morlighem M, *et al.* 2014. Improved representation of East Antarctic surface mass balance in a regional atmospheric climate model. *J Glaciol* **60**: 761–70.

Wickham H. 2017. tidyverse: Easily Install and Load the “Tidyverse.”

Xin M and Zhou P. 2007. Mrakia psychrophila sp. nov., a new species isolated from Antarctic soil. *J Zhejiang Univ Sci B* **8**: 260–5.

Zawierucha K, Marshall CJ, Wharton D, and Janko K. 2019. A nematode in the mist: Scottnema lindsayae is the only soil metazoan in remote antarctic deserts, at greater densities with altitude. *Polar Res* **38**: 1–12.

Zhang E, Thibaut LM, Terauds A, *et al.* 2020. Lifting the veil on arid-to-hyperarid Antarctic soil microbiomes: a tale of two oases. *Microbiome* **8**: 37.

# Figure Captions:

**Figure 1**: Sampling area. All sampling locations are marked with a crosshair. Heat shading (at map scale) indicates density of 18S Amplicon Sequence Variants (*sensu* Callahan *et al.* 2017) determined to be significantly influenced by substrate qualities as available. Base layers compiled by the Norwegian Polar Institute and distributed in the Quantarctica package. Visit http://www.quantarctica.org/. Base layers courtesy of the SCAR Antarctic Digital Database, © 1993­–2015 Scientific Committee on Antarctic Research; The National Snow and Ice Data Centre, University of Colorado, Boulder; NASA, Visible Earth Team, http://visibleearth.nasa.gov/; Australian Antarctic Division, © Commonwealth of Australia 2006.

**Figure 2**: Counts of Amplicon sequence Variants for phyla deemed significantly influenced by substrate composition. Species exemplifying phyla as highlighted in the text, from top to bottom: *Mrakia frigida, Chloroidium angustoellipsoideum, Dileptus jones, Scottnema lindsayae* and *Mesobiotus furciger*.

**Figure 3:** Subset of phyla with distributions significantly correlated with analyzed environmental predictors (also compare Table 1 and Supporting Information). Left panels: Proportions of bootstrap (n = 10 000) samples with non- zero estimates and delineation of 95% confidence levels. Right panels: Confidence intervals for estimates (significant predictors should not include 0).

**Table 1**: Numerical summary of significant coefficient estimates for each phylum as obtained through lasso logistic regression.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | 95% CI Coefficient | | 95% CI Odds ratio | | Bootstrap % not zero |
| phylum | predictor | lower | upper | lower | upper |  |
| Basidiomycota | Dolomite | 0 | 1.32 | 1 | -3.70 | 0.93 |
|  | PH | -1.54 | 0.46 | 0.21 | -0.63 | 1.00 |
| Chlorophytes | MM\* | -1.32 | -0.10 | 0.26 | 0.90 | 0.99 |
|  | PH | -1.28 | -0.10 | 0.28 | 0.91 | 0.99 |
| Ciliophora | Garnets | -2.07 | -0.11 | 0.13 | 0.90 | 0.99 |
|  | MM | -1.22 | 0.00 | 0.29 | 1.00 | 0.93 |
|  | Sulphur | -3.14 | 0.00 | 0.04 | 1.00 | 0.85 |
| Nematodes | Cond | -2.17 | 0.00 | 0.11 | 1.00 | 0.99 |
|  | MM | -2.10 | -0.26 | 0.12 | 0.77 | 0.99 |
| Tardigrades | PH | -1.42 | 0.00 | 0.23 | 1.00 | 0.95 |