Check for updates

Antarctic biodiversity predictions through substrate qualities and environmental DNA

Paul Czechowski[1*], Michel de Lange[2,3], Michael Knapp[1], Aleks Terauds[4], and Mark I Stevens[5,6]

Antarctic conservation science is crucial for enhancing Antarctic policy and understanding alterations to terrestrial Antarctic biodiversity. Antarctic conservation will have limited long-term impacts in the absence of large-scale biodiversity data, but if such data were available, it is likely to improve environmental protection regimes. To enable the prediction of Antarctic biodiversity across continental spatial scales through proxy variables, in the absence of baseline surveys, we linked Antarctic substrate-derived environmental DNA (eDNA) sequence data from the remote Antarctic Prince Charles Mountains to a selected range of concomitantly collected measurements of substrate properties. We achieved this through application of a statistical method commonly used in machine learning. Our analysis indicated that neutral substrate pH, low conductivity, and certain substrate minerals are important predictors of the presence of basidiomycetes, chlorophytes, ciliophorans, nematodes, and tardigrades. A bootstrapped regression revealed how variations in the identified substrate parameters influence probabilities of detecting eukaryote phyla across vast and remote areas of Antarctica. We believe that our work will improve future taxon distribution modeling and aid in developing more targeted surveys of biodiversity conducted under logistically challenging conditions.

Front Ecol Environ 2022; 20(10): 550–557, doi:10.1002/fee.2560

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 island-like terrestrial habitats, soil-like substrates, lakes, and cryoconite holes (small depressions formed by melting of snow or ice below radiation-absorbing dark-colored dust or soot) (Convey *et al.* 2014; Chown *et al.* 2015). Anthropogenic threats, including climate change, pollution, and the introduction of invasive species, among others, imperil Antarctic biodiversity. Mitigation of these threats will rely on implementation of well-tailored management strategies across the continent's bioregions (Coetzee *et al.* 2017).

Effective continental-scale conservation management requires concomitant continental-scale data (Wauchope *et al.* 2019). However, knowledge of terrestrial Antarctic biodiversity remains limited because logistical difficulties exacerbated by harsh environmental conditions, along with funding constraints, impede research in Antarctica's ice-free areas. Environmental DNA (eDNA) analysis, despite shortcomings, represents one of the more practical and economical options for continental-wide surveys of terrestrial Antarctic biodiversity, given logistical challenges (Czechowski *et al.* 2017). Comparable large-scale, systematic approaches to protect soil biodiversity are required globally but are often limited to charismatic taxa, such as those found in Arctic regions (Gillespie *et al.* 2020).

In this study, we linked commonly measured substrate properties to the cryptic eukaryotic biodiversity of terrestrial Antarctic ice-free regions. Soil nutrient status is the most important attribute of biodiverse soils (Geisen *et al.* 2019), and corresponding key variables can be, and are, routinely measured economically. We analyzed molecular data (eDNA) from an extremely remote Antarctic terrestrial region to clarify relationships between substrate properties and the presence of eukaryotic phyla. We believe that such an approach will be useful for predicting biodiversity encompassing a wide taxonomic spectrum across extensive areas of the Antarctic and especially for identifying regions worthy of lower-level taxonomic biodiversity surveys, possibly realized through the use of "barcoding" via mitochondrial DNA (such as with mitochondrial cytochrome oxidase 1) or logistically more challenging field biodiversity surveys.

The Prince Charles Mountains (PCMs), the most remote terrestrial area in eastern Antarctica, were first recorded by the US Operation Highjump (a US Navy initiative to establish an Antarctic research base conducted in 1946–1947) and mapped in greater detail during subsequent Australian (1954–1961) and Russian (1983–1991) expeditions. In 2011, we obtained eDNA samples from substrates collected throughout the PCMs and measured geochemical and mineral properties. Previously, Czechowski *et al.* (2016b) focused on invertebrates as the primary substrate-inhabiting metazoans and discovered major changes in their distribution over salinity gradients, as known from other Antarctic areas and taxa (Bottos

[1]*University of Otago, Department of Anatomy, Dunedin, New Zealand* [*]*(p.czechowski@otago.ac.nz);* [2]*University of Otago, Department of Biostatistics, Dunedin, New Zealand;* [3]*Pacific Edge Ltd, Centre for Innovation, Dunedin, New Zealand;* [4]*Australian Antarctic Division, Department of Agriculture, Water and the Environment, Kingston, Australia;* [5]*Securing Antarctica's Environmental Future, South Australian Museum, Adelaide, Australia;* [6]*School of Biological Sciences, The University of Adelaide, Adelaide, Australia*

© 2022 The Authors. *Frontiers in Ecology and the Environment* published by Wiley Periodicals LLC on behalf of The Ecological Society of America.

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al. 2020). Here, we expanded our analyses of environmental variables to all eukaryotic phyla, thereby exploring approaches of inferring biodiversity presence that could be applied across the entirety of ice-free Antarctica. Beyond phylum-level surveys, our technique could be applied using other genetic markers and predictors to link future smaller scaled conservation projects anywhere in terrestrial Antarctica, aid taxon distribution modeling, and contribute toward improving conservation management strategies across Antarctic bioregions.

Methods

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

To infer climatic conditions in the PCMs, we used raster layers from Quantarctica v3 (https://www.npolar.no/quantarctica; Matsuoka *et al*. 2021) encoding annual mean precipitation (mm), wind speed (m s^{-1} at 10 m above the ground surface) and mean annual temperature (°C at 2 m above the ground surface, as this was the only temperature data available via Quantarctica). We disaggregated the layer rasterization from 35 kilometers per pixel (km px^{-1}) to 1 km px^{-1} through bilinear interpolation. We then extracted median values for the three variables from a 20-km buffer surrounding each sampling location (WebFigure 1).

As predictor data for the presence of eukaryotic phyla in substrates, geochemical composition (ammonium [NH^{4+}], carbon [C], soil density [ρ], nitrate [NO^{3-}], soil pH in water [pH_{H2O}], soil pH in calcium chloride [pH_{CaCl2}], phosphorus [P], potassium [K], sulfur [S], and soil texture) was analyzed by the agricultural soil testing service Australian Precision Ag Laboratory (www.apal.com.au). Measurements below detection levels were excluded to yield data completeness of at least 96.7% (WebPanel 1). The final analysis included K, S, ρ, and pH_{CaCl2} (pH_{H2O} was excluded as co-linear and texture was excluded as categorical). As additional predictors, substrate mineral compositions were considered through integration of X-ray diffraction spectra of quartz, calcite, feldspar, titanite, pyroxene/amphibole/garnet, micas, dolomite and kaolin/chlorite, and chlorite (Czechowski *et al*. 2016b). We handled the sum-to-unity constraint of the mineral compositions by excluding quartz, the most common mineral, from further analysis. As further predictors for most locations (MM: n = 26, ME: n = 69, LT: n = 57), we included unpublished measurements of soil-substrate adenosine 5′-triphosphate (ATP) (eg Conklin and Macgregor 1972), obtained with a Clean-Trace Luminometer (3M; Maplewood, Minnesota), and slope measurements. Prior to regression, all predictors were standardized to zero mean and unit variance. Predictor densities are provided in WebFigure 2.

Biological response data were prepared in QIIME 2020-2 (Bolyen *et al*. 2019) and R v4.0.0 (R Core Development Team 2019) from raw sequence data generated as described in Czechowski *et al*. (2016b, 2017). Briefly, 125 base pair (bp) eukaryotic 18*S* rDNA polymerase chain reaction (PCR) products (with 85 bp target region) were previously amplified using the primers "Euk1391f" and "EukBr" (Caporaso *et al*. 2012), which were established specifically for eukaryotic microbial surveying (Thompson *et al*. 2017). Following accepted recommendations, PCRs were conducted in triplicate, with each replicate carrying identical barcodes. The resulting eDNA libraries were combined for sequencing across two Illumina MiSeq runs (WebFigure 3). We redefined amplicon sequence variants (ASVs) (Callahan *et al*. 2017) from those data with QIIME: after pre-filtering (Phred score ≥ 25), we trimmed read pairs with Cutadapt v1.18 (Martin 2011), and denoised using DADA2 v1.6.0 (Callahan *et al*. 2016). We retained merged reads with an expected error value less than 3 that were not deemed chimeric.

Due to the shortness and slow evolution of the employed 18*S* marker, we elected to conduct our analyses at the phylum level, and to use species-level assignments solely to verify data credibility. Accordingly, we designed the retrieval of taxonomic annotations for our Antarctic DNA sequences to yield reliable species identifications in cases where Antarctic reference data were available, while still returning higher taxonomic (eg phylum-level) identifications when closely matching reference data were unavailable. Doing so enabled inclusion of a larger number of Antarctic sequences into our statistical analysis at the phylum level, but rendered species-level identifications as potentially unreliable, requiring verification at the alignment level. We identified eukaryotic sequences among our reads with a recent local copy (April 2020) of the entire National Center for Biotechnology Information (NCBI) nucleotide collection using the Basic Local Assignment Search Tool (BLAST) v2.10.0+. Taxonomic assignments were retrieved from reference sequences *at least* 50% identical to queries, with an assignment significance threshold (*e* value) of 10^{-10}, considering only matches with at least 90% coverage and excluding environmental sequences (*evalue* $1e^{-10}$, *max_hsps 5*, *max_target_seqs 5*, *qcov_hsp_perc 90*, and *perc_identity 50*). For each Antarctic search query, we used the highest bit score among all NCBI-returned sequences for that query to choose the final taxonomic assignment. Subsequently, we used the R package *decontam* (Davis *et al*. 2018) to remove putatively contaminating reads, and likewise subtracted all sequences and taxa in negative controls from field samples. Because we focused on eukaryotes, all reads identified as non-eukaryotic were discarded (WebFigure 4). Post-filtering, we evaluated whether sufficient data were retained for subsequent analysis, both by bootstrapping simulated accumulation curves (n = 1000 per

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

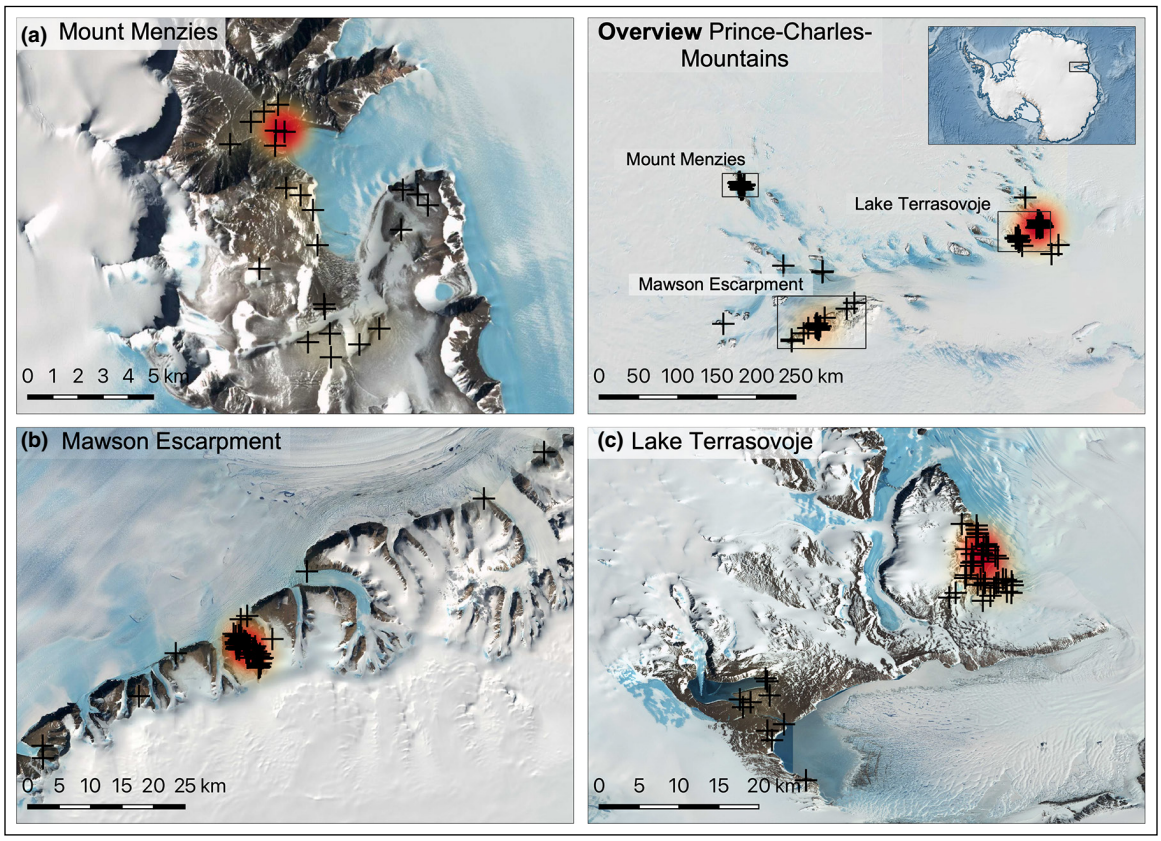

Figure 1. Sampling area in the Prince Charles Mountains, in eastern Antarctica. (a) Mount Menzies (MM), (b) the Mawson Escarpment (ME), and (c) Lake Terrasovoje (LT). All sampling locations are marked with a crosshair. Heat shading (at map scale) indicates density of 18*S* amplicon sequence variants (ASVs) (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 (www.npolar.no/quantarctica). Base layers courtesy of Scientific Committee on Antarctic Research (SCAR) Antarctic Digital Database (© SCAR, 1993–2015); The National Snow and Ice Data Centre, University of Colorado, Boulder; the National Aeronautics and Space Administration's (NASA's) Visible Earth Team (http://visibleearth.nasa.gov); and the Australian Antarctic Division (© Commonwealth of Australia, 2006).

sample) and by analyzing phylum accumulation per sample with accumulating reads, as an absorbing Markov chain.

Using the *lasso* technique (Tibshirani 1996) of the R package *glmnet* (Friedman *et al.* 2010), we regressed each phylum present in at least 12 samples against the aforementioned predictors (WebFigure 5). In regressions, we disregarded sequence read abundances as meaningless due to inherent constraints of amplicon sequencing (Czechowski *et al.* 2017), analyzed presences instead, and used the most biodiverse of all locations (Czechowski *et al.* 2016b) as a reference location, such that predictor effects at MM and ME are reported as relative to those at LT. We initially retrieved the active set (variables not set to zero) estimated by *lasso*, repeated the regression of phylum presence against 1000 randomly chosen sample-sets of predictors, and calculated the number of times each variable was estimated to be non-zero: variables were considered significant if reported as non-zero more than 950 times. Accordingly, 95% non-parametric bootstrap confidence intervals (CIs) were also calculated for our estimates (ie 5% significance level). We did not adjust for multiple comparisons.

Hoping to find evidence of localized Antarctic invertebrate occurrences (Convey *et al.* 2014), we obtained putative species-level assignments among phyla significantly influenced by environmental predictors (see below) by querying the Global Biodiversity Information Facility (GBIF; www.gbif.org), iNaturalist (www.inaturalist.org), and Biodiversity Information Serving Our Nation (BISON; www.gbif.us, formerly bison.usgs.gov) databases with the R package *spocc* (see WebPanel 1 for detailed methods).

Results

Taking into consideration the coarse raster resolution and model-like character of the climate data, annual mean climate at MM was the coldest (–32 ± 0.3°C) and windiest (10.2 ± 0.05 ms^{-1}) of the three locations, but with intermediate precipitation (86 ± 1 mm) (WebFigure 1), whereas ME had the least precipitation (55.3 ± 7 mm), comparatively low wind speeds (5.4 ± 0.5 ms^{-1}), and slightly higher temperatures than MM (–28.4 ± 0.6°C). Closest to the coast and largely exposed, LT exhibited the highest precipitation

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(136 ± 16 mm) and temperature (–24.1 ± 1.6°C) of the surveyed areas, along with variable but moderate winds (5.5 ± 1.7 ms^{-1}). Our chosen climatic variables correlated strongly with the sampling locations, and to improve predictive power we excluded the climatic variables from further considerations. Instead, we interpreted the statistical effect of location (below) to be a function of annual mean climatic variables.

Retention of eukaryotes in field-derived samples after filtering yielded 2,285,773 reads across 145 samples, derived from 16,524,031 unfiltered sequences (WebTable 2). Per-sample mean coverage was 9450 reads (minimum [min]: 2, median: 2379, maximum [max]: 86,804). ASV mean coverage after filtering was 2984 reads (min: 2, median: 132, max: 207,718; WebFigure 6). Collectively after filtering, 766 ASVs were assigned to 495 species across 25 phyla (WebTable 3). Accumulation curve analysis confirmed sufficient sample coverage, and Markov chain analysis deemed eight samples (5%) potentially undersequenced (2 MM, 4 ME, 2 LT). Most prevalent phyla (and among those, the most prevalent species) by coverage were Ascomycota (*Acanthothecis fontana*), chlorophytes (*Coccomyxa* sp), Basidiomycota (*Mrakia frigida*), ciliophorans (*Pseudochilodonopsis quadrivacuolata*), Nematoda (*Scottnema lindsayae*), Rotifera (*Embata laticeps*), and Tardigrada (*Mesobiotus furciger*). All taxonomic assignments listed here aligned with reference data without gaps at full coverage and a bit score of 154.6, apart from a bit score of 145.6 for *P quadrivacuolata*.

Five phyla (26 classes, 59 orders, 100 families, 173 species) distributions across the PCMs were significantly correlated with the considered soil predictors (Figure 2; WebPanel 1). Those taxa were defined by 265 ASVs across 1,210,855 sequences and 142 samples (23 MM, 64 ME, 55 LT). Per-sample mean coverage was 9460 (min: 2, median: 3863, max: 84,892) and per-ASV mean coverage was 4596, (min: 2, median: 157, max: 128,358; WebFigure 6).

For each predictor significantly correlating with a phylum's presence (WebFigure 7), we report the expected effect on phylum presence corresponding to one standard deviation (σ) increase of the predictor from its mean (μ), with all other variables held at mean μ. Key significant results included: (1) low levels of Basidiomycota (62 putative species assignments; Figure 2a) in high pH environments (μ = 7.15, σ = 0.88, E[present $_{\mu}$] = 0.6 and E[present $_{\mu+1\sigma}$] = 0.4), and a

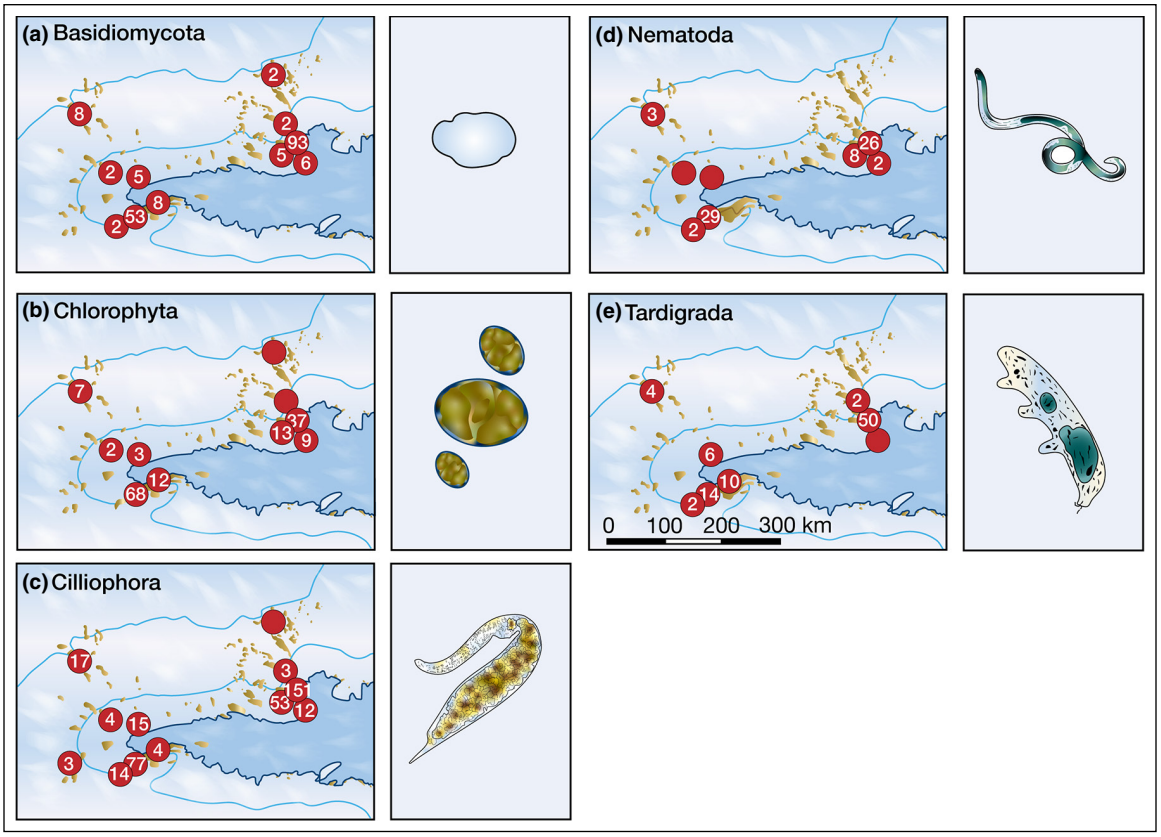

Figure 2. ASV counts for significant phyla within the sampling area (left) and artistic renderings of taxonomic examples (right). Each red circle indicates a unique ASV observation; to avoid overplotting the counts of more than one unique ASV in proximity are indicated by numbers within red circles. (a) *Mrakia frigida* – perfect alignment, possibly present in the Antarctic (Xin and Zhou 2007); (b) *Chloroidium angustoellipsoideum* – perfect alignment, same genus as *Chloroidium antarcticum* (Darienko *et al*. 2018); (c) *Dileptus jonesi* – 97.6% identity, Antarctic distribution unconfirmed; (d) *Scottnema lindsayae* and (e) *Mesobiotus furciger* – both perfectly aligned and present in the Antarctic (Velasco-Castrillón *et al*. 2014a). Base layers courtesy of SCAR Antarctic Digital Database (© SCAR, 1993–2015); The National Snow and Ice Data Centre, University of Colorado, Boulder; NASA's Visible Earth Team (http://visibleearth.nasa.gov); and the Australian Antarctic Division (© Commonwealth of Australia, 2006).

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

strong positive relationship of this phylum with dolomite (μ = 0.025%, σ = 0.05%, E[present $_{\mu+1\sigma}$] = 0.7); (2) very low levels of chlorophytes (47 species; Figure 2b) at MM plausibly attributable to harsh environmental conditions (see WebFigure 1) (E[present $_{LT}$] = 0.61 and E[present $_{MM}$] = 0.32, including more alkaline substrates, E[present $_{\mu+1\sigma}$] = 0.46); (3) very low levels of ciliophorans (47 species; Figure 2c) at MM (E[present $_{LT}$] = 0.70 and E[present $_{MM}$] = 0.39), in sulfur-rich substrates (μ = 528 mg kg $^{-1}$, σ = 1410 mg kg $^{-1}$, E[present $_{\mu+1\sigma}$] = 0.61), and in areas relatively rich in pyroxene, amphibole, or garnet (μ = 4%, σ = 4%, E[present $_{\mu+1\sigma}$] = 0.52); (4) very low levels of nematodes (eight species; Figure 2d) at MM (E[present $_{LT}$] = 0.47 and E[present $_{MM}$] = 0.28), and in highly conductive substrates (μ = 0.55 decisiemens per meter [dS m^{-1}], σ = 1.07 dS m^{-1}, E[present $_{\mu+1\sigma}$] = 0.35); and (5) very low levels of tardigrades (nine species; Figure 2e) in alkaline substrates (E[present $_{\mu}$] = 0.22, E[present $_{\mu+1\sigma}$] = 0.14).

Observed fractions of non-zero coefficients are provided in Table 1 and in WebPanel 1 (95% non-parametric bootstrap CIs for non-zero estimates are also in WebPanel 1). Directions of all predictor effects on all analyzed taxa presences, including non-significant effects, are shown in WebPanel 1.

For 66 of the 173 putative species assignments, 778 georeferenced records could be obtained (of those, 65% derived from GBIF, 27% from iNaturalist, and 7% from BISON). Of the obtained 123 locations, 4% were in Africa, 1.6% were in Antarctica, 13% were in Asia, 32% were in Europe, 21% were in North America, and 10% were in South America (WebPanel 1). The sole species recorded for Antarctica south of the polar circle (south of 66.56°S) was the nematode *S lindsayae*. Observations north of the polar circle (north of 66.56°S) included Basidiomycota (*Gloiocephala aquatica*, *Stereum rugosum*, *M frigida*, *Rhodotorula mucilaginosa*), chlorophytes (*Haematococcus lacustris*, *Oophila amblystomatis*), and ciliophorans (*Furgasonia blochmanni*, *Chilodonella acuta*, *Tachysoma pellionellum*). Refer to WebTable 3 for alignment qualities.

Discussion

Our work demonstrates two key technologies useful for performing baseline biodiversity surveys across large spatial scales in extremely remote environments: (1) robust statistics (such as *lasso*), often used in machine-learning algorithms (Muthukrishnan and Rohini 2016) and (2) biodiversity information derived from eDNA (Czechowski *et al.* 2017). To the best of our knowledge, our work is the first in associating eDNA data to environmental predictors using *lasso* to yield accurate detection probabilities for taxonomic groups, also in Antarctica. Thus, we present an analytical framework to identify areas for targeted species-level biodiversity surveys, using other markers, or predictors for Antarctica, and possibly for other locations that are difficult to access.

Our expanded analyses of the original data (Czechowski *et al.* 2016b) make use of new eDNA sequence processing algorithms (Callahan *et al.* 2016, 2017), along with more extensive reference databases for taxonomic assignment and new algorithms available with R. Our results align with earlier findings relating the distribution of eukaryotes to their environment in the PCMs and Antarctica (Czechowski *et al.* 2016a,b; Bottos *et al.* 2020), but improve the accuracy of those findings for five phyla.

A key strength of our analyses is the relatively easy retrieval of survey data encompassing many phyla (probably including many cryptic and unknown species) across many samples. Conversely, a weakness of the employed 18*S* marker is its

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

		95% CI coefficient		95% CI odds ratio		
Phylum	Predictor	Lower	Upper	Lower	Upper	Proportion of bootstrap replicates not zero
Basidiomycota	Dolomite	0	1.32	1	3.74	0.93
	PH	−1.54	−0.46	0.21	0.63	1.00
Chlorophyta	MM	−1.32	−0.10	0.27	0.90	0.99
	PH	−1.28	−0.10	0.28	0.90	0.99
Ciliophora	Garnet	−2.07	−0.11	0.13	0.90	0.99
	MM	−1.22	0.00	0.30	1.00	0.93
	Sulfur	−3.14	0.00	0.04	1.00	0.85
Nematoda	Conductivity	−2.17	0.00	0.11	1.00	0.99
	MM	−2.10	−0.26	0.12	0.77	0.99
Tardigrada	PH	−1.42	0.00	0.24	1.00	0.95

Notes: Confidence intervals (CIs) for the coefficients are provided on the logit scale. Units of the coefficient CI limits in logistic regression are the reciprocal of the unit of the predictor. Exponentiation of coefficient CI will result in the shown CIs for the odds ratios. Odds ratios are unitless: if they are greater than 1, then the predictor has a positive influence on the probability of finding the given taxon. Mount Menzies abbreviated as "MM".

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

limited ability to discern many distinct sequence variants at low taxonomic levels (eg at the species level). Regardless, identification of species with likely Antarctic occurrence, such as the nematode *S lindsayae* and the tardigrade *M furciger*, through the use of a relatively short and highly conserved pair of primers, highlights the ability of eDNA to retrieve species occurrence records, provided that sufficient sequence data are available for taxonomic assignment. Consequently, we believe that eDNA analysis should be the method of choice for obtaining biodiversity data from Antarctica, particularly when many samples are to be analyzed, but other markers and eDNA analysis techniques are needed to investigate fine-scaled endemism and to obtain better taxonomic resolution.

Georeferencing our putative species assignments with publicly accessible data had limited success, as expected (Cameron *et al*. 2018). The limitations of reference databases became obvious when species known to occur in the Antarctic, such as *Acutuncus antarcticus* (WebPanel 1), identified here through a perfect alignment (bit score 154.6), were absent from the reference databases, and only 38% of all putative Antarctic species assigned by us were georeferenced at all. High occurrence prevalence in North America and Europe not only indicates sampling bias in the GBIF, iNaturalist, and BISON databases, alongside substantial weaknesses of publicly accessible global biodiversity data concerning cryptic eukaryote species, but also supports our choice of merely conducting a phylum-level analysis despite overall well-matching species-level alignments.

Eukaryotic distribution patterns reported in previous Antarctic studies provide context for our observations from the PCMs. The rarity of chlorophytes, ciliophorans, and the otherwise ubiquitous nematodes at MM in relation to the two other, lower altitude, more northerly locations (ME, LT) seem to confirm trends of increasing eukaryotic richness and diversity with decreasing latitude and altitude (Czechowski *et al*. 2016a; Thompson *et al*. 2020; Zhang *et al*. 2020). However, such patterns are not always evident at the scales investigated here; rather, Antarctic biodiversity can be surprisingly regionalized (Convey *et al*. 2014), and our analysis revealed surprisingly high eukaryotic diversity to unexpectedly occur even in the harshest environments, such as local ice–soil substrate boundaries at MM (Figures 1a and 2). The absence of ciliophorans from sulfur-rich substrates and of nematodes from highly conductive soil interstices matches the findings from previous work employing non-eDNA approaches, in which distribution patterns were reported to be shaped by age-related salt accumulation at the surface–air interface of frozen soils (Velasco-Castrillón *et al*. 2014b; Lee *et al*. 2019).

In the absence of other predictors, the results of our analysis emphasize the importance of neutral substrate pH, low conductivity, and key minerals (dolomite, pyroxene, amphibole, or garnet) for predicting high eukaryotic density in Antarctic substrates and corroborate the negative influence of substrate alkalinity on Antarctic Basidiomycota reported by Arenz and Blanchette (2011). Bioregionalization notwithstanding, distance to coast appears to be a suitable proxy variable negatively related to the presence of chlorophytes and ciliophorans, supporting previous findings (Thompson *et al*. 2020). In addition, soil alkalinity and substrate concentrations of sulfur, pyroxene, amphibole, or garnet also appear to constrain chlorophyte and ciliophoran distributions. Among nematodes, our results (ie perfect alignment between our Antarctic 18*S* sequence from MM and an annotated reference sequence) indicate that *S lindsayae* could occur in high altitude and high latitude environments such as MM, but in such environments would be influenced by the species' general indifference (rather than affinity; compare with Zawierucha *et al*. 2019) 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). Finally, we confirm the negative association between tardigrade occurrence and alkaline substrates previously observed in Victoria Land (Smykla *et al*. 2018).

Antarctic ice-free areas exhibiting high annual mean precipitation, low wind speeds, and relatively high temperatures, with substrates possessing neutral pH, low conductivity, abundance in dolomite, and scarcity in pyroxene, amphibole, or garnet, are likely to be highly biodiverse and should harbor candidates for focused conservation management. Furthermore, locations with more extreme environmental conditions may harbor endemic relic fauna equally warranting protection (Convey *et al*. 2014). Our results align with observations in other (including polar and alpine) ecosystems, where soil pH was an important factor determining bacterial and fungal community composition (Siciliano *et al*. 2014; Bottos *et al*. 2020). At the same time, Antarctic soil ecosystems are relatively simple and are assumed to largely lack complex biotic interactions, although such interactions may be more common in coastal terrestrial ecosystems (Velasco-Castrillón *et al*. 2014b; Lee *et al*. 2019). Consequently, the observed distribution patterns of soil eukaryotes (particularly at MM) are likely predominantly shaped by abiotic factors and would be gradually more influenced by limited biotic interactions, lower latitude substrates, or more coastal substrates (ME, LT).

Conclusions

Our analysis highlights the utility of environmental molecular data and predictive analysis algorithms to detect the presence of eukaryotes by means of relatively easily measured soil predictors, which can be combined with readily available climate data. Rather than recognizing trends, our analytical technique provides accurate detection probabilities for Basidiomycota, chlorophytes, nematodes, and tardigrades in relation to bedrock mineral composition, pH, conductivity, sulfur content, and, arguably, overall harshness of environmental conditions. These relationships, as quantified here, enable more precise modeling of phylum distributions over large spatial scales. Our approach may be used to identify regions worthy of species-level biodiversity surveys, possibly employing faster evolving molecular markers, other eDNA analysis techniques, or logistically more challenging

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

morphologic biodiversity assessments. We believe our approach to be valuable to better inform both Antarctic biogeography and delineation of conservation areas.

Acknowledgements

PC was supported by The University of Adelaide through an International Post-Graduate Research Scholarship through the Royal Society of South Australia and by the University of Otago. MIS received Australia Antarctic Science funding from The Australian Antarctic Division, science project 2355; AT was supported by Australia Antarctic Science project 4296; A Cooper and MIS received funding for this project through Australian Research Council linkage grant LP0991985 and ARC SRIEAS grant SR200100005 (MIS); MIS and PC received funding from the Sir Mark Mitchell Foundation and the University of Otago; and MK was funded by the University of Otago. We thank A Corvino, F Shanhun, N Morgan, J Scarrow, and T Williams for help in the Prince Charles Mountains; R McPhee for the artistic contribution to Figure 2; and D Wall for helpful comments on an earlier manuscript draft. Open access publishing facilitated by University of Otago, as part of the Wiley – University of Otago agreement via the Council of Australian University Librarians.

Data Availability Statement

Some data were included in previous publications, which are cited in the text. All data were provided as in-confidence for peer review and have been revised during peer review. Data associated with this work are located at https://doi.org/10.5281/zenodo.4579840. External occurrence data are additionally registered via https://doi.org/10.15468/dd.rkjads. Code is available at https://github.com/macrobiotus/pcm-eukaryotes- and https://github.com/OldMortality/eukaryotes.

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* **37**: 852–57.

Bottos EM, Laughlin DC, Herbold CW, *et al*. 2020. Abiotic factors influence patterns of bacterial diversity and community composition in the Dry Valleys of Antarctica. *FEMS Microbiol Ecol* **96**: fiaa042.

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**: 581–83.

Cameron EK, Martins IS, Lavelle P, *et al*. 2018. Global gaps in soil biodiversity data. *Nat Ecol Evol* **2**: 1042–43.

Caporaso JG, Lauber CL, Walters WA, *et al*. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**: 1621–24.

Chown SL, Clarke A, Fraser CI, *et al*. 2015. The changing form of Antarctic biodiversity. *Nature* **522**: 431–38.

Coetzee BWT, Convey P, and Chown SL. 2017. Expanding the protected area network in Antarctica is urgent and readily achievable. *Conserv Lett* **10**: 670–80.

Conklin AR and Macgregor AN. 1972. Soil adenosine triphosphate: extraction, recovery and half-life. *B Environ Contam Tox* **7**: 296–300.

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

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. *Roy 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 DM, 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*. 2020. Circumpolar terrestrial arthropod monitoring: a review of ongoing activities, opportunities and challenges, with a focus on spiders. *Ambio* **49**: 704–17.

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**: 10–12.

Matsuoka K, Skoglund A, Roth G, *et al*. 2021. Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands. *Environ Modell Softw* **140**: 105015.

Muthukrishnan R and Rohini R. 2016. LASSO: a feature selection technique in predictive modeling for machine learning. In: 2016 Institute of Electrical and Electronics Engineers (IEEE) International Conference on Advances in Computer Applications; 24 Oct 2016; Coimbatore, India. New York, NY: IEEE.

R Core Development Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

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.

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

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.

Thompson LR, Sanders JG, McDonald D, *et al*. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**: 457–63.

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

Velasco-Castrillón A, Gibson JAE, and Stevens MI. 2014a. A review of current Antarctic limno–terrestrial microfauna. *Polar Biol* **37**: 1517–31.

Velasco-Castrillón A, Schultz MB, Colombo F, *et al*. 2014b. Distribution and diversity of soil microfauna from East Antarctica: assessing the link between biotic and abiotic factors. *PLoS ONE* **9**: e87529.

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

Xin M and Zhou P. 2007. *Mrakia psychrophila* sp nov, a new species isolated from Antarctic soil. *J Zhejiang Univ-Sc B* **8**: 260–65.

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**: 3494.

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.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Supporting Information

Additional, web-only material may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/fee.2560/suppinfo

FrontiersEcoPics

A story behind what's on the menu

For some species with a nonselective diet, which seem to eat everything, the story behind their menu selection warrants much investigation. Omnivores have a generalist diet governed by competitive pressures and resource availability. The degree of meat consumption by omnivores informs their trophic position within the community of sympatric species. To the extent that competition determines diet choice by omnivores, in regions experiencing declines in apex mammalian carnivores, which usually regulate ungulate prey populations, carnivory by omnivores may increase. In Ghana, preliminary evidence from stable isotope analysis suggested a trophic ascension in baboons (*Papio* spp) where populations of African lions (*Panthera leo*) have been extirpated (*Ecological and conservation implications of mesopredator release*. In: *Trophic cascades: predators, prey, and the changing dynamics of nature*. Washington, DC: Island Press; 2010). This photograph of ungulate carnivory by an olive baboon (*Papio anubis*) provides some evidence of such an ascension. The image – which was captured during an extensive camera trap survey investigating carnivores within the W-Arly-Pendjari (WAP) protected area complex in Burkina Faso, West Africa – depicts a neonatal ungulate that is normally consumed by lions (*Conserv Lett* 2019; doi.org/10.1111/conl.12667). Although small and largely fragmented populations of lions remain throughout West Africa, research and conservation efforts along with financial support for effective protected area management remain paramount. Future investigations need to determine the degree of dietary overlap between lions and baboons as well as whether carnivory in baboons is highest in areas with reduced or extirpated lion populations. As African lion populations continue to decline, so does their regulatory capacity. As a result, new niches become available for more abundant species like baboons to begin the classic trophic cascade.

Nyeema C Harris
Applied Wildlife Ecology Lab,
Yale School of the Environment, Yale University,
New Haven, CT
doi:10.1002/fee.2577

15409309, 2022, 10, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.2560 by University of Adelaide Alumni, Wiley Online Library on [20/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License