# Shotgun metagenomics reveal a diverse assemblage of protists in a model Antarctic soil ecosystem

Applied and Environmental Microbiology

Thompson, Andrew; Geisen, Stefan; Adams, Byron J.

https://doi.org/10.1111/1462-2920.15198

environmental microbiology

Environmental Microbiology (2020) **22**(11), 4620–4632 doi:10.1111/1462-2920.15198

# Shotgun metagenomics reveal a diverse assemblage of protists in a model Antarctic soil ecosystem

**Andrew R. Thompson**[1*] **Stefan Geisen**[2,3] **and Byron J. Adams**[1,4]

[1]*Department of Biology, Brigham Young University, Provo, UT.*

[2]*Department of Terrestrial Ecology, Netherlands Institute of Ecology, Wageningen, 6708 PB, The Netherlands.*

[3]*Laboratory of Nematology, Wageningen University, Wageningen, 6700AA, The Netherlands.*

[4]*Monte L. Bean Life Science Museum, Brigham Young University, Provo, UT.*

## Summary

**The soils of the McMurdo Dry Valleys (MDV) of Antarctica are established models for understanding fundamental processes in soil ecosystem functioning (e.g. ecological tipping points, community structuring and nutrient cycling) because the extreme physical environment drastically reduces biodiversity and ecological complexity. Understanding the functioning of MDV soils requires in-depth knowledge of the diversity of MDV soil species. Protists, which contribute significantly to soil ecosystem functioning worldwide, remain poorly characterized in the MDV. To better assess the diversity of MDV protists, we performed shotgun metagenomics on 18 sites representing a variety of landscape features and edaphic variables. Our results show MDV soil protists are diverse at both the genus (155 of 281 eukaryote genera) and family (120) levels, but comprise only 6% of eukaryotic reads. Protists are structured by moisture, total N and distance from the local coast and possess limited richness in arid (< 5% moisture) and at high elevation sites, known drivers of communities in the MDV. High relative diversity and broad distribution of protists in our study promotes these organisms as key members of MDV soil microbiomes and the MDV as a useful system for understanding the contribution of soil protists to the structure of soil microbiomes.**

Received 23 April, 2020; revised 4 August, 2020; accepted 12 August, 2020. *For correspondence. E-mail andrew.thompson8956@gmail.com



## Introduction

High taxonomic diversity in soil ecosystems complicates efforts to characterize the relationship between soil biodiversity and ecosystem processes (Bardgett, 2002; Adams *et al*., 2006; Nielsen *et al*., 2011; Chakraborty *et al*., 2012). The soils of Antarctica's largest ice-free region, the McMurdo Dry Valleys (MDV), are extremely cold, arid, saline and oligotrophic. These factors have limited biotic diversity and trophic complexity and provide a unique model system for exploring physical controls on the activity, structure and assembly of biological communities, and the relationship between biodiversity and ecosystem functioning (Virginia and Wall, 1999; Barrett *et al*., 2004; 2006; Wall, 2005; Adams *et al*., 2006). Charles Elton proposed to use the relatively simple Arctic terrestrial systems as models for community ecology in more complex ecosystems in a similar way in the early 20th century (Elton, 1927). The McMurdo long-term ecological research (MCM LTER) programme aims to characterize the MDV system in order to unravel the relationship between abiotic factors, biodiversity and soil ecosystem functioning (Virginia and Wall, 1999; Wall, 2005; Adams *et al*., 2006; MCMLTER, 2019). Much research on MDV soil communities over the last several decades has focused on linking population dynamics of indigenous nematodes with ecosystem-level processes (Freckman and Virginia, 1997; Doran *et al*., 2002; Ayres *et al*., 2008), while other research has investigated the diversity and physical drivers of soil metazoa, fungi and bacteria (Courtright *et al*., 2001; Adams *et al*., 2006; Buelow *et al*., 2016). Relatively little has been done to explore the composition and abiotic drivers of soil protists in this system (Bamforth *et al*., 2005; Adams *et al*., 2006), despite their importance as primary producers and regulators of soil microbiomes worldwide (Crotty *et al*., 2012; Wilkinson *et al*., 2012; Geisen *et al*., 2018).

Protists are a phylogenetically and functionally diverse group of usually unicellular eukaryotes that are widespread in soil ecosystems (Adl *et al*., 2019; Oliverio *et al*., 2020). Phagotrophic protists include most members of Ciliophora, Cercozoa and Amoebozoa, and some members in other clades (e.g. Discoba and Stramenopiles) (Adl *et al*., 2019). Previous studies on phagotrophic protists in the MDV suggest their taxonomic diversity is an order of magnitude


**Fig 1.** Map of study area. The McMurdo Dry Valleys are located at roughly S 77° E 162° in Southern Victoria Land and open towards the Ross Sea to the East.
Samples represent a variety of MDV soil habitats and landscape features that exist in the MDV. Samples in green (Φ) possessed visible organic matter (e.g. moss and algae); samples in black were coarse to fine mineral soils with varying degrees of visible moisture.

lower than non-Antarctic sites (Foissner, 1996; Bates *et al*., 2013). Phototrophic protists (eukaryotic algae) include Chlorophyta and most Ochrophyta (Stramenopiles) and are better documented in the MDV than phagotrophic protists (Adams *et al*., 2006).

As one of the driest deserts on Earth (Doran *et al*., 2010), MDV soils are commonly categorized in terms of moisture (Freckman and Virginia, 1998): moist sites (> 5% soil moisture content) are infrequent and occur near ephemeral ponds or streams; arid sites (< 5%) dominate the landscape, accounting for > 95% of the MDV surface (Burkins *et al*., 2001). Additionally, the MDV have been subdivided into three climate zones (coastal thaw, inland mixed and upland stable) to reflect the positive correlation between environmental parameters with elevation and distance from the Ross sea (Marchant and Head, 2007). Moist sites likely contain the greatest diversity and biomass but have not been systematically studied for phagotrophic protists. Arid sites are dominated by flagellate and amoeba morphospecies (Bamforth *et al*., 2005), but possess a patchy distribution of very few Ciliophora (≤ 3 species per ~10–50 g soil examined) (Foissner, 1996; Fell *et al*., 2006). No study exists that incorporates the MDV climate zones into its analyses of local protist biogeography (Thompson, in review).

We sequenced 18 shotgun metagenomes and measured soil physicochemical parameters and environmental variables across gradients of moisture and the MDV climate zones to better understand the composition and abiotic drivers of MDV soil protists. In comparison to amplicon- and cultivation-based approaches, shotgun metagenomics more evenly covers genomic content and better preserves information on community structure (Guo *et al*., 2016).

## Results

### *Phagotrophic protists*

Phagotrophic protists (90 genera from 71 families) make up 58% of total protist genera and 32% of total eukaryote

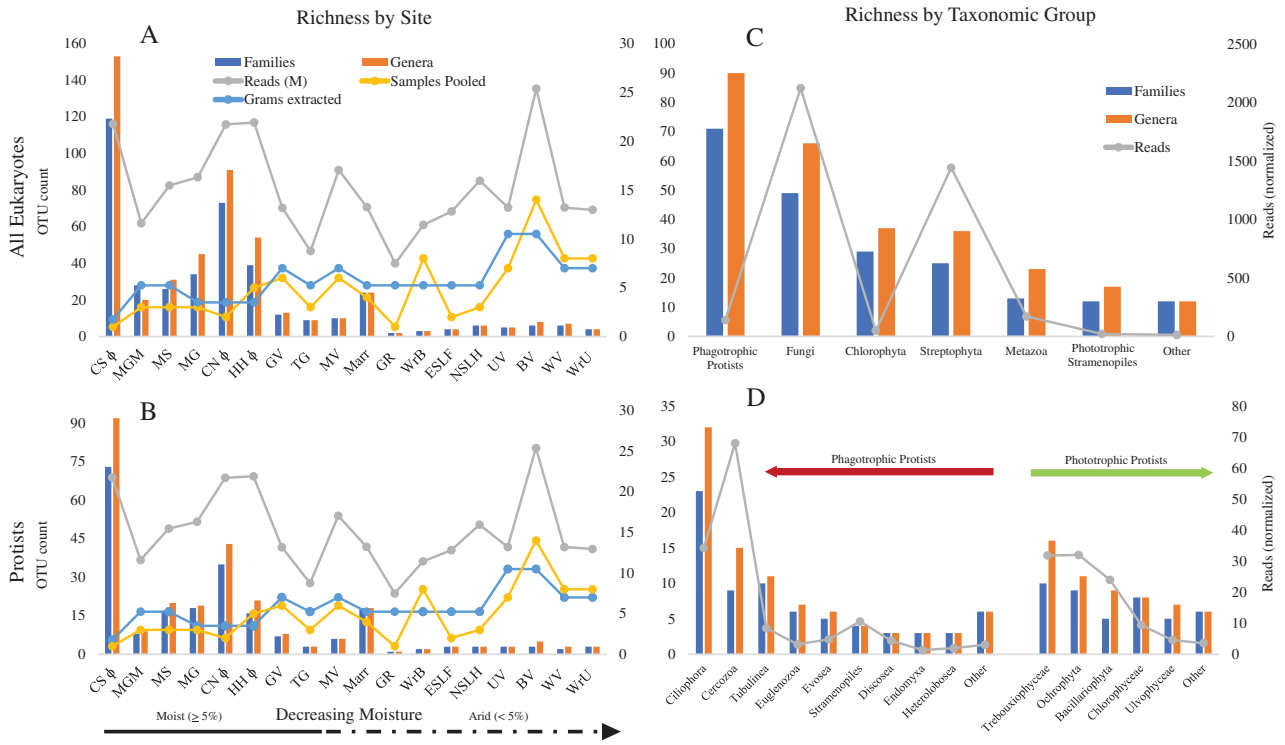


**Fig 2.** OTU richness by site and taxonomic group.
A,B. Alpha diversity for all eukaryotes (A) and protists (B) by site, ordered by decreasing moisture. The line below (B) indicates direction of decreasing moisture and also the transition point from moist (>5%; solid line) to arid (<5%; dashed line). Counts were made using read abundance data normalized to account for read depth.
C,D. Alpha diversity is also plotted by taxonomic and functional groups (where convenient) for all eukaryotes (C) and for protist phyla (D). Phagotrophic protists (C) include Ciliophora, Colpodellidae, Cercozoa, Tubulinea, Evosea, Discosea, Euglenozoa, Heterolobosea, Cryptista, Vampyrellida, Choanaflagellata, Ancyromonadida, Apusomonadida, Telonemia, some Stramenopiles and some Ochrophyta. Phagotrophic and phototrophic protist counts are separated (D) for convenience (see Table S4).
Whole shotgun metagenome read count (A,B) or total processed read count by group (C,D) is plotted (grey line). Number of samples pooled (yellow) and total grams of soil extracted (blue) is also plotted for each site (A,B). Reads, samples and grams soil all correspond to the right *y*-axis on their respective graph. Sites in A and B are ordered from higher to lower moisture content, left to right.

taxa, but only 4% of total eukaryote reads (Fig. 2C). Most sites (14) have fewer than 10 phagotrophic protist genera and only Canada Stream (CS; 61 genera) has more than 20 genera (Table 1). Cercozoa are the most abundant group of phagotrophic protists by read count (49%), are relatively diverse phylogenetically (15 genera from 9 families) and are the most widely distributed eukaryotes, occurring in every soil site but one (Figs 2D and 3C). Two taxa (an unidentified Sandonidae and the genus *Rhogostoma*) occur in 11 and 10 soil sites and account for 20% and 12% of phagotrophic protist reads respectively (Table S4). A majority of cercozoan diversity belongs to these two groups: 59% of cercozoan reads are classified as four genera within Sandonidae and 24% are assigned to *Rhogostoma*. The only site from which a cercozoan read is absent is lower Wright Valley (WrB), although Northside Lake Hoare (NSLH), Towle Glacier (TG), Miers Valley (MV), University Valley (UV), Wall Valley (WV), Mount Gran (GR) and Benson Glacier/Flatiron (MG) each have only a single raw read.

Ciliophora are the most taxonomically rich phylum of phagotrophic protists (32 genera from 23 families) and the second most abundant (24% of total phagotrophic protist reads); however, each genus is present in low abundance (Fig. 2D). The most abundant ciliate operational taxonomic units (OTUs) belong to unidentified genera in the Oxytrichidae (9%), Euplotidae (7%) and Chilodonellidae (6%), while the latter are the most widely distributed (Table S4). The supergroup Amoebozoa (13%), phagotrophic Stramenopiles (8%) and Discoba (4%) account for the majority of the remaining phagotrophic protist diversity and read abundance. Amoebozoa is represented by phyla Discosea (4 genera, 3 families), Evosea (6 genera, 5 families) and Tubulinea (10 genera, 9 families) (Fig. 2D), while *Spumella* accounts for 88% of the phagotrophic Stramenopile reads. Discoba includes 10 genera (7 Euglenozoa and 3 Heterolobosea) from 9 families (6 Euglenozoa and 3 Heterolobosea).

Several reads are assigned to eukaryotes that have an unresolved relationship within Eukarya, including a member of the Opalozoan MAST-12C group, a *Fabomonas* sp. (Ancyromonadida), an *Hatena* sp. (Kathablepharidacea) and a Telonemia species, a lineage that is so morphologically unique that a sister group was only recently identified (Strassert *et al*., 2019). The latter two taxa are present only in the CS biocrust sample (Table S4) and each taxon is represented by only a single read.

### *Phototrophic, parasitic and mixotrophic protists*

Phototrophic protists (certain Stramenopiles and Chlorophyta) have low relative abundance (1.8% of total eukaryote reads) but high taxonomic diversity (54 genera from 41 families) compared to non-vascular plants (36% of total eukaryote reads; 37 genera from 26 families) (Fig. 2C). Chlorophyta (37 genera from 29 families) are more evenly distributed than Stramenopiles (17 genera from 12 families), occurring in 12 of 18 sites (Fig. 3B). The two most abundant Stramenopiles are diatoms [*Hantzschia* (43%) and *Nitzschia* (18%)] while an Ochrophyte [*Heterococcus* (7%)] is the most widely distributed (4 of 18 sites). *Trebouxia* and *Chloroidium* are the most abundant Chlorophyta (9% and 8% respectively), and an unidentified Chlamydomonadales (7%) is the most widely distributed, occurring in four sites. We found four genera of parasites, the latter three occurring exclusively in CS: an unidentified Gregarine, a single *Dermocystidium* sp., an unidentified Thraustochytriceae and *Pythium* (Table S4). We identified only two protist reads as belonging to mixotrophic genera, an *Amphitrema* sp. (Stramenopiles) and an *Hatena* sp. (Katablepharidophyta).

### *Other eukaryotes*

We recovered sequences assigned to 126 genera from 88 families of plants, fungi and metazoa (Fig. 2A). Fungi are the most taxonomically diverse eukaryote group and the most abundant in terms of normalized reads (52% of total reads from soil sites) (Fig. 2C). Fungal diversity is extremely uneven across our sites, with the majority of fungal reads (97%) assigned to just six genera: *Glaciozyma* (53%), *Rhodotorula* (3%), an unidentified Microbotryomycetes (7%), *Sporobolomyces* (5%) and *Sporidiobolus* (2%). Each of these taxa occurred almost exclusively in one site [Hjorth Hill (HH); 99.4% of reads) and few fungal reads occurred outside of HH (3%). *Glaciozyma* sp., probably *Glaciozyma antarctica* Fell, Statzell, Hunter and Phaff, 1969, is the most abundant taxon in our study: it dominated HH (55% of HH reads) but less than 1% of *Glaciozyma* reads occurred outside of HH.

Non-vascular plants are the next most abundant eukaryotic group (36%) after fungi and consist predominately of Streptophyta (36 genera from 25 families), although a single genus of Rhodophyta, from class Bangiophyceae, is present (Fig. 2C, Table S4). Streptophyta reads occur primarily (99.8%) in moist sites (≥ 5% moisture) – CS, Cliff Nunatak (CN), HH, MG, Mount Murray (MGM) (Fig. 3B). Reads most similar to members of the *Bryum* genus (30%) and to an unidentified Embryophyta (35%) are the most abundant Streptophytes.

Metazoa account for 4% of eukaryotic reads (Fig. 2C). We recovered 23 genera from 13 families, including a single read belonging to a *Dermonoton* feather mite (Fig. 2C, Table S4). The most abundant metazoan phylum is Rotifera (54% of metazoan reads), followed by Tardigrada (38%). The most abundant identified Rotifer genus is *Adineta* (Adinetidae), and the most abundant Tardigrade genus is *Acutuncus* (Hypsibiidae). Reads assigned to the three major nematode genera in the

**Table 1.** Site, sequencing and processing stats. Metaxa2, with its custom Silva hmm profile and reference database, was used to extract and classify prokaryotic 16S rRNA.

| | Metaxa2 Silva hmm profile | | | | | | Barrnap hmm profile + $PR^2$ db | | | OTUs | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample name | All SSU (reads) | All SSU (%) | Bacteria (reads) | Bacteria (% 16S) | Archaea (reads) | Archaea (% 16S) | 18S (reads) | 18S (%) | 18S (> 200 bp, raw) | All Eukarya | Phagotrophic Protists | Phototrophic Protists |
| Canada Stream Φ | 12,674 | 0.06 | 9357 | 74 | 21 | 0.2 | 3296 | 26.0 | 2417 | 153 | 61 | 21 |
| Mount Murray | 5426 | 0.05 | 5141 | 95 | 22 | 0.4 | 263 | 4.8 | 188 | 28 | 4 | 5 |
| Mount Suess | 8281 | 0.05 | 8113 | 98 | 71 | 0.9 | 97 | 1.2 | 64 | 31 | 8 | 10 |
| Benson Glacier/Flatiron | 8091 | 0.05 | 7770 | 96 | 125 | 1.5 | 196 | 2.4 | 131 | 45 | 9 | 8 |
| Cliff Nunatak Φ | 12,422 | 0.06 | 9699 | 78 | 19 | 0.2 | 2704 | 21.8 | 1720 | 91 | 19 | 21 |
| Hjorth Hill Φ | 28,490 | 0.13 | 18,267 | 64 | 5 | 0.0 | 10,218 | 35.9 | 6488 | 54 | 14 | 7 |
| Garwood Valley | 6178 | 0.05 | 6139 | 99 | 15 | 0.2 | 24 | 0.4 | 18 | 13 | 6 | 2 |
| Towle Glacier | 4004 | 0.05 | 3962 | 99 | 18 | 0.4 | 24 | 0.6 | 17 | 9 | 2 | 1 |
| Miers Valley | 7648 | 0.04 | 7612 | 100 | 14 | 0.2 | 22 | 0.3 | 12 | 10 | 5 | 1 |
| Marr Pond | 7246 | 0.05 | 7181 | 99 | 2 | 0.0 | 63 | 0.9 | 36 | 24 | 12 | 5 |
| Mount Gran | 3318 | 0.04 | 3310 | 100 | 5 | 0.2 | 3 | 0.1 | 2 | 2 | 1 | 0 |
| Lower Wright Valley | 6063 | 0.05 | 6042 | 100 | 6 | 0.1 | 15 | 0.2 | 3 | 3 | 1 | 1 |
| East Side Lake Fryxell | 6080 | 0.05 | 6063 | 100 | 10 | 0.2 | 7 | 0.1 | 4 | 4 | 2 | 1 |
| North Side Lake Hoare | 6498 | 0.04 | 6470 | 100 | 17 | 0.3 | 11 | 0.2 | 6 | 6 | 3 | 0 |
| University Valley | 4916 | 0.04 | 4903 | 100 | 3 | 0.1 | 10 | 0.2 | 7 | 5 | 2 | 1 |
| Beacon Valley | 12,034 | 0.05 | 12,012 | 100 | 2 | 0.0 | 20 | 0.2 | 13 | 8 | 5 | 0 |
| Wall Valley | 5846 | 0.04 | 5825 | 100 | 0 | 0.0 | 21 | 0.4 | 16 | 7 | 3 | 0 |
| Upper Wright Valley | 5736 | 0.04 | 5702 | 99 | 24 | 0.4 | 10 | 0.2 | 6 | 4 | 3 | 0 |
| Mean | 8386 | 0.05 | 7420 | 94.39 | 21.06 | 0.29 | 945 | 5.32 | 619.33 | 28 | 9 | 5 |
| Total | 150,951 | 1 | 133,568 | 88 | 379 | 0.25 | 17,004 | 11 | 11,148 | 281 | 90 | 54 |

The barrnap eukaryote hmm profile was used to extract eukaryotic 18S rRNA and the protist ribosomal database ($PR^2$) was used to classify them. Unfiltered eukaryotic 18S rRNA were summed with prokaryotic 16S rRNA to estimate the relative proportion of SSU in the full metagenomes. Φ indicates sites with visible organic matter. Sites are ordered from higher to lower moisture content, top to bottom.

 

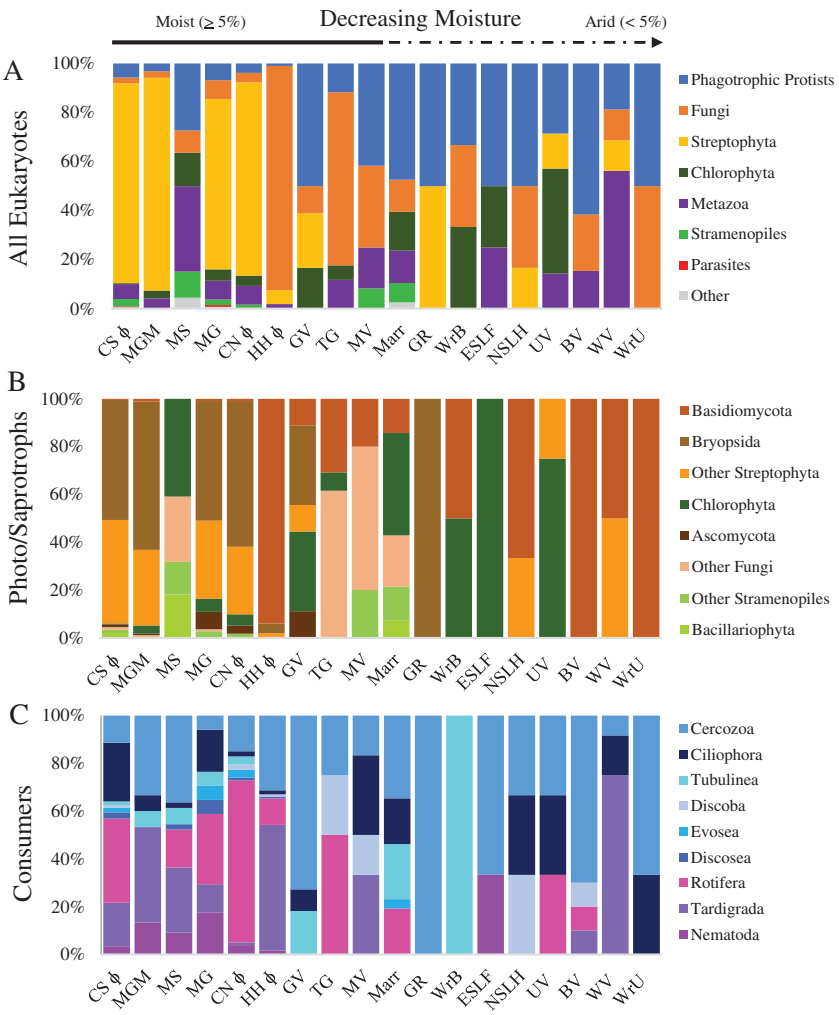


**Fig 3.** Relative abundance using normalized reads by site.
A–C. Relative abundance for All Eukaryotes (A); phototrophs/saprotophs, browns and greens; (B) and consumers, in blues and purples (C). Metazoa are in purples, select phagotrophic protist groups in blues (C), Streptophyta in yellows and Fungi in oranges (B).
Sites are ordered the same in each graph, from higher to lower moisture content, left to right. The line above (A) indicates direction of decreasing moisture and also the transition point from moist (> 5%; solid line) to arid (< 5%; dashed line).

MDV [*Scottnema* (Cephalobidae), *Plectus* (Plectidae) and *Eudorylaimus* (Dorylaimidae)] (Adams *et al*., 2006) accounted for only 6% of metazoan reads and are absent from 11 of 17 sites (Fig. 3C).

### *Diversity across sites*

Individual site richness ranges from 1 to 92 genera (Fig. 2B) and protist richness in moist (≥ 5% moisture content) or coastal climate zone sites is qualitatively greater than arid (< 5% moisture content) or inland and upland zones sites respectively (Fig. 4A). Relative community composition across moisture and climate zone thresholds appear similar, except for a notable increase in the relative proportion of Cercozoa in arid and inland/upland sites, and a lack of parasites outside of moist and coastal sites (Fig. 4B). Differences in richness between climate zones do not appear to be significant, but richness between moist and arid sites is significantly different for phototrophic protists and marginally significant for all protists (Welch's two sample student *T* test: $p = 0.02885$; $p = 0.06242$ respectively) (Table S5). When considering eukaryotic communities as a whole, moist and inland soils tend to intermingle, however, a cluster comprising CN, CS, MGM, MG and HH appears distinct from the other sites (Fig. 5A). Arid sites tend to cluster away from moist sites for all protists (Fig. 5B), but the trends are obscure when examining phototrophic and phagotrophic protists separately (Fig. S2AB).

### *Environmental drivers*

Moisture and total N are significant drivers of phagotrophic protists [permutational multivariate analysis of variance (PERMANOVA): moisture $r^2 = 0.09932$, $p = 0.038$; total N $r^2 = 0.06797$, $p = 0.017$] but not of phototrophic protists (PERMANOVA: moisture $r^2 = 0.09696$, $p = 0.352$; total N $r^2 = 0.09$, $p = 0.467$) (Table S6AB). Moisture, total N and distance to coast are significant and accounted for 30.6% of variation in all protists (PERMANOVA: moisture $r^2 = 0.10741$, $p = 0.005$; total N $r^2 = 0.1053$, $p = 0.013$; distance

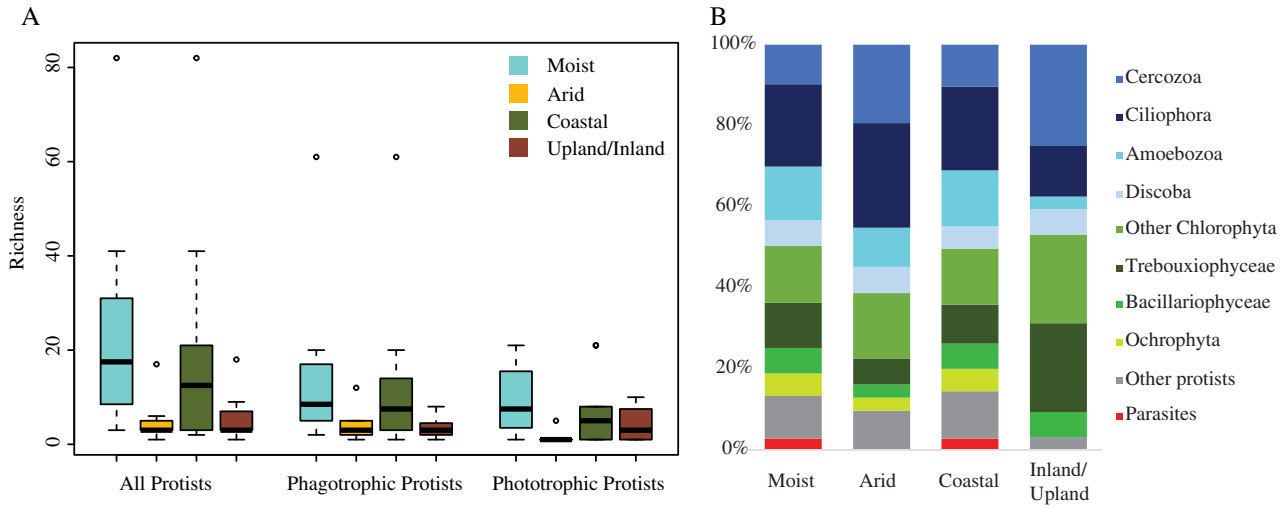


**Fig 4.** Comparison of richness across moisture and climate zone categories.
A. Boxplots comparing richness between moist (≥ 5%), arid (< 5%), coastal and a combined upland and inland geographic zones for groupings of all protists, phagotrophic protists and phototrophic protists.
B. Relative community composition of major taxonomic groups of all protists (supergroups, phyla and classes) for phagotrophic protists (blues), phototrophic protists (greens), parasites (red) and others (grey), by moisture and climate zone categories.

to coast $r^2 = 0.0943$, $p = 0.034$) (Table S6C). Moisture, pH, % clay, total N and distance to coast are significant drivers of for all eukaryotes and account for 5.3% of the variation between our samples (PERMANOVA: moisture $r^2 = 0.12597$, $p = 0.001$; pH $r^2 = 0.08458$, $p = 0.045$; % clay $r^2 = 0.11423$, $p = 0.003$; distance to coast $r^2 = 0.09186$, $p = 0.025$) (Table S6D).

## Discussion

Protists are taxonomically diverse at the genus and family level, showing that protist richness in our study is not artificially inflated due to intraspecific variation in the rDNA gene. Although diverse, protists account for only 6% of all eukaryotic reads (Fig. 2C). The disproportionate representation of protist reads in the data set is likely

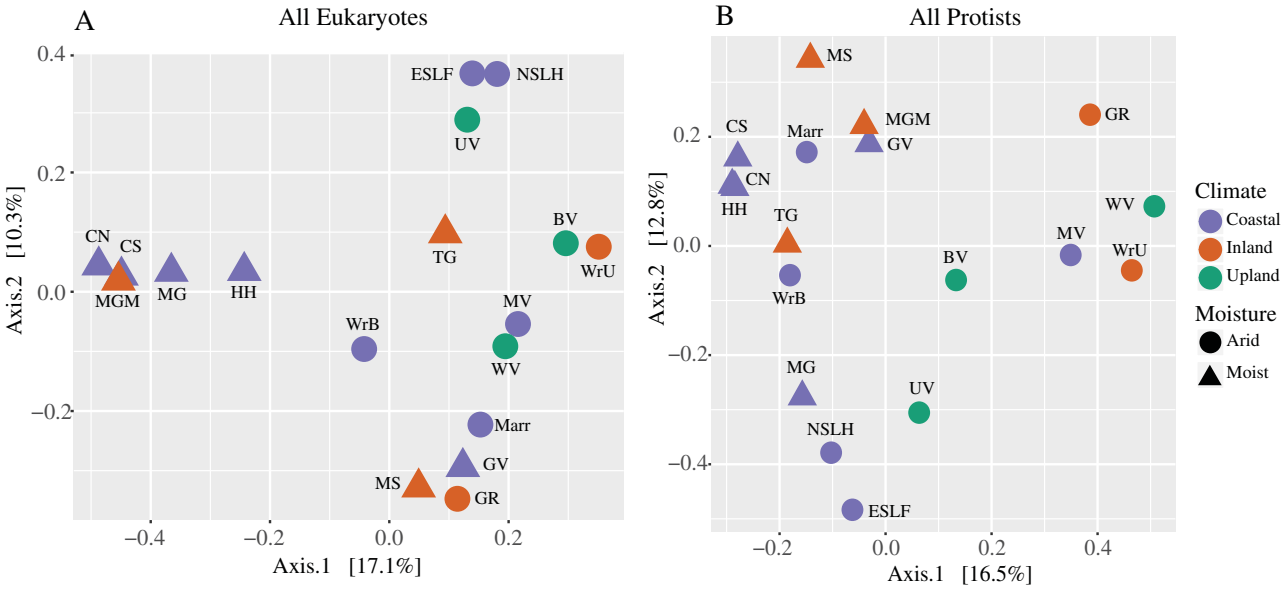


**Fig 5.** Site community composition comparisons.
A,B. PCoAs were performed with Bray-Curtis distance matrices and show distribution of sites according to community composition for (A) all eukaryotes and (B) all protists. Site symbols denote site moisture category (shape) and climate zone (shape).

explained by multicellular eukaryote (i.e. the metazoa, plants and fungi in the data set) individuals possessing several orders of magnitude more cells than a unicellular protist. Thus, low protist read count likely reflects lower protist biomass rather than lower protist abundance relative to other eukaryotes (Schenk *et al*., 2019). Certain protist morphogroups (flagellates and amoeba) are already known to be more abundant than metazoa in the MDV (Bamforth *et al*., 2005).

The true diversity of protists in these soils is likely to be higher than our study shows: our sequencing depth is insufficient for all sites (Fig. S1), our sampling of MDV habitats is not comprehensive, and our methods bias against novel taxa and the lower genomic content of eukaryotes compared with prokaryotes (Fierer, 2017; Mahé *et al*., 2017). However, an unexpectedly high sequencing depth for Beacon Valley (BV) (Table S3) still resulted in a eukaryotic OTU count similar to other upland climate zone sites (e.g. WV and UV), suggesting deeper sequencing in at least the upland climate zones may not capture additional eukaryotic taxa.

Ciliophora possessed the highest number of genera, but each genus occurred in low relative abundance, suggesting there is a high degree of either functional diversity or redundancy among Ciliophora in the MDV (Fig. 2D). Cercozoa are the most widely distributed of any eukaryote group, occurring in every site except for lower WrB (Fig. 3C). As the MDV are dominated by arid soils (Burkins *et al*., 2001), cercozoan taxa may therefore be the dominant eukaryotic consumers. The dominance of Ciliophora and Cercozoa and the prevalence of Sandonidae and *Rhogostoma* in our data reflect global patterns of soil protist diversity (Oliverio *et al*., 2020). That these trends hold in an environment as extreme as the MDV suggests that (i) similar processes structure soil protists in environments as different as cold deserts and tropical soils, (ii) at least some of these processes may function independently of the physical environment and (iii) Ciliophora and Cercozoa are more diverse or possess a higher proportion of ecological generalists than other protist phyla. Future research that identifies Antarctic Sandonidae, *Rhogostoma* species and other phagotrophic protists as either physiologically specialized endemics or cosmopolitan generalists will clarify the apparent similarities between extreme and temperate soil communities.

Our results confirm the richness and prevalence of Chlorophyta and diatoms (Stramenopiles) in terrestrial MDV environments (Fig. 3B) noted by others (Cavacini, 2001; Fumanti and Cavacini, 2005; Adams *et al*., 2006). Putative parasites are rare in our data (Fig. 3A), which contrasts with their dominance in other soil data sets (Bates *et al*., 2013; 2018; Geisen *et al*., 2015; Mahé *et al*., 2017), but is explained by the paucity of higher order hosts (microarthropods and plants) that occur in these soils (Adams *et al*., 2006; Barrett *et al*., 2006). The patchy distribution of eukaryotic parasites, phototrophs and predators (Shaw *et al*., 2018) indicates these soils could be used to investigate the biological and physical controls governing trophic structuring and energy transfer in terrestrial environments (Leibold *et al*., 1997; Barbier and Loreau, 2019). Moisture is one of the main drivers of soil protist diversity (Bates *et al*., 2013; Geisen *et al*., 2014; Oliverio *et al*., 2020), and this trend is supported by our analysis. The differences in richness between moist and arid sites is more pronounced than differences between climate zones (Fig. 4A), although this may be a statistical anomaly owing to low sample count per group. Increasing elevation and distance from the coast are correlated with changes in edaphic parameters in the MDV (Marchant and Head, 2007), suggesting the influence of landscape variables on MDV protists is ultimately conflated with that of soil moisture. Community composition at higher taxonomic levels remains relatively stable across important MDV environmental gradients (Fig. 4B).

## Conclusions

Using polymerase chain reaction-free metagenomics and a regional-scale sampling scheme, this study demonstrates that protists are a highly diverse and integral component of a broad variety of MDV soil ecosystems. Next steps include investigating the ecological role prominent protists play in this ecosystem and to what degree their activity contributes to ecosystem functioning and processes in the MDV. Metatranscriptomics, in vitro experiments on isolated strains, and stable isotope probes can be used to begin addressing these questions, but methodologies must account for high habitat heterogeneity and low eukaryote abundance by broadening sampling schemes and increasing soil volume for analyses. Characterizing the interactions these organisms have with other members of the soil community will improve the accuracy and applicability of the MDV soil model to soil ecosystems in non-Antarctic environments.

## Experimental procedures

### *Sampling methods and locations*

Our sampling design aimed to capture the high heterogeneity of MDV soil parameters (i.e. salinity, moisture, organic matter and temperature) (Barrett *et al*., 2004) and comprises 87 soil samples from 18 sites in 11 lake basins in 9 valleys across latitudinal, moisture and elevation gradients and a range of distances to the Ross Sea (Fig. 1, Table S1). Site names and acronyms are included together throughout this paper for clarity in relating the

 

text to the figures. In cases where a site is referenced multiple times in the same subsection, acronyms will be used exclusively after the first instance. Sites in Taylor valley, the main focus area of the MCM LTER research program, are eastside Lake Fryxell (ESLF), HH, near Marr Pond (Marr), NSLH and CS. Notably, CS is considered to be one of the most productive terrestrial sites in the MDV (McKnight *et al*., 1999), and UV is one of the most inhospitable terrestrial habitats on Earth (Goordial *et al*., 2016). Samples from ESLF, NSLH, BV, UV, GR and WV were typical mineral soils with no visible moisture or organic matter. Soils from three sites had visible organic matter, noted hereafter with Φ: HH (chunks of moss), CN (bits of algae) and CS (biocrust). Samples were collected during the 2014, 2015, 2016 and 2017 field seasons following the sampling procedure outlined in Freckman and Virginia (1993). Briefly, sterilized scoops were used to remove the top 10 cm (where possible) of soil into sterile Whirl-Pak® bags. Soils were placed in insulated coolers and returned to McMurdo Station via helicopter at ambient air temperature (~ −5°C) and gradually cooled over a week to −20°C for long-term storage. Prior to subsequent analyses, samples were thawed over a week's time to 10°C and subsampled. Working with psychrotolerant organisms necessitates freezing samples to prevent biological activity, but crossing the boundary from liquid to solid water can be highly stressful and may potentially alter communities (Knox *et al*., 2015). Therefore, samples were frozen and thawed as infrequently and slowly as practical throughout the course of this study.

### *Environmental parameters*

Subsamples were submitted to the Environmental Analytics Lab at Brigham Young University to measure moisture, pH, electrical conductivity, total N, total C and total P, $NO_3$-N, C:N ratio and texture of each individual soil sample, except for CS (Table S1). To measure moisture content, samples were weighed before and after drying overnight at 105°C. The pH was measured according to Rhodes (1982), electrical conductivity using a RC-16C Conductivity Bridge (Beckman Instruments, Brea, CA), $NO_3$-N following Sims and Jackson (1971), total C and N on a TruSpec CN Determinator (LECO Instruments, St. Joseph, MI) following McGeegan and Naylor (1988), texture following Day (1965) and extracted total P with 0.5 M sodium bicarbonate following Olsen *et al*. (1954). Elevation (m), distance to coast (km) and aspect (°) were measured using the Antarctic reference elevation model of antarctica (REMA) explorer (Howat *et al*., 2019). Environmental parameter values were averaged from individual samples (Table S1) for each site (except CS) to perform our analyses (Table S2). Sites were labelled as arid if they possessed an average moisture content < 5% and as moist if ≥ 5% (Burkins *et al*., 2001). Climate zones were determined for 13 of 18 sites using Fig. 5 of Fountain *et al*. (2014) but could not be determined for the remaining 5 sites (MS, TG, GR, MGM and CN) because the figure did not include those regions. Instead, we designated climate zones for the remaining 5 sites by comparing their elevations with the elevations of the 13 sites for which a climate zone could be directly determined.

### *DNA extraction and sequencing*

Prior to DNA extraction, 5 g of each sample belonging to the same site was pooled using sterile techniques in order to obtain 18 metagenomes. Whole genomic DNA was extracted from each of these representative samples using the DNeasy PowerSoil Kit (Qiagen) following a modified protocol recommended by Qiagen for use with soils with extremely low DNA content. Briefly, 1.8–2 g of soil instead of 0.5 g was used per reaction; solution C3 (100 μl instead of 200 μl) was added immediately after solution C2 (100 μl instead of 200 μl) without an intervening incubation; and half the volume of eluate (50 μl instead of 100 μl) was incubated at room temperature on the surface of the filter for 1 min prior to elution. DNA was extracted from each subsample until enough DNA (≥ 2 μg) had been recovered to avoid the need for an amplification step during library preparation, which would have potentially complicated interpretations of community structure. HH, MG, CS and CN samples had noticeably higher DNA content than the other sites. Therefore, only two extractions (4 g) were performed on the high DNA samples while 12–20 g of material was used for the other samples (Table S3).

The concentration and quality of extracted DNA were assessed using a Qubit 4 Fluorometer (Invitrogen; Q33226) and a NanoDrop 2000C (Thermo Fisher) respectively. Shotgun metagenome libraries were made with the NEBNext Ultra II DNA Prep kit (New England Biolabs) with custom primers from Integrated DNA Technologies and sequenced over one and one-half lanes on an Illumina HiSeq 2500 in the Rapid Run mode with read lengths set to 2 × 250 bp and a total insert length of 500 bp. Because communities from the mesic habitats of CS, CN and HH were likely to harbor significantly more taxa than dry mineral soils (Cavacini, 2001; Adams *et al*., 2006; Wall, 2007), DNA from mesic sites was loaded onto the lanes with twice the concentration of other sites to increase coverage of genomic diversity (Table S3). Sequencing over two lanes resulted in two sets of paired-end files for each metagenome library, with each set containing a single right read file and a single left read file. To create a larger data set for analysis, corresponding read files were concatenated for each corresponding library: for example, right reads for the HH metagenome library on

lane 1 were concatenated with right reads from the HH metagenome library on lane 2. FastQC (Andrews, 2010) was used to determine where to trim reads and trimming was done using Trimmomatic with the following settings LEADING 2, TRAILING 2, SLIDINGWINDOW 4:15 (default), MINLEN 30 (Bolger *et al*., 2014). Paired-end reads that passed trimming (Table S3) were merged using FLASH with the following parameters: -m 10 (min overlap) (default) -M blank (max-overlap) -x 0.20 (max mismatch density) -r 250 (average read length) -f 500 (fragment length) -s 50 (Magoč and Salzberg, 2011) (Table S3). Unassembled forward and reverse reads from each sample were uploaded to MG-RAST (Wilke *et al*., 2015), under the project name SVLSoil18Proj090218.

### *Taxonomic assignment*

The Metaxa2.2 default SILVA database, *E*-value cutoff (–*E*) of 1e-5, –-allow_single_domain 1e-5.0 and –N 1, was used to extract and assign taxonomy to bacterial and archaeal 16S rRNA sequences(Bengtsson-Palme *et al*., 2015). Sequences with reliability scores lower than 80 (1639 sequences), shorter than 200 bp (35,407 sequences), and with per cent identity lower than 80% (38 sequences) were subsequently removed. Sequenced libraries were searched for eukaryotic rDNA using nhmmer (Wheeler and Eddy, 2013) with the eukaryote hmm profile developed for the rRNA prediction software Barrnap and an *e*-value cutoff of 1e-5 (Seemann, 2018). Recovered eukaryote SSU sequences were converted to fasta format using the esl-reformat miniapp provided with the hmmer software package, version 3.2.1 (Eddy, 2018); these sequences were aligned against the $PR^2$ database, version 4.10.0 (Guillou *et al*., 2013) using BLASTn v2.7.1 + (Camacho *et al*., 2009). For each query sequence, the hit with the highest bit score and lowest *E*-value was retained as the taxonomic assignment. Sequences shorter than 200 bp were removed. The eukaryote SSU sequences were then aligned against the NCBI nt database using BLASTn v2.7.1 (Camacho *et al*., 2009; NCBI, 2018) to confirm taxonomic assignments. Higher-order taxonomy for each eukaryotic assignment was checked against Mycobank, Integrative Taxonomic Information System (ITIS), Algaebase and the most recent classification revision for eukaryotes (Robert *et al*., 2013; Adl *et al*., 2019; Guiry and Guiry, 2018; ITIS, 2018). The paucity of eukaryotic reads in our libraries precluded assembly and the creation of traditional OTUs (Guo *et al*., 2016); instead, OTU analogs were obtained by grouping sequences with the same assigned genus (Table S4). Abundance counts for each site were normalized using the total counts method (Pereira *et al*., 2018) (Table S3).

Extracted 18S rRNA sequences were analysed with the phyloseq version 1.30.0 (McMurdie and Holmes, 2013) and vegan version 2.3-5 (Oksanen *et al*., 2016) packages in R version 3.2.2 (R Core Development Team, 2016). Rarefaction curves were generated to visualize sequencing depth (Fig. S1). Permutational multivariate analysis of variance (PERMANOVA) and principle coordinate analyses (PCoAs) were run on normalized abundance counts using Bray-Curtis distance matrices. To examine phototrophic protists alone, five sites were removed from the analysis due to an absence of phototrophic protist counts: WV, Upper Wright Valley (WrU), BV, GR and NSLH. To compare observed OTU richness between moist and arid sites and sites within different climate zones, we multiplied abundance counts by 10 and rounded the result to create discrete counts without losing sites with low counts (i.e. GR, BV, UV, WrB, WrU, ESLF, NSLH and MV). Thus, 10 and 8 sites were categorized as moist and arid, respectively, and 10, 5, and 3 sites were categorized as coastal thaw, inland mixed and stable upland respectively (Table S2). Sites from stable upland and inland mixed climate zones were combined to increase statistical power. Broad trophic functional categories – consumer, parasite, phototroph, saprotoph, mixotroph and unsure – were assigned to all eukaryote genera based on the literature (Table S4).

## Acknowledgements

This research was funded by the National Science Foundation Grant #OPP-1637708 to Byron J. Adams and is a contribution to the McMurdo Dry Valleys Long Term Ecological Research (LTER) program. Additional support was provided by the Monte L. Bean Life Science Museum and the Department of Biology, Brigham Young University. Stefan Geisen acknowledges funding from the Netherlands Organization for Scientific Research (NWO-VENI, 016.Veni.181.078). Andrew R. Thompson would like to thank Coltin Romney and Hailey Olsen for their assistance in data generation; Zach Aanderud and Natasha Griffin for their help with certain preliminary analyses; and Stuart Bamforth for early and critical discussions on protist diversity in the dry valleys. Our paper was significantly improved by the constructive criticisms of two reviewers, Christophe Seppey and David M. Wilkinson, and we would like to thank them for their time and effort.

## Conflict of interest

The authors declare no conflict of interest.

## References

Adams, B.J., Bardgett, R.D., Ayres, E., Wall, D.H., Aislabie, J., Bamforth, S., *et al*. (2006) Diversity and distribution of Victoria land biota. *Soil Biol Biochem* **38**: 3003–3018.

Adl, S.M., Bass, D., Lane, C.E., Lukes, J., Schoch, C.L., Smirnov, A., *et al*. (2019) Revisions to the classification,

nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology* **66**: 4–119. https://doi.org/10.1111/jeu.12691.

Andrews, S. 2010. FastQC: a quality control for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ayres, E., Nkem, J.N., Wall, D.H., Adams, B.J., Barrett, J.E., Broos, E.J., *et al*. (2008) Effects of human trampling on populations of soil Fauna in the McMurdo dry valleys, Antarctica. *Conserv Biol* **22**: 1544–1551.

Bamforth, S.S., Wall, D.H., and Virginia, R.A. (2005) Distribution and diversity of soil protozoa in the McMurdo dry valleys of Antarctica. *Polar Biol* **28**: 756–762.

Barbier, M., and Loreau, M. (2019) Pyramids and cascades: a synthesis of food chain functioning and stability. *Ecol Lett* **22**: 405–419.

Bardgett, R.D. (2002) Causes and consequences of biological diversity in soil. *Zoology (Jena)* **105**: 367–374.

Barrett, J.E., Virginia, R.A., Hopkins, D.W., Aislabie, J., Bargagli, R., Bockheim, J.G., *et al*. (2006) Terrestrial ecosystem processes of Victoria land, Antarctica. *Soil Biol Biochem* **38**: 3019–3034.

Barrett, J.E., Virginia, R.A., Wall, D.H., Parsons, A.H., Powers, L.E., and Burkins, M.B. (2004) Variation in biochemistry and soil biodiversity across spatial scales in a polar desert ecosystem. *Ecology* **85**: 3105–3118.

Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R., and Fierer, N. (2013) Global biogeography of highly diverse protistan communities in soil. *ISME J* **7**: 652–659.

Bengtsson-Palme, J., Hartmann, M., Eriksson, K.M., Pal, C., Thorell, K., Larsson, D.G., and Nilsson, R.H. (2015) METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour* **15**: 1403–1414.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Buelow, H.N., Winter, A.S., Van Horn, D.J., Barrett, J.E., Gooseff, M.N., Schwartz, E., and Takacs-Vesbach, C.D. (2016) Microbial community responses to increased water and organic matter in the arid soils of the McMurdo dry valleys, Antarctica. *Front Microbiol* **7**: 1040.

Burkins, M.B., Virginia, R.A., and Wall, D.H. (2001) Organic carbon cycling in Taylor Valley, Antarctica: quantifying soil reservoirs and soil respiration. *Glob Chang Biol* **7**: 113–125.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf* **10**: 421.

Cavacini, P. (2001) Soil algae from northern Victoria land (Antarctica). *Polar Biosci* **14**: 45–60.

Chakraborty, S., Pangga, I.B., and Roper, M.M. (2012) Climate change and multitrophic interactions in soil: the primacy of plants and functional domains. *Glob Chang Biol* **18**: 2111–2125.

Courtright, E.M., Wall, D.H., and Virginia, R.A. (2001) Determining habitat suitability for soil invertebrates in an extreme environment: the McMurdo dry valleys, Antarctica. *Antarct Sci* **13**: 9–17.

Crotty, F.V., Adl, S.M., Blackshaw, R.P., and Murray, P.J. (2012) Protozoan pulses unveil their pivotal position within the soil food web. *Microb Ecol* **63**: 905–918.

Day, P.R. (1965) Particle fractionation and particle size analysis. In *Methods of Soil Analysis. Part 1. Agronomy*. Madison, WI: American Society of Agronomy, pp. 545–566.

Doran, P.T., Lyons, W.B., and McKnight, D.M. (2010) *Life in Antarctic Deserts and Other Cold Dry Environments: Astrobiological Analogs*. Cambridge: Cambridge University Press.

Doran, P.T., Priscu, J.C., Lyons, W.B., Walsh, J.E., Fountain, A.G., McKnight, D.M., *et al*. (2002) Antarctic climate cooling and terrestrial ecosystem response. *Lett Nat* **415**: 517–520.

Eddy, S. R. 2018). *HMMER: Biosequence Analysis Using Profile Hidden Markov Models*. http://hmmer.org/.

Elton, C.S. (1927) *Animal Ecology*. London: Sidgwick and Jackson.

Fell, J.W., Scorzetti, G., Connell, L., and Craig, S. (2006) Biodiversity of micro-eukaryotes in Antarctic Dry Valley soils with < 5% soil moisture. *Soil Biol Biochem* **38**: 3107–3119.

Fierer, N. (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol* **15**: 579–590.

Foissner, W. (1996) Faunistics, taxonomy and ecology of moss and soil ciliates (Protozoa, Ciliophora) from Antarctica, with description of new species, including *Pleuroplitoides smithi* gen. n., sp. n. **35 2**: 95–123.

Fountain, A.G., Levy, J.S., Gooseff, M.N., and Van Horn, D. (2014) The McMurdo dry valleys: a landscape on the threshold of change. *Geomorphology* **225**: 25–35.

Freckman, D.W., and Virginia, R.A. (1993) Extraction of nematodes from Dry Valley Antarctic soils. *Polar Biol* **13**: 483–487.

Freckman, D.W., and Virginia, R.A. (1997) Low-diversity Antarctic soil nematode communities: distribution and response to disturbance. *Ecology* **78**: 363–369.

Freckman, D.W., and Virginia, R.A. (1998) Soil biodiversity and community structure in the Mcmurdo dry valleys, Antarctica. In *Ecosystem Dynamics in a Polar Desert: The Mcmurdo Dry Valleys*, Priscu, J. (ed). Antarctica, pp. 323–335). American Geophysical Union. https://dx.doi.org/10.1029/AR072p0323.

Fumanti, B., and Cavacini, P. (2005) *ANTADATA: a Database on Biogeography of Non-Marine Algae in Continental Antarctica*. CD-ROM: Casa Editrice Università "La Sapienza", Rome.

Geisen, S., Bandow, C., Römbke, J., and Bonkowski, M. (2014) Soil water availability strongly alters the community composition of soil protists. *Pedobiologia* **57**: 205–213.

Geisen, S., Mitchell, E.A.D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., *et al*. (2018) Soil protists: a fertile frontier in soil biology research. *FEMS Microbiol Rev* **42**: 293–323.

Geisen, S., Tveit, A.T., Clark, I.M., Richter, A., Svenning, M. M., Bonkowski, M., and Urich, T. (2015) Metatranscriptomic census of active protists in soils. *ISME J* **9**: 2178–2190.

Goordial, J., Davila, A., Lacelle, D., Pollard, W., Marinova, M.M., Greer, C.W., *et al*. (2016) Nearing the cold-arid

limits of microbial life in permafrost of an upper dry valley, Antarctica. *ISME J* **10**: 1613–1624.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., *et al*. (2013) The Protist ribosomal reference database ($PR^2$): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41**: D597–D604.

Guiry, M.D., and Guiry, G.M. (2018) *AlgaeBase*. Galway: World-wide Electronic Publication, National University of Ireland.

Guo, J., Cole, J.R., Zhang, Q., Brown, C.T., and Tiedje, J.M. (2016) Microbial community analysis with ribosomal gene fragments from shotgun metagenomes. *Appl Environ Microbiol* **82**: 157–166.

Howat, I.M., Porter, C., Smith, B.E., Noh, M.J., and Morin, P. (2019) The reference elevation model of Antarctica. *Cryosphere* **13**: 665–674.

ITIS. 2018. *Integrative Taxonomic Information System*. http://www.itis.gov.

Knox, M.A., Wall, D.H., Virginia, R.A., Vandegehuchte, M.L., Gil, I.S., and Adams, B.J. (2015) Impact of diurnal freeze–thaw cycles on the soil nematode *Scottnema lindsayae* in Taylor Valley, Antarctica. *Polar Biol* **39**: 583–592.

Leibold, M.A., Chase, J.M., Shurin, J.B., and Downing, A.L. (1997) Species turnover and the regulation of trophic structure. *Ann Rev Ecol Syst* **28**: 467–494.

Magoč, T., and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., *et al*. (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol* **1**: 0091.

Marchant, D.R., and Head, J.W. (2007) Antarctic dry valleys: microclimate zonation, variable geomorphic processes, and implications for assessing climate change on Mars. *Icarus* **192**: 187–222.

McGeegan, S.L., and Naylor, D.V. (1988) Automated instrumental analysis of carbon and nitrogen in plant and soil samples. *Commun Soil Sci Plant Anal* **19**: 493–505.

McKnight, D., Niyogi, D., Alger, A., Bomblies, A., Conovitz, P., and Tate, C. (1999) Dry valley streams in Antarctica: ecosystems waiting for water. *Bioscience* **49**: 985.

MCMLTER. 2019. *McMurdo Dry Valleys LTER*. http://mcm.lternet.edu/.

McMurdie, P.J., and Holmes, S. (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

NCBI Resource Coordinators. (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **46**: D8–D13.

Nielsen, U.N., Ayres, E., Wall, D.H., and Bardgett, R.D. (2011) Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity–function relationships. *Eur J Soil Sci* **62**: 105–116.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al.* (2016) *Vegan: community ecology package. R package version 2.3-5.* https://CRAN.R-project.org/package=vegan.

Oliverio, A.M., Geisen, S., Delgado-Baquerizo, M., Maestre, F.T., Turner, B.L., and Fierer, N. (2020) The global-scale distributions of soil protists and their contributions to belowground systems. *Sci Adv* **6**: eaax8787.

Olsen, S.R., Cole C.V., Watanabe F.S., and Dean L.A. (1954) Estimation of available phosphorus in soil by extraction with sodium bicarbonate. *USDA Circ*., **939**: 1–19.

Pereira, M.B., Wallroth, M., Jonsson, V., and Kristiansson, E. (2018) Comparison of normalization methods for the analysis of metagenomic gene abundance data. *BMC Genomics* **19**: 274.

R Core Development Team. (2016) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R foundation for Statistical Computing.

Rhodes, J.D. (1982) Soluble salts. In *Methods of Soil Analysis Part 2*, Page, A.L. (ed). Madison, WI: American Society of Agronomy, Inc., pp. 167–179.

Robert, V., Vu, D., Amor, A.B.H., van de Wiele, N., Brouwer, C., Jabas, B., *et al*. (2013) MycoBank gearing up for new horizons. *IMA Fungus* **4**: 371–379.

Schenk, J., Geisen, S., Kleinboelting, N., and Traunspurger, W. (2019) Metabarcoding data allow for reliable biomass estimates in the most abundant animals on earth. *Metabarcoding Metagenomics* **3**: e46704. https://mbmg.pensoft.net/article/46704/list/8/.

Seemann, T. (2018) *Barrnap 0.9: Rapid Ribosomal RNA Prediction*. https://github.com/tseemann/barrnap.

Shaw, E.A., Adams, B., Barrett, J., Lyons, W., Virginia, R., and Wall, D. (2018) Stable C and N isotope ratios reveal soil food web structure and identify the nematode Eudorylaimus antarcticus as an omnivore–predator in Taylor Valley, Antarctica. *Polar Biology* **41**: 1013–1018. https://doi.org/10.1007/s00300-017-2243-8.

Sims, J.R., and Jackson, G.D. (1971) Rapid analysis of soil nitrate with chromotropic acid. *Soil Sci Soc Am J* **35**: 603–606.

Strassert, J.F.H., Jamy, M., Mylnikov, A.P., Tikhonenkov, D.V., and Burki, F. (2019) New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. *bioRxiv*, **36**: 757–765.

Virginia, R.A., and Wall, D.H. (1999) How soils structure communities in the Antarctic dry valleys. *Bioscience* **49**: 973–983.

Wall, D.H. (2005) Biodiversity and ecosystem functioning in terrestrial habitats of Antarctica. *Antarct Sci* **17**: 523–531.

Wall, D.H. (2007) Global change tipping points: above- and below-ground biotic interactions in a low diversity ecosystem. *Philos Trans R Soc Lond B Biol Sci* **362**: 2291–2306.

Wheeler, T.J., and Eddy, S.R. (2013) Nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**: 2487–2489.

Wilke, A., Bischof, J., Harrison, T., Brettin, T., D'Souza, M., Gerlach, W., *et al*. (2015) A RESTful API for accessing microbial community data for MG-RAST. *PLoS Comput Biol* **11**: e1004008.

Wilkinson, D.M., Creevy, A.L., and Valentine, J. (2012) The past, present and future of soil protist ecology. *Acta Protozool* **51**: 189–199.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix S1** Supplemental materials.

**Fig. S1**: Rarefaction curves of whole eukaryotic community.

**Fig. S2**: Site community composition comparisons.

**Table S5**: Welch's two sample student T-test on sites categorized by moisture and climate zones

**Table S6**: Output tables for PERMANOVA

**Appendix S2** Supporting Information Tables S1–S4.

**Table S1**: Individual Sample Info.

**Table S2**: Averages and standard deviations for environmental variables for all sites.

**Table S3**: Sequencing and processing statistics.

**Table S4**: OTUs.

**Appendix S3** Sequences master list.