Biotic interactions are an unexpected yet critical control on the complexity of an abiotically driven polar ecosystem

Charles K. Lee et al.*

Abiotic and biotic factors control ecosystem biodiversity, but their relative contributions remain unclear. The ultraoligotrophic ecosystem of the Antarctic Dry Valleys, a simple yet highly heterogeneous ecosystem, is a natural laboratory well-suited for resolving the abiotic and biotic controls of community structure. We undertook a multidisciplinary investigation to capture ecologically relevant biotic and abiotic attributes of more than 500 sites in the Dry Valleys, encompassing observed landscape heterogeneities across more than 200 km². Using richness of autotrophic and heterotrophic taxa as a proxy for functional complexity, we linked measured variables in a parsimonious yet comprehensive structural equation model that explained significant variations in biological complexity and identified landscape-scale and fine-scale abiotic factors as the primary drivers of diversity. However, the inclusion of linkages among functional groups was essential for constructing the best-fitting model. Our findings support the notion that biotic interactions make crucial contributions even in an extremely simple ecosystem.
Understanding how ecosystems self-organize at landscape scales has long been a formidable challenge in ecology since the trophic complexity of most ecosystems obscures the relative contributions of the biotic and abiotic factors regulating biological diversity. Given the fundamental effects of biodiversity on ecosystem function, a critical task is to resolve the relative importance of three sets of ecological factors that drive community structure: abiotic environmental filtering, dispersal limitation in space, and biotic interactions (e.g., competition, mutualism, and trophic relationships). Through and spatially explicit descriptions of these ecosystem drivers are required for this task, but the complexity of most ecosystems creates enormous logistical obstacles.

Biotic interactions, including those among higher eukaryotes and those between higher eukaryotes and microorganisms, have long been recognized as important drivers of ecosystem structure and function. However, attempts to capture biotic interactions at the ecosystem level have often been restricted both by sampling approaches and/or the expertise of individual investigators, and studies have largely focused on testing hypotheses associated with pre-identified biotic interactions or ecosystem components. A comprehensive investigation of abiotic and biotic interactions within an ecosystem therefore requires a sampling design that is consistent across all major biological groups present. It also requires an explicitly interdisciplinary and comprehensive approach for data collection and analysis of both abiotic and biotic variables.

For microorganisms (bacteria, archaea, and unicellular fungi), culture-independent characterization using molecular genetic techniques is widely recognized as the most consistent and sensitive approach, whereas conventional surveys remain the most reliable and practical approach for larger invertebrates and higher animals and plants in terrestrial environments. For abiotic variables and some major macroecological features (e.g., primary productivity), geographic information system (GIS) has become an essential tool for collecting information across spatial scales. The integration of GIS and remote sensing technologies (e.g., satellites) can provide spatially explicit environmental information for entire geographic regions. Specifically, the availability of high-resolution data layers from sources such as the Landsat 7 and MODIS satellites facilitates complete and consistent descriptions of environmental conditions (e.g., surface temperature and snow coverage) at the landscape scale. Using these descriptions in conjunction with information on bedrock geology and geomorphology, it is now feasible to carry out systematic landscape-scale surveys that capture heterogeneities in abiotic conditions within an ecosystem. Additionally, all information collected within a GIS-enabled sampling framework is spatially explicit and enables thorough examinations of dispersal limitation effects across multiple spatial scales.

Despite the advances in methodologies, disentangling the relative roles of abiotic and biotic controls on the complexity in terrestrial ecosystems is still a major challenge in ecology. We propose that extreme ecosystems offer a natural laboratory to reduce this complexity while representing its major features. Here, we offer an analysis of the controls on the biological complexity of a region of the McMurdo Dry Valleys of Antarctica. Located between the Polar Plateau and the Ross Sea (Fig. 1), the McMurdo Dry Valleys (hereinafter the Dry Valleys) are the largest contiguous ice-free area on the Antarctic continent and subject to some of the most extreme conditions of any terrestrial habitat on Earth, which severely constrain the range of biota present. Vascular plants and vertebrates are entirely absent, and soils are predominantly ultraoligotrophic, hyperarid, and often hypersaline. Consequentially, abiotic factors are widely regarded as the primary force shaping the ecology of Dry Valley soils. The extreme environmental conditions and lack of evidence for critical biotic interactions have made ecologists hypothesize that these ecosystems are fundamentally constrained by abiotic factors and host some of the simplest trophic structures on Earth. Structural equation modelling (SEM) and biotic interactions were significantly nested (nestedness temperature = 17.56, P = 0.010), and the richness of multicellular taxa was strongly correlated (r = −0.99, P = 0.001) with the first NMS axis of multicellular community composition. Nematodes were by far the most frequently observed organism, occurring in 80% of the sampled tiles, followed by hypoliths (40%), rotifers (32%), lichens (25%), tardigrades (23%), springtails (19%), cyanobacterial mats (15%), mites (12%), and mosses (11%).

**Results**

**Correlations and nestedness of measured variables.** Our data indicated that richness (see Methods for definitions) was related to community composition in each of the three groups of organisms examined (i.e., multicellular taxa, cyanobacteria, and fungi), which was to be expected given the low species richness of each group in the analyzed system. Multicellular taxon communities were significantly nested (nestedness temperature = 17.56, P = 0.010), and the richness of multicellular taxa was strongly correlated (r = −0.99, P = 0.001) with the first NMS axis of multicellular community composition. Nematodes were by far the most frequently observed organism, occurring in 80% of the sampled tiles, followed by hypoliths (40%), rotifers (32%), lichens (25%), tardigrades (23%), springtails (19%), cyanobacterial mats (15%), mites (12%), and mosses (11%).

Cyanobacterial
communities were significantly nested (nestedness temperature = 2.01, \(P = 0.010\)), and cyanobacterial richness was correlated with both NMS axes (axis 1: \(r = -0.75\), axis 2: \(r = -0.66\), \(P = 0.001\)) derived from the cyanobacterial community matrix. Fungal communities were also significantly nested (nestedness temperature = 2.61, \(P = 0.010\)), and fungal richness was correlated with both NMS axes (axis 1: \(r = 0.61\), axis 2: \(r = -0.79\), \(P = 0.003\)) derived from the fungal community matrix.

**Structural equation models.** The initial \textit{a priori} SEM (Supplementary Figure 2) did not fit the data well (Comparative Fit Index [CFI] = 0.706, \(\chi^2 = 394.397\), df = 43, \(P < 0.0001\)). This initial model only included pathways that were well supported by previous empirical studies at the time the model was fit. This initial result made it clear that our \textit{a priori} expectations were missing important relationships within the ecosystem that were not well known in the literature. Consequently, to obtain a model with an implied covariance structure that matched the observed data well, we made two modifications. First, we identified missing pathways that contributed to poor model-fit by inspecting the residual covariance matrix and added as few of these theoretically plausible pathways as possible (e.g., direct pathways from abiotic variables to biotic variables). Second, we removed non-significant pathways. The final model (Fig. 2a) fits the data well (CFI = 0.996, \(\chi^2 = 45.018\), df = 35, \(P = 0.1196\)) and represents the most parsimonious model possible. Each pathway in the model is significant (the standardized path coefficients can be interpreted as partial correlation coefficients). The model explains between 30 and 40% of the variance in the richness variables (i.e., Multicellular Taxa S, Cyano S, and Fungal S), which are strongly correlated with community composition, as described above. Importantly, soil properties do not clearly mediate the effects of topography and climate on biotic diversity, and topographic and climate variables have many important direct pathways to biota. Attempts to trim the model by removing select pathways substantially reduced the goodness-of-fit of the model, so each pathway is important. This model is particularly valuable and robust because it explicitly accounts for spatial patterns that do not depend on environmental variables (see Methods). For example, some areas could be richer in species simply because they are located in regions that receive a higher supply of immigrants supporting local populations even when conditions are not favorable. The \textquote{total effects} (i.e., sums of direct and indirect effects) of each environmental variable on each biotic variable (Multicellular Taxa S, Cyano S, and Fungal S) indicate that elevation, slope, aspect, distance to coast, and wetness index all have significant total effects on richness and composition of multicellular taxon and microbial assemblages (Supplementary Table 1).

**Importance of biotic interactions.** It is also important to note that the positive links among the richness of multicellular taxon and microbial assemblages were essential to the model;
removing these pathways yielded very poor model-fit indices. The chosen directions of these pathways were guided by both empirical data and theory. There were eight possible combinations of directed paths among three variables, and after arriving at the final model (Fig. 2a), we tested all eight combinations to evaluate the sensitivity to the directions of these pathways. Four of these eight models yielded poor-fitting models ($P < 0.05$, specific models not shown), and each of these poor models included a pathway from multicellular taxa to fungi, which is strong evidence against that particular pathway. However, the other four models were indistinguishable from a model-fitting perspective (all $P > 0.05$, specific models not shown), and so we relied on theory to specify the direction of these pathways. Ecological theory supports pathways from cyanobacterial richness to fungal richness and from cyanobacterial richness to multicellular taxon richness, given the foundational contribution of these autotrophic single-celled organisms to this extreme ecosystem. Importantly, the positive covariance among the biota is not simply due to similar responses to abiotic conditions because each group responds individually to the sets of abiotic variables in the model (Fig. 2a). This implies that processes other than abiotic filtering drive the positive covariance among the three groups of biota.

**Relative contributions of abiotic and biotic factors.** Overall, environmental filtering imposed the strongest net effects on biotic richness (Table 1). Spatial processes were the second most important set of richness drivers, with nearly the same magnitude...
of effect as environmental filtering for cyanobacteria (Table 1). Biotic interactions were important in determining fungal and multicellular taxon richness, and their impact on multicellular taxa was comparable to that of spatial processes (Table 1). Finally, the SEM was used to generate spatially explicit predictions of biodiversity across the study area (Fig. 2b–d) to demonstrate its potential as a tool for understanding the spatial heterogeneity of soil biota in the Dry Valleys for both scientific investigation and environmental management.

**Discussion**

Earlier investigators had suggested that the species richness of Antarctic terrestrial vegetation south of 72°S is determined by local conditions. Our data and model show the prominence of abiotic drivers (in particular total soil N, soil wetness index, elevation, and distance to the coast) and support the general view that abiotic factors are the most important ecological filter in extreme environments. However, there is a new and large amount of variance in the species richness of major functional groups and their reciprocal correlations that is not accounted for by abiotic factors. Notably, soil ATP level (a proxy for biomass) was not significantly correlated with any of the other variables measured and removed from the final model (Fig. 2a). In the absence of other practical measures of biotic variables, we believe that species richness (which is significantly correlated with composition, see Methods) effectively captures biotic variables within this study.

Our model shows that the spatial autocorrelation vectors (i.e., independent of variation in abiotic factors) are the second strongest correlate of richness (Table 1). The spatial patterns accounted for by these autocorrelation vectors can be caused by both unmeasured biological processes (e.g., dispersal limitation) and legacy effects (e.g., historic distribution of glaciers and pro-glacial lakes). It is very unlikely that these spatial patterns are caused by major unmeasured environmental variables, given the quantity and quality of the environmental measurements collected in this study. It is possible that an unmeasured soil attribute (e.g., soil bulk density, water-holding capacity) could account for some of the patterns observed. However, given the variety of biotic variables captured in this study, it is unlikely that any single unmeasured abiotic variable will show strong correlation with measured biotic influences. Legacy effects are important in establishing and maintaining ice-free refugia for terrestrial biota, and the importance of the spatial vectors in the SEM thus potentially supports a role for legacies linked to glaciation geomorphology in shaping distributions of biota in the Dry Valleys.

Our model also shows that, besides spatial vectors and abiotic factors, a notable amount of variance in the species richness of each group is explained by correlations between biotic groups. As we further explain below, we hypothesize that these correlations reflect variation in the biological complexity of the ecosystem. Specifically, our model highlights some of the major linkages in the Dry Valleys ecosystem: in this system, cyanobacteria provide the energetic foundation of food webs, and their species richness does not appear to be influenced by the richness of fungi and multicellular taxa (Table 1). However, fungal richness was highest where cyanobacterial richness was high, and multicellular taxon richness was highest where both cyanobacterial and fungal richness was high (Fig. 2b–d), highlighting the fundamental importance of autotrophic cyanobacteria as the primary producers in this extreme ecosystem. We are confident that the positive covariance among the groups of organisms considered here is not confounded by covariation with abiotic conditions because the model explicitly allowed each group to respond to a unique combination of abiotic variables. Conversely, the set of correlations that link the species richness of the three groups are essential to the fit of the model; removing them from the model produces models that fit the data very poorly.

The Dry Valleys are arguably the simplest large-scale (4500 km² of ice-free area) ecosystem on the Earth. Therefore, a small increase in the richness of any of the three major groups in this system may imply a disproportionate increase in the biotic complexity of the system because every added species can bring in a new set of interactions between the three functional groups. This is consistent with the trophic theory of island biogeography, which is particularly relevant to systems such as the Dry Valleys because they experience dispersal limitation and disconnection between local communities. Specifically, trophic constraints (i.e., species need their resource to establish successfully) alter immigration and extinction dynamics, which ultimately determine species richness. In the Dry Valleys, food webs are particularly isolated compared to other soil food webs, which should reduce the recruitment of lower trophic level species for their consumers. The low connectivity of the food webs in the Dry Valleys is thus expected to translate into lower immigration and higher extinction rates. This is expected to create high spatial and temporal variability in species composition and richness, and contributes to food webs dominated by generalist primary consumers with very few secondary consumers. Overall, the patterns of species richness we observed are consistent with these dynamics and suggest that the diversity of the primary producers plays a central role in driving the diversity of the other organisms. A further implication is that understanding the drivers of microbial diversity will be central to predicting higher trophic level responses to environmental change, which is happening at marked rates in polar regions.

The empirical SEM (Fig. 2a) incorporated all measured factors, including soil physicochemical properties that cannot be obtained through remote sensing. Therefore, an additional SEM was derived using only unstandardized coefficients associated with factors that are obtainable through remote sensing and GIS (e.g., wetness index, temperature, elevation, aspect, distance to the coast, and slope) (Supplementary Figure 3). In the future, this “predictive” SEM can be used to make spatially explicit predictions of biodiversity across the entire Dry Valley landscape. Ultimately, the model and its future version can be used to support the development of best management practices for this unique ecosystem protected by the Antarctic Treaty System (http://www.ats.aq).

In conclusion, we found that abiotic factors such as soil temperature and topography had important direct effects on richness as well as indirect effects mediated through physicochemical soil properties (Fig. 2a). However, contrary to our expectations, we also found that the correlations between the functional groups and spatial autocorrelation in the variation of the richness of the

---

**Table 1 Net effects of various parameters on biological richness**

| Parameter                  | Abiotic | Spatial | Biotic |
|----------------------------|---------|---------|--------|
| Cyanobacteria              | 0.45    | 0.40    | 0      |
| Fungi                      | 0.42    | 0.34    | 0.20   |
| Multicellular Taxa         | 0.39    | 0.21    | 0.21   |

Net effects of abiotic environmental filters, spatial processes, and biotic interactions on cyanobacterial, fungal, and multicellular taxon richness. Effects were calculated using composite variables within the SEM and represent the absolute standardized path coefficients (ranging from 0 to 1).
functional groups are a major determinant of the biological diversity of the system. This result suggests that biotic factors are an underestimated control on the complexity of the Dry Valley ecosystem and raises the question of whether biotic interactions and processes have been similarly underappreciated in other simple and/or extreme ecosystems. Furthermore, our findings highlight the fundamental importance of incorporating biotic factors and spatial constraints when forecasting community responses to changing environmental conditions. This has direct relevance to more complex ecosystems where biotic interactions play a markedly greater role in shaping community structure and ecosystem functioning.

Methods

Study area. Approximately 0.4% of Antarctica is permanently ice-free, and the main ice-free areas are the Antarctic Peninsula, the McMurdo Dry Valleys, and various mountains and nunataks along the Transantarctic Mountains. Of these, the McMurdo Dry Valleys contain the largest contiguous ice-free areas (~4500 km²) and have been the focus of terrestrial biology research on the continent for the past 50 years. The McMurdo Dry Valleys are situated in southern Victoria Land along the western coast of McMurdo Sound (between 160°–164°E and 76°–78°S) and contain markedly complex surface geology and topography that result in highly heterogeneous physicochemical conditions in soils across the landscape. The area chosen for this study comprises 220 km² of largely ice-free terrain that includes Garwood, Marshall, and Miers Valleys as well as Shangri-La, an area west of Marshall Valley and enclosed by Joyce Glacier, Mt. Pams, and Mt. Lama (Fig. 1).

In addition to the extreme cold (mean annual air temperature of approximately −20 °C), the McMurdo Dry Valleys are characterized by strong winds, extreme aridity (precipitation of <10 cm per year water equivalent), and lack of appreciable solar input for much of the year. Despite the extreme selective pressure, the Dry Valleys appear to sustain a functional but simple ecosystem comprised of prokaryotes, invertebrate fauna, and non-vascular flora. Cyanobacteria (both aquatic and edaphic) appear to be the main primary producers, although important photosynthetic activity occurs in lithic communities (i.e., endoliths, hypoliths, and chasmoidaliths) as well as mosses and lichens.

The invertebrate fauna consists of the microarthropods Collembolla (i.e., springtails) and Acari (i.e., mites), as well as a range of microinvertebrates including nematodes, tardigrades, and rotifers. Nematodes are the dominant invertebrate taxon across much of the landscape, and their distribution and abundance primarily correlate with the presence of liquid water, pH, salinity, and inorganic carbon. Taxonomic diversity for nematodes is low (five species), but abundances can be as high as hot desert soils. Bacteria (both aquatic and edaphic) appear to be the main primary producers, although important photosynthetic activity occurs in lithic communities (i.e., endoliths, hypoliths, and chasmoidaliths) as well as mosses and lichens.

Table 2 Landscape-scale variables captured by nzTABS

| Category                  | Variables                                    |
|---------------------------|----------------------------------------------|
| Remote Sensing and GIS    | Elevation, Slope, Aspect                      |
| Bedrock Geology           | Snow/Ice/Water Presence, Distance to the Coast |
| Glacial Geomorphology     | Soil Surface Temperature, Wetness Index      |
| Biology                   | Lichen and Moss (Abundance and Size)         |
| Geochemistry              | Endolith and Hypolith (Abundance)            |
|                           | Cyanobacterial Mat (Abundance and Size)      |
|                           | Invertebrates (Abundance and Taxonomy)       |
|                           | ATP Level                                    |
|                           | Bacterial Richness (ARISA)                   |
|                           | Fungal Richness (ARISA)                      |
|                           | pH                                           |
|                           | Conductivity                                 |
|                           | Water Activity (A_w)                         |
|                           | Total Soil Moisture Content                  |
|                           | Total Soil C & N                             |

Table 2: Landscape-scale variables captured by nzTABS

Variables used for tile delineation

DNA-based analysis of microbial communities. After quantification and quality check (Supplementary Methods), extracted DNA samples were used for molecular fingerprinting of bacterial (total and cyanobacteria-only) and fungal communities based on automated ribosomal intergenic spacer analysis (ARISA). Briefly, the intergenic spacer between the 16S and 23S rRNA genes of the bacterial/cyanobacterial ribosomal operon and the intergenic spacer between the 18S and 23S rRNA genes of the fungal ribosomal operon were amplified using PCR for each sample (Supplementary Methods).

ARISA fragment length profiles (Supplementary Data 1) were analyzed using an in-house pipeline (a combination of Applied Biosystem Peak Scanner and corresponding tile and representative of the geographic and geologic attributes for the tile. At each sampling location (GPS coordinates and elevation were recorded), the top 10 cm (top 2 cm for prokaryotes) of soil was collected aseptically using a trowel from multiple spots within a 1 m² area for the following subsamples (Fig. 3): bulk soil (~400 g) with large pebbles (~2 cm diameter) removed aseptically and homogenized in a sterile 42 oz. Whirl-Pak; soil (~20 g) for moisture content measurement, subsampled from homogenized bulk soil into a sterile 15 mL centrifuge tube sealed with Parafilm; soil (~300 g) for microinvertebrate count, stored in a sterile 18 oz. Whirl-Pak (pebbles not removed to minimize disturbance).

A microarthropod survey (i.e., springtails and mites) was carried out by examining the underside of small (5–10 cm dia.), flat (<2 cm thickness), and preferably dark rocks within a 20 m radius of the soil sampling location for 10 min. The number and types of microarthropods observed were recorded, and the organisms were collected using an aspirator and preserved in a vial containing 100% ethanol for later analyses. A survey of vegetation (i.e., lichens, mosses, algae, and cyanobacterial mats and lithic communities) was carried out along a transect (20 m long and 2 m wide, 40 m²) adjacent to the soil sampling location.

Vegetation presence was recorded quantitatively in 100 cm² units for each taxon, and the numbers of observed lithic communities were recorded. All our activities were conducted in accordance with the McMurdo Dry Valley Antarctic Specially Managed Area manual and were deemed to have limited and transient impact to the environment according to the Preliminary Environmental Evaluation from the New Zealand Ministry of Foreign Affairs and Trade. All soil samples collected are stored at −60 °C or −80 °C at the International Centre for Terrestrial Antarctic Research at the University of Waikato.

Soil samples were subsequently aliquoted and analyzed for total ATP, pH, conductivity, water activity (A_w), total moisture content, microinvertebrate (i.e., nematodes, tardigrades, and rotifers) richness and abundance, and organic carbon and total nitrogen content (Fig. 3 and Supplementary Methods), as well as used for bulk DNA extraction (Supplementary Methods). All extracted DNA samples are available from the Antarctic Genetic Archive (AGAr, https://ictar.aa/antarctic-genetic-archive/) at the University of Waikato.
custom R and Python scripts, see Supplementary Data 2) that examines all peaks between 100 and 1200 base pairs for cyanobacterial electropherograms and 100 and 1400 base pairs for fungal electropherograms57. Peaks in these size ranges that made up greater than 0.3% of all peaks over 30 relative fluorescence units in each electropherogram were accepted as true peaks. The total number of true peaks was taken as a measure of taxon richness for each sample. Peaks within one base pair of one another were binned for the purpose of comparing electropherograms between samples.

ARISA was used to measure richness due to its proven ability to detect and discern diversity of edaphic cyanobacteria signals in the Dry Valleys over 165 rRNA gene PCR amplicons27,57 and its proven ability to capture fungal diversity patterns in Dry Valley soils in a consistent and cost-effective manner.

Environmental metadata. A number of key environmental attributes were derived from satellite imagery and the DEM, including surface soil temperature, a topographically derived “wetness index,” and distance to the coast. Soil surface temperatures were obtained from Landsat 7 ETM+ using band 6 (at 60 m resolution), which captured the up-welling thermal infrared spectrum (in the 10.4–12.5 μm band). Landsat 7-derived temperature data corresponding to locations of forty-five on-the-ground temperature loggers (DS1921G iButtons, Maxim Integrated, San Jose, CA) were compared with records from the iButtons, and significantly the shortest distance was calculated as the Euclidean distance (in meters) defined by cells with zero elevation in the DEM. Specifically, the shortest distance was determined by the perpendicular from the coastline to the sampling point. After quality control (removal of tiles with missing or questionable information, such as unintended duplicates and incorrect GPS location), data for 490 samples were included in the analysis.

Data analysis. A broad suite of geological, geographical, geochemical, hydrological, and biological variables (Table 2 and Supplementary Data 3) were collected and evaluated to derive the most parsimonious set of predictors of biodiversity in our study area. Biodiversity is represented by the richness of key autotrophic and heterotrophic groups and the presence of known taxa. Specifically, species richness of cyanobacteria and fungi was estimated using the number of ribosomal intergenic spacer length-polymorphic fragments observed from community fingerprinting analyses. These intergenic spacers exhibit length polymorphism across species and even at the intra-specific level, and the length profiles of PCR fragments are therefore indicative of the diversity and abundance of microbial communities. We note that these techniques do not resolve richness at a consistent taxonomic level; however, given that they can both over- and under-estimate species-level richness, we do not believe the results were influenced by systematic biases. Taxon richness for multicellular taxa was represented by the number of the following supraspecific taxa present in a sample: nematodes, rotifers, tardigrades, springtails, mites, cyanobacterial mats, mosses, lichens, and hypolithic consortia. These taxonomic groups also correspond to distinct trophic functional groups in the system. Specifically, the animals are all primary consumers of both bacteria and fungi, cyanobacteria are the main primary producers besides mosses, and fungi represent the major microbial decomposer group. Given the very low number of metazoan (Supplementary Figure 1), cyanobacterial57, and fungal species in the Dry Valleys, relatively small increase in the species richness of each compartment may imply a marked increase in the complexity of the system in terms of increased number of ecological interactions. To verify that inferences made from patterns in species richness apply similarly to community composition, richness was correlated with community composition in all three groups of organisms (i.e., cyanobacteria, fungi, and multicellular taxa) based on an analysis of nestedness (R script available upon request). Nestedness occurs when species-poor communities are generally subsets of species-rich communities, and when rare species tend to only occur in species-rich communities. The nestedness of each of the three community matrices was evaluated by calculating their respective ‘temperatures’, which determine whether species-poor communities are subsets of species-rich ones. The “temperatures” were calculated using the “nesttemp” function64 in the “vegan” library of R65, and their significance was assessed via permutation using the “ocecosim” function. The relationship between richness and non-metric multidimensional scaling (NMS) ordinations (based on Bray–Curtis similarity) of community composition (obtained using the “metaMDS” function in “vegan”)65 was quantified using correlation analysis.

Overall, all these preliminary analyses supported the assumption that in the specific system analyzed in this work, species richness of major functional groups is
the best metric to describe the richness of each group as well as the correlations between groups and the relationship between biota and abiotic factors. Cyanobacterial richness, rather than total bacterial richness, was included in our analysis for the following reasons. First, including both would effectively be “double-counting” since total bacterial richness includes cyanobacteria as well. Second, cyanobacteria are arguably the most critical group of bacteria, given their large proportional input to primary production in this extreme environment. Finally, cyanobacterial richness was significantly and positively correlated with total bacterial richness ($r = 0.31, P < 0.0001$), so knowing the richness of one group provides reasonable estimates about the richness of the other.

Biological communities closer in space are likely to be more similar in species richness and community composition. However, historical population- and landscape-level processes that are relatively independent of environmental conditions can also be a driver of Antarctic biodiversity. Thus, environmentally independent spatial variables were computed to account for spatial patterns linked to intrinsic population- and landscape-level processes, such as dispersal limitation or source-sink dynamics (see Supplementary Methods). Competition- and predation-related direct biotic interactions were not explicitly considered due to limited evidence for such interactions among Dry Valley biota.

To represent spatial patterns driven by intrinsic population- and community-level processes (e.g., limited dispersal), environmentally independent spatial variables were obtained as follows. First, optimal (in terms of describing spatial autocorrelation) combinations of Principal Coordinates of Neighbor Matrices (PCNM) were calculated. To explicitly model spatial patterns that are independent of environmental gradients, the PCNMs were regressed against all environmental variables to allow extraction of the residuals (aka “spatial residuals”). The “spatial residuals” were then used in a linear regression model to predict the three biotic richness variables, and their predicted values were derived. This was followed by a principal component analysis (PCA) on these predicted values, allowing spatial patterns to be summarized in the multivariate distribution of the three biotic richness variables. The first two components ($s1$ and $s2$) accounted for 90% of the environmentally independent spatial patterns. Finally, the net effect of spatial variation ($s1 + s2$) was captured through the use of a composite variable (diamond shape). These two spatial vectors thus account for all the spatial variation that is not explainable in terms of measured biotic and abiotic variables. This variation also implicitly accounts for the effects of spatial variation in unmeasured variables, which contribute to autocorrelation in measured variables.

Structural equation modelling (SEM) with composite latent variables was used to determine the relative importance of abiotic conditions, biotic interactions, and spatial patterns due to population- and community-level processes. Based upon previous work known at the time the model was fitted, as an a priori SEM of biodiversity was built, in which geographic properties and surface temperature (summer average) are mediated through the effects of soil properties and indirectly influence the richness of cyanobacteria (which positively correlates with total bacterial richness as described above), fungi, and multicellular taxa (Supplementary Figure 2). To identify variables to be included in the a priori model, the entire set of predictors was evaluated to determine which variables were most likely to be important for reasons of parsimony, thereby eliminating soil age, geology, soil C, and conductivity.

To derive a final model with good fit to the data from the a priori SEM, non-significant pathways were removed, and theoretically justifiable pathways were added that were deemed to be important through inspecting the residual covariance matrices and modification indices. The relationship between biotic richness and geology was analyzed using ANOVA, and the covariance matrices and model fit indicated that the net effects of three constructs (abiotic environmental, biotic interactions, and spatial landscape-level processes) that are relatively independent of environmental conditions also implicitly account for the effects of spatial variation in unmeasured variables, and their predicted values were derived. This was followed by a principal component analysis (PCA) on these predicted values, allowing spatial patterns to be summarized in the multivariate distribution of the three biotic richness variables. The first two components ($s1$ and $s2$) accounted for 90% of the environmentally independent spatial patterns. Finally, the net effect of spatial variation ($s1 + s2$) was captured through the use of a composite variable (diamond shape). These two spatial vectors thus account for all the spatial variation that is not explainable in terms of measured biotic and abiotic variables. This variation also implicitly accounts for the effects of spatial variation in unmeasured variables, which contribute to autocorrelation in measured variables.

Structural equation modelling (SEM) with composite latent variables was used to determine the relative importance of abiotic conditions, biotic interactions, and spatial patterns due to population- and community-level processes. Based upon previous work known at the time the model was fitted, as an a priori SEM of biodiversity was built, in which geographic properties and surface temperature (summer average) are mediated through the effects of soil properties and indirectly influence the richness of cyanobacteria (which positively correlates with total bacterial richness as described above), fungi, and multicellular taxa (Supplementary Figure 2). To identify variables to be included in the a priori model, the entire set of predictors was evaluated to determine which variables were most likely to be important for reasons of parsimony, thereby eliminating soil age, geology, soil C, and conductivity.

To derive a final model with good fit to the data from the a priori SEM, non-significant pathways were removed, and theoretically justifiable pathways were added that were deemed to be important through inspecting the residual covariance matrices and modification indices. The relationship between biotic richness and geology was analyzed using ANOVA, and the final SEM was found to explain more variation than geology alone. The relationship between biotic richness and evenness on soil microbial community diversity and function. Plant Soil 338, 483–495 (2010).

de Vries, F. T. et al. Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities. FEMS Microbiol. Ecol. 15, 1230–1239 (2012).

Cardinale, B. J. et al. Biodiversity loss and its impact on humanity. Nature 486, 59–67 (2012).

Borcard, D. & Legendre, P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Model 153, 51–68 (2002).

van der Heijden, M. G. A., Bardgett, R. D. & van Straalen, N. M. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol. Lett. 11, 296–310 (2008).

Bardgett, R. D. & Wardle, D. A. Aboveground-Belowground Linkages. (Oxford University Press, Oxford, 2010).

Loreau, M. Linking biodiversity and ecosystems: towards a unifying ecological theory. Philos. Trans. Roy. Soc. B 365, 49–60 (2010).

Nielsen, U. N., Opler, G. H. R., Campbell, C. D., Burleson, D. F. R. P. & van der Wal, R. The influence of vegetation type, soil properties and precipitation on the composition of soil mite and microbial communities at the landscape scale. J. Biogeogr. 37, 1317–1328 (2010).

Caruso, T. et al. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. ISME J. 5, 1406–1413 (2011).

Bru, D. et al. Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. ISME J. 5, 532–542 (2011).

Zimmerman, N. B. & Vitousek, P. M. Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. Proc. Natl Acad. Sci. USA 109, 13022–13027 (2012).

Hugenholz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. 180, 4765–4774 (1998).

Boyd, D. S. & Foody, G. M. An overview of recent remote sensing and GIS-based research in ecological informatics. Ecol. Inform. 6, 25–36 (2011).

Foody, G. M. GIS biodiversity applications. Prog. Phys. Geog. 32, 223–235 (2008).

Bindschadler, R. et al. The Landsat image mosaic of Antarctica. Remote Sens Environ. 112, 4214–4226 (2008).

Convey, P. et al. The spatial structure of Antarctic biodiversity. Ecol. Monogr. 84, 203–244 (2014).

Gray, C. S., McDonald, I. R., Barrett, J. E. & Cowan, D. A. On the rocks: the microbiology of Antarctic Dry Valley soils. Nat. Rev. Microbiol. 8, 129–138 (2010).

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

Adams, R. J. et al. Diversity and distribution of Victoria Land biota. Soil Biol. Biochem. 38, 3003–3018 (2006).

Sokol, E. R., Herbold, C. W., Lee, K. C., Cary, S. C. & Barrett, J. E. Local and regional influences over soil microbial metacommunities in the Transantarctic Mountains. Ecosphere 4, 136 (2013).

Pointing, S. B. et al. Highly specialized microbial diversity in hyper-arid polar desert. Proc. Natl Acad. Sci. USA 106, 19964–19969 (2009).

Poage, M., Barrett, J. E., Virginia, R. A. & Wall, D. H. The influence of soil geochemistry on nematode distribution, McMurdo Dry Valleys, Antarctica. Arct. Antarct. Alp. Res. 40, 119–128 (2008).

Lee, C. K., Barbier, B. A., Bottos, E. M., McDonald, I. R. & Cary, S. C. The Inter-Valley Soil Comparative Survey: the ecology of Dry Valley edaphic microbial communities. ISME J. 6, 1046–1057 (2012).

Hogg, I. D. et al. Biotic interactions in Antarctic terrestrial ecosystems: are they a factor? Soil. Biol. Biochem. 38, 3035–3040 (2006).

Barrett, J. E. et al. Terrestrial ecosystem processes of Victoria Land, Antarctica. Soil Biol. Biochem. 38, 3019–3034 (2006).

Gracie, J. R. & Bollen, K. A. Representing general theoretical concepts in structural equation models: the role of composite variables. Environ. Ecol. Stat. 15, 191–213 (2008).

Gracie, J. R., Adler, P. B., Harpole, W. S., Borer, E. T. & Seabloom, E. W. Causal networks clarify productivity-richness interrelations, bivariate plots do not. Funct. Ecol. 28, 787 (2014).

Gracie, J. R. & Bollen, K. A. Interpreting the results from multiple regression and structural equation models. Bull. Ecol. Soc. Am. 86, 283–295 (2005).
Charles K. Lee\textsuperscript{1,2}, Daniel C. Laughlin\textsuperscript{1,2,18}, Eric M. Bottos\textsuperscript{1,2,19}, Tancredi Caruso\textsuperscript{2,3}, Kurt Joy\textsuperscript{1,2}, John E. Barrett\textsuperscript{2,4}, Lars Brabyn\textsuperscript{2,5}, Uffe N. Nielsen\textsuperscript{6}, Byron J. Adams\textsuperscript{2,7}, Diana H. Wall\textsuperscript{2,8}, David W. Hopkins\textsuperscript{2,9}, Stephen B. Pointing\textsuperscript{2,10}, Ian R. McDonald\textsuperscript{1,2}, Don A. Cowan\textsuperscript{2,11}, Jonathan C. Banks\textsuperscript{1,2,20}, Glen A. Stichbury\textsuperscript{2,12}, Irfon Jones\textsuperscript{13}, Peyman Zawar-Reza\textsuperscript{14}, Marwan Katurji\textsuperscript{14}, Ian D. Hogg\textsuperscript{1,2,21}, Ashley D. Sparrow\textsuperscript{15}, Bryan C. Storey\textsuperscript{2,13}, T. G. Allan Green\textsuperscript{1,2,16} & S. Craig Cary\textsuperscript{1,2,17}

\textsuperscript{1}School of Science, University of Waikato, Hamilton 3240, New Zealand. \textsuperscript{2}International Centre for Terrestrial Antarctic Research, University of Waikato, Hamilton 3240, New Zealand. \textsuperscript{3}School of Biological Sciences and Institute for Global Food Security, Queen’s University Belfast, Belfast BT7 1NN, UK. \textsuperscript{4}Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA. \textsuperscript{5}School of Social Sciences, University of Waikato, Hamilton 3240, New Zealand. \textsuperscript{6}Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW 2751, Australia. \textsuperscript{7}Department of Biology, Evolutionary Ecology Laboratories, and Monte L. Bean Museum, Brigham Young University, Provo, UT 84602, USA. \textsuperscript{8}Department of Biology & School of Global Environmental Sustainability, Colorado State University, Fort Collins, CO 80523, USA. \textsuperscript{9}SRUC – Scotland’s Rural College, Edinburgh EH9 3JG, UK. \textsuperscript{10}Yale-NUS College and Department of Biological Sciences, National University of Singapore, Singapore 138527, Singapore. \textsuperscript{11}Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0002, South Africa. \textsuperscript{12}Environmental Research Institute, University of Waikato, Hamilton 3240, New Zealand. \textsuperscript{13}Gateway Antarctica, University of Canterbury, Christchurch 8041, New Zealand. \textsuperscript{14}Centre for Atmospheric Research, Department of Geography, University of Canterbury, Christchurch 8041, New Zealand. \textsuperscript{15}CSIRO Ecosystem Sciences, Alice Springs, NT 0870, Australia. \textsuperscript{16}Departamento de Biología Vegetal II, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid 28040, Spain. \textsuperscript{17}College of Earth and Ocean Sciences, University of Delaware, Newark, DE 19958, USA. \textsuperscript{18}Present address: Department of Botany, University of Wyoming, Laramie, WY 82071, USA. \textsuperscript{19}Present address: Department of Biology, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada. \textsuperscript{20}Present address: Cawthron Institute, Nelson 7010, New Zealand. \textsuperscript{21}Present address: Polar Knowledge Canada, Canadian High Arctic Research Station, Cambridge, Bay X0B 0C0, Nunavut, Canada. These authors contributed equally: Charles K. Lee, Daniel C. Laughlin.