ECOLOGY

Polar lake microbiomes have distinct evolutionary histories

Bjorn Tytgat[1]†, Elie Verleyen[1]†, Maxime Sweetlove[1]†, Koen Van den Berge[2]‡, Eveline Pinseel[1,3]§, Dominic A. Hodgson[4,5], Steven L. Chown[6], Koen Sabbe[1], Annick Wilmotte[7], Anne Willems[8], The Polar Lake Sampling Consortium||, Wim Vyverman[1]†*

Toward the poles, life on land is increasingly dominated by microorganisms, yet the evolutionary origin of polar microbiomes remains poorly understood. Here, we use metabarcoding of Arctic, sub-Antarctic, and Antarctic lacustrine benthic microbial communities to test the hypothesis that high-latitude microbiomes are recruited from a globally dispersing species pool through environmental selection. We demonstrate that taxonomic overlap between the regions is limited within most phyla, even at higher-order taxonomic levels, with unique deep-branching phylogenetic clades being present in each region. We show that local and regional taxon richness and net diversification rate of regionally restricted taxa differ substantially between polar regions in both microeukaryotic and bacterial biota. This suggests that long-term evolutionary divergence resulting from low interhemispheric dispersal and diversification in isolation has been a prominent process shaping present-day polar lake microbiomes. Our findings illuminate the distinctive biogeography of polar lake ecosystems and underscore that conservation efforts should include their unique microbiota.

INTRODUCTION

Toward the poles, life on land is increasingly dominated by microorganisms, which perform critical ecosystem functions (*1*, *2*) and, like their macroscopic counterparts, show adaptations to extreme physical stresses experienced at high latitudes (*3*). However, it remains unclear if polar microbiomes and macroscopic organisms also share similar evolutionary trajectories. In macroscopic biota, biodiversity patterns differ strongly between the Arctic and Antarctic (*4*). Despite experiencing similar environmental conditions, terrestrial food webs in Antarctica are less diverse than those in the Arctic. Moreover, Antarctic food webs are disharmonic (*5*), in that they lack several key groups such as land-based mammals and, apart from the northern part of the Antarctic Peninsula and the sub-Antarctic islands, vascular plants (*4*, *6*). Whereas circum-Arctic distributions are more common among terrestrial plants and animals (*7*, *8*), terrestrial macrobiota in the Antarctic mostly show strong geographical population divergence (*9*), often with high levels of regional endemism (*3*). This has been attributed to the long-term geographic isolation and persistence of biota in scattered glacial refugia in Antarctica and on the isolated sub-Antarctic islands (*10*–*14*). Antarctica has been experiencing glacial conditions since the Late Eocene [ca. 35 million years (Ma)] and nearly complete glaciation since the mid-Miocene (ca. 13 Ma). As a result, multiple psychrophilic and psychrotolerant endemic clades of macrobiota evolved in long-term isolation in the Antarctic and on the sub-Antarctic islands (*3*, *4*). This evolution in long-term isolation, and past and ongoing dispersal limitation between ice-free regions, also resulted in unique haplotypes being restricted to particular ice-free regions and even single valley systems, and hence high levels of population genetic diversity in several invertebrate groups (*15*, *16*). In the Arctic, extensive continental ice sheets only started to develop intermittently during the Pliocene-Pleistocene transition (ca. 3.2 Ma) (*17*). This relatively recent and highly dynamic glaciation history of the Arctic, combined with the continuity of land masses which allowed for highly effective dispersal and gene flow, likely constrained evolutionary time for in situ speciation of cold-adapted taxa. Although less well studied than terrestrial animals, lacustrine macroinvertebrates also show a strong biogeographic differentiation between the Arctic, sub-Antarctic, and Antarctica (*18*).

In contrast to macrobiota, it has since long been hypothesized that polar microbiota consist of cold-adapted species recruited from a globally dispersing species pool (*19*–*21*). This hypothesis stems from the long-standing belief that microorganisms are able to disperse globally and that their local occurrence mainly depends on environmental selection (*22*). The assertion, however, that decreasing body size invariably scales with increasing dispersal ability (*23*, *24*) generally ignores the deeply divergent evolutionary histories and biology of major microbial groups (*25*) [but see (*26*)]. Recent work elsewhere has suggested that some microbiota may be regionally restricted and much more spatially structured than previously thought, while others have a global distribution (*27*, *28*). It also remains largely unknown if diversification rates in microorganisms are taxon-specific and vary in time or regionally. A recent analysis of multiple sequencing datasets covering >400,000 bacterial lineages has shown that recent global bacterial speciation and extinction rates are roughly stable, with a net increase in the

[1]Laboratory of Protistology and Aquatic Ecology, Ghent University, Gent, Belgium. [2]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium. [3]Meise Botanic Garden, Meise, Belgium. [4]British Antarctic Survey, Natural Environment Research Council, Cambridge, UK. [5]Department of Geography, Durham University, Durham, UK. [6]Securing Antarctica's Environmental Future, School of Biological Sciences, Monash University, Melbourne, VIC, Australia. [7]InBio-Centre for Protein Engineering, University of Liège, Liège, Belgium. [8]Laboratory of Microbiology, Ghent University, Gent, Belgium.
*Corresponding author. Email: wim.vyverman@ugent.be
†These authors contributed equally to this work.
‡Present address: Janssen R&D, Beerse, Belgium.
§Present address: Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA.
||The Polar Lake Sampling consortium authors and affiliations are listed in the Supplementary Materials.

number of species (*29*), confirming previous observations (*30*). However, this does not preclude the possibility of differences in speciation and extinction rates between and within taxonomic groups (*29*). Order of magnitude differences in speciation rates between different clades of eukaryotes and bacteria (*31*), and even within clades (*30*), have been reported. In marine fish (*32*) and angiosperms (*33*), diversification rates are most elevated in ecosystems with unsaturated niche space (*34*), but it is unknown to what extent this also applies to microbial communities, which are often regarded to be characterized by high functional redundancy (*35*). Given the past extinction events, the isolated nature of ice-free regions in Antarctica and the sub-Antarctic islands, and consequently low microbial species diversity, niches are likely to be unsaturated in these regions. We can hence expect an increased net diversification rate (i.e., the net accumulation of new "genotypes") for Antarctic and sub-Antarctic microorganisms compared to the Arctic, which should be reflected in a relatively higher amount of operational taxonomic units (OTUs) at a high taxonomic resolution.

Here, we apply molecular tools to investigate biogeographic patterns and differences in the evolutionary history among the major bacterial and microeukaryote phyla between the polar regions. Our analyses are based on an extensive inventory of benthic lacustrine biota in the polar regions (Fig. 1, B and D, and table S1). These organisms typically form structurally complex communities covering the bottom of lakes (Fig. 2) (*36*), representing hotspots of biodiversity and biological activity in an otherwise harsh and generally unproductive terrestrial environment (*2*). Dispersal of these benthic organisms is shown to occur through several mechanisms, such as aerosols (*37*) or as detached microbial mats (*38*) and wind dispersal (*39*), and can enter lakes directly, via catchment runoff, or through hydrological connectivity between lakes via glaciers, ice-sheets, and associated cryoconites (*37*). Long-distance dispersal of microorganisms is thought to be common (*28*), and while certain species have been found to be present in both polar regions (*40*), it is unclear to what extent this occurs. Here, we investigate the hypothesis of the existence of a global cold-adapted or cold-tolerant microbial species pool with a shared evolutionary history which resulted from recurrent long-distance dispersal and high gene flow between the polar regions.

We apply metabarcoding of variable regions of the 16*S* (Bacteria) and 18*S* (Eukarya) *ribosomal RNA* (*rRNA*) genes and use three clustering algorithms [UPARSE (*41*), DADA2 (*42*), and Swarm (*43*)] to assess differences in region and taxon-specific diversity at lower taxonomic levels (e.g., intraspecies diversity). We then compare the number of OTUs at different similarity levels between regions and use their ratios in a smooth function as a proxy for detecting differences in net diversification rates between regions. In combination with phylogenetic analyses of amplicon sequence variants (ASVs), this allows us to assess regional differentiation between microbiota at deeper phylogenetic levels and to reveal differences in the evolutionary history among phyla between the polar regions.

RESULTS AND DISCUSSION

Polar microbiomes are biogeographically structured

Canonical Analysis of Principal Coordinates (CAP) of presence-absence OTU, ASV, and Swarm phylotype data consistently revealed strong biogeographical structuring into Arctic, sub-Antarctic, and Antarctic regions for both eukaryotes and bacteria at the domain level (fig. S1), as well as for all major microbial phyla analyzed separately (Fig. 3A and table S2).

Both geographic distance and environmental variables were correlated with the bacterial and eukaryotic community composition and these results were congruent between clustering algorithms (table S3). The effects of both sets of predictors were comparable for eukaryotes (Euk_{OTU97} $r_{ENV|GEO}$ = 0.40, P = 0.001; Euk_{OTU97} $r_{GEO|ENV}$ = 0.44, P = 0.001), while correlations with environmental variables were relatively more important than geographic distance in bacteria (Bac_{OTU97} $r_{ENV|GEO}$ = 0.43, P = 0.001; Bac_{OTU97} $r_{GEO|ENV}$ = 0.25, P < 0.001). Although part of the proportion uniquely explained by geographic distance might be related to unmeasured environmental variables, geographic distance was also substantially correlated with beta diversity patterns in prokaryotic and eukaryotic microbial communities in lakes, rivers, and marine systems suggesting that their dispersal is not unlimited (*23*, *44*–*46*). The proportion of shared sequences between the three regions was low and decreased with increasing taxonomic resolution according to the algorithm used to cluster the amplicon sequences. Only 13.7, 3.5, and 2.2% of the eukaryotic and 12.1, 1.7, and 1.0% of the bacterial OTUs, phylotypes, and ASVs, respectively, were present in all three regions, resulting in high correct classification rates (CCRs), which are a measure of the differentiation between a priori defined regions in the CAP analysis. As expected, these patterns were observed in the metazoan sequence data (Fig. 3A), but in the case of microbiota, they were clearly inconsistent with the hypothesis of a shared microbial species pool between polar regions. This finding was further evidenced by permutational multivariate analyses of variance (PERMANOVA) of the prokaryote and eukaryote datasets (fig. S1 and table S4). Furthermore, in both microeukaryotes and Bacteria, the CCR remained virtually constant for the different similarity cutoff levels used to define OTUs, apart from the two lowest similarity cutoff levels in Bacteria (Fig. 3B). This suggests deep phylogenetic divergences of polar microbiota, with many clades being restricted to one of the three biogeographic regions.

Phylogenies built from the ASV dataset validated the presence of regionally unique phylogenetic clades in all major phyla (figs. S2A and S3A). We used the consenTRAIT algorithm (*47*) to calculate the mean trait depth (τ_D) at which these clades were restricted to a particular region, thus treating "geographic region" as a trait. We only considered clades consisting of two or more ASVs that were all restricted to the same region, excluding singletons. ConsenTRAIT calculates τ_D of a clade as the distance of its tips to the clade's root, which is subsequently averaged across all clades in the phylogeny. We can then calculate the average *rRNA* identity of tips within a cluster, by multiplying τ_D by 2 and subsequently subtracting from 1. For several bacterial phyla, including Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, and Proteobacteria, τ_D was significantly larger than expected by chance (fig. S2B), indicating strong phylogenetic conservatism of geographic region in bacterial phyla. For the eukaryotes, a significant τ_D was only observed in some regions for the Ciliophora, Ochrophyta, and Streptophyta (fig. S3B). Nevertheless, the analysis indicated that the *rRNA* identity of the eukaryotes was within the same range as the bacteria: average values of τ_D roughly centered on 0.01 for both, which corresponds with 98% *rRNA* similarity. In addition, maximum values of τ_D centered on 0.05, which equates to 90% *rRNA* similarity (figs. S2C and S3C). Although these differences in sequence divergence

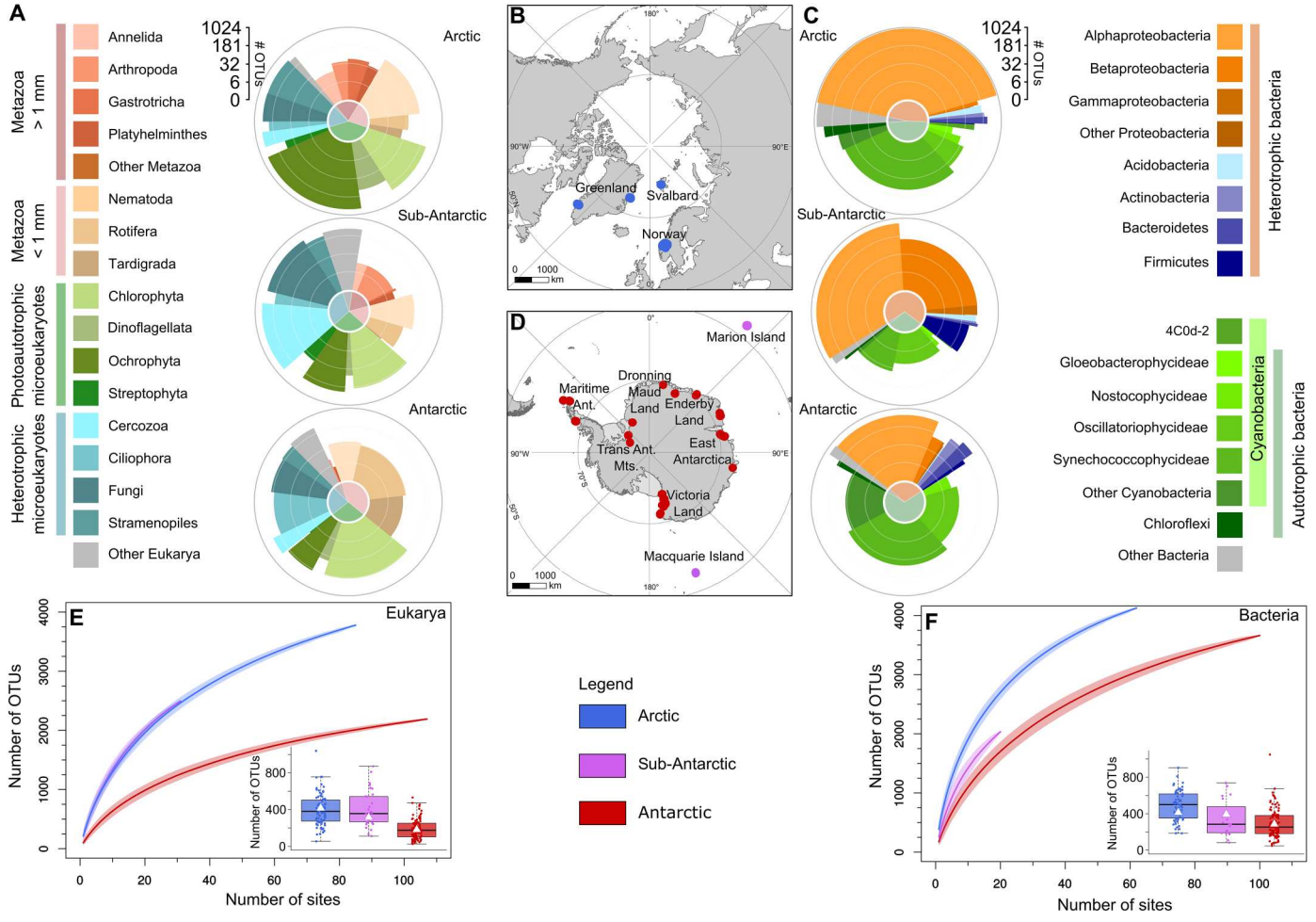

Fig. 1. Bipolar differences in microbial diversity and freshwater food web structure. (**A** and **C**) Rose plots of the average regional community composition for the major (i.e., >1% of the reads) phyla identified in the lakes for (A) eukaryotes and (C) bacteria. Taxa are grouped according to broadly defined functional groups (i.e., Metazoa > 1 mm, Metazoa < 1 mm, photoautotrophic microeukaryotes, heterotrophic microeukaryotes, heterotrophic bacteria, and autotrophic bacteria). For Cyanobacteria, Proteobacteria, and large Metazoa (>1 mm), classes representing >0.1% of the reads are also shown. Phyla representing less than 1% of the reads of either dataset are binned into "other Eukarya" and "other Bacteria," respectively. The arcs of the rose plot segments represent the mean proportions of the taxa. The height (along the radius) of the segments denotes the OTU richness (log_2-scaled). The inner pie chart within each rose plot shows the relative abundances of the functional groups. (**B** and **D**) Maps showing the 216 studied lakes (depicted by circles: blue = Arctic, magenta = sub-Antarctic, red = Antarctic) displayed on a polar projection of the (B) Northern and (D) Southern Hemispheres. (**E** and **F**) Rarefaction curves showing the total regional richness (means ± SD) based on rarefied samples, and boxplot inserts showing the distribution of the per-sample richness and the mean richness (white triangles) for the separate regions for Eukarya (E) and Bacteria (F), respectively.

are seemingly small, both 16*S* and 18*S rRNA* evolve relatively slowly. In bacteria, a single mutation in full-length 16*S rRNA* takes 0.7 Ma to 3 Ma, meaning that 1% divergence in 16*S rRNA* is expected to equate to ~10 Ma to 50 Ma of evolution (*48*). Similar rates have been calculated for microeukaryotes: e.g., a 1124–base pair (bp) region of diatom 18*S rRNA*, including the *V4* and *V9* regions, evolves at a rate of about one substitution every 3.6 Ma (*40*). Given that most substitutions will occur in the hypervariable loop regions, including *V1* to *V3* 16*S* and *V4* 18*S rRNA*, the actual mutation rates of our amplicons will be higher than the above-listed values. Yet, small differences (e.g., 99% similarity) in 16*S* and 18*S rRNA* amplicons likely still equate to several hundred thousand or even millions of years of evolution. Furthermore, given that closely related sister species of microeukaryotes are not always differentiated in 18*S rRNA* (*40*), here, unrecognized species-level diversity might be hiding within ASVs. Thus, consenTRAIT's mean and maximum trait depth values indicate that many of the phylogenetic clusters that are confined to one of the polar regions (figs. S2A and S3A) might have been restricted to these regions for a considerable amount of time, almost certainly predating the last glacial maximum, and possibly the Quaternary. Clearly, the phylogenetic analyses confirm the CAP results that indicated deep phylogenetic divergences of polar microbiota (Fig. 3B). Note, however, that our analysis did not include temperate regions in the Northern and Southern Hemispheres, so it is possible that some of these restricted clades show wider geographic distributions outside of the polar regions, such as is known for Arctic testate amoebae which are part of Holarctic radiation (*49*). Nevertheless, our data are in line with the handful of available phylogenetic datasets of Antarctic biota that support endemism at deep phylogenetic levels. For

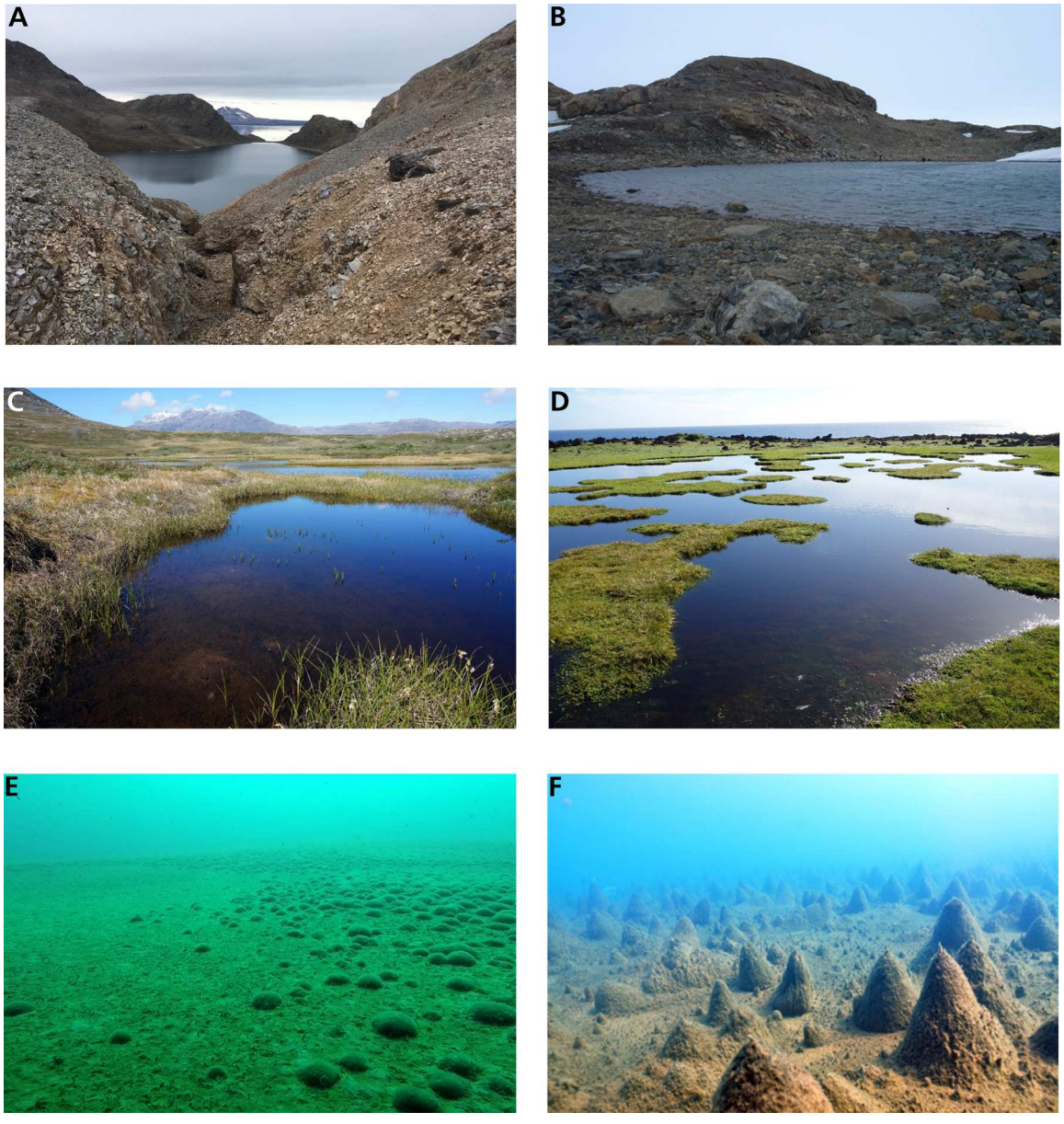

Fig. 2. Polar lake catchments and microbial consortia. (**A** to **D**) Examples and comparison of typical lake and pond catchments in polar deserts in (A) Svalbard, High Arctic and (B) Schirmacher Oasis, East Antarctica, and low latitude wet tundra in (C) Nuuk, Greenland and (D) sub-Antarctic Marion Island. (**E** and **F**) High latitude benthic lacustrine microbial mats in (E) Svalbard and (F) Skarvsnes, East Antarctica. [Photographs courtesy of J. Elster (A), E. Verleyen (B), K. Sabbe (C), D. A. Hodgson (D), D. Velazquez (E), and S. Kudoh (F).]

example, distinct Antarctic chlorophyte lineages have been shown to have estimated ages between 17 Ma and 84 Ma and probably diverged around the opening of the Drake Passage (30 Ma to 45 Ma), with evidence of even older lineages (65 Ma to 100 Ma) (*50*). In another study, the radiation of a soil diatom was found to be confined to the Antarctic and sub-Antarctic regions for ~6 Ma (*40*).

The high CCR of the samples analyzed at even low similarity cutoff levels (Fig. 3B) also reflects the virtual absence of several higher-order taxa in the Antarctic, including most animal classes as well as vascular plants, with the exception of two species in the Peninsula, and the dominance of other groups in the Arctic, such as arthropods and annelids. This deep divergence between the different regions further contributes to pronounced differences in the diversity (composition and species richness) and relative abundance of most food web compartments (Fig. 1, A and C, and fig. S4). Antarctic food webs are impoverished compared with the Arctic and sub-Antarctic. This is true both in terms of overall local and regional OTU richness at the domain level and for 9 out of the 16 major phyla (Fig. 1, E and F, figs. S4 and S5, and table S5), as well as in the relative abundance of functional guilds and phylogenetic clades (Fig. 1, A and C, and figs. S6 and S7). The general lack of large metazoan (>1 mm in size) OTUs, including Platyhelminthes, Annelida, and Gastrotricha (Fig. 1A and fig. S6B), supports earlier reports of reduced abundance of these organisms in Antarctic lakes

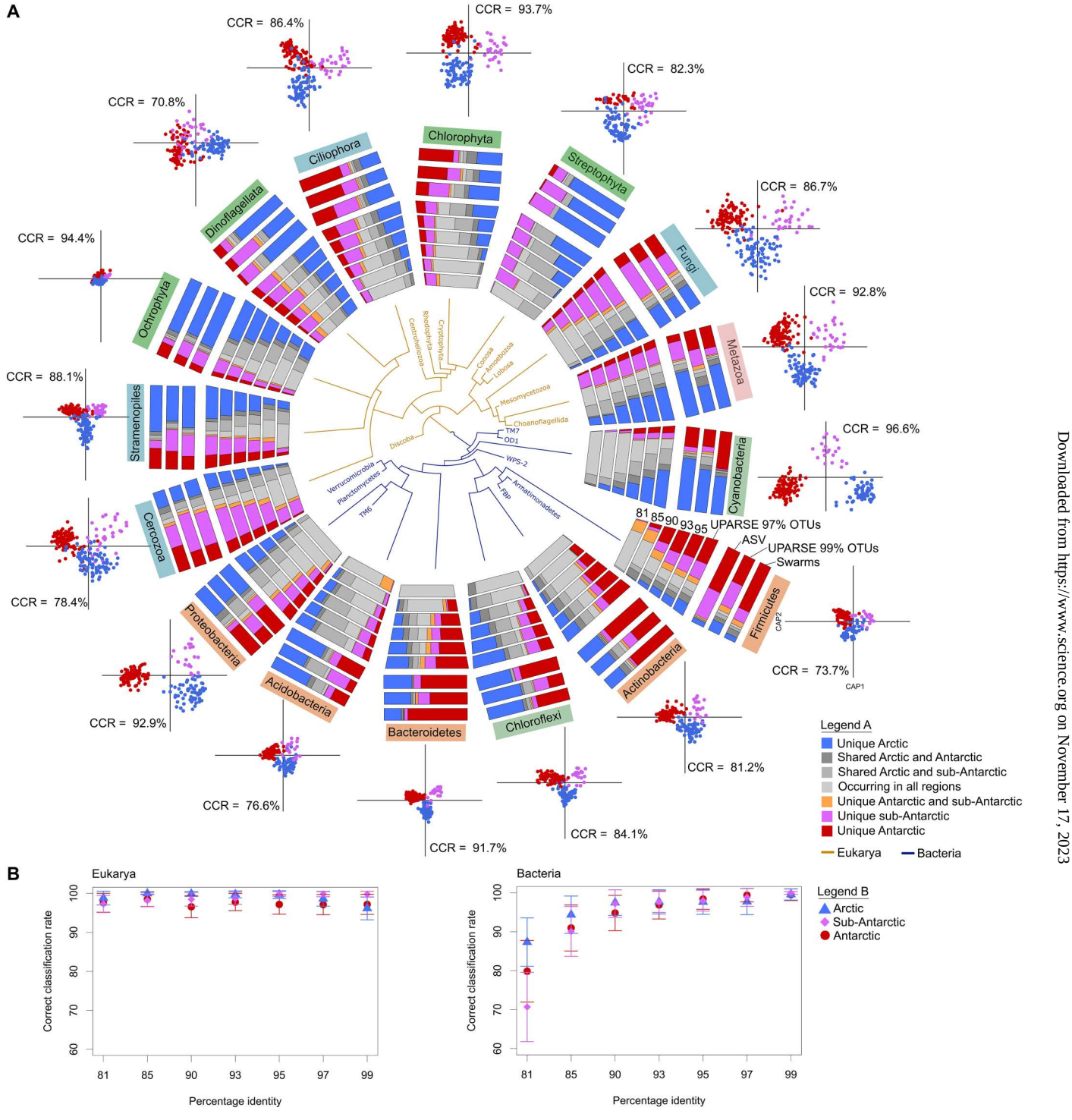

Fig. 3. The proportion of unique and shared OTUs, ASVs, and phylotypes and biogeographic clustering for the main taxonomic groups of lacustrine microorganisms in the polar regions. (**A**) The phylogenetic clades (phylum level) that represent at least 0.1% of the sequences are shown in the synthesized cladogram, with clade branches that represent at least 1% of the total number of sequences terminating in the center figure (branch lengths are meaningless). The bar plots show the proportion of unique (i.e., restricted to one single region) or shared OTUs for the different sequence similarity cutoffs (percentage identity) used with the UPARSE algorithm. Bars toward the plot center represent lower cutoff values (81% for the innermost bar). The three outermost bars correspond to phylotypes, UPARSE OTUs clustered at 99% sequence identity, and ASVs, respectively. The clustering according to the CAP analysis for each major phylum at the OTU 97% cutoff level is added on the outside, including the overall correct classification rate (CCR), representing how well samples are clustered into the three biogeographic regions (see table S2). The colors highlighting the phylum names denote the functional guild to which they belong as used in Fig. 1. (**B**) The CCR (means ± SD) at the domain level for Eukarya and Bacteria for each region across all OTU sequence similarity cutoff levels showing the strong biogeographic clustering into the three regions even at higher-order taxonomic levels.

(*18*). This low abundance of metazoans and hence a lack of strong top-down control by grazing results in low levels of bioturbation of sediments and biofilms (*51*, *52*) in Antarctic lakes, which promotes the formation of perennial microbial mats. These are physically structured by filamentous Cyanobacteria and Chlorophyta (Figs. 2 and 1, A and C) and mainly grazed by heterotrophic microorganisms (ciliates) and stress-tolerant microinvertebrates (Tardigrada and Rotifera) (*53*, *54*) and, predominantly in the Maritime Antarctic and sub-Antarctic, small crustaceans (*55*, *56*). By contrast, Arctic benthic biofilms and microbial mats (Fig. 2) are generally dominated by fast-growing and grazing tolerant ochrophytes and unicellular and filamentous Cyanobacteria, as previously observed in Canadian High Arctic lakes (*57*). The dominance of Cercozoa and Fungi among microeukaryotic heterotrophs in the sub-Antarctic lakes is likely related to the high organic matter influx from lake catchments and suggests an important role for allochthonous carbon, fuelling food webs in these lakes (*58*).

Net diversification patterns are pronounced in Antarctic microeukaryotes

Clustering reads at different similarity levels will generally show an increase in the observed number of OTUs with increasing similarity. It can thus be expected that there are more OTUs at 99% similarity (i.e., binning all reads that have at least 99% nucleotide similarity) than at 97% similarity. This also implies that there are likely more OTUs shared between regions at lower similarity levels (cf. higher-order taxonomic ranks such as class and order), while at high similarity levels (i.e., species and subspecies), the proportion of OTUs unique for a region will increase. We therefore used the relative increase in OTUs unique for a region at, e.g., 99% compared to 97% as a measure for regional taxon diversification, and we refer to this as the "amplicon richness ratio." Although there might be differences between and even within clades, we expect that the average background net diversification for a certain taxon is equal between regions (*29*, *30*), especially when dispersal between regions takes place. When the relative increase for a region is higher than expected, we assume positive net diversification (e.g., radiation). To test this, we modeled the divergence in regionally unique taxa as a smooth function of the number of regionally unique UPARSE OTUs clustered at the different sequence similarity thresholds by fitting a Poisson generalized additive model (GAM) for every major phylum (Fig. 4), accounting for differences in sequencing depth and number of samples per region (see Materials and Methods for a detailed description; see also note S3 for a discussion on the regionally balanced control dataset). This showed that the increase in the proportion of regionally unique taxa with increasing taxonomic resolution varied strongly among phyla as can also be observed in Fig. 3A.

Specifically, the model revealed that the relative increase in the number of regionally unique OTUs at the 99% similarity level compared with regionally unique OTUs at the 97% cutoff level (i.e., amplicon richness ratio) is significantly higher for Antarctica compared with the Arctic in the bacterial phyla Cyanobacteria and Firmicutes, as well as in all eukaryotic clades except for Dinoflagellata and Streptophyta. By contrast, the increase in the amplicon richness ratio is significantly higher in the Arctic compared to Antarctica only for Streptophyta and Acidobacteria. In Fungi, Ciliophora, Chlorophyta, and the bacterial phyla Actinobacteria and Cyanobacteria, the amplicon richness ratio is significantly higher in Antarctica compared to sub-Antarctica. In none of the eukaryotic and bacterial clades, except for Actinobacteria, is this richness ratio significantly higher in the Arctic compared with sub-Antarctica (Figs. 3A and 4), while it is significantly higher for Metazoa, Fungi, Cercozoa, and Chlorophyta in the sub-Antarctic compared with the Arctic. These patterns suggest that the potential for (recent) positive net diversification and radiations vary strongly among microbial clades in both domains of life, as well as between regions. The higher amplicon richness ratios in all but two microeukaryotic phyla in Antarctica compared with the Arctic are consistent with the patterns observed in Metazoa in our dataset (Fig. 3A), as well as with earlier reports of high population divergences observed in Antarctic invertebrate species, such as mites, tardigrades, and springtails (*9*, *15*, *59*). In addition, the absence of unlimited dispersal within Antarctica, and thus the potential for population-level divergence, has been shown in cyanobacteria (*60*), chlorophytes (*50*), and diatoms (*61*). Similarly, evolution in isolation, likely characterized by elevated diversification rates, has been shown to drive speciation in the soil diatom *Pinnularia borealis* (*40*).

Polar microbial biogeographical patterns are congruent with those of macroorganisms

The strong biogeographic structuring of benthic polar lake microbiomes in terms of taxonomic turnover, local and regional diversity patterns, and food web structure is concordant with biogeographic schemes recognized in terrestrial macrobiota (*4*, *62*). These patterns are robust at the scale of this study and are also supported by an extensive analysis of fossil and recent diatom floras (*63*). Notably, the disharmonic nature of Antarctic lacustrine food webs mirrors biodiversity patterns of marine (in)vertebrates in the Southern Ocean where some globally important groups are largely underrepresented, whereas others have undergone strong diversification and display unusual forms of adaptation (*64*).

An unexpected finding was the consistently higher amplicon richness ratios in lineages that are restricted to Antarctica in 75% of the investigated microeukaryotic phyla, as well as Metazoa, Cyanobacteria, and Firmicutes compared with the Arctic (Fig. 4). Although the evolutionary histories of most phyla remain poorly documented and information on their dispersal ability is almost completely lacking, these findings are consistent with increased opportunity for allopatric divergence, and the hypothesis of a long-term episode of regional extinction of lake biota (*65*) in Antarctica. This extinction wave was initiated since the onset of the mid-Cenozoic cooling (*17*) and paralleled by the adaptation and diversification of surviving microbiota on a continent subject to intensifying habitat fragmentation due to expanding continental ice sheets (*17*, *63*). On the basis of fossil evidence, repeated glacial-interglacial cycles in Antarctica might have acted as a species pump for diatoms, by triggering range shifts and isolation of populations in glacial refugia, followed by vicariant speciation (*63*). Even during interglacials, suitable habitats for microbiota were few and far between, hampering effective gene flow that would have homogenized populations. While this hypothesis does not preclude recolonization of the Antarctic from lower latitudes (*40*) or dispersal within and between regions (*66*), our dataset indicates that allopatric divergence is likely a major driver for diversification across not only diatoms, but most Antarctic microbial lineages, and comparable to what is observed in macrobiota, such as Antarctic mosses (*67*) and springtails (*66*). In addition, incomplete niche-filling following

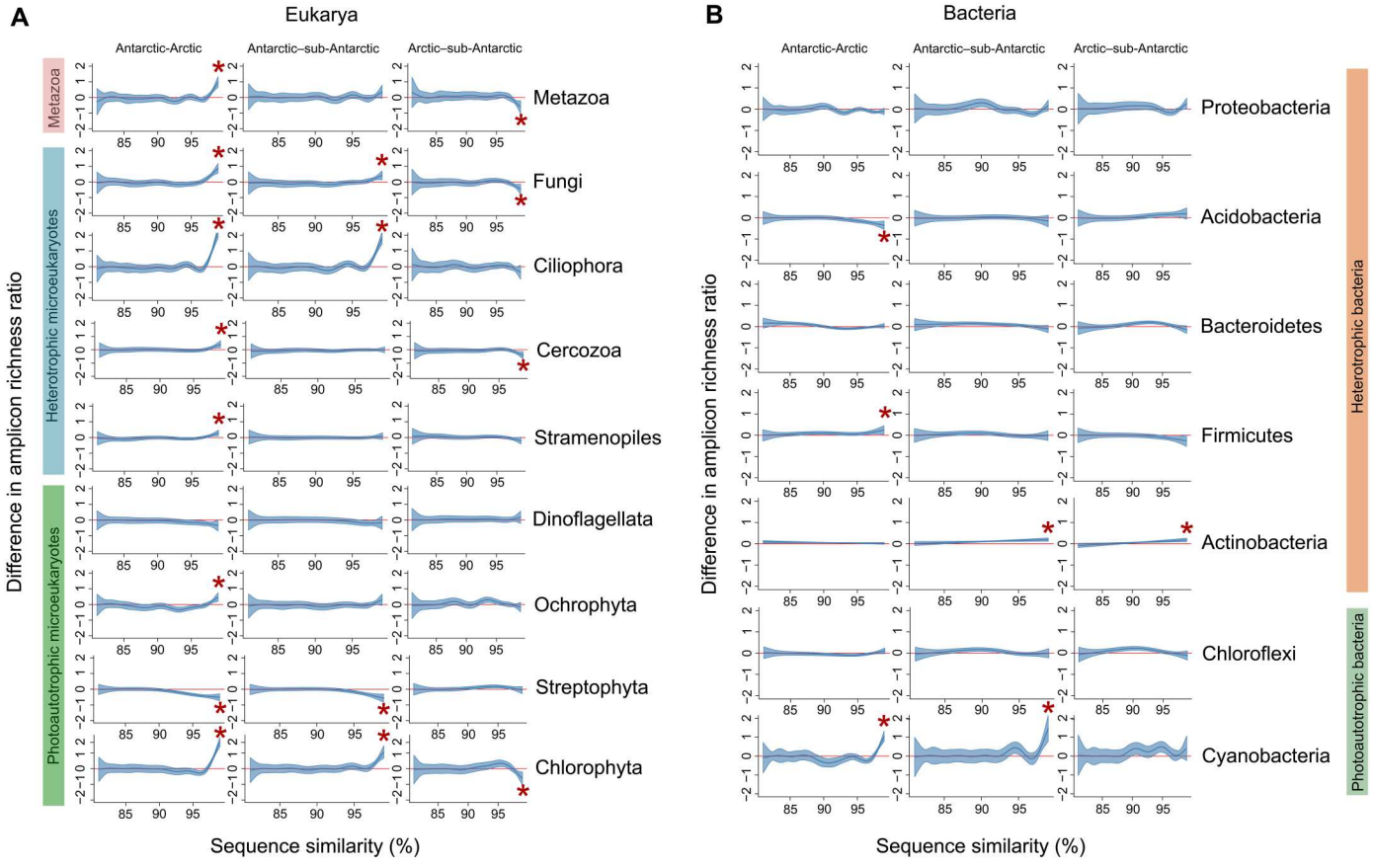

Fig. 4. Modeled lineage through time plots showing the smooth function of the regionally unique UPARSE OTUs clustered at the different thresholds. The differences in amplicon richness ratios (i.e., the relative increase in the number of OTUs unique for a region at a higher similarity level compared to those at the preceding similarity level) for the 16 major eukaryotic (**A**) and bacterial (**B**) phyla are shown as a pairwise comparison between regions. Differences in this richness ratio of unique taxa between regions are significant (*) when the confidence bands do not include "zero" (horizontal red line). If the difference is positive (confidence band, >0), then there is a higher amplicon richness ratio (or net diversification rate) for the left-hand region in the column title, while the richness ratio is higher for the right-hand region when the difference is negative (confidence band, <0). Phyla are grouped according to the different functional guilds. Functional group sidebar colors as in Fig. 1.

regional extinctions may act as a powerful driver for in situ diversification and the increased accumulation of genetic novelty in microorganisms in species-poor versus more diverse communities, as has previously been observed in macroscopic organisms such as polar marine fish and tropical songbirds (*32*, *34*, *68*, *69*). Moreover, in both marine mollusks, and beetles that are specialized in the extreme (sub-)Antarctic environment, widespread habitat restructuring in response to climate shifts, combined with novel ecological opportunities, has been shown to drive diversification (*70*, *71*), and our data indicate that this is similarly true for microbial lineages in the Antarctic and sub-Antarctic islands.

Together, our study provides strong evidence that the biogeography and evolutionary history of polar microbial phyla are influenced by both the geological and climatological history of the polar regions, as well as by their biology and life histories. The depauperate nature of contemporary Antarctic and sub-Antarctic microbiomes not only raises important questions regarding their functional saturation and adaptive potential under future climate scenarios but also stresses their vulnerability to non-native invasions. This underscores the urgent need for the inclusion of microbial taxa in management plans for the polar regions (*10*, *72*, *73*).

MATERIALS AND METHODS

Sample collection and datasets

All analyses in this study are based on samples from 216 lakes collected during the course of field expeditions by coordination of national and international (paleo)limnological research programs in three biogeographic regions, namely, Antarctica, the sub-Antarctic, and the Arctic (table S1) following a standardized protocol (*74*, *75*). In general, visible microbial mats were targeted for sampling in the littoral zone (between ±20- and 50-cm depth) using a sterilized spatula. Deeper parts of the euphotic zone were sampled using Glew surface sediment corers. The upper 1 cm of the core was consecutively aseptically removed using a sterilized spatula. All samples were immediately stored in the dark at −20°C until DNA extraction.

In the Northern Hemisphere, study sites from the (sub-)Arctic included 34 Norwegian high-altitude mountain lakes, 16 lakes in southwest Greenland, 23 lakes in northeast Greenland, and 13 lakes in Svalbard (Fig. 1B and table S1). In the Southern Hemisphere, samples were available from Antarctica (n = 114), and the sub-Antarctic Macquarie (n = 13) and Marion (n = 18) Islands (Fig. 1D and table S1).

Environmental data

In situ measurements of specific conductance and pH were available for 208 lakes. Month-averaged (1990–2013) air temperature measurements were obtained from the CRUTEM4 database (*76*, *77*) (version 4-2015-06) using records for the nearest station within a 500-km radius. When no temperature data were available (Transantarctic Mountains), an approximation was made on the basis of monthly averaged satellite data (1985–2005) from the NASA Earth Observations database (NEO; NEO-team), which used data provided by the MODIS Land Science Team (MODIS Land Science Team, 2017; https://lpdaac.usgs.gov). Monthly averaged (1983–2005) data of insolation incidents on a horizontal surface (solar irradiance, in kWh m^{-2} day^{-1}) were obtained from NASA Surface Meteorology and Solar Energy (SSE; release 6.0) (*78*) at a one-by-one degree resolution.

DNA extraction, PCR amplification, and sequencing

Genomic DNA was extracted from homogenized subsamples (1.5 to 3 mg wet weight) using a phenol-chloroform–based protocol with ethanol precipitation (*79*), preceded by a coagulation step to remove extracellular proteins and DNA (*80*). The *V4* variable region of the 18*S rRNA* gene of eukaryotes was targeted with the universal primers TAReuk454FWD1 (5′-CCA GCA SCY GCG GTA ATT CC-3′) and TAReukREV3 (5′-ACT TTC GTT CTT GAT YRA-3′) (*81*). For Bacteria, the *V1-V3* region of the 16*S rRNA* gene was targeted with the universal primer pair pA (5′-AGA GTT TGA TCC TGG CTC AG-3′) (*82*) and BKL1 (5′-GTA TTA CCG CGG CTG CTG GCA-3′) (*83*). The polymerase chain reaction (PCR) was performed in duplicate per sample to minimize potential biases. Each reaction contained 2.5 μl of PCR reaction buffer (High-Fidelity PCR system, Roche), 2.5 μl of 2 mM of each deoxynucleotide triphosphate (Life Technologies Inc.), 0.25 μl of FastStart High fidelity Taq polymerase (Roche Inc.), 2 μl of each primer, 1 μl of DNA template, and 14.75 μl of sterile high-performance liquid chromatography–grade water with a final volume of 25 μl. For the 18*S rRNA* libraries, the PCR program consisted of an initial DNA denaturation step at 96°C for 5 min, followed by 35 touchdown cycles of denaturation at 96°C for 1 min, annealing at 57° to 52°C for 1 min, and elongation at 72°C for 3 min. The final elongation was extended to 20 min to reduce the number of chimeric sequences (*84*). For bacteria, initial denaturation was done for 3 min at 95°C, followed by 27 cycles with 30-s denaturation at 95°C, 45-s annealing at 55°C, elongation at 72°C for 3 min, and a final elongation of 10 min. After pooling PCR duplicates, products were purified with Agencourt AMPure XP beads (Beckman Coulter Inc.). The amplicon libraries were barcoded using the NEXTERA XT tags (Illumina Inc.) using a 12-cycle version of the amplification PCR, after which they were pooled equimolarly and sequenced on the Illumina MiSeq platform (v3 reagent kit for 2 × 300 bp paired-end reads). Cluster density was reduced by spiking with 20% PhiX DNA (Illumina Inc.), which has been shown to increase the overall quality of the sequencing run (*85*). Sequencing was performed by BaseClear B.V. (Leiden, Netherlands).

Sequence quality control and technical reproducibility

Each run with either eukaryotic or bacterial samples contained two replicates of a positive control sample (mock community), respectively, containing 16 eukaryote or 21 bacterial strains (note S1), which were grown from pure isolates. These mock samples allowed for checking of overall run quality and inter-run reproducibility. In addition, randomly chosen samples were replicated within and between runs to assess technical reproducibility (note S2). The first run of each primer set also included a blank sample composed of negative PCR controls.

The reproducibility and robustness of the patterns were assessed with a control sequencing run (further referred to as "control run" or "control dataset") with a regionally balanced design, based on the DNA extracts of 29 Arctic, 28 sub-Antarctic, and 29 Antarctic samples that were randomly selected from the list of samples sequenced in the other runs (table S1). This was done for both eukaryotes and bacteria, respectively, and sequenced independently at Edinburgh Genomics (Edinburgh, UK). All major analyses were performed on this control dataset and are shown in the supplementary data to confirm the patterns observed in the entire dataset.

Sequence data processing, clustering, and taxonomic assignment

The main and control datasets (note S3) were processed into traditionally used OTUs using the well-established UPARSE (*41*) algorithm, while two additional clustering algorithms based on either phylotypes [SWARM v2.1.12 (*43*)] or ASVs [DADA2 (*42*)] were applied to confirm the biogeographic patterns observed in the OTU datasets. OTUs and Swarm phylotypes cluster together closely related sequences based on sequence similarity (OTUs) or on a limited number of nucleotide differences (Swarm phylotypes). Recent approaches, such as DADA2, apply a noise correction algorithm using a per-sequencing run-based error model to identify true biological sequences, resulting in ASVs (*42*), also called exact sequence variants, resolving to single-nucleotide differences. Consequently, ASVs are thought to generally result in a higher (taxonomic) resolution.

Paired-end reads were assembled using PEAR (*86*) version 0.9.1, with minimum and maximum paired-read lengths of 360 and 500 bp for eukaryotes and 450 and 570 bp for bacteria, respectively. Reads with more than two mismatches with the forward or reverse primers were excluded from the dataset. The first and last four bases of the remaining reads were trimmed to account for their on average lower quality. Reads were further processed using the USEARCH (*87*) (version 8.0.1632) pipeline with a maximum expected error set at 1 for both datasets. De novo chimera detection was performed simultaneously when clustering the reads into OTUs using UPARSE (*41*) at the default 97% similarity level, excluding true singleton reads (i.e., occurring only once in the entire dataset) as clustering seeds. To create an OTU-by-sample matrix, quality-filtered reads were mapped to OTUs with the *usearch_global* command, setting the *maxrejects* and *maxaccepts* to respectively 800 and 20. The clustering was repeated using similarity levels of 81, 85, 90, 93, 95, and 99%. These similarity cutoffs were used to broadly approach the main taxonomic levels between phylum and (sub) species (*88*, *89*). It is important to emphasize that these values are used as an approximation and that these cutoffs may represent different ranks depending on the focal taxon, gene, and target region of a gene (*88*).

Swarm v2.1.12 was applied with default settings to the quality-filtered reads of the main dataset (i.e., before the OTU clustering step). The Swarm algorithm uses iterative single linkage with a local clustering threshold, followed by a phase that uses the internal abundance structures of clusters to split interlinked OTUs, resulting

in fine-scaled molecular OTUs. To differentiate between UPARSE-generated OTUs and Swarm OTUs, we refer to the latter as phylotypes.

ASVs were created using DADA2 v1.14.0. Primers were removed a priori using the FASTX trimmer tool (http://hannonlab.cshl.edu/fastx_toolkit/), removing the primer lengths from the 5′ side of both the forward and reverse reads. DADA2 parameter settings were *maxEE* set to 3 for the forward and set to 6 for the reverse reads; *truncQ* = 2, *maxN* = 0, and a *truncLen* of 275 and 265 for, respectively, the forward and reverse 16*S rRNA* reads; and 250 and 210 for, respectively, the forward and reverse 18*S rRNA* reads. An error model was calculated per run using 10^8 bases per sequencing direction for error learning after which the DADA2 algorithm was applied. All runs were then merged and bimeras were removed using the *removeBimeraDenovo* function.

Taxonomy was assigned with the Ribosomal Database Project naïve Bayesian classifier algorithm (*90*) implemented in Mothur (*91*) for each OTU, phylotype representative sequence, and ASV, using a bootstrap value of 80%. The PR^2 database (*92*) (version gb_203) was used as a reference template for eukaryotic taxonomic assignment and the Greengenes database (*93*) (version 13_5) for bacteria.

Data postprocessing and filtering

The evaluation of the positive controls on mock samples (note S1) required removing OTUs that were represented by less than three reads to optimize the data quality at the different OTU-clustering similarities. Furthermore, OTUs that were present in only one sample were additionally removed, as well as nontarget groups [Archaea, chloroplasts, and mitochondria in the 16*S rRNA* dataset, and higher plants (Embryophyceae) and craniate animals in the 18*S rRNA* dataset]. Samples with less than 4500 sequences in the main dataset were not included in further analyses due to undersampling based on species accumulation curves (fig. S8, A and B), which were calculated using the *rarecurve* function with default settings in the *vegan* v. 2.5-2 R package (*94*). Respectively, 223 and 182 samples were retained for eukaryotes and bacteria after these quality controls. The same filtering was applied for the phylotype and ASV data. For the ASV dataset, 4500 could be retained as the per-sample minimum read threshold for both domains, resulting in 218 and 182 samples for both eukaryotes and bacteria, respectively. For the Swarm dataset, 4500 reads per sample were used for eukaryotes, while lowering the threshold to 1000 reads was needed for bacteria to retain a comparable number of samples with the other cluster methods, resulting in respectively 221 and 190 samples. This conservative approach resulted in 6924 and 6795 OTUs, 6073 and 10,591 ASVs, and 24,844 and 22,751 phylotypes for Eukarya and Bacteria, respectively, with a total of 12,693,595 and 3,243,458, 12,042,783 and 7,492,766, and 9,151,145 and 1,319,563 reads remaining for the respective domains and clustering approaches.

Differences in biodiversity, biogeography, and the amplicon richness ratios between the three regions were assessed for the whole eukaryotic or bacterial dataset (mentioned explicitly if subsampled or not) as well as for the phyla representing more than 1% of the total reads within either the Eukarya or the Bacteria datasets. These major phyla were (in order of the phylogenetic relationships, counterclockwise as shown in Fig. 3A) Metazoa, Fungi, Streptophyta, Chlorophyta, Ciliophora, Dinoflagellata, Ochrophyta (pigmented Stramenopiles), nonpigmented Stramenopiles (formerly Stramenopiles X in the PR^2 database), and Cercozoa for the Eukarya, and Proteobacteria, Acidobacteria, Bacteroidetes, Chloroflexi, Actinobacteria, Firmicutes, and Cyanobacteria for the Bacteria. Note that Metazoa and Fungi are here considered to be phyla following the PR^2 database taxonomy but generally are considered to be kingdoms. All downstream statistical analyses were executed in the R statistical language environment (*95*) (version 3.5.1) and implemented in the RStudio IDE (version 1.1.383).

Biogeographic regionalization and correlation with environmental variables and geographic distance

Canonical Analysis of Principal Coordinates (CAP) (*96*) with the CAPdiscrim function in the R package BiodiversityR (*97*) v2.11.3 was used to test to what extent the samples were clustering according to the predefined three main biogeographic regions by calculating the CCR, both at the domain level, as well as for each of the major phyla. CAPdiscrim parameter *m* was set to 0 to automatically calculate the number of principal coordinates analysis (PCoA) axes that provide the best distinction between the groups, with a default maximum of 10 axes investigated. Only the first two axes were plotted for the separate major phyla and the two domains (Fig. 3A and fig. S1). A PCoA was performed on a Bray-Curtis dissimilarity matrix followed by a linear discriminant analysis. The presence-absence data were used because this is the most conservative approach to detecting potential patterns: Giving occurrences equal weight will increase inter-regional clustering, attenuating (and hence underestimating) biogeographical patterns. Samples were subsampled to 4500 reads (1000 reads for the Swarm bacteria dataset) and datasets were standardized to account for differences in the number of samples between the regions, which otherwise might inflate regional clustering patterns. This was done by taking a random subset of samples equal to the lowest number of samples with the taxon present in any of the three regions, or the lowest number of samples for the domain-level analysis. This procedure was iterated 1000 times to calculate the mean coordinates of the individual samples and the mean regional and overall CCR (Fig. 3A, fig. S1, and table S2). To test the consistency of the biogeographic patterns at different taxonomic levels, we clustered reads using a range of similarity cutoffs (81, 85, 90, 93, 95, 97, and 99%) in UPARSE. The CAP analysis was repeated for each different similarity cutoff dataset and the overall CCR for each cutoff is shown for both domains in Fig. 3B, which enabled us to assess phylogenetic divergence at higher-order taxonomic levels. To further corroborate the differentiation at lower taxonomic levels, we performed a phylogenetic analysis on the individual main phyla of both bacteria and eukaryotes using the DADA2 ASVs since these represent unique biological entities with a resolution as low as 1 nucleotide (*42*). ASV reads were extracted per phylum and aligned with Clustal Omega v.1.2.4 (*98*) using 10 iterations, including three outgroup sequences from the closest related major group based on the cladogram in Fig. 3 (e.g., Proteobacteria for Acidobacteria). Highly variable regions (i.e., loop regions) were visually assessed and removed to prevent issues with misaligned regions. Phylogenetic trees were subsequently obtained with IQ-TREE (*99*) using a generalized time reversible (GTR) substitution model and 1000 ultrafast bootstraps (*100*). IQ-TREE created a consensus tree for visual representation (figs. S2A and S3A) as well as 1000 bootstrapped trees for each phylum which were used for further analysis (see below). All

phylogenetic trees were rooted and the outgroups were removed in R using the packages ape v.5.7.1 (*101*), Treetools v.1.9.1 (*102*), and phytools v.1.5.1 (*103*). Consensus trees were displayed using GGTREE (*104*), incorporating branch lengths. Branches were colored according to the occurrence of the tip ASV: red for uniquely occurring in Antarctica, blue when uniquely occurring in the Arctic, magenta for unique sub-Antarctic ASVs, and ASVs shared between any two or more regions are displayed in gray (figs. S2A and S3A). Clades were colored accordingly while branches shared the same state, switching to gray when internal nodes were no longer uniquely combining ASVs from a single region.

For each phylum, 100 randomly selected bootstrapped trees were imported in castor v.1.7.8 (*105*) to calculate the consenTRAIT metric (*47*). We were interested in testing the mean and maximum trait depths for clades that were uniquely restricted to any of the geographic regions and therefore considered geographic region to be a trait. Since castor does not allow more than two traits (i.e., presence/absence), the analysis was run four times for each tree, respectively distinguishing between ASVs that were distributed uniquely in (i) the Arctic, (ii) the Antarctic, (iii) the sub-Antarctic, and (iv) the Antarctic plus sub-Antarctic (= presence of the trait) versus any other distributions (= absence of the trait). Castor was run using a minimum fraction of 1, 1000 permutations, and with singletons excluded, given that singletons are inherently not part of a clade with a unique distribution.

The initial sets of explanatory variables for the partial Mantel test included lake pH and conductivity as environmental variables; annual solar irradiance, variability in monthly mean solar irradiance, summer mean solar irradiance, winter mean solar irradiance, mean annual temperature, and variability in monthly mean temperature as contemporary climate and energy-related predictors, as well as the geographical factors longitude and latitude. Pearson's correlations were calculated to assess multicollinearity. Setting a threshold of Pearson's $r > 0.8$, climate- and energy-related variables were reduced to retain only mean annual temperature and mean annual solar irradiance. Conductivity was $\log_e$-transformed to better conform to Gaussian distributions. All variables were centered to zero mean and unit variance. These variables were then used to test the effects of geographic or environmental distance on the community composition using partial Mantel tests. This allows to test the correlation between two distance matrices (i.e., the community matrix and either one of the environmental or geographic datasets) while controlling for variation in the third distance matrix (respectively, the geographic or environmental dataset). Partial Mantel tests were performed using the *partial.mantel* function in *vegan*, with the energy-related and environmental variables, combined as "environmental variables," converted to Euclidian distances, while the geographic variables longitude and latitude were transformed as haversine distances. The community matrix consisted of Bray-Curtis dissimilarities of either presence-absence or Hellinger-transformed rarefied community datasets.

Statistical support for the proposed biogeographic zones was examined and validated using (nonparametric) PERMANOVA using distance matrices. This analysis compares multivariate species means between different sample categories and uses the *adonis* function of the *vegan* package. We implemented the default method, which is based on first identifying the relevant centroids followed by the calculation of the squared deviations using the $n \times n$ outer product matrix. Significance tests were done by calculating pseudo-F ratios based on sequential sums of squares from permutations of the raw data ($n = 9999$) and corrected according to Benjamini and Hochberg (*106*).

OTU, phylotype, and ASV richness and community composition

Regional species accumulation curves for both domains (Fig. 1, E and F) were calculated using the *specaccum* function in R package *vegan* using 100 permutations. The mean regional richness and relative abundance for each major phylum were calculated by aggregating all samples within a region and subsampling 100 times to 2,000,000 reads for Eukarya and 200,000 reads for Bacteria (Fig. 1, A and C). For comparison, additional species accumulation curves were calculated using a balanced number of samples per region (i.e., subsampled datasets to the lowest number of samples in any region, being the sub-Antarctic) and iterated 100 times applying both a with and without sample replacement strategy. In each iteration, the selected samples were randomly subsampled to 4500 reads (fig. S9). The taxa comprising less than 1% of the reads were binned in either "other Eukarya" or "other Bacteria." In addition, the three most abundant phyla (Metazoa, Proteobacteria, and Cyanobacteria) were split into their respective classes (following the PR^2 database) that made up >0.1% of the reads. The remaining classes were binned in "other Metazoa," "other Proteobacteria," and "other Cyanobacteria" (Fig. 1 and figs. S6 and S7). The lower number to subsample to in the Bacteria dataset is due to the relatively low number of reads for the sub-Antarctic ($n = 210{,}583$). Unclassified reads at the domain level were excluded from this analysis. The mean regional richness was $\log_2$-transformed for visualization (Fig. 1, A and C).

To assess the differences in mean local OTU richness at the domain level between the three major regions, a negative binomial generalized linear model (Eukarya) or quasi-Poisson generalized linear model (Bacteria) was applied (*107*), thereby accounting for differences in the (log) number of sequences and evenness (defined as the Shannon index divided by the natural logarithm of the richness) between samples. Different distributions for the two clades were applied since the mean-variance relationship of the data for the Eukarya corresponded to the negative binomial distribution, while for Bacteria, this conformed to a quasi-Poisson distribution, following Ver Hoef and Boveng (*107*). We assessed the effect of region conditional on sequencing depth and evenness using type III ANOVA tests. Model-based estimates of the OTU richness are shown as boxplot inserts in Fig. 1 (E and F) and in fig. S10 for the two other clustering algorithms. The model deviance (analogous to R^2 in linear models) explained by the region was calculated as the difference in explained deviance between a model with and without the region as a fixed effect in the model. A post hoc analysis was performed using function *emmeans* in package *emmeans* v.1.4.5 (*108*) using the Tukey method for a P value adjustment to investigate pairwise differences between regions. The same approach was applied to the individual major phyla in the OTU dataset (fig. S3 and table S5), where the distribution was determined separately.

Lineage through time assessment and detection of amplicon richness ratio per region

To assess regional differences in amplicon richness ratios, we calculated the proportion of OTUs unique to (i.e., defined as only

occurring in one particular region) or shared between (i.e., occurring in at least two regions) the different regions for each major phylum based on the different OTU similarity cutoffs used for CAP, as well as for the ASV and phylotype data (Fig. 3A). ASVs, phylotypes, and OTUs clustered at the 99% similarity level provide a finer taxonomic resolution (*42*, *43*, *109*, *110*) and therefore represent recent diversification events. When the proportions of these high-resolution taxa uniquely occurring in each region are compared with proportions of region-specific OTUs clustered at the 97% similarity level, this gives an approximation of in situ amplicon divergence in each of the three regions. Hence, assessing OTU biodiversity over different sequence clustering similarity levels provides a proxy for the historical evolutionary radiation in a taxonomical group in each region [i.e., "lineage-through-time" plots (*88*)]. The occurrence of OTUs at lower similarity levels unique for a region suggests the existence of higher-order taxonomic clades that either evolved in that, or even a neighboring, region or have since been lost in other regions. This analysis is based on data subsampled to 4500 reads in 20 randomly selected samples per region and iterated 100 times to minimize the effect of varying sampling depth, and subsequent conversion to presence-absence data.

These lineage-through-time plots (Fig. 3A) were confirmed by a model-based approach (Fig. 4). For every major phylum, we calculated the number of OTUs for every sample uniquely occurring in each of the three regions, hereafter referred to as the unique biodiversity. We modeled the unique biodiversity as a smooth function of sequence similarity using a Poisson GAM to account for the count nature of biodiversity estimates. We allowed for an interaction between the clustering percentage and the three regions (Arctic, sub-Antarctic, and Antarctic), effectively allowing the biodiversity patterns to vary between the different regions. We also added a fixed sample effect to account for the correlation of the unique biodiversity levels across clustering percentages within a sample. Comparing unique biodiversity across clustering percentages by conditioning on the sample has the additional advantage that between-sample normalization, e.g., for sequencing depth, is unnecessary since the comparisons are performed within every sample. Using the GAM, we assessed differences in sequence divergence rate in unique lineages between regions. We approximated the divergence rate by estimating first derivatives using finite differencing techniques on the smoothers across their range, which amounts to estimating pointwise slopes of the smoother. Comparing first derivatives hence amounts to comparing amplicon divergence rates in unique taxa between regions. Since entire smooth curves are being compared, pointwise confidence intervals are too liberal and we needed to account for the fact that we compared curves across an entire range, rather than at a particular point. We therefore constructed simultaneous confidence bands for the differences in the first derivatives to show the estimation uncertainty along the entire two-dimensional fit of the curve. Hence, we recast the GAM as a generalized linear mixed model (GLMM) and simulated from the multivariate normal distribution of the GLMM model parameters to estimate simultaneous confidence bands (*111*). Using the 95% simultaneous confidence bands, a smoother is significantly different from zero at a global 5% significance level as soon as its confidence band does not include zero (Fig. 4).

To demonstrate the principle of this approach (Fig. 4), an artificial toy dataset was generated (fig. S11), representing three regions (region 1, region 2, and region 3) with 100 samples each. We simulated a case where a region effect was present in "region 1" and "region 2" compared to "region 3," i.e., there is a stronger increase in the ratio between the number of unique OTUs from two sequential "cutoffs" compared to "region 3," but not between the former two although a difference in the absolute number of OTUs is present (fig. S11A). We assumed an exponential function for the increase in the number of unique OTUs (λ = 0.30, 0.20, and 0.15 for region 1, region 2, and region 3, respectively). At each cutoff, the number of unique OTUs was drawn from a Poisson distribution for each sample, based on the value of the exponential function at each cutoff for the specific region. Since we assumed an exponential function, smoothing would not be necessary since log-transforming it will make it perfectly linear. We therefore added random variation to approach a more realistic distribution. This generated dataset was then incorporated into the smooth function described above. Results are displayed in fig. S11B.

We demonstrated the existence of distinct biogeographical regions in both bacteria and microeukaryotes from polar regions using a high-throughput sequencing approach. These patterns are robust and comparable between the different pipelines used, despite differences in the number of OTUs, ASVs, or Swarm phylotypes. This suggests that variation is induced through, e.g., sequencing errors, or simply because of true biological variation (for example, the presence of closely related strains or gene copies) and how algorithms deal with them. The robustness of the patterns however indicates that the constructed representative reads are based on "true biological sequences" that are largely restricted to one or few regions. Despite this, it is clear that Swarm performed relatively poorly, with a highly inflated richness (up to nearly 40 times more phylotypes in the mock communities than expected), generating many singletons. UPARSE OTUs and DADA2 ASVs were more comparable and approached mock community richness much better. Patterns in the dataset were very similar and show that conclusions from past OTU-based studies are still valid, provided that sufficient stringency in data processing was applied. Clustering approaches also have the advantage that different similarity cutoffs can be used, targeting different approximate taxonomic levels.

Supplementary Materials

This PDF file includes:

Notes S1 to S4
Figs. S1 to S11
Legend for table S1
Tables S2 to S5
References

Other Supplementary Material for this manuscript includes the following:

Table S1

REFERENCES AND NOTES

1. M. Ji, C. Greening, I. Vanwonterghem, C. R. Carere, S. K. Bay, J. A. Steen, K. Montgomery, T. Lines, J. Beardall, J. van Dorst, I. Snape, M. B. Stott, P. Hugenholtz, B. C. Ferrari, Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* **552**, 400–403 (2017).
2. R. Cavicchioli, Microbial ecology of Antarctic aquatic systems. *Nat. Rev. Microbiol.* **13**, 691–706 (2015).

3. P. Convey, S. L. Chown, A. Clarke, D. K. A. Barnes, S. Bokhorst, V. Cummings, H. W. Ducklow, F. Frati, T. G. A. Green, S. Gordon, H. J. Griffiths, C. Howard-Williams, A. H. L. Huiskes, J. Laybourn-Parry, W. B. Lyons, A. McMinn, S. A. Morley, L. S. Peck, A. Quesada, S. A. Robinson, S. Schiaparelli, D. H. Wall, The spatial structure of Antarctic biodiversity. *Ecol. Monogr.* **84**, 203–244 (2014).
4. S. B. Pointing, B. Büdel, P. Convey, L. N. Gillman, C. Körner, S. Leuzinger, W. F. Vincent, Biogeography of photoautotrophs in the high polar biome. *Front. Plant Sci.* **6**, 1–12 (2015).
5. M. Cantonati, S. Poikane, C. M. Pringle, L. E. Stevens, E. Turak, J. Heino, J. S. Richardson, R. Bolpagni, A. Borrini, N. Cid, M. Čtvrtlíková, D. M. P. Galassi, M. Hájek, I. Hawes, Z. Levkov, L. Naselli-Flores, A. A. Saber, M. Di Cicco, B. Fiasca, P. B. Hamilton, J. Kubečka, S. Segadelli, P. Znachor, Characteristics, main impacts, and stewardship of natural and artificial freshwater environments: Consequences for biodiversity conservation. *Water (Switzerland)* **12**, 260 (2020).
6. H. Meltofte, T. Barry, D. Berteaux, H. Bültmann, J. S. Christiansen, J. A. Cook, A. Dahlberg, F. J. A. Daniëls, D. Ehrich, J. Fjeldså, F. Fridriksson, B. Ganter, A. J. Gaston, L. J. Gillespie, L. Grenoble, E. P. Hoberg, I. D. Hodkinson, H. P. Huntington, R. A. Ims, A. B. Josefson, S. J. Kutz, S. L. Kuzmin, K. L. Laidre, D. R. Lassuy, P. N. Lewis, C. Lovejoy, C. Michel, V. Mokievsky, T. Mustonen, D. C. Payer, M. Poulin, D. G. Reid, J. D. Reist, D. F. Tessler, F. J. Wrona, *Arctic Biodiversity Assessment: Status and trends in Arctic biodiversity - Scientific report (2013)*, www.caff.is/assessment-series/arctic-biodiversity-assessment/233-arctic-biodiversity-assessment-2013.
7. Q. Wang, J. Liu, G. A. Allen, Y. Ma, W. Yue, K. L. Marr, R. J. Abbott, Arctic plant origins and early formation of circumarctic distributions: A case study of the mountain sorrel, Oxyria digyna. *New Phytol.* **209**, 343–353 (2016).
8. R. J. Abbott, H. P. Comes, Evolution in the Arctic: A phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (Purple saxifrage). *New Phytol.* **161**, 211–224 (2004).
9. H. P. Baird, C. Janion-Scheepers, M. I. Stevens, R. I. Leihy, S. L. Chown, The ecological biogeography of indigenous and introduced Antarctic springtails. *J. Biogeogr.* **46**, 1959–1973 (2019).
10. S. L. Chown, A. Clarke, C. I. Fraser, S. C. Cary, K. L. Moon, M. A. McGeoch, The changing form of Antarctic biodiversity. *Nature* **522**, 431–438 (2015).
11. S. L. Terauds, F. Chown, H. J. Morgan, D. J. Peat, H. Watts, P. Keys, D. M. Convey, D. M. Bergstrom, Conservation biogeography of the Antarctic. *Divers. Distrib.* **18**, 726–741 (2012).
12. C. I. Fraser, R. Nikula, D. E. Ruzzante, J. M. Waters, Poleward bound: Biological impacts of Southern Hemisphere glaciation. *Trends Ecol. Evol.* **27**, 462–471 (2012).
13. P. J. A. Pugh, P. Convey, Surviving out in the cold: Antarctic endemic invertebrates and their refugia. *J. Biogeogr.* **35**, 2176–2186 (2008).
14. P. Convey, J. A. E. Gibson, C. D. Hillenbrand, D. A. Hodgson, P. J. A. Pugh, J. L. Smellie, M. I. Stevens, Antarctic terrestrial life—Challenging the history of the frozen continent? *Biol. Rev.* **83**, 103–117 (2008).
15. G. E. Collins, I. D. Hogg, P. Convey, A. D. Barnes, I. R. McDonald, Spatial and temporal scales matter when assessing the species and genetic diversity of springtails (Collembola) in Antarctica. *Front. Ecol. Evol.* **7**, 1–18 (2019).
16. A. Velasco-Castrillón, S. J. McInnes, M. B. Schultz, M. Arróniz-Crespo, C. A. D'Haese, J. A. E. Gibson, B. J. Adams, T. J. Page, A. D. Austin, S. J. B. Cooper, M. I. Stevens, Mitochondrial DNA analyses reveal widespread tardigrade diversity in Antarctica. *Invertebr. Syst.* **29**, 578–590 (2015).
17. J. C. Zachos, N. J. Shackleton, J. S. Revenaugh, H. Pälike, B. P. Flower, Climate response to orbital forcing across the Oligocene-Miocene boundary. *Science* **292**, 274–278 (2001).
18. H. J. G. Dartnall, The freshwater fauna of the south polar region: A 140-year review. *Pap. Proc. R. Soc. Tasmania* **151**, 19–57 (2017).
19. J. Kleinteich, F. Hildebrand, M. Bahram, A. Y. Voigt, S. A. Wood, A. D. Jungblut, F. C. Küpper, A. Quesada, A. Camacho, D. A. Pearce, P. Convey, W. F. Vincent, C. Zarfl, P. Bork, D. R. Dietrich, Pole-to-pole connections: Similarities between Arctic and Antarctic microbiomes and their vulnerability to environmental change. *Front. Ecol. Evol.* **5**, 137 (2017).
20. B. J. Finlay, K. Clarke, Ubiquitous dispersal of microbial species. *Nature* **400**, 828 (1999).
21. B. J. Finlay, Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063 (2002).
22. J. Davison, M. Moora, M. Öpik, A. Adholeya, L. Ainsaar, A. Bâ, S. Burla, A. G. Diedhiou, I. Hiiesalu, T. Jairus, N. C. Johnson, A. Kane, K. Koorem, M. Kochar, C. Ndiaye, M. Pärtel, S. Reier, R. Singh, M. Vasar, M. Zobel, Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **349**, 970–973 (2015).
23. C. de Vargas, S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, C. Berney, N. Le Bescot, I. Probert, M. Carmichael, J. Poulain, S. Romac, S. Colin, J.-M. Aury, L. Bittner, S. Chaffron, M. Dunthorn, S. Engelen, O. Flegontova, L. Guidi, A. Horák, O. Jaillon, G. Lima-Mendez, J. Lukeš, S. Malviya, R. Morard, M. Mulot, E. Scalco, R. Siano, F. Vincent, A. Zingone, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, S. G. Acinas, P. Bork, C. Bowler, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, F. Not, H. Ogata, S. Pesant, J. Raes, M. E. Sieracki, S. Speich, L. Stemmann, S. Sunagawa, J. Weissenbach, P. Wincker, E. Karsenti, E. Boss, M. Follows, L. Karp-Boss, U. Krzic, E. G. Reynaud, C. Sardet, M. B. Sullivan, D. Velayoudon, Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
24. L. Zinger, P. Taberlet, H. Schimann, A. Bonin, F. Boyer, M. De Barba, P. Gaucher, L. Gielly, C. Giguet-Covex, A. Iribar, M. Réjou-Méchain, G. Rayé, D. Rioux, V. Schilling, B. Tymen, J. Viers, C. Zouiten, W. Thuiller, E. Coissac, J. Chave, Body size determines soil community assembly in a tropical forest. *Mol. Ecol.* **28**, 528–543 (2019).
25. F. Burki, A. J. Roger, M. W. Brown, A. G. B. Simpson, The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
26. M. Bahram, F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, J. Huerta-Cepas, M. H. Medema, M. R. Maltz, S. Mundra, P. A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, P. Bork, Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
27. C. J. van der Gast, Microbial biogeography: The end of the ubiquitous dispersal hypothesis? *Environ. Microbiol.* **17**, 544–546 (2015).
28. S. Louca, The rates of global bacterial and archaeal dispersal. *ISME J.* **16**, 159–167 (2022).
29. S. Louca, P. M. Shih, M. W. Pennell, W. W. Fischer, L. W. Parfrey, M. Doebeli, Bacterial diversification through geological time. *Nat. Ecol. Evol.* **2**, 1458–1467 (2018).
30. J. Marin, F. U. Battistuzzi, A. C. Brown, S. B. Hedges, The Timetree of Prokaryotes: New insights into their evolution and speciation. *Mol. Biol. Evol.* **34**, 437–446 (2017).
31. J. P. Scholl, J. J. Wiens, Diversification rates and species richness across the Tree of Life. *Proc. R. Soc. B Biol. Sci.* **283**, 20161334 (2016).
32. D. L. Rabosky, J. Chang, P. O. Title, P. F. Cowman, L. Sallan, M. Friedman, K. Kaschner, C. Garilao, T. J. Near, M. Coll, M. E. Alfaro, An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395 (2018).
33. J. Igea, A. J. Tanentzap, Angiosperm speciation cools down in the tropics. *Ecol. Lett.* **23**, 692–700 (2020).
34. D. Schluter, M. W. Pennell, Speciation gradients and the distribution of biodiversity. *Nature* **546**, 48–55 (2017).
35. S. Louca, M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli, L. W. Parfrey, Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2**, 936–943 (2018).
36. T. J. Mackey, D. Y. Sumner, I. Hawes, A. D. Jungblut, J. Lawrence, S. Leidman, B. Allen, Increased mud deposition reduces stromatolite complexity. *Geology* **45**, 663–666 (2017).
37. K. Weisleitner, A. Perras, C. Moissl-Eichinger, D. T. Andersen, B. Sattler, Source environments of the microbiome in perennially ice-covered Lake Untersee, Antarctica. *Front. Microbiol.* **10**, 1019 (2019).
38. M. W. Van Goethem, T. P. Makhalanyane, A. Valverde, S. C. Cary, D. A. Cowan, Characterization of bacterial communities in lithobionts and soil niches from Victoria Valley, Antarctica. *FEMS Microbiol. Ecol.* **92**, fiw051 (2016).
39. N. O. Schulte, A. L. Khan, E. W. Smith, A. Zoumplis, D. Kaul, A. E. Allen, B. J. Adams, D. M. McKnight, Blowin' in the wind: Dispersal, structure, and metacommunity dynamics of aeolian diatoms in the McMurdo Sound region, Antarctica. *J. Phycol.* **58**, 36–54 (2022).
40. E. Pinseel, S. B. Janssens, E. Verleyen, P. Vanormelingen, T. J. Kohler, E. M. Biersma, K. Sabbe, B. Van de Vijver, W. Vyverman, Global radiation in a rare biosphere soil diatom. *Nat. Commun.* **11**, 2382 (2020).
41. R. C. Edgar, UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
42. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
43. F. Mahé, T. Rognes, C. Quince, C. de Vargas, M. Dunthorn, Swarm v2: Highly-scalable and high-resolution amplicon clustering. *PeerJ.* **3**, e1420 (2015).
44. T. De Bie, L. De Meester, L. Brendonck, K. Martens, B. Goddeeris, D. Ercken, H. Hampel, L. Denys, L. Vanhecke, K. Van der Gucht, J. Van Wichelen, W. Vyverman, S. A. J. Declerck, Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecol. Lett.* **15**, 740–747 (2012).
45. K. V. Cook, C. Li, H. Cai, L. R. Krumholz, K. D. Hambright, H. W. Paerl, M. M. Steffen, A. E. Wilson, M. A. Burford, H. Grossart, D. P. Hamilton, H. Jiang, A. Sukenik, D. Latour, E. I. Meyer, J. Padisák, B. Qin, R. M. Zamor, G. Zhu, The global *Microcystis* interactome. *Limnol. Oceanogr.* **65**, S194–S207 (2020).
46. J. Soininen, A. Teittinen, Fifteen important questions in the spatial ecology of diatoms. *Freshw. Biol.* **64**, 2071–2083 (2019).

47. A. C. Martiny, K. Treseder, G. Pusch, Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* **7**, 830–838 (2013).
48. C. H. Kuo, H. Ochman, Inferring clocks when lacking rocks: The variable rates of molecular evolution in bacteria. *Biol. Direct* **4**, 35 (2009).
49. D. Singer, E. A. D. Mitchell, R. J. Payne, Q. Blandenier, C. Duckert, L. D. Fernández, B. Fournier, C. E. Hernández, G. Granath, H. Rydin, L. Bragazza, N. G. Koronatova, I. Goia, L. I. Harris, K. Kajukało, A. Kosakyan, M. Lamentowicz, N. P. Kosykh, K. Vellak, E. Lara, Dispersal limitations and historical factors determine the biogeography of specialized terrestrial protists. *Mol. Ecol.* **28**, 3089–3100 (2019).
50. A. De Wever, F. Leliaert, E. Verleyen, P. Vanormelingen, K. Van der Gucht, D. A. Hodgson, K. Sabbe, W. Vyverman, Hidden levels of phylodiversity in Antarctic green algae: Further evidence for the existence of glacial refugia. *Proc. Biol. Sci.* **276**, 3591–3599 (2009).
51. A. D. Jungblut, W. F. Vincent, C. Lovejoy, Eukaryotes in Arctic and Antarctic cyanobacterial mats. *FEMS Microbiol. Ecol.* **82**, 416–428 (2012).
52. J. Laybourn-parry, H. J. Marchant, P. Brown, The plankton of a large oligotrophic freshwater Antarctic lake. *J. Plankton Res.* **13**, 1137–1149 (1991).
53. K. C. McKenna, D. L. Moorhead, E. C. Roberts, J. Laybourn-Parry, Simulated patterns of carbon flow in the pelagic food web of Lake Fryxell, Antarctica: Little evidence of top-down control. *Ecol. Model.* **192**, 457–472 (2006).
54. J. Laybourn-Parry, D. A. Pearce, The biodiversity and ecology of Antarctic lakes: Models for evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 2273–2289 (2007).
55. J. Laybourn-Parry, Survival mechanisms in Antarctic lakes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357**, 863–869 (2002).
56. P. J. A. Pugh, H. J. G. Dartnall, S. J. McInnes, The non-marine Crustacea of Antarctica and the Islands of the Southern Ocean: Biodiversity and biogeography. *J. Nat. Hist.* **36**, 1047–1103 (2002).
57. V. Mohit, A. Culley, C. Lovejoy, F. Bouchard, W. F. Vincent, Hidden biofilms in a far northern lake and implications for the changing Arctic. *NPJ Biofilms Microbiomes* **3**, 17 (2017).
58. W. F. Vincent, J. Laybourn-Parry, *Polar Lakes and Rivers* (Oxford Univ. Press, 2008).
59. B. J. van Vuuren, J. E. Lee, P. Convey, S. L. Chown, Conservation implications of spatial genetic structure in two species of oribatid mites from the Antarctic Peninsula and the Scotia Arc. *Antarct. Sci.* **30**, 105–114 (2018).
60. A. Taton, S. Grubisic, P. Balthasart, D. A. Hodgson, J. Laybourn-Parry, A. Wilmotte, Biogeographical distribution and ecological ranges of benthic cyanobacteria in East Antarctic lakes. *FEMS Microbiol. Ecol.* **57**, 272–289 (2006).
61. E. Verleyen, B. Van de Vijver, B. Tytgat, E. Pinseel, D. A. Hodgson, K. Kopalová, S. L. Chown, E. Van Ranst, S. Imura, S. Kudoh, W. Van Nieuwenhuyze, K. Sabbe, W. Vyverman, Diatoms define a novel freshwater biogeography of the Antarctic. *Ecography* **44**, 548–560 (2021).
62. S. L. Chown, P. Convey, Antarctic entomology. *Annu. Rev. Entomol.* **61**, 119–137 (2016).
63. E. Pinseel, B. Van de Vijver, A. P. Wolfe, M. Harper, D. Antoniades, A. C. Ashworth, L. Ector, A. R. Lewis, B. Perren, D. A. Hodgson, K. Sabbe, E. Verleyen, W. Vyverman, Extinction of austral diatoms in response to large-scale climate dynamics in Antarctica. *Sci. Adv.* **7**, eabh3233 (2021).
64. B. David, T. Saucèdes, *Biodiversity of the Southern Ocean* (Elsevier, 2015).
65. A. R. Lewis, D. R. Marchant, A. C. Ashworth, L. Hedenäs, S. R. Hemming, J. V. Johnson, M. J. Leng, M. L. Machlus, A. E. Newton, J. I. Raine, J. K. Willenbring, M. Williams, A. P. Wolfe, Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10676–10680 (2008).
66. A. McGaughran, A. Terauds, P. Convey, C. I. Fraser, Genome-wide SNP data reveal improved evidence for Antarctic glacial refugia and dispersal of terrestrial invertebrates. *Mol. Ecol.* **28**, 4941–4957 (2019).
67. M. Saługa, R. Ochyra, M. Ronikier, Phylogeographical breaks and limited connectivity among multiple refugia in a pan-Antarctic moss species. *J. Biogeogr.* **49**, 1991–2004 (2022).
68. M. Pontarp, L. Bunnefeld, J. S. Cabral, R. S. Etienne, S. A. Fritz, R. Gillespie, C. H. Graham, O. Hagen, F. Hartig, S. Huang, R. Jansson, O. Maliet, T. Münkemüller, L. Pellissier, T. F. Rangel, D. Storch, T. Wiegand, A. H. Hurlbert, The latitudinal diversity gradient: Novel understanding through mechanistic eco-evolutionary models. *Trends Ecol. Evol.* **34**, 211–223 (2019).
69. T. D. Price, D. M. Hooper, C. D. Buchanan, U. S. Johansson, D. T. Tietze, P. Alström, U. Olsson, M. Ghosh-Harihar, F. Ishtiaq, S. K. Gupta, J. Martens, B. Harr, P. Singh, D. Mohan, Niche filling slows the diversification of Himalayan songbirds. *Nature* **509**, 222–225 (2014).
70. J. A. Crame, Key stages in the evolution of the Antarctic marine fauna. *J. Biogeogr.* **45**, 986–994 (2018).
71. H. P. Baird, S. Shin, R. G. Oberprieler, M. Hullé, P. Vernon, K. L. Moon, R. H. Adams, D. D. McKenna, S. L. Chown, Fifty million years of beetle evolution along the Antarctic Polar Front. *Proc. Natl. Acad. Sci. U.S.A.* **118**, (2021).
72. K. A. Hughes, D. A. Cowan, A. Wilmotte, Protection of Antarctic microbial communities – 'out of sight, out of mind'. *Front. Microbiol.* **6**, 151 (2015).
73. I. Hawes, C. Howard-Williams, N. Gilbert, K. A. Hughes, P. Convey, A. Quesada, The need for increased protection of Antarctica's inland waters. *Antarct. Sci.* **35**, 64–88 (2023).
74. K. Sabbe, D. A. Hodgson, E. Verleyen, A. Taton, A. Wilmotte, K. Vanhoutte, W. Vyverman, Salinity, depth and the structure and composition of microbial mats in continental Antarctic lakes. *Freshw. Biol.* **49**, 296–319 (2004).
75. D. A. Hodgson, W. Vyverman, E. Verleyen, K. Sabbe, P. R. Leavitt, A. Taton, A. H. Squier, B. J. Keely, Environmental factors influencing the pigment composition of in situ benthic microbial communities in east Antarctic lakes. *Aquat. Microb. Ecol.* **37**, 247–263 (2004).
76. T. J. Osborn, P. D. Jones, The CRUTEM4 land-surface air temperature data set: Construction, previous versions and dissemination via Google Earth. *Earth Syst. Sci. Data.* **6**, 61–68 (2014).
77. P. D. Jones, D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, C. P. Morice, Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res. Atmos.* **117**, (2012).
78. P. W. Stackhouse, J. M. Kusterer, "Surface meteorology and solar energy (SSE)," *NASA Tech Rep.* (Release 6.0, ID 20080012200) (2008).
79. G. Zwart, W. D. Hiorns, B. A. Methé, M. P. van Agterveld, R. Huismans, S. C. Nold, J. P. Zehr, H. J. Laanbroek, Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. *Syst. Appl. Microbiol.* **21**, 546–556 (1998).
80. C. Corinaldesi, R. Danovaro, A. Dell'Anno, Simultaneous recovery of extracellular and intracellular DNA suitable for molecular studies from marine sediments. *Appl. Environ. Microbiol.* **71**, 46–50 (2005).
81. T. Stoeck, D. Bass, M. Nebel, R. Christen, M. D. M. Jones, H. W. Breiner, T. A. Richards, Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
82. U. Edwards, T. Rogall, H. Blöcker, M. Emde, E. C. Böttger, Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* **17**, 7843–7853 (1989).
83. I. Cleenwerck, N. Camu, K. Engelbeen, T. De Winter, K. Vandemeulebroecke, P. De Vos, L. De Vuyst, *Acetobacter ghanensis* sp. nov., a novel acetic acid bacterium isolated from traditional heap fermentations of Ghanaian cocoa beans. *Int. J. Syst. Evol. Microbiol.* **57**, 1647–1652 (2007).
84. A. Engelbrektson, V. Kunin, K. C. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman, P. Hugenholtz, Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J.* **4**, 642–647 (2010).
85. J. J. Kozich, S. L. Westcott, N. T. Baxter, S. K. Highlander, P. D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
86. K. Zhang, T. Kobert, A. Flouri, PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
87. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
88. P. D. Schloss, J. Handelsman, Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**, 1501–1506 (2005).
89. N. Youssef, B. L. Steidley, M. S. Elshahed, Novel high-rank phylogenetic lineages within a sulfur spring (Zodletone Spring, Oklahoma), revealed using a combined pyrosequencing-Sanger approach. *Appl. Environ. Microbiol.* **78**, 2677–2688 (2012).
90. Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
91. P. D. Schloss, S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, C. F. Weber, Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
92. L. Guillou, D. Bachar, S. Audic, D. Bass, C. Berney, L. Bittner, C. Boutte, G. Burgaud, C. de Vargas, J. Decelle, J. del Campo, J. R. Dolan, M. Dunthorn, B. Edvardsen, M. Holzmann, W. H. C. F. Kooistra, E. Lara, N. Le Bescot, R. Logares, F. Mahé, R. Massana, M. Montresor, R. Morard, F. Not, J. Pawlowski, I. Probert, A.-L. Sauvadet, R. Siano, T. Stoeck, D. Vaulot, P. Zimmermann, R. Christen, The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604 (2012).
93. T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G. L. Andersen, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).

94. J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H.Stevens, H. Wagner, *vegan: Community Ecology Package* (2018); R package version 2.5-2.
95. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017); www.R-project.org/.
96. M. J. Anderson, T. J. Willis, Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. *Ecology* **84**, 511–525 (2003).
97. R. Kindt, *BiodiversityR: Package for Community Ecology and Suitability Analysis* (2019); R package version 2.11-3.
98. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
99. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, E. Teeling, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
100. D. T. Hoang, O. Chernomor, A. Von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
101. E. Paradis, K. Schliep, Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
102. M. R. Smith, *TreeTools: Create, Modify and Analyse Phylogenetic Trees* (2019); R package version 1.9.
103. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
104. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Y. Lam, GGTREE: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
105. S. Louca, M. Doebeli, Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
106. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B.* **57**, 289–300 (1995).
107. J. M. Ver Hoef, P. L. Boveng, Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology* **88**, 2766–2772 (2007).
108. R. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means* (2019); R package version 1.8.9.
109. K. Tapolczai, V. Vasselon, A. Bouchez, C. Stenger-Kovács, J. Padisák, F. Rimet, The impact of OTU sequence similarity threshold on diatom-based bioassessment: A case study of the rivers of Mayotte (France, Indian Ocean). *Ecol. Evol.* **9**, 166–179 (2019).
110. S. I. Glassman, J. B. H. Martiny, Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere* **3**, e00148-18 (2018).
111. D. Ruppert, M. P. Wand, R. J. Carroll, *Semiparametric Regression* (Cambridge Univ. Press, 2010).
112. M. Schirmer, R. D'Amore, U. Z. Ijaz, N. Hall, C. Quince, Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics.* **17**, 125 (2016).

Acknowledgments: We thank S. D'hondt for executing the DNA extractions and helping with the library preparation and E. Van de Vyver for composing the eukaryotic mock community. We wish to thank the Australian Antarctic Division, the British Antarctic Survey, the Japanese Institute for Polar Research, the South African National Antarctic Program, the Dirección Nacional del Antártico (DNA) and Instituto Antártico Argentino (IAA), the International Polar Foundation (IPF), the Programma Nazionale di Ricerche in Antartide (PNRA), and the Czech Centre for Polar Ecology for providing logistic and/or financial support during the sampling campaigns. In addition, we would like to thank all the people from the Lagos group involved in the Vega Island fieldwork. We also wish to thank D. Obbels, J. Gibson, and P. Vanormelingen for sampling. Lastly, we also wish to thank the reviewers for their valuable comments to improve the manuscript. **Funding:** This work was supported by Belgian Science Policy Office (BelSPo) project CCAMBIO (SD/BA/03); BelSPo project SAFRED (BR/154/A6/SAFRED); EU Interact project Mibipol; Belgian American Educational Foundation (BAEF) (to K.V.d.B.); Research Foundation - Flanders (FWO) no. 1S 418 16 N (to K.V.d.B.); FWO Postdoctoral research grant no. 1246220 N (to K.V.d.B.); Research Foundation – Flanders (FWO) no. 1104315 N (to E.P.); Research Foundation – Flanders (FWO) no. 1104317 N (to E.P.); Research Foundation – Flanders (FWO) no. 1221323 N (to E.P.); Simons Foundation, grant no. 725407 (to E.P.); FRS-FNRS (A.Wilm.); NERC core funding to the BAS 'Biodiversity, Evolution and Adaptation' Team (to P.C.); Charles University Research Centre program no. 204069 (to K.K.); Ministry of Education, Youth and Sports of the Czech Republic grant CENAKVA LM2018099 (O.S.); European Regional Development Fund–Project No. CZ.02.1.01/0.0/0.0/16_025/0007370 (O.S.); Kurchatov Center of Genome Research Agreement #075-15-2019-1659 (Z.N.); The Czech Ministry of Education (MŠMT in Czech); The government fund of the Czech Republic; European Social Fund grant LM2010009; European Social Fund grant CZ.1.07/2.2.00/28.0190; European Social Fund grant RVO67985939; and the Italian National Programme for Research in Antarctica (PNRA) (to R.B). **Author contributions:** W.V. and E.V. conceived, developed, and coordinated the study. E.V., W.V., D.A.H., K.S., S.L.C., and E.P. supervised and organized the sampling campaigns together with members of the Polar Lake Sampling Consortium. B.T., M.S., and E.P. performed the laboratory, bioinformatics, and statistical analyses under the supervision of W.V., E.V., A.Wilm., and A.Will. K.V.d.B. provided code for the LTT and GLM analyses. E.P. provided code for the consenTRAIT analysis. W.V., E.V., B.T., and M.S. wrote the manuscript, and all other coauthors commented on the manuscript and contributed substantially to revisions. B.T. made all the figures, and M.S. contributed to Fig. 3. **Competing interests:** S.L.C. declares that he was the president of the Scientific Committee on Antarctic Research from 2016 to 2021. The other authors declare that they have no competing interests. **Data and materials availability:** The raw Illumina 18*S* and 16*S rRNA* reads are available from the NCBI Sequence Read Archive under the BioProject accession number PRJNA1008438. Full alignments and alignments with loop regions removed used for the concenTRAIT analyses are available from the Mendeley Data repository (DOI: 10.17632/bn8vg9m5nt.1). High-resolution images of supplementary figures are also available at this repository. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The R code to model the mean local richness using GLM and the lineage through time plots through GAM modeling, including the toy dataset code, are also available at the Mendeley Data repository (DOI: 10.17632/bn8vg9m5nt.1).

Submitted 30 September 2022
Accepted 19 October 2023
Published 17 November 2023
10.1126/sciadv.ade7130

ScienceAdvances

Polar lake microbiomes have distinct evolutionary histories

Bjorn Tytgat, Elie Verleyen, Maxime Sweetlove, Koen Van den Berge, Eveline Pinseel, Dominic A. Hodgson, Steven L. Chown, Koen Sabbe, Annick Wilmotte, Anne Willems, The Polar Lake Sampling Consortium, and Wim Vyverman

Sci. Adv. **9** (46), eade7130. DOI: 10.1126/sciadv.ade7130