Article

Estimating biodiversity across the tree of life on Mount Everest's southern flank with environmental DNA

GLOBAL: Alpine areas >4500m occupy <3% of total land surface

MT. EVEREST: Water samples collected on south slopes at 4500–5500m for eDNA analysis

Collect water

Capture and sequence DNA

Bioinformatic analysis

Diversity of Taxa

Everest samples contain 1/6 of known taxonomic orders

Marisa C.W. Lim, Anton Seimon, Batya Nightingale, ..., Sandra Elvin, Aurora C. Elmore, Tracie A. Seimon

tseimon@wcs.org

Highlights

First comprehensive eDNA biodiversity survey conducted on Earth's highest mountain

One-sixth (16%) of global orders detected are >4,500m on the south flank of Everest

Identified 187 unique orders from 36 phyla across the six kingdoms

Metabarcoding and WGS approaches provide distinct yet complementary information

iScience

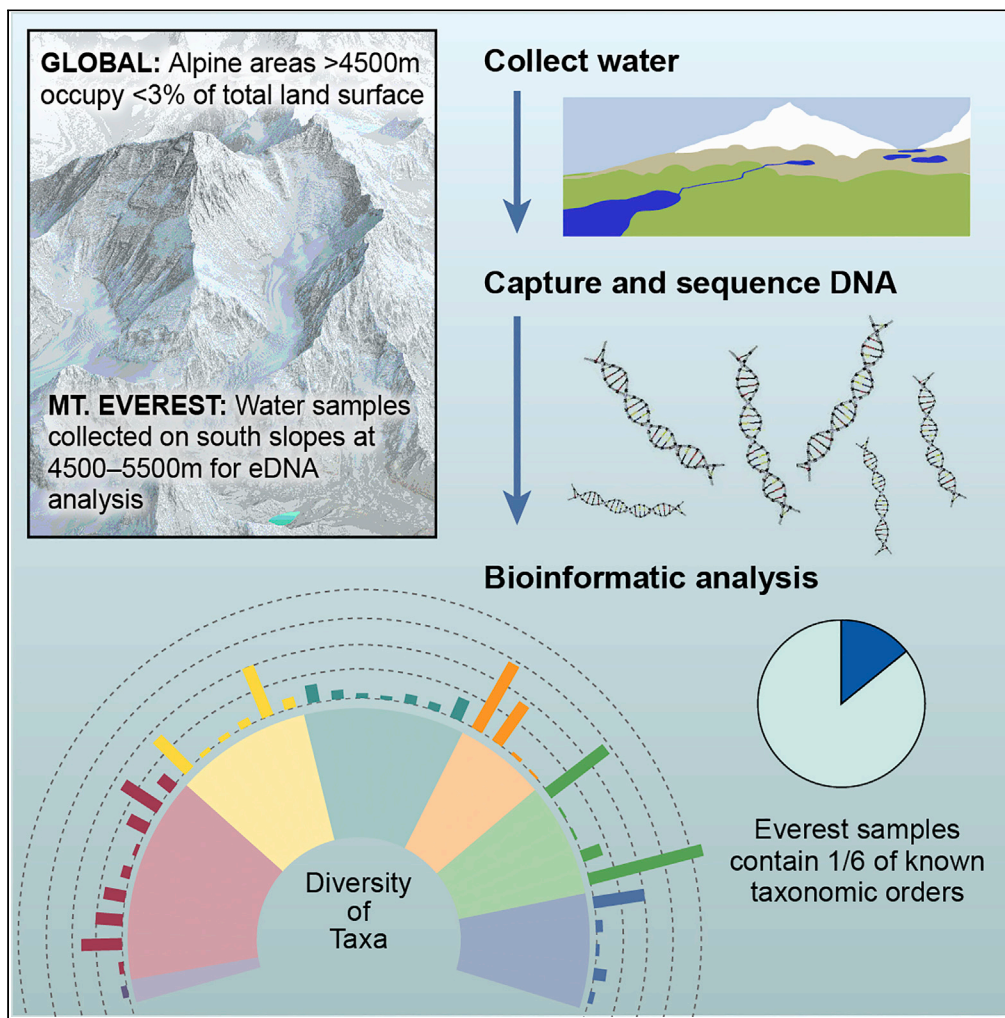## Article

Estimating biodiversity across the tree of life on Mount Everest's southern flank with environmental DNA

Marisa C.W. Lim,[1] Anton Seimon,[2] Batya Nightingale,[1] Charles C.Y. Xu,[3] Stephan R.P. Halloy,[4] Adam J. Solon,[5] Nicholas B. Dragone,[5] Steven K. Schmidt,[5] Alex Tait,[6] Sandra Elvin,[6] Aurora C. Elmore,[6,7] and Tracie A. Seimon[1,8,*]

SUMMARY

Species composition in high-alpine ecosystems is a useful indicator for monitoring climatic and environmental changes at the upper limits of habitable environments. We used environmental DNA (eDNA) analysis to document the breadth of high-alpine biodiversity present on Earth's highest mountain, Mt. Everest (8,849 m a.s.l.) in Nepal's Khumbu region. In April-May 2019, we collected eDNA from ten ponds and streams between 4,500 m and 5,500 m. Using multiple sequencing and bioinformatic approaches, we identified taxa from 36 phyla and 187 potential orders across the Tree of Life in Mt. Everest's high-alpine and aeolian ecosystem. These organisms, all recorded above 4,500 m—an elevational belt comprising <3% of Earth's land surface—represents ∼16% of global taxonomic order estimates. Our eDNA inventory will aid future high-Himalayan biomonitoring and retrospective molecular studies to assess changes over time as climate-driven warming, glacial melt, and anthropogenic influences reshape this rapidly transforming world-renowned ecosystem.

INTRODUCTION

High-elevation alpine, or high-alpine, and aeolian ecosystems are living laboratories that offer opportunities to study life forms adapted to extreme environmental conditions. The alpine zone exists above the tree line and contains an array of flowering plants and shrub species (Swan, 1992). The aeolian zone reaches beyond the range of flowering plants and shrubs at the uppermost reaches of the biosphere. The aeolian zone is composed of exposed rock and snow and may contain glacier-fed pools and streams. This zone is populated with organisms such as mosses, and their nutrients are carried up and deposited by the wind (Swan, 1992). Upward range extensions and contractions in response to ongoing climatic warming in these ecosystems have been documented in a number of taxonomic groups including microbial communities, plants, insects, and vertebrates such as amphibians, mammals, and birds (Grabherr et al., 2001; Moritz et al., 2008; Pauli and Halloy, 2019; Seimon et al., 2006, 2017). This broad spectrum of taxa can be utilized as indicators for long-term monitoring and studying biodiversity changes over time, and in turn, can inform how mountainous ecosystems may change as global temperatures rise. The establishment of baseline knowledge of species diversity and distributions coupled with continued monitoring is needed to document these changes.

The Upper Khumbu region of Nepal is an ideal place to study high-alpine and aeolian biodiversity, particularly around Earth's highest mountain, Mt. Everest (known locally as Sagarmatha and Qomolangma; 8,849 m above sea level (a.s.l.); all elevations hereafter are a.s.l.). The southern flank of the Everest massif contains some of the highest lakes in the Himalayan range, while the subtropical latitude (∼28° N) and placement within the Hindu Kush Himalaya (HKH) mountain system yields precipitation from both warm-season monsoonal flow and cold-season extratropical disturbances. Research to document animal life above 5,000 m began as early as 1924 with R.W.G. Hingston (Hingston, 1925). Since then, seminal work conducted in the Himalayas include biogeographical research with a focus on biodiversity, ecosystem services, and ecology of insects, plants, and microbial species (Dobremez, 1976; Kumar et al., 2019; Mani, 1978; Miehe, 1991; Schmidt et al., 2011; Swan, 1963, 1990; Troll, 1972; Xu et al., 2019). Despite this work,

[1]Wildlife Conservation Society, Zoological Health Program, Bronx Zoo, Bronx, NY 10460, USA

[2]Appalachian State University, Department of Geography and Planning, Boone, NC 28608, USA

[3]McGill University, Redpath Museum and Department of Biology, Montreal, Quebec H3A 0G4, Canada

[4]Ministry for Primary Industries, Wellington 6011, New Zealand

[5]University of Colorado, Department of Ecology and Evolutionary Biology, Boulder, CO 80309, USA

[6]National Geographic Society, Washington, D.C. 20036, USA

[7]National Oceanic and Atmospheric Administration, Silver Spring, MD 20910, USA

[8]Lead contact

*Correspondence: tseimon@wcs.org

https://doi.org/10.1016/j.isci.2022.104848

knowledge of the full breadth of biodiversity present in and around the high-alpine and aeolian zones has been limited owing to the physical challenges of conducting fieldwork in these remote areas. Research on species inhabiting these extreme environments, and their long-term response to climate change is even more limited. To date, few studies have focused on documenting the taxonomic diversity in and around the Mt. Everest region based on DNA sequence approaches. With the exception of microbial and some plant studies, baseline genetic information or species references—particularly for eukaryotic life—in this region are lacking (Gurung et al., 2009; Liu et al., 2006a, 2006b, 2007; Valcárcel et al., 2017; Villaverde et al., 2015).

High-alpine regions are important models for studying the effects of climate change because they host specialized species communities that have adapted to the dynamic habitat and landscape-driven changes over hundreds of thousands of years including advancing and receding glaciers, warming and cooling temperature, shifting water resources, and a high level of spatial and temporal heterogeneity (Pauli and Halloy, 2019; Seimon et al., 2006, 2017). This dynamic landscape of alpine water bodies also presents the opportunity to study ephemeral or recently formed species communities. Long-term biodiversity monitoring can inform communities within the Khumbu region how their environment is changing in response to the warming regional climate. Critical baseline information is needed to document what taxa are present now to measure future change. Our study is motivated by the question: What is the breadth of biodiversity in the uppermost reaches of the biosphere on Earth's highest mountain, and could application of advanced biomonitoring technologies such as environmental DNA (eDNA) be used to establish a baseline inventory of the diversity of life residing there?

Whereas traditional biodiversity surveys are onerous and require several months of fieldwork, eDNA sample collection, combined with DNA metabarcoding and other sequencing methodologies, offers a more accessible, rapid, and comprehensive approach to increasing survey capacity for assessing biodiversity in aquatic environments (Garlapati et al., 2019; Taberlet et al., 2012; Thomsen and Willerslev, 2015). To assess the breadth of biodiversity present in high-Himalayan aquatic ecosystems, we set out to sample some of the highest ponds and streams to compile a baseline genomic library of organisms living in high-alpine and aeolian environments. Such data can help track trends in biodiversity as the environment and water resources change over time. We show how the combined strengths of two genetic approaches to cataloging biodiversity—DNA metabarcoding and whole genome shotgun sequencing—can improve our ability to detect a breadth of taxonomic diversity. We also discuss how the scientific community can refine the eDNA approach and the resulting genomic datasets to monitor changes in organism diversity and distribution, define taxonomy, reconstruct genomes, and supplement water quality biomonitoring and ecosystem health assessments.

RESULTS

Area of study

The Khumbu region is an area undergoing marked environmental changes in response to a warming climate. We investigated proglacial environments undergoing these changes, and how they relate to organism distribution. All sampled lakes, except the Lake South of Nuptse site, lie downstream of receding glaciers and are likely of recent origin. Specific site locations and descriptions are provided in Figure 1 and Table S1, respectively. To better understand the distribution and lake morphology, we compared composite satellite imagery taken in October and November of 2016 and 2001, to images taken in December 1984, 1969, and 1962 (Figure S1). All sampled lakes were identified in 1969 imagery. These lakes were also discernible in the poorly resolved 1962 image (i.e., Above EBC, Kala Pattar Lake 2, Kongma La Lakes 1 and 2, and Lake South of Nuptse), indicating that these lakes have existed for a minimum of 52-58 years. Four of the seven lakes (Kala Pattar Lakes 1 and 2, Lake Above EBC, and Kongma La Lake 1) have bright coloration from glacial silt, indicating that they are directly fed by glacier melt water to the present day. Current deglaciation trends in our field region have been examined in detail (King et al., 2020; Potocki et al., 2022).

Taxonomic classification of organisms

Ten eDNA samples were extracted, processed, and analyzed using two DNA sequencing approaches: 1. Whole genome shotgun (WGS) sequencing and 2. DNA metabarcoding targeting the mitochondrial cytochrome *c* oxidase subunit 1 (CO1) gene. High-quality metabarcoding and WGS paired-end sequence reads were obtained for all samples (Table S2). We focused organism identification on higher taxonomic levels
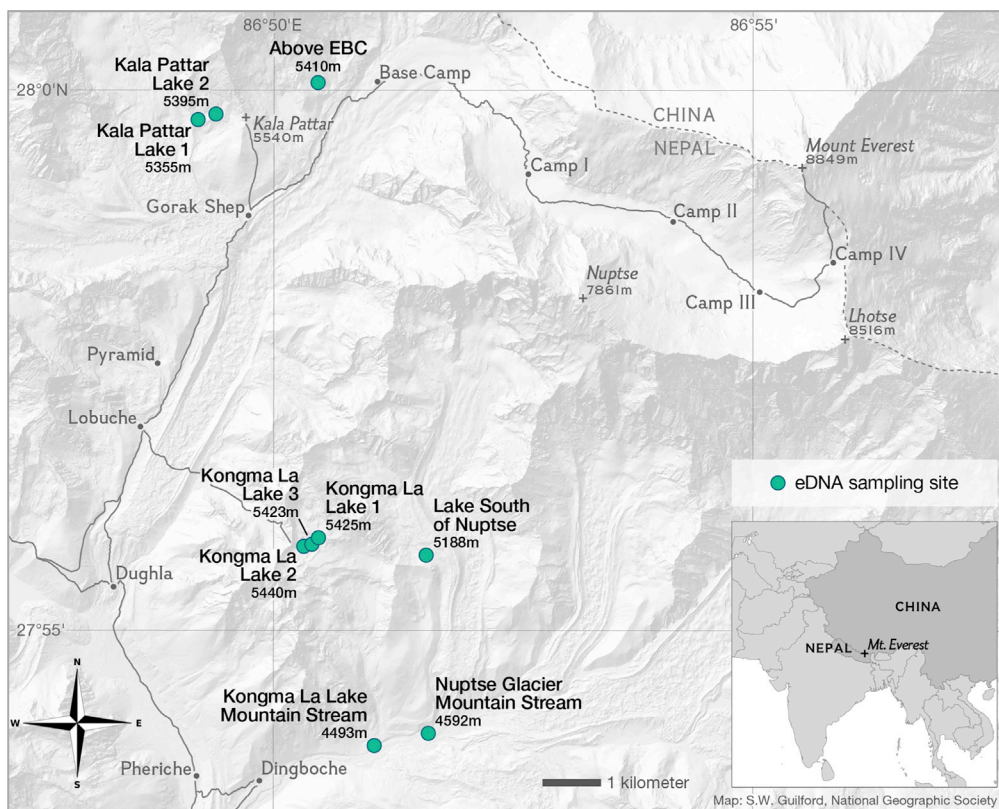
Figure 1. Map of Mount Everest's southern flank, Khumbu region, Nepal, identifying sampling sites and major water catchments visited for eDNA surveys. Scale bar = 1 kilometer.

(kingdom, phylum, order) because species-level reference genetic sequence data are limited from the Mt. Everest region of Nepal.

To assess bacterial diversity, WGS reads were analyzed using SingleM (Woodcroft, n.d.) and the Greengenes microbial reference database (DeSantis et al., 2006; Kumar et al., 2019). Bacterial community structure differed between sites. More sequences from the Lake Above EBC, Kala Pattar Lakes 1 and 2, and Kongma La Lake one samples were identified to bacterial order compared to the other sample sites (Figure 2). The bacterial community profiles of the nearby Kala Pattar Lake one and two sites were similar and both contained bacterial sequences predominantly matching the orders Betaproteobacteriales and Flavobacteriales (Figures 2A and 2B). Despite being further apart and separated by the Khumbu glacier, the overall bacterial profiles from Kala Pattar Lake two and Kongma La Lake one outlet sites were also similar, and both contained sequences from order Cytophagales. The Lake Above EBC contained the highest number of taxonomic orders (27), with sequences identified predominantly to order Opitutales. Kongma La Lake three and Lake South of Nuptse had the lowest number of bacterial orders (3) and contained predominantly Betaproteobacteriales and Sphingomonadales, respectively. Overall, 40 orders from nine phyla of bacteria were detected from all sites combined. Asymptotic regression model analysis of the average number of duplicate samples required to achieve saturation (for the number of unique taxonomic orders detected) shows we detected at least 95% of total orders estimated able to be detected using this method (Figure 2C).

To assess eukaryotic and cyanobacterial microbial diversity, WGS reads were analyzed by Genewiz using Kraken (Wood and Salzberg, 2014) and the RefSeq database (O'Leary et al., 2016). We found that Kongma La Lakes 1 and 2, as well as the Kongma La Mountain Stream and Nuptse Glacier Mountain Stream had the highest abundance of orders, comprising predominantly cyanobacteria (Nostocales, Chroococcales, and Oscillatoriales) and organisms from the phylum Apicomplexa (Eucoccidiorida) (Figure 3). Very few reads could be classified as viruses or archaea, and most of the virus-assigned reads were bacteriophage viruses
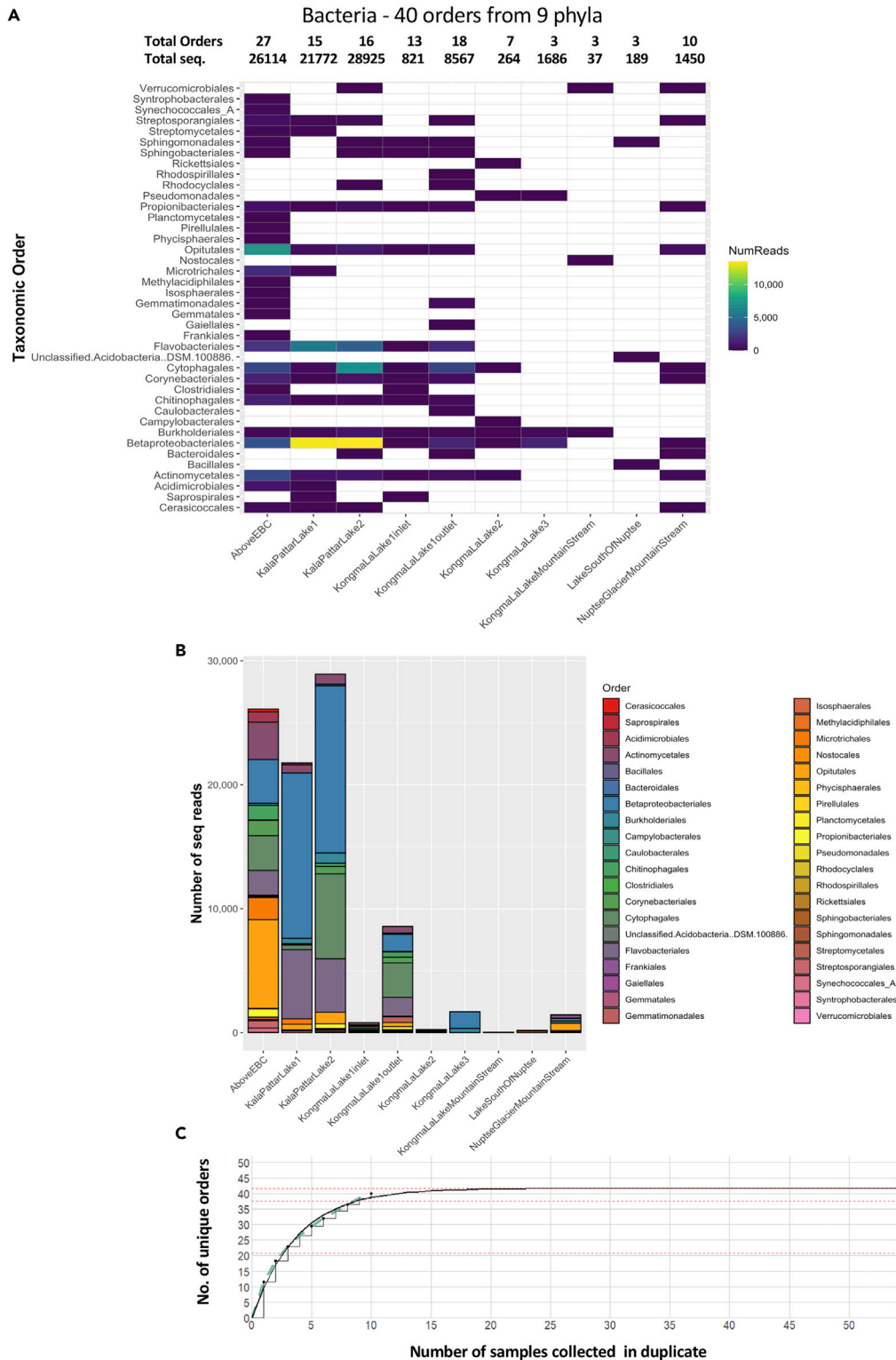
Figure 2. Diversity of bacterial microbes from WGS data
(A and B) Shown is a heatmap (A) and bar plot (B) of the number of sequence reads of bacteria identified and classified to taxonomic order using SingleM and the Greengenes database from each site.

Figure 2. *Continued*

(C) Asymptotic regression model of the average number of duplicate samples required to achieve saturation in the number of unique taxonomic orders detected. The three red dashed lines represent the median (21), 90th percentile (38), and total number (42) of unique orders that could be detected using this method.

of the order Caudovirales (Figure 3). Kala Pattar Lake 1, Kongma La Lake one outlet, and both Kongma La Mountain Stream and Nuptse Glacier Mountain Stream had bacteriophage sequences present. Overall, 41 orders from 15 phyla of microbial eukaryotes, cyanobacteria, and bacteriophage viruses, were identified from all sites combined. Asymptotic regression model analysis of the average number of duplicate samples required to achieve saturation (for the number of unique taxonomic orders detected) shows that saturation was achieved with respect to the total number of orders estimated able to be detected using this method (Figure 3C).

### Metabarcoding and taxonomic classification targeting the CO1 gene

DNA metabarcoding analysis was performed targeting the CO1 gene, and again we focused our investigation on higher taxonomic classifications such as phylum, class, and order. 571 contig sequences were 70-85% similar to GenBank references, and 158 contigs were >90% similar. After increasing filter stringency to >80% identity, we analyzed a total of 667 mitochondrial CO1 contig sequences. From this metabarcoding dataset, we recovered a range of taxa from four kingdoms: Animalia, Chromista, Fungi, and Plantae. The complete results showing all matches from our metabarcoding analysis are available in Data S4. We found that Kala Pattar Lake one and Kongma La Lake two had the highest diversity, with the majority of their sequences classified as fungi (Umbelopsidales and Sporidiobolales). Organisms from the phylum Arthropoda (Ephemeroptera and Odonata) were found only in sites east of the Khumbu Glacier including the Nuptse Glacier Mountain Stream (Figure 4). Contigs from seven of the ten sites had potential matches to aquatic invertebrate species such as mayflies (order Ephemeroptera) and dragonflies or damselflies (order Odonata). From Kala Pattar Lake 1 (TS17), we identified a 310 base pair (bp) contig matching 100% to the reference for *Fujientomon dicestum* (order Protura), a hexapod of the family Fujientomidae, which has only two known species, found in China and Japan (Dell'Ampio et al., 2011). Samples from Kongma La Lake 2 (TS24) and Kongma La Lake Mountain Stream (TS31) sites each had a 310bp contig with 100% match to CO1 sequences from several species of the order Cupressales, which also includes the genus *Juniperus*. Owing to the highly conserved nature of CO1 for Cupressales, we were unable to confirm the species of these particular sequences.

Fifteen orders from nine phyla were identified by metabarcoding from all sites combined. Asymptotic regression model analysis of the average number of duplicate samples required to achieve saturation in the number of unique taxonomic orders detected shows we detected at least 83% of the total orders estimated to be detected using this method (Figure 4C).

Using the whole genome shotgun database to extract and identify eukaryotic taxa using reference genomes

Identification of eukaryotes from metagenomic WGS samples is challenging because they comprise a tiny fraction of the dataset and there is limited genomic reference material. To experiment with the detection ability for eukaryotic organisms in the WGS dataset and provide targets for future metabarcoding, we mapped reads to a broad taxonomic range of reference genomes from species expected to inhabit the region (Figure 5, and Table S3). From this analysis, we generated 7,070 contig sequences, and after filtering for >85% identity (4,889 sequences), we identified 115 potential orders from 21 phyla (Figures 5A and 5B) (Data S5). The majority of the sequences were represented by the orders Arthropoda (80.1% of listed entities) and Streptophyta (9.1% of listed entities). A range of arthropod orders was present in the dataset, including stoneflies (order Plecoptera), mayflies (order Ephemeroptera), butterflies, or moths (order Lepidoptera), true bugs (order Hemiptera), true flies (order Diptera), and beetles (order Coleoptera). Other eukaryote animal phyla included: Rotifera (70-90% sequence similarity from the *Adineta*, *Habrotrocha*, *Macrotrachela*, *Philodina*, and *Rotaria* genera), Tardigrada (93.6% sequence similarity and 98% query coverage to the *Hypsibius dujardini*), Chordata (97% match to domestic dog, *Canis familiaris* and 100% match to Tibetan snowcock, *Tetraogallus tibetanus*). Algal phyla were represented by Chlorophyta and Rhodophyta. In our dataset, the orders in phyla Streptophyta included Pinopsida (gymnosperms), Magnoliopsida (flowering plants), Bryopsida (mosses and liverworts), and Polypodiopsida (ferns). In Magnoliopsida, the dominant orders are Fagales, Asterales, and Asparagales; the remainder of the sequences match

A Microbial eukaryotes, cyanobacteria, fungi and viruses: 41 orders from 15 phyla

B

C
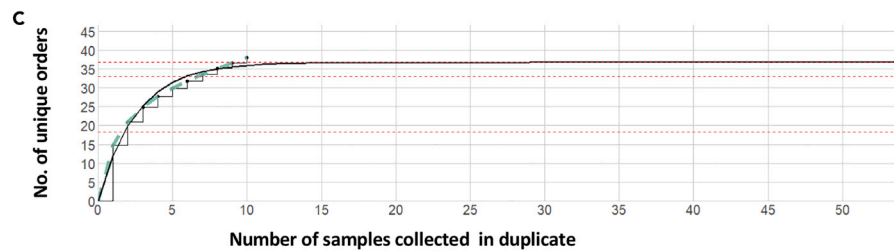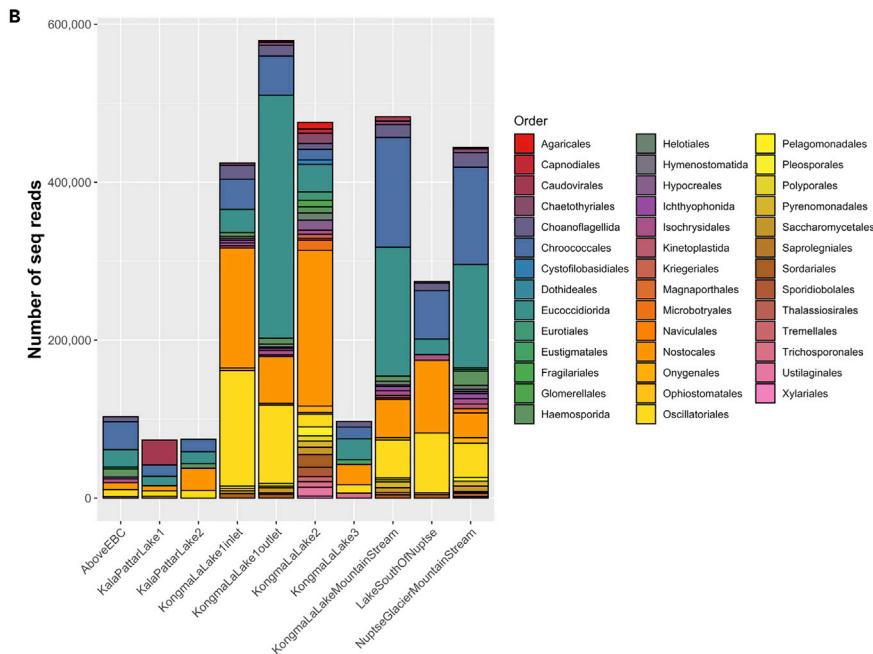
Figure 3. Diversity of single cellular eukaryotes, cyanobacteria, fungi, and viruses using Kraken analysis from WGS data

(A and B) Shown is a heatmap (A) and bar plot (B) of the number of sequence reads of microbial eukaryotes, cyanobacteria, and viruses identified and classified to taxonomic order using Kraken and the RefSeq database from each site.

(C) Asymptotic regression model of the average number of duplicate samples required to achieve saturation in the number of unique taxonomic orders detected. The three red dashed lines represent the median (18), 90th percentile (33), and total number (37) of unique orders that could be detected using this method.

references to nine other orders (Caryophyllales, Ericales, Fabales, Lamiales, Liliales, Poales, Ranunculales, Rosales and Sapindales).

In Pinopsida, three orders are represented: Cupressales, Pinales, and Araucariales. Consistent with results from the metabarcoding data, we detected sequences from the order Cupressales in the WGS data with >99% sequence similarity to *Juniperus squamata*, *Juniperus indica*, and *Juniperus recurva* at four sites: Kongma La Lake 2, Lake South of Nuptse, Kongma La Mountain Stream, and Nuptse Glacier Mountain Stream. The latter two sites were dominated by *Juniperus cf. indica* shrubs that were observed growing around the streams we sampled from. Owing to the conserved nature of the genes detected in our dataset, and the incomplete reference sequence for *J. indica*, identification to the species level could not be confirmed.
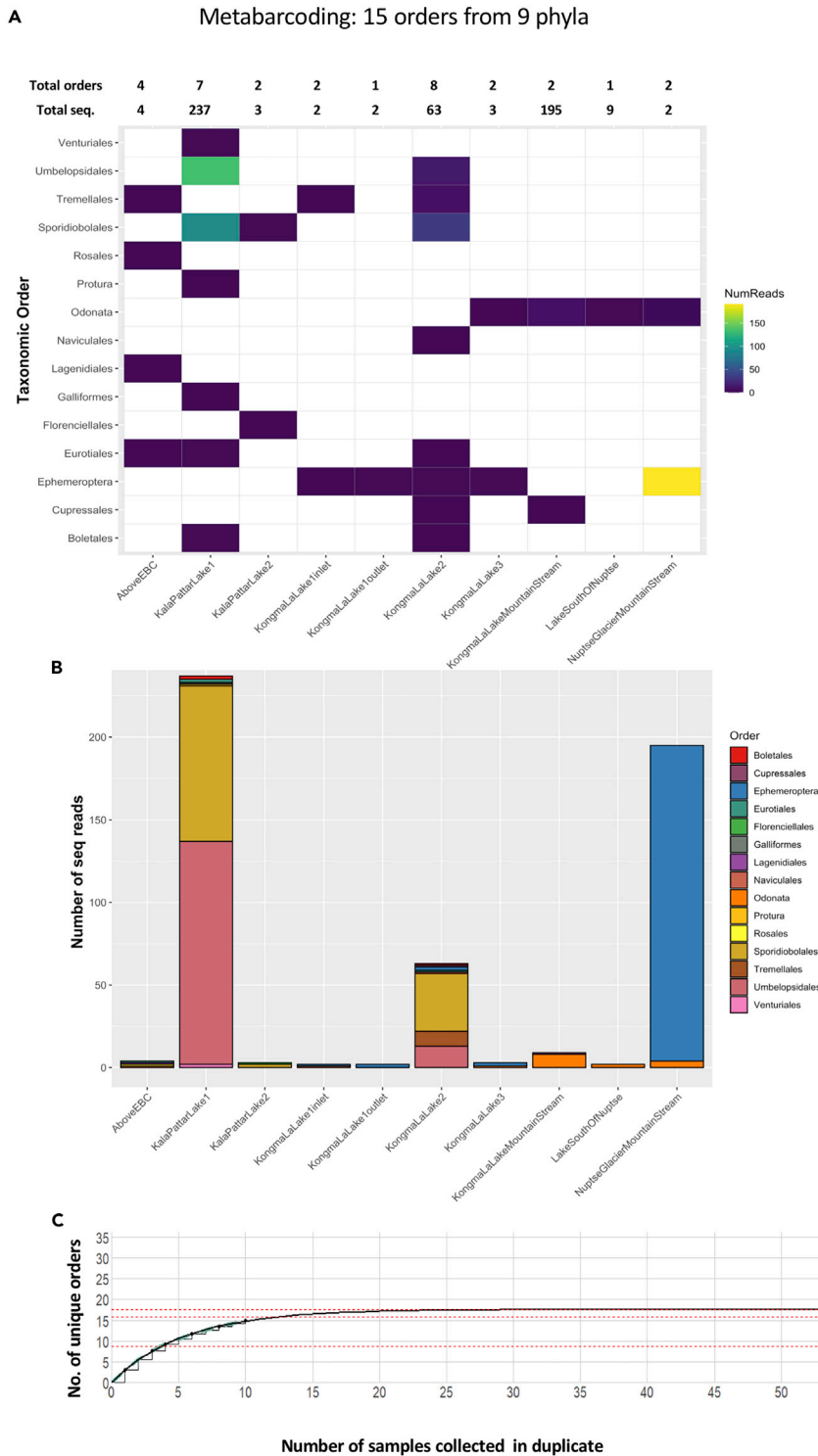
Sequence similarities to each reference sequence can be found in Data S5. Asymptotic regression model analysis of the average number of duplicate samples required to achieve saturation in the number of unique taxonomic orders detected shows we detected at least 79% of total orders that are estimated able to be detected using this method (Figure 5C). When modeling the distribution in the number of unique orders identified using up to nine reference genomes, asymptotic regression model analysis indicates that at least 23 different reference genomes would be needed to achieve 90% saturation using this method (Figure S2B).

By combining the above results, we detected a total of 187 orders from 36 phyla across seven kingdoms from ten sites and 20 L of water from the Khumbu region of Mount Everest (Figure 6). The number of orders detected across samples and sites, as well as across analysis methods (Figure S3) is presented in Table S4 and Data S6.

DISCUSSION

Catalogs of organism assemblages from the upper reaches of the biosphere in high-alpine and aeolian zones are informative for investigations on extremophile life and lay the foundation for further research about how species are adapted to living at the elevational limits of what is considered habitable. For organisms to survive and live at these extreme elevations (4,500-5,500 m, abutting the aeolian nival zone), they have to be adapted to living in environmental conditions that include hypoxic low pressure of the mid-troposphere (~50% of mean sea level pressure at 5,400 m), low CO_2, intense solar irradiation, low absolute humidity, diurnal freeze-thaw cycles, sub-freezing temperatures reached nocturnally, as well as deep snow and ice that can persist for half the year (Halloy, 1989; Pauli and Halloy, 2019; Spehn et al., 2010). Biodiversity inventories are still being developed for the HKH ranges, and new species are still being discovered. For example, from 1998 to 2008, 353 new species were discovered in the Eastern Himalayas alone, including 242 plants, 16 amphibians, 16 reptiles, 14 fish, two birds, two mammals, and 61 invertebrates (World Wildlife Fund, 2012). Our study is the first to apply environmental DNA sequencing methods to assess the breadth of diversity across the Tree of Life in the high-alpine ecosystem of the HKH ranges and Mt. Everest, where scientists have documented some of the world's highest plants and animals (Halloy, 1989; Swan, 1963, 1990). However, there are two fundamental questions raised by our study and approach. One is methodological: How do the eDNA and bioinformatic analysis methods perform in terms of identifying the local biota (i.e. taxonomic richness and composition)? Having constructed a snapshot of biodiversity present during expedition surveys, the second question asks: How does the estimated biodiversity at the HKH sites compare to other high-alpine aquatic sites around the world? As there are no eDNA studies across the Tree of Life that are directly comparable to high-alpine aquatic environments, this question cannot be answered until more data become available and rigorous meta-analysis is conducted across studies (King et al., 2010; Liu et al., 2006a, 2006b, 2009; Schmidt et al., 2011, 2018; Seimon et al., 2006; Vimercati et al., 2019) for available taxa. We, therefore, discuss observations from existing research and share lessons learned to help standardize eDNA collection and analysis approaches for future surveys.
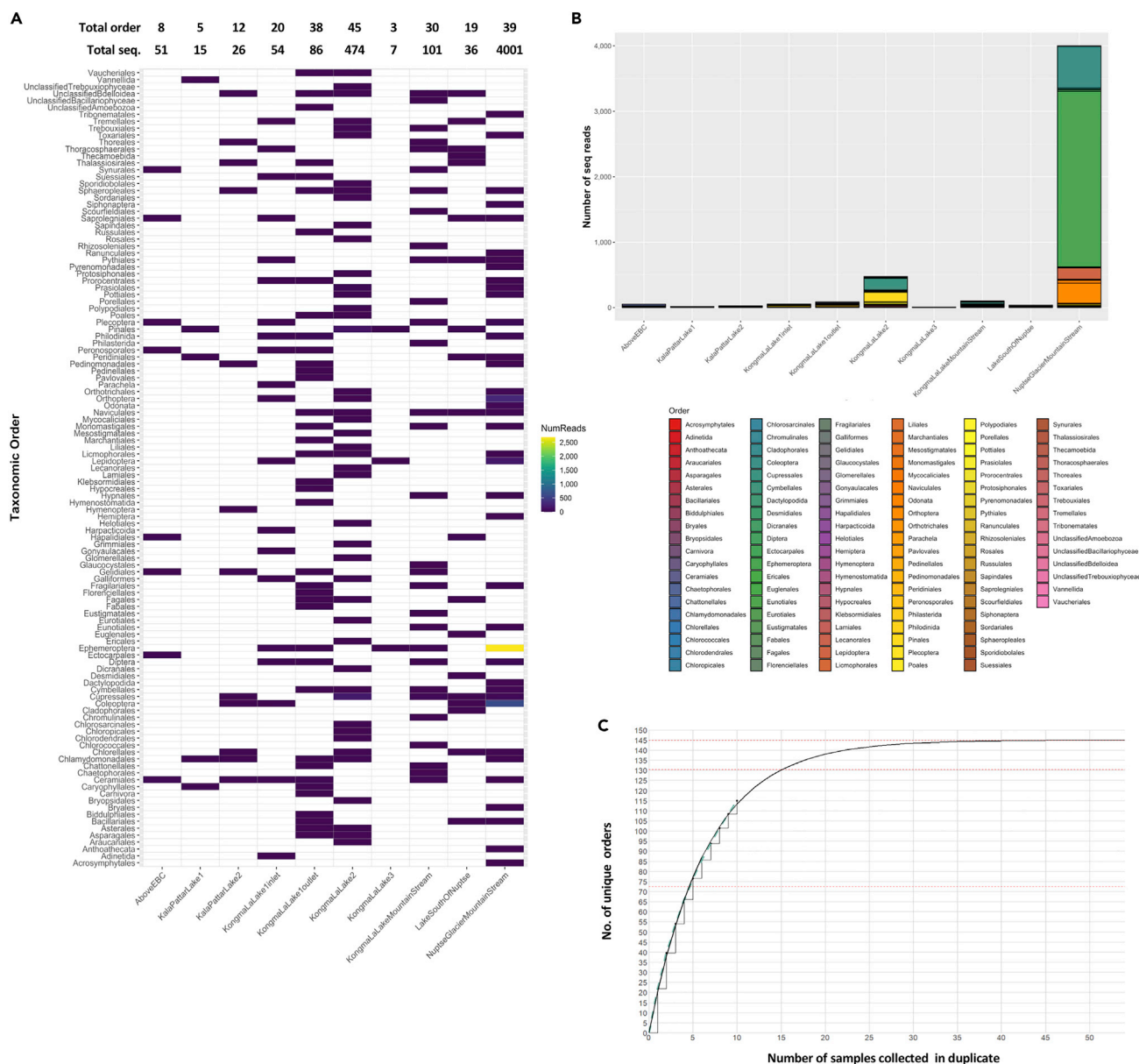
Figure 4. Diversity organisms recovered using metabarcoding targeting the CO1 gene

(A and B) Shown is a heatmap (A) and bar plot (B) of the number of contigs obtained from primers targeting the mitochondrial CO1 gene classified to taxonomic order from each site.

(C) Asymptotic regression model of the average number of duplicate samples required to achieve saturation in the number of unique taxonomic orders detected. The three red dashed lines represent the median (8), 90th percentile (16), and total number (18) of unique orders that could be detected using this method.

Figure 5. Diversity of organisms recovered using WGS data and reference sequences

(A and B) Shown is a heatmap (A) and bar plot (B) of the number of sequence reads classified to taxonomic order per site from the WGS dataset. WCS data from each sample were mapped to an array of reference sequences, and resulting contigs and sequences were identified based on the BLASTn search to the NCBI reference database. The data were filtered to include BLAST results with >85% sequence similarity to GenBank references, >90% query coverage, and >200bp alignment length between query and NCBI reference.

(C) Asymptotic regression model of the average number of duplicate samples required to achieve saturation in the number of unique taxonomic orders detected. The three red dashed lines represent the median (73), 90th percentile (130), and total number (145) of unique orders that could be detected using this method.

Estimated diversity of organisms above 4,500m in the Mt. Everest region

In total, we identified 187 unique orders from 36 phyla in our survey of the Mt. Everest region above 4,500 m (Data S6, and Figure 6). When compared to the estimated 1,150 global total orders (Mora et al., 2011), our data suggest that the high-alpine ecosystem on Mt. Everest contains at least 16.3% of global taxonomic
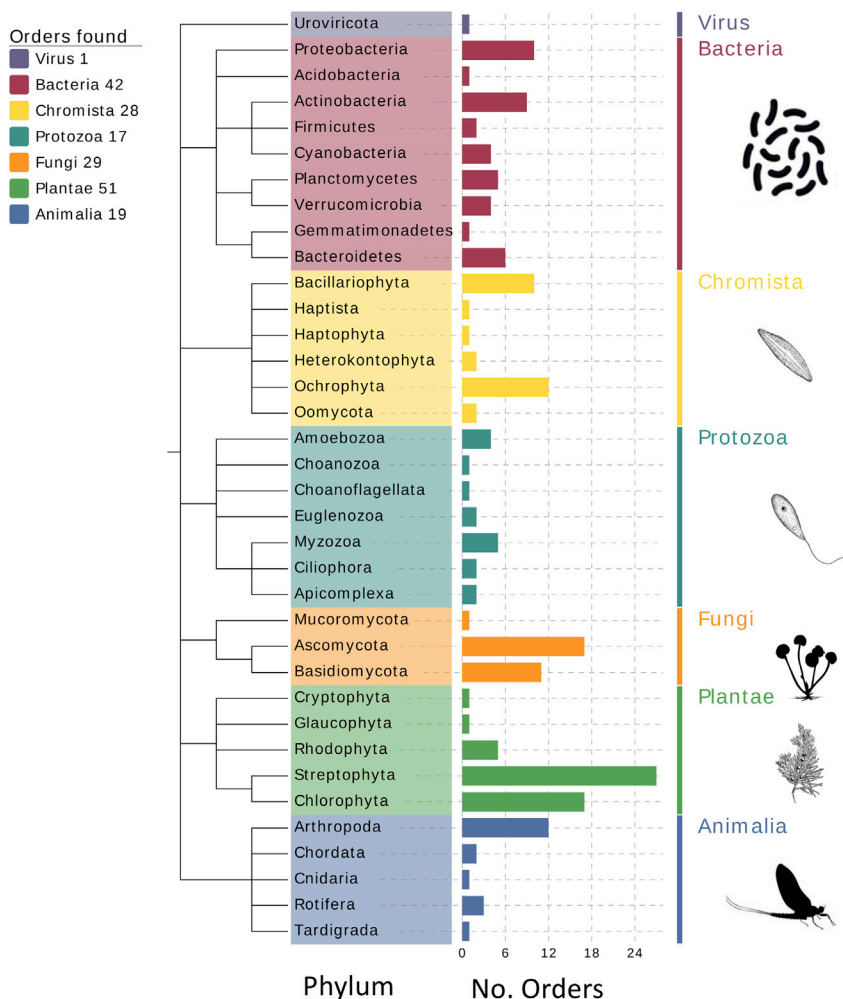
Figure 6. Organisms detected across the Tree of Life from a high-alpine ecosystem
Number of Orders and Phyla recovered across seven kingdoms from all samples combined in the high-alpine elevational bracket (4,500 - 5,500 m) on Mount Everest's southern flank, Khumbu region, Nepal.

orders. Our analysis identified: 19 of an estimated 500 Animalia orders; 28 of 115 Chromista orders; 29 of 160 Fungi orders; 51 of 200 Plantae orders (majority were algae); 17 of 60 Protozoa orders; and 43 orders of Bacteria. Importantly, caution must be exercised when interpreting identified taxa from eDNA. For example, some of the plant orders we identified through eDNA, such as Pinales, are not known to exist above the tree line. The eDNA captured is presumably from pollen carried upslope by wind and deposited in the water, which would lead to an overestimation in some orders (notably plants that travel as pollen) at high-alpine sampling sites. Nevertheless, based on our modeling of the order richness curves for the bioinformatic methodologies used to assess the number of organisms present, the estimate of 187 orders is likely an under-representation and additional targeted analysis of our data may reveal more taxa. Further analyses incorporating additional metabarcoding approaches to assess arthropod, bacterial, and viral diversity will provide us with a more holistic assessment of the community structure and ecological composition of species at a more detailed level among the different sites.

Differences in organism assemblages between the west and east side of the Khumbu Glacier
Based on WGS data analysis, we found the sites west of the Khumbu Glacier were richer overall in bacteria than sites on the east side. Conversely, the eastern sites were richer in microbial eukaryotic organisms (Eucoccidiorida), cyanobacteria (Nostocales, Chroococcales, and Oscillatoriales), and fungi (Ascomycota) than the lakes on the west side. *Nostoc spp.* and other cyanobacteria are known for colonizing biological

soil crusts during the glacial retreat and surviving freeze-thaw cycles and cold temperatures at high elevations (Schmidt and Vimercati, 2019). Our results are also consistent with earlier studies showing that cyanobacteria including *Chroococcus minutus* (order Chroococcales) and *Anabaena cf. laxa* (order Nostocales) exist in the Kongma La lake complex (Rai et al., 2010). Mining of the WGS data also revealed a higher abundance of eukaryotic reads from the Kongma La complex and the Lake South of Nuptse sites, which also included diatoms (Bacillariophyta), plants (Streptophyta), insects (Arthropoda), and rotifers (Rotifera).

Such differences likely relate to differences in the age of the lakes after they were formed through deglaciation. Satellite images spanning 1962-2016 show that the Lake Above EBC and Kala Pattar lakes on the west side of the glacier emerged from melting ice as early as 1962, and are still fed by glacial melt and sparsely vegetated around their margins. In contrast, the eastern lakes were already well downstream of receding glacial ice by 1962 imagery, and presently have abundant plant growth around their margins. These observations offer strong indications that, among the lakes we sampled, the eastern lakes are older than those west of the Khumbu Glacier, and thus have had more time for more complex organism assemblages to develop. Similar successional patterns have been previously identified with fungi, nematodes, and other microeukaryotes along increasing soil age of a receding glacier in the Peruvian Andes (5,000 m) (Hu et al., 2021). Rates of richness increase for all eukaryotes from that study were higher than bacteria (although in contrast to this study bacteria did increase in richness as well), and the presence of plants, and metazoans other than nematodes, were only found in the oldest soils (Hu et al., 2021).

Combination of DNA sequencing and analysis approaches yields more taxa

DNA metabarcoding has been used to survey organismal biodiversity from a variety of aquatic habitats, including coastal and marine environments, as well as freshwater ponds, and lakes (Garlapati et al., 2019; Lacoursière-Roussel et al., 2018; Majaneva et al., 2018; Stat et al., 2017). Primers are chosen to target the amplification and sequencing of specific taxonomic groups. Although we were targeting vertebrate and invertebrate taxa, these primers were successful in amplifying other types of eukaryotes. Species detection using metabarcoding primers is limited by primer bias and insufficient cross-reactivity across distantly related organisms. Applying DNA metabarcoding across the Tree of Life also requires the use of multiple primer sets that target taxonomic groups across phyla, ensuring the gene targets chosen are appropriate for assigning taxonomic rank. In contrast, the WGS method is capable of capturing a greater breadth of taxonomic (i.e., micro and macro-organisms) and genomic (i.e., genes to genomes) information that can also be used for genomic reconstruction and functional gene analysis. A trade-off is that the WGS dataset's larger, complex sequence composition requires intensive bioinformatic analyses and reference databases for accurate processing and classification. Sequence abundance (as a proxy for taxon abundance) is not directly comparable between datasets owing to methodological differences; however, the combination of these methods provided a more comprehensive snapshot of the detectable biodiversity present at our sites. In our study, DNA metabarcoding yielded 15 orders from nine phyla. By mapping the WGS data to reference genomes, we increased our detection to 115 potential orders from 21 phyla, of which 110 orders were unique to the WGS dataset and not found in the metabarcoding dataset. This finding was not too surprising given our metabarcoding primers were designed to target a subset of taxa (vertebrates and invertebrates), while the non-targeted WGS approach could detect a broader spectrum of bacteria, eukaryotic microorganisms, plants, and fungi.

Our dataset contributes to the available genetic sequences of organisms from the Khumbu region and helps to refine targets for future surveys using eDNA, as well as comparison to genetic material from voucher specimens collected from this region and in the HKH ranges more broadly. For example, *Juniperus sp.* sequences were recovered with 99-100% identity using both metabarcoding and chloroplast reference mapping to WGS data. Although confirmation of the species level could not be determined, *J. recurva* and *J. indica* are both known to be present in Sagarmatha National Park (Byers, 2005). Among the species of *Juniperus* documented in the Himalaya, *J. recurva* is widely distributed between 2,500 and 4,724 m (Cubey, 2018; Li et al., 2013), and *J. indica* has been documented between 3,600 and 5,050 m in elevation (Farjon, 2013, n.d.). *J. indica* (above 4,050 m) and *J. recurva* (3,800-4,050 m) have been found specifically in Sagarmatha National Park (Jha and Khanal, 2010). It is likely that the eDNA being picked up in our water samples originated from pollen deposited by wind or other transport modes (i.e., by animals). We also found matches to the plant orders—Fagales and Sapindales (woody flowering plants and trees), and Polypodiales (ferns) —further indicating our water sampling method was able to detect shrubs and other plant pollen likely deposited upslope in high-alpine lakes by aeolian transport prior to cold season ice cover formation.

We identified sequences 100% identical to *T. tibetanus* (Tibetan snowcock), a species found in high-altitude regions of the Himalayas that we observed visually during fieldwork. We also detected sequences >96% identical to *Baetis sp.* (order Ephemeroptera). Mayflies are an ancient order of insects with evolutionary origins dating more than 300 million years; they are found worldwide with about 3,700 currently recognized species (Jacobus et al., 2019). Previous studies have also found mayflies from the family Baetidae in the Gokyo wetlands (4,700-5,000 m) in Sagarmatha National Park (Sharma et al., 2010).

Biogeographic considerations of taxonomic richness

The alpine zone above the tree line occupies <3% of Earth's surface, and species richness is generally thought to decrease with rising elevation from sea level to high-alpine zones, with the exception of taxonomic groups that have evolved to live in insular environments and have higher diversity before decreasing near permanent snow cover or the nival zone (Körner and Spehn, 2002; Pauli and Halloy, 2019). Our results reveal a wide breadth of biodiversity among taxonomic groups present on this one mountain within the HKH range. This raises the question: Are the number of organisms detected at the highest water bodies near Mt. Everest more diverse than would be expected? For example, would a similar sampling regime and methodology detect more or fewer taxa in other high-alpine environments or protected areas at different latitudes and elevations? To our knowledge, there are currently no comparable eDNA studies that summarize diversity across the Tree of Life in high-alpine environments; other studies have applied eDNA analysis to assess the breadth of biodiversity in ocean environments and in river systems (Deiner et al., 2016; Stat et al., 2017; Zhang et al., 2020). As more data become available from future eDNA studies that assess biodiversity across the Tree of Life, we will be able to assemble more complete datasets for meta-analysis and cross-comparison. Taxonomy-free approaches such as OTU and ASV analysis, when standardized, could also be a useful strategy when applied for routine biomonitoring (Apothéloz-Perret-Gentil et al., 2017).

The promise of the eDNA sequencing approach is its potential to be used across diverse, under-studied, and difficult-to-reach regions of the world such as the high-alpine aquatic environments we sampled. eDNA sequencing will likely become a cost-effective approach for comparing biodiversity from similar ecosystems across the world by applying standardized collection and bioinformatic protocols. The absence of standardized methodologies for establishing the effective domain size sampled, such that species-area relationships can be meaningfully established, currently complicates ecosystem biodiversity comparisons. In our high-alpine case, each water body sampled has both an upstream watershed catchment and a much broader "windshed," from which aeolian transport may introduce additional organic material such as leaves, pollen, and other windborne debris. The Global Observation Initiative in Alpine Environments (GLORIA) network—including the GLORIA sites set up during the 2019 Everest Expedition (Mayewski et al., 2020)—is one such effort to standardize long-term vegetation survey protocols in alpine environments (GLORIA-Coordination, 2015). Eventually, it will be possible to compare vegetation survey results from the Mt. Everest region to those at the other GLORIA sites, which are now set up on six continents. Given the challenges of performing fieldwork in these remote locations, we hope that continued efforts toward eDNA methods standardization, such as the GLORIA network, will provide field-friendly options for cross-comparisons of biodiversity around the globe.

There are many future applications for the sequence data generated in our study. The data can be further mined to reconstruct microbial genomes and to extrapolate information about species composition or ecosystem properties through functional gene analysis. Our catalog of present taxa can also be compared to archived reference samples including historical voucher plants (including fossil pollen) and animal specimens, plant and microbial remains from lake sediment, and ice core samples collected from the region. Finally, our dataset can be used in combination with future surveys to track changes in organism diversity and distribution with ongoing climate warming. For example, the environmental sensitivity of some mayfly genera within the order Ephemeroptera make them good indicator species for biomonitoring water quality and assessing ecosystem health (Brittain, 2008; Jacobus et al., 2019). Although species-level monitoring is utilized for the purpose of understanding climate and environmental change, monitoring patterns of change at higher classification levels (order, family, and genus) over time and across different ecosystems may provide additional avenues for assessing ecosystem health and environmental change.

The taxon counts registered by our study, based on the eDNA analysis of just 20 L of water obtained from a handful of water bodies in the alpine to aeolian transition on the southern slope of Mt. Everest, encompass

187 unique orders (16.3% of global orders) from 36 phyla across the six kingdoms, and included at least one bacteriophage virus order. These findings register Mt. Everest as iconic for reasons beyond its superlative height and enduring fascination to humanity: It should be recognized as a target for sustained long-term biodiversity monitoring of high-alpine taxa to complement bioclimatic monitoring and climate change impact assessments. Combining sequencing approaches helped to expand our inventory of biodiversity in the Mt. Everest region, but it remains incomplete. By providing this open-source dataset to the research community, we hope to contribute toward the continued efforts to build up molecular resources to study and track the shifts in biodiversity of Earth's highest mountain.

Limitations of the study

An important consideration for future studies is the timing of sample collections for peak biological activity. Our water sample collections were performed while lakes were covered in seasonal ice. Field studies repeating an eDNA approach would, therefore, benefit from repeated sampling conducted after the spring melt, during or shortly after the summer monsoon, and also during mid-winter to catalog seasonal changes in biodiversity and eDNA detection over the annual cycle.

The taxonomic resolution from DNA metabarcoding is limited by the choice and specificity of primers to distinguish among taxa of interest. The use of additional metabarcoding primers that better capture specific taxonomic groups may have revealed additional taxa in our dataset (Stat et al., 2017). The primers we used were designed to be broadly reactive toward vertebrates and invertebrates, but picked up some taxa including fungi and plants. Certain eukaryotic taxa—including Tardigrada, Rotifera, Amebozoa, Oomycota, and several phyla from the Chromista and Plantae kingdoms—were missed in the metabarcoding, and were instead picked up by mining the WGS database.

For all sequencing approaches, it was challenging to confidently determine taxonomic identity below order level across kingdoms, due in part to the relatively low sequence read abundance of non-bacterial taxa in these samples and the aforementioned lack of available reference DNA sequence data from this region. Additionally, high-throughput amplicon sequencing yields DNA sequences that may be too short to resolve taxonomic differences at the genus and species level. Furthermore, compared to microbial reference sequence databases (i.e., Greengenes and RefSeq), there is not yet a publicly available or "complete" genomic database for eukaryotic taxa. Taxonomic classification with such a database would also require much more computationally and time-intensive analyses, as eukaryotes have larger genomes than microbial species. For these reasons, we experimented with mapping to mitochondrial and chloroplast reference genomes, both to test its utility and to mine the WGS dataset for eukaryote organisms. Although we chose reference genomes from a broad range of taxonomic groups that included vertebrates (birds and reptiles), invertebrates (arthropods, rotifers, and tardigrades), and plants (Juniper) (Table S3 and Figure S2A), the WGS eukaryote reference mapping analysis workflow is limited to detecting taxa based on the reference mitochondrial and chloroplast genomes chosen and have room for improvement. When modeling the order richness by the number of duplicate samples collected or by the number of reference genomes used, it is evident that the expected number of taxa is underestimated and has not reached saturation. This is an important consideration for how future researchers might refine this approach to search for specific target eukaryotic taxa. Also, given that using nine reference genomes took up to 48 hours of effort (or six work days) per reference to analyze (not including the time required to back-check and perform quality control analysis of the results), the cost-benefit of applying such resources, if available, is a significant consideration when using this approach. Thus, it remains an open challenge to identify eukaryotic taxa more broadly from WGS eDNA samples, and this is why some studies have focused on using DNA metabarcoding (Stat et al., 2017).

The sequencing approaches we used have distinct utility. DNA metabarcoding offers a focused survey of specific taxa, which may be more practical for repeated monitoring over time or future studies of metaphylogeography using sequence variants (Antich et al., 2021). In contrast, WGS data can be used for functional gene analysis, genome reconstruction, as well as for assessing diversity across all kingdoms (Stat et al., 2017). Although estimating taxon abundance from aquatic eDNA sequence data is another challenge for use in applied conservation research, (Beng and Corlett, 2020) the ability to detect the presence of taxa in eDNA samples is sufficient for laying the foundations for long-term biodiversity monitoring studies. Hopefully, this study will provide the baseline data with which future sequences can be compared and built upon, along with information for refining sequencing, bioinformatics methods, and marker specificity in a more targeted approach.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHODS DETAILS
  - Historical imagery
  - Sample and field data collection
  - DNA extraction and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatics
  - Species richness curves
  - Defining OTU's for heatmap analysis

AUTHOR CONTRIBUTIONS

Conceived and designed the study: TAS, AS. Performed the fieldwork or experiments: TAS, AS. Analyzed the data: MCWL, TAS, SH, CX, AS, AJS, NBD, SS, and BN. Provided data: AT. Wrote or helped edit the article: TAS, MCWL, AS, NBD, SH, SE, SS, AJS, ACE, BN, AT.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Andrews, S. (2019). FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Antich, A., Palacin, C., Wangensteen, O.S., and Turon, X. (2021). To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. BMC Bioinf. *22*, 177.

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., and Pawlowski, J. (2017). Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Mol. Ecol. Resour. *17*, 1231–1242.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.

Beng, K.C., and Corlett, R.T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. Biodivers. Conserv. *29*, 2089–2121.

Brittain, J.E. (2008). In Mayflies, Biodiversity and Climate Change, F.R. Hauer, J.A. Stanford, and

R.L. Newell, eds. (University of California Press), pp. 1–14.

Bushnell, B. (2015). BBmerge: BBMap Short-read Aligner, and Other Bioinformatics Tools. https://sourceforge.net/projects/bbmap/.

Byers, A. (2005). Contemporary human impacts on alpine ecosystems in the Sagarmatha (Mt. Everest) national Park, Khumbu, Nepal. Ann. Assoc. Am. Geogr. 95, 112–140.

Chamberlain, S., and Szocs, E. (2013). Taxize - taxonomic search and retrieval. F1000Res. 2, 191.

Crampton-Platt, A., Timmermans, M.J.T.N., Gimmel, M.L., Kutty, S.N., Cockerill, T.D., Vun Khen, C., and Vogler, A.P. (2015). Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. Mol. Biol. Evol. 32, 2302–2316.

Cubey, R. (2018). Occurrence dataset. R. Bot. Gard. Edinburgh Herb.

Deiner, K., Fronhofer, E.A., Mächler, E., Walser, J.-C., and Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nat. Commun. 7, 12544.

Dell'Ampio, E., Szucsich, N.U., and Pass, G. (2011). Protura and molecular phylogenetics: status quo of a young love. Soil Org 83, 347–358.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072.

Dobremez, J. (1976). Le Népal - Ecologie et Biogéographie. Paris (Éditions du Centre national de la recherche scientifique (CNRE)).

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Farjon, A. (2013). Juniperus indica. The IUCN Red List of Threatened Species 2013: e.T42238A2965473. 2013-1.RLTS.T42238A2965473.en. https://www.gbif.org/dataset/82a53efc-f762-11e1-a439-00145eb45e9a.

Farjon, A., n.d. Global mountain biodiversity assessment - GMBA. The himalayan uplands plant database (HUP version 1). Occurrence dataset [WWW Document]. https://www.gbif.org/dataset/82a53efc-f762-11e1-a439-00145eb45e9a.

Federhen, S. (2012). The NCBI taxonomy database. Nucleic Acids Res. 40, 136–143.

Garlapati, D., Charankumar, B., Ramu, K., Madeswaran, P., and Ramana Murthy, M.V. (2019). A Review on the applications and recent advances in environmental DNA (eDNA) metagenomics. Rev. Environ. Sci. Biotechnol. 18, 389–411.

Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., Hallwachs, W., and Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through

DNA metasystematics. Proc. Natl. Acad. Sci. USA 111, 8007–8012.

GLORIA-Coordination (2015). The GLORIA Field Manual – Standard Multi-Summit Approach, Supplementary Methods and Extra Approaches, Fifth Edition (Austrian Academy of Sciences & University of Natural Resources and Life Sciences).

Grabherr, G., Gottfried, M., and Pauli, H. (2001). High mountain environment as indicator of global change. In Advances in Global Change Research, G. Visconti, M. Beniston, E.D. Iannorelli, and D. Barba, eds. (Springer), pp. 331–345.

Guillaume, M., and Kingsford, C. (2011). Jellyfish: a fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770.

Gurung, T.D., Sherpa, C., Agrawal, V.P., and Lekhak, B. (2009). Isolation and characterization of antibacterial actinomycetes from soil samples of Kalapatthar, Mount Everest Region. Nepal J. Sci. Technol. 10, 173–182.

Hajibabaei, M., Porter, T.M., Wright, M., and Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS One 14, e0220953.

Hajibabaei, M., Spall, J.L., Shokralla, S., and van Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. BMC Ecol. 12, 28.

Halloy, S.R.P. (1989). Altitudinal limits of life in subtropical mountains: what do we know? Pac. Sci. 43, 170–184.

Hingston, R.W.G. (1925). Animal life at high altitudes. Geogr. J. 65, 185–195.

Hu, W., Schmidt, S.K., Sommers, P., Darcy, J.L., and Porazinska, D.L. (2021). Multiple-trophic patterns of primary succession following retreat of a high-elevation glacier. Ecosphere 12, e03400.

Jacobus, L.M., Macadam, C.R., and Sartori, M. (2019). Mayflies (Ephemeroptera) and their contributions to ecosystem services. Insects 10, 170.

P.K. Jha, and I.P. Khanal, eds. (2010). Contemporary research in Sagarmatha (Mt. Everest) Region, Nepal. Nepal Academy of Science and Technology, Khumaltar, Lalitpur, pp. 173–188.

King, A.J., Karki, D., Nagy, L., Racoviteanu, A., and Schmidt, S.K. (2010). Microbial biomass and activity in high elevation (>5100 meters) soils from the annapurna and Sagarmatha regions of the Nepalese Himalayas. Himal. J. Sci. 6, 11–18.

King, O., Bhattacharya, A., Ghuffar, S., Tait, A., Guilford, S., Elmore, A.C., and Bolch, T. (2020). Six decades of glacier mass changes around Mt. Everest are revealed by historical and contemporary images. One Earth 3, 608–620.

Körner, C., and Spehn, E.M. (2002). Mountain Biodiversity: A Global Assessment (Boca Raton: Parthenon Pub. Group).

Kraken taxonomic sequence classification system. Version 1.0 operating manual. https://ccb.jhu.edu/software/kraken/MANUAL.html#confidence-scoring.

Kumar, S., Suyal, D.C., Yadav, A., Shouche, Y., and Goel, R. (2019). Microbial diversity and soil physiochemical characteristic of higher altitude. PLoS One 14, e0213844.

Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E.K., Archambault, P., Deiner, K., Lodge, D.M., Hernandez, C., Leduc, N., and Bernatchez, L. (2018). eDNA metabarcoding as a new surveillance approach for coastal arctic biodiversity. Ecol. Evol. 8, 7763–7777.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv:1303.3997v2 [q-bio.GN].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome project data processing subgroup (2009). The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 26, 2078–2079.

Li, N., Farjon, A., Christian, T., and Rushforth, K. (2013). Juniperus recurva. The IUCN red list of threatened species.

Liu, Y., Yao, T., Jiao, N., Kang, S., Xu, B., Zeng, Y., Huang, S., and Liu, X. (2009). Bacterial diversity in the snow over Tibetan Plateau glaciers. Extremophiles 13, 411–423.

Liu, Y., Yao, T., Jiao, N., Kang, S., Zeng, Y., and Huang, S. (2006a). Microbial community structure in moraine lakes and glacial meltwaters, Mount Everest. FEMS Microbiol. Lett. 265, 98–105.

Liu, Y., Yao, T., Kang, S., Jiao, N., Zeng, Y., Huang, S., and Luo, T. (2007). Microbial community structure in major habitats above 6000 m on Mount Everest. Chin. Sci. Bull. 52, 2350–2357.

Liu, Y., Yao, T., Kang, S., Jiao, N., Zeng, Y., Shi, Y., Luo, T., Jing, Z., and Huang, S. (2006b). Seasonal variation of snow microbial community structure in the east Rongbuk Glacier, Mt. Everest. Sci. Bull. 51, 1476–1486.

Lorenz, J.G., Jackson, W.E., Beck, J.C., and Hanner, R. (2005). The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 1869–1877.

Majaneva, M., Diserud, O.H., Eagle, S.H.C., Boström, E., Hajibabaei, M., and Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. Sci. Rep. 8, 4682.

Mani, M.S. (1978). Ecology and Phytogeography of High Altitude Plants of the Northwest Himalaya (Oxford & IBH Publishing Co.).

Mayewski, P.A., Gajurel, A.P., Elvin, S., Athans, P., Dinley, T., Sherpa, P., Elmore, A.C., Ghimire, J., Perry, L.B., Matthews, T., et al. (2020). Pushing climate change science to the roof of the world. One Earth 3, 556–560.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P.

(2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. *6*, 610–618.

Miehe, G. (1991). Der Himalaya, eine multizonale Gebirgsregion. Fischer, Stuttgart., Ökologie d (Stuttgart: Fischer).

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B. (2011). How many species are there on earth and in the ocean? PLoS Biol. *9*, e1001127.

Moritz, C., Patton, J.L., Conroy, C.J., Parra, J.L., White, G.C., and Beissinger, S.R. (2008). Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. Science *322*, 261–264.

Motulsky, H.J., and Christopoulos, A. (2004). Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting (Oxford, U.K: Oxford University Press).

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) data-base at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Pauli, H., and Halloy, S.R.P. (2019). High Mountain Ecosystems under Climate Change (OXFORD Res. Encycl. Clim. Sci. Oxford University Press USA).

Potocki, M., Mayewski, P.A., Matthews, T., Perry, L.B., Schwikowski, M., Tait, A.M., Korotkikh, E., Clifford, H., Kang, S., Sherpa, T.C., et al. (2022). Mt. Everest's highest glacier is a sentinel for accelerating ice loss. npj Clim. Atmos. Sci. *5*, 7.

Rai, S., Rai, R., and Jha, S. (2010). Cyanobacteria of Nepal: a checklist with distribution. Our Nat. *8*, 336–354.

Schmidt, S.K., Gendron, E.M.S., Vincent, K., Solon, A.J., Sommers, P., Schubert, Z.R., Vimercati, L., Porazinska, D.L., Darcy, J.L., and Sowell, P. (2018). Life at extreme elevations on atacama volcanoes: the closest thing to mars on earth? Antonie Leeuwenhoek *111*, 1389–1401.

Schmidt, S.K., Lynch, R.C., King, A.J., Karki, D., Robeson, M.S., Nagy, L., Williams, M.W., Mitter, M.S., and Freeman, K.R. (2011). Phylogeography of microbial phototrophs in the dry valleys of the high Himalayas and Antarctica. Proc. Biol. Sci. *278*, 702–708.

Schmidt, S.K., and Vimercati, L. (2019). Growth of cyanobacterial soil crusts during diurnal freeze-thaw cycles. J. Microbiol. *57*, 243–251.

Seimon, T.A., Seimon, A., Daszak, P., Halloy, S.R., Schloegel, L.M., Aguilar, C.A., Sowell, P., Hyatt, A.D., Konecky, B., and E Simmons, J. (2006). Upward range extension of Andean anurans and chytridiomycosis to extreme elevations in response to tropical deglaciation. Global Change Biol. *13*, 288–299.

Seimon, T.A., Seimon, A., Yager, K., Reider, K., Delgado, A., Sowell, P., Tupayachi, A., Konecky, B., McAloose, D., and Halloy, S. (2017). Long-term monitoring of tropical alpine habitat change, Andean anurans, and chytrid fungus in the Cordillera Vilcanota, Peru: results from a decade of study. Ecol. Evol. *7*, 1527–1540.

Sharma, C.M., Sharma, S., Gurung, S., Jüttner, I., Bajracharya, R.M., and Shrestha, N. (2010). Ecological studies within the Gokyo wetlands, Sagarmatha national Park, Nepal. In Contemporary Research in Sagarmatha (Mt. Everest) Region, Nepal. Nepal Academy of Science and Technology, Khumaltar, Lalitpur, P. Jha and I. Khanal, eds..

Spehn, E., Rudmann-Maurer, K., and Körner, C. (2010). Mountain biodiversity and global change. In Global Mountain Biodiversity Assessment (GMBA) of DIVERSITAS, Basel, Switzerland, E.M. Spehn, K. Rudmann-Maurer, C. Körner, and D. Maselli, eds..

Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey, E.S., and Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Sci. Rep. *7*, 12240.

Swan, L.W. (1992). The aeolian biome. Bioscience *42*, 262–270.

Swan, L.W. (1990). The highest life. Himal. J. *46*, 125–133.

Swan, L.W. (1963). Aeolian zone. Science *140*, 77–78.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. *21*, 2045–2050.

Thomsen, P.F., and Willerslev, E. (2015). Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. Biol. Conserv. *183*, 4–18.

Townzen, J.S., Brower, A.V.Z., and Judd, D.D. (2008). Identification of mosquito bloodmeals using mitochondrial cytochrome oxidase subunit I and cytochrome b gene sequences. Med. Vet. Entomol. *22*, 386–393.

Troll, C. (1972). Geoecology of the high-mountain regions of eurasia: proceedings of the international geographical union commission on high-altitude geoecology. In Symposium of the International Geographical Union Commission on High-Altitude Geoecology, Mainz, DE, 20-22 November 1969.

Uchida, N., Kubota, K., Aita, S., and Kazama, S. (2020). Aquatic insect community structure revealed by eDNA metabarcoding derives indices for environmental assessment. PeerJ *8*, e9176.

Valcárcel, V., Guzmán, B., Medina, N.G., Vargas, P., and Wen, J. (2017). Phylogenetic and paleobotanical evidence for late Miocene diversification of the Tertiary subtropical lineage of ivies (Hedera L., Araliaceae). BMC Evol. Biol. *17*, 146.

Villaverde, T., Escudero, M., Luceño, M., and Martín-Bravo, S. (2015). Long-distance dispersal during the middle–late Pleistocene explains the bipolar disjunction of Carex maritima (Cyperaceae). J. Biogeogr. *42*, 1820–1831.

Vimercati, L., Darcy, J.L., and Schmidt, S.K. (2019). The Disappearing periglacial ecosystem atop Mt. Kilimanjaro supports both cosmopolitan and endemic microbial communities. Sci. Rep. *9*, 10676.

Washburn, B. (1989). Mapping Mount Everest. Am. Acad. Arts Sci. Bull. *42*, 29–44.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. *15*, R46.

Woodcroft, B.. SingleM is written by Ben Woodcroft (@wwood) at the Australian Centre for Ecogenomics (UQ) and is licensed under GPL3 or later. v0.13.2. https://github.com/wwood/singlem.

World Wildlife Fund (2012). The Eastern Himalayas: Where Worlds Collide.

Xu, J., Badola, R., Chettri, N., Chaudhary, R.P., Zomer, R., Pokhrel, B., Hussain, S.A., Pradhan, S., and Pradhan, R. (2019). Sustaining biodiversity and ecosystem services in the Hindu Kush Himalaya. In The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People, P. Wester, A. Mishra, A. Mukherji, and A.B. Shrestha, eds. (Springer International Publishing, Cham).

Zhang, Y., Pavlovska, M., Stoica, E., Prekrasna, I., Yang, J., Slobodnik, J., Zhang, X., and Dykyi, E. (2020). Holistic pelagic biodiversity monitoring of the Black Sea via eDNA metabarcoding approach: from bacteria to marine mammals. Environ. Int. *135*, 105307.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
eDNA samples TS13-33 (see Table S1, Document S1)	This paper	N/A
Critical commercial assays		
QIAamp DNA Mini Kit	Qiagen Inc.	Cat# 51304
Pre-assembled filter-pak cartridges	Smith-Root Inc.	https://www.smith-root.com/edna/edna-filters
47mm 1.0 um pore filters	GE Healthcare Whatman	Cat# 7190-004
AmpliTaq Gold™ 360 Master Mix	Applied Biosystems™	Cat# 4398790
ExoSAP-IT Express	Applied Biosystems™	Cat# 75001.200.UL
Qubit dsDNA BR assay kit	Thermo Fisher Scientific	Cat# Q32850
EZ-Amplicon NGS deep sequencing	Genewiz Inc.	https://www.genewiz.com/Public/Services/Next-Generation-Sequencing/Amplicon-Sequencing-Services/Amplicon-EZ
Whole Genome Sequencing	Genewiz Inc.	https://www.genewiz.com/en/Public/Services/Next-Generation-Sequencing/Whole-Genome-Sequencing
Whole Genome Metagenomics	Genewiz Inc.	https://www.genewiz.com/en/Public/Services/Next-Generation-Sequencing/Metagenomics-Solutions
Deposited data		
RefSeq v76	O'Leary et al. 2016	https://www.ncbi.nlm.nih.gov/refseq/about/
Greengenes 16S	DeSantis et al. 2006	https://greengenes.secondgenome.com/
NCBI database	Federhen 2012	https://www.ncbi.nlm.nih.gov/nucleotide/
Raw data (https://www.ncbi.nlm.nih.gov/sra)	This paper	[BioProject PRJNA629845]: [Accessions: SRR11700438- SRR11700405]
Oligonucleotides		
Primer B-Forward primer (5′-CCI GAYATRGCITTYCCICG-3′)	IDT Technologies	N/A
Primer B-ForwardMOD primer (5′-C CIGAYATRGCITTYCCNCG-3′)	IDT Technologies	N/A
Primer R5-Reverse primer (5′-GTRA TIGCICCIGCIARIACIGG-3′)	IDT Technologies	N/A
Primer DegHCOI2198Mod (5′-TA AACTTCAGGGTGACCAAARAAYC AGAA-3′)	IDT Technologies	N/A
Primer CO1_long - 6597 (5′-AAGA ATCAGAATARGTGTTG-3′)	IDT Technologies	N/A
Adaptor ligation forward 5′-ACACTC TTTCCCTACACGACGCTCTTCCGATCT-3′	Genewiz Inc.	https://web.genewiz.com/amplicon-ez-faq?_ga=2.102081345.1700467531.1653082033-1572836208.1653082033#Submission
Adaptor ligation reverse 5′-GACTG GAGTTCAGACGTGTGCTCTTCCGATCT-3′	Genewiz Inc.	https://web.genewiz.com/amplicon-ez-faq?_ga=2.102081345.1700467531.1653082033-1572836208.1653082033#Submission
Software and algorithms		
FastQC v0.11.8	Andrews, 2019	https://www.bioinformatics.babraham.ac.uk/projects
MultiQC v1.8	Ewels et al. 2016	https://multiqc.info/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Kraken v0.10.5-beta	Wood and Salzberg, 2014	http://ccb.jhu.edu/software/kraken/
Jellyfish v1.1.1	Guillaume and Kingsford, 2011	http://www.cbcb.umd.edu/software/jellyfish/jellyfish-manual.html
SingleM v0.13.2	Woodcroft	https://github.com/wwood/singlem
Geneious Prime v2019.2.3	Geneious	https://www.geneious.com/
BBDuk	Bushnell 2015	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbduk-guide/
BLASTn 2.6.0+	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Taxize v0.9.92	Chamberlain and Szocs, 2013	https://docs.ropensci.org/taxize/
BWA-MEM v0.7.15-r1140	Li, 2013	https://github.com/lh3/bwa
SAMtools v1.9	Li et al., 2009	http://samtools.sourceforge.net/
SPAdes v3.10.1	Bankevich et al., 2012	http://cab.spbu.ru/software/spades/
BBmerge v38.22	Bushnell 2015	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmerge-guide/
Other		
Code for data analysis	This paper	https://github.com/marisalim/Everest-eDNA-biodiversity-survey (https://doi.org/10.5281/zenodo.5149298)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to Dr. Tracie Seimon, Wildlife Conservation Society, Zoological Health Program, 2300 Southern Boulevard, Bronx, NY, USA; tseimon@wcs.org.

Materials availability

This study did not generate new materials or reagents.

Data and code availability

- Raw sequence read data have been deposited at the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject: PRJNA629845 (accessions: SRR11700438- SRR11700405) and are publicly available.

- All original code has been deposited at https://github.com/marisalim/Everest-eDNA-biodiversity-survey under https://doi.org/10.5281/zenodo.5149298 and is publicly available.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact (tseimon@wcs.org) upon request.

METHODS DETAILS

Historical imagery

The images were sourced as follows: Corona KH-4 taken the 15 Dec, 1962 at a resolution of 7.6 m; Corona KH-4A taken the 18 Dec, 1969 at 5.2 m; Aerial Photography (Washburn), Wild RC-10 camera taken 20 Dec 1984 at 0.5m; IKONOS-2 taken 29 Nov, 2001 at 0.8 m, and WorldView-2 taken 29 Oct, 2016 at 0.5 m.

The Corona KH-4 and KH-4a images come from aerial photographs collected by the U.S. government from 1962-1969 using photographic cameras on satellites. They were used for reconnaissance and to produce maps for the U.S intelligence agencies and were declassified in 1995. The 1984 images come from high resolution (0.5 m ground resolution) aerial photographs acquired by Bradford Washburn (Washburn, 1989), using a Wild RC-10 camera, on behalf of the National Geographic Society. The Ikonos-2 and WorldView-2 satellite images come from the digital archive imagery collections of DigitalGlobe, now part of Maxar. All images were processed using Esri ArcGIS Pro software image analysis tools for monochrome or natural-color image presentation.

Sample and field data collection

Field data collections from some of the highest elevation waterbodies on Mt. Everest were performed in April through May 2019 as part of the National Geographic and Rolex Perpetual Planet Everest Expedition (hereafter 2019 Everest Expedition) (Mayewski et al., 2020). Ten sampling sites for capturing eDNA were selected from alpine lakes and streams between 4,500 m and 5,500 m within four sub-watersheds feeding into the Dudh Kosi River on the southern flank of the Everest massif (Figure 1). Sites were chosen based on accessibility in catchments where anthropogenic disturbance should be relatively minor both east and west of the heavily disturbed Khumbu Valley trekking route to Everest Base Camp (EBC). We collected duplicate 1 liter samples for filtration from: three lakes on the west side of the Khumbu Glacier (Lake Above EBC, Kala Pattar Lakes 1-2); four lakes on its east side (Kongma La Lakes 1-3, Lake South of Nuptse); and two streams (Nuptse Glacier Mountain Stream, Kongma La Lake Mountain Stream). The Kala Pattar Lakes and the Lake Above EBC sites occupy small cirques beneath hanging glaciers on the south and southeast faces of Pumori (7,161 m), and are situated approximately 10 km west and 8.4 km west-northwest, respectively, of the summit of Mt. Everest. About 10 km southwest of the Mt. Everest summit, the Kongma La Lake complex lies downstream of the Kongma Glacier, which at present consists of remnant ice blocks detached and ablating from the cirque headwall. The Lake South of Nuptse site is located about 9.25 km southwest of the Mt. Everest summit on the southern slope of Nuptse (7,861 m), between a large lateral moraine of the Nuptse Glacier and an unglacierized rock ridge. Cold conditions during fieldwork delayed the seasonal melt-out of these high-alpine lakes, so samples were taken from holes chipped through ice or from partially melted lake margins. During the sampling season, nocturnal temperatures frequently fell below -10°C; along with recurrent snowfalls, this contributed to slow springtime green-up of alpine vegetation.

Duplicate 1 L samples (for a total of 20) were collected in sterile Nasco Whirl-Pak Easy-To-Close polyethylene sample bags (1064.7 mL) and filtered within 2 hours back at camp or at the water source (weather dependent). All samples were filtered using single-use ANDe pre-assembled filter-pak cartridges containing a 47 mm cellulose nitrite 1.0 µm pore filter (Smith-Root Inc, Vancouver WA, USA), attached to Thermo Scientific™ Nalgene™ Reusable Filter Holders with a 1 L receiver and a Nalgene® hand vacuum pump with gauge (30 cc, with tubing) (Thermo Fisher Scientific, MA, USA). We used the 1.0 µm filters, as we knew the ponds around Mt. Everest have a high degree of glacial silt which would clog smaller filters. For ponds that were frozen with no liquid water access, we used an ice tool to break a small hole in the top layer of ice and inserted the filter cartridge several centimeters into the water to filter one liter of water per cartridge. The filter was then quickly removed from the cartridge with a sterile pair of single-use forceps and transferred to a sterile 50 mL conical tube with 0.5 mL of 95% molecular grade ethanol to preserve DNA on the filters. The filters were subsequently air-dried and stored at ambient temperature until they could be processed.

DNA extraction and sequencing

DNA was extracted from the filters using the QIAamp DNA Mini Kit (QIAGEN Inc., CA, USA). QIAGEN extraction procedures followed the manufacturer's protocol and the final extract was eluted twice with 50 µL DNase/RNase-free water into group A and B aliquots. DNA concentration for all aliquots was measured by Nanodrop (Thermo Fisher Scientific, MA, USA). Samples were prepared for two sequencing approaches: (1) DNA metabarcoding with cytochrome oxidase 1 (CO1) primers; and (2) whole genome shotgun (WGS) metagenomic sequencing. The higher DNA concentration Group A aliquot was sent to Genewiz for WGS sequencing (Genewiz, Inc. South Plainfield, New Jersey 07080, USA). There, the DNA samples were re-quantified by Qubit and treated with RNase A to eliminate RNase contamination. DNA concentration ranged between 10 and 850 ng/sample. We also included two control samples, TS34 (a blank filter) and TS36 (a blank DNA extract without a filter), from which, compared to the water samples, DNA was not detectable by Qubit and there was insufficient DNA for analysis. The sequence libraries for WGS were sequenced on an Illumina HiSeq 4000 with 150bp paired-end reads.

Conventional PCR targeting the mitochondrial CO1 region was performed with 3 µL of DNA extracts from the group B aliquots. Degenerate primers were purchased from IDT Technologies, Coralville, IA, USA. We tested several published primers on a sequence alignment of 131 vertebrates and invertebrates that are known to occur in the Everest region. Primers were chosen based on the maximum product size feasible (between 150 - 500bp) for the Amplicon EZ NGS platform (Genewiz, Inc. South Plainfield, New Jersey 07,080, USA). For amplification of 358bp of the CO1 gene from invertebrates (macroinvertebrates and arthropods), 0.5 µL of 100 µM B-Forward primer (5′-CCIGAYATRGCITTYCCICG-3′) (Hajibabaei et al., 2012, 2019; Majaneva et al., 2018), 0.5 µL of 100 µM B-ForwardMOD primer (5′-CCIGAYATRGCITTYCCNCG-3′), and 1.0 µL

of 100 μM of R5-Reverse primer (5′-GTRATIGCICCIGCIARIACIGG-3′) (Gibson et al., 2014) were added to a 50 μL reaction containing 3 μL of DNA template, 20 μL of DNase/RNase-free water, and 25 μL of AmpliTaq Gold 360 Master Mix (2X concentration) (Thermo Fisher Scientific, MA, USA). For amplification of 353 to 469bp of the CO1 gene from vertebrate species, 0.5 μL of 100 μM B-Forward primer (5′-CCIGAYATRGCIT TYCCICG-3′), 0.5 μL of 100 μM B-ForwardMOD primer (5′-CCIGAYATRGCITTYCCNCG-3′), 0.5 μL of 100 μM of a modified version of reverse primers DegHCOI2198Mod (5′-TAAACTTCAGGGTGACCAAARAAYCA GAA-3′) (Lorenz et al., 2005), and 0.5 μL of 100 μM of reverse primer CO1_long - 6597 (5′-AAGAATCAGAA TARGTGTTG-3′) (Townzen et al., 2008) were added to a 50 μL reaction containing 3 μL of DNA template, 20 μL of DNase/RNase-free water, and 25 μL of AmpliTaq Gold 360 Master Mix (2X concentration). Samples were amplified by conventional PCR using the following cycling conditions: 95°C for 5 min, followed by 50 cycles of 94°C for 30 seconds, 45 °C for 45 seconds, and 72°C for 45 seconds, followed by a final extension step of 72°C for 5 minutes. Positive PCR products from vertebrate and invertebrate PCR reactions (amplified separately) were combined for each sample and enzymatically treated with ExoSAP-IT® (Affymetrix; Santa Clara, California 95051, USA), and normalized to 20 ng/μL using Qubit dsDNA BR assay kit (Thermo Fisher Scientific, MA, USA). PCR products were sent to Genewiz for adaptor ligation (forward sequencing read: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′, reverse sequencing read: 5′-GACTGGAGTTCA GACGTGTGCTCTTCCGATCT-3′) and metabarcoding sequencing using their EZ-Amplicon NGS deep sequencing service on an Illumina MiSeq with 250bp paired-end reads (Genewiz, Inc. South Plainfield, New Jersey 07,080, USA). Raw data has been uploaded to the NCBI Short Read Archive under Bioproject ID: PRJNA629845.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatics

For both DNA metabarcoding and WGS datasets, sequence adapters were removed by Genewiz and read quality control was assessed with FastQC v0.11.8 (Andrews, 2019) and summarized with MultiQC v1.8.(Ewels et al., 2016) The respective pipelines used to analyze each dataset are presented in Figure S3.

Two software pipelines were used to classify microbial taxa from the WGS reads. In the Genewiz pipeline, WGS reads were taxonomically classified using Kraken v0.10.5-beta to align shotgun reads to their custom k-mer indexed database, built on ~60,000 genomes (archaea, bacteria, viruses, protozoa, and fungi) compiled from RefSeq v76 using Jellyfish v1.1.1 for k-mer indexing (Figure S3B, Data S2) (Guillaume and Kingsford, 2011; O'Leary et al., 2016; Wood and Salzberg, 2014). Classification results were then normalized according to genome size with Genewiz custom scripts prior to downstream analyses. Kraken classifications were summarized for cyanobacteria and eukaryotic microorganisms. For the Kraken analysis, we used a threshold of 0, which corresponds to 95.43% precision for genus rank identification (Sensitivity 77.32%, F score 85.43) (Kraken taxonomic sequence classification system, n.d.) Because genus-level rank has high precision with Kraken, we felt confident using this threshold for assessing order-level rank.

For bacterial classifications, we used SingleM v0.13.2 (Woodcroft, n.d.) to classify bacteria by order and phylum with both the default set of single-copy ribosomal protein genes and the Greengenes 16S database (McDonald et al., 2012). We generated operational taxonomic unit (OTU) tables for each sample with the "singlem pipe" command, and combined the results from the ribosomal and Greengenes analyses. The OTU tables were filtered (Data S1) to remove repeated taxonomic identifications between markers and hits with fewer than 10 reads. We also removed hits to eukaryotic taxa per author recommendation that SingleM was not designed to assess the eukaryotic community.

There are several available pipelines for taxonomic classification of metabarcoding datasets. However, due to the limited resolution of classification in our dataset and difficulty in validating taxon identifications for our data using other methods, we developed a pipeline that allowed us to directly verify contig sequence quality and taxon identity both visually and statistically (Figure S3C). Using Geneious Prime v2019.2.3, metabarcoding reads were trimmed to remove reads with an average Phred quality score <20 and read length <150bp with BBDuk (Bushnell, 2015). Next, paired reads were merged with the 'Merge Paired Reads' option and primers were removed (allowing up to 5 mismatches). After trimming, reads >200bp were kept for *de novo* assembly using the Geneious assembler. *De novo* assembly parameters were set to 'Custom Sensitivity' with the following settings: a maximum of 1% gaps per read, a minimum overlap of 100bp, ignore words repeated more than 200 times (ensure each contig will comprise only closely related sequences), a maximum of 2% mismatches per read, and a minimum overlap identity of 98%. These

parameters were selected to ensure contigs were comprised of highly similar sequences. Contigs with quality score <90% (%HQ metric) were also removed. To classify taxon identity of the contig sequences, we compared the resulting contigs against the full nucleotide NCBI database with BLASTn 2.6.0+ (Altschul et al., 1990; Federhen, 2012).

The BLAST search was restricted to the top 10 hits with >70% sequence similarity and >80% query coverage with respect to the NCBI reference. Taxonomic information (phylum, class, and order) from NCBI was added with the R package taxize v0.9.92 based on genus and species matches from the BLAST results (Chamberlain and Szocs, 2013). For downstream analysis, we removed short contigs (<200bp) and increased sequence similarity stringency (>80%) to reduce non-specific taxon identifications. We also removed off-target contigs that were not mitochondrial genes, as well as potential contaminant sequences that were identified as human DNA. In the BLAST analysis of metabarcoding data, we obtained classifications for 15,833 contigs, made up of 358,810 reads. Out of these reads, 626 (making up 20 contigs) matched 90-100% to a human or primate reference sequence from GenBank from the following samples: TS13, 153 reads; TS14, 257; TS15, 97; TS16, 5; TS18, 2; TS19, 34; TS20, 2; TS21, 4; TS23, 3; TS24, 55; TS25, 8; TS26, 2; TS27, 4. This makes up a small proportion of the classifiable reads (0.17%). These 20 contigs were removed from downstream analysis, leaving 15,813 contig taxon identifications made up of 358,184 reads. There was subsequent filtering to remove off-target non-mitochondrial sequences and matches of greater than or equal to 80% sequence similarity. We did not determine the source of the human DNA, but it could be from lab contamination or due to sample collection at sites where people were present. Finally, contigs represented by one read were removed (singletons), such that all remaining contigs used for downstream analysis were represented by >2 reads/contig, had >80% sequence similarity to the NCBI reference, had >80% query coverage matching the NCBI reference sequence, and were >200bp long (Data S4). For analyzing the metabarcoding COI sequences, we chose 80% as the threshold based on an internal validation study of COI sequence similarity thresholds at the intraspecific (99%), intragenus (85%), intrafamily (83%), and intra-order (80%) levels for Arthropoda including Ephemeroptera, Plecoptera, and Trichoptera (Uchida et al., 2020).

To extract eukaryotic taxa from the WGS dataset, we mapped WGS reads to reference genomes as a rough method of reducing off-target sequences and especially removing bacterial sequences (Figure S3D). Since eukaryote DNA is already in low abundance in this dataset, we mapped to mitochondrial or chloroplast reference genomes as mitochondrial and chloroplast DNA are more abundant than nuclear DNA. We chose nine references from NCBI for a variety of animal and plant species that span the range of eukaryotic kingdoms to analyze against each sample. Note that a tenth reference (NC_023939.1) was also used as a test to compare how two closely related references (NC_023939.1 and NC_040970.1, *Tetraogallus tibetanus* and *Gallus gallus*) behaved in pulling out targets from a test sample (TS20) (Table S3).

Read mapping to reference genomes was conducted with BWA-MEM v0.7.15-r1140 (Li, n.d.). After mapping, SAM files were converted to FASTQ files with SAMtools v1.9 (Li et al., 2009) and mapped reads were *de novo* assembled with SPAdes v3.10.1(Bankevich et al., 2012) to form longer contigs. We checked taxon identity via BLAST search against the full NCBI nucleotide database. BLAST results were limited to the top 10 hits per contig with >70% sequence similarity and >90% query coverage with respect to the NCBI reference sequence (Data S5). If SPAdes failed to form contigs, we conducted the BLAST search on merged and unmerged reads that were processed with BBMerge v38.22 (Bushnell, 2015). For reads mapped to the chloroplast reference genome (*Juniperus recurva*), we removed off-target non-chloroplast gene hits. We also removed possible contaminants, including any matches to human DNA. For quantifying the phyla and orders detected using this data, we further sorted the BLAST data table for >85% sequence similarity to count unique phyla and orders (Data S6, Total phyla and order detected). Mapping the WGS data to reference mitogenomes is an exploratory approach we applied to "fish out" additional DNA sequences of potential eukaryotic origin. A similar mitogenomics approach was used to identify beetle species by mapping WGS data to assembled mitogenomes associated with COI barcodes (Crampton-Platt et al., 2015). Because this method relies on WGS data mapped to complete mitochondrial references, there are limitations to this approach. Thresholds are likely to be different between taxonomic groups and cannot be assumed to be the same for non-barcode regions. We tested various sequence similarity thresholds using this method and found that using lower thresholds (70–80%) introduced implausible order identities. A threshold of 85% eliminated these outliers, leaving identifications more consistent with metabarcoding results and with species that were observed in our sampling area vicinity. Thus, we found 85%

served as an effective minimum threshold for preventing type I errors, based on manual checking of sequences extracted for this environment and study site.

Species richness curves

To assess how many potential orders could be detected by using eukaryotic reference genomes to initially map WGS reads, we plotted a richness curve by modeling a non-linear asymptotic regression using the mean number of orders detected with all possible combinations of the 1-9 reference genomes we utilized in this study using a custom R script (Motulsky and Christopoulos, 2004). The position of the asymptote is an estimate of how many total potential orders could be present within our WGS data, but this is likely an over-estimate as each additional genome is likely to decrease the average phylogenetic distance between all reference genomes, which could result in fewer additional sequences mapped to each additional reference genome.

To assess how many potential orders could be detected using a certain method across the region, we plotted richness curves by modeling asymptotic non-linear regressions using the mean number of orders detected when using all possible combinations of the duplicates and non-merged duplicates from this study. Results from merged duplicates are shown for all asymptotic regressions plots in each figure.

Defining OTU's for heatmap analysis

We generated operational taxonomic unit (OTU) tables for each dataset (Data S1, Data S2, Data S3, Data S5). The OTU counts for replicate site samples were combined, but raw data are available in Supplementary data files. We defined OTUs by taxonomic order, as this was the lowest taxonomic rank with the best resolution and representation across both datasets.

The WGS dataset taxon identifications for microbial eukaryotes are based on Kraken sequence read classifications from different regions of the genome. The outputs of the Kraken reports were used to derive the number of sequences per taxa in each sample to generate the OTU table.

For the metabarcoding dataset, OTUs were defined from CO1 contigs identified to taxonomic order. We chose taxonomic order to be consistent with the WGS dataset and because contig species identifications are based on having >85% sequence similarity to the reference sequence. This threshold is not sufficient to confidently determine family, genus or species identity. The OTU table was generated by counting the number of contigs identified to a given order per sample.