Microbial diversity in extreme environments

Wen-Sheng Shup^{1,3} and Li-Nan Huang^{2,3}

Abstract | A wide array of microorganisms, including many novel, phylogenetically deeply rooted taxa, survive and thrive in extreme environments. These unique and reduced-complexity ecosystems offer a tremendous opportunity for studying the structure, function and evolution of natural microbial communities. Marker gene surveys have resolved patterns and ecological drivers of these extremophile assemblages, revealing a vast uncultured microbial diversity and the often predominance of archaea in the most extreme conditions. New omics studies have uncovered linkages between community function and environmental variables, and have enabled discovery and genomic characterization of major new lineages that substantially expand microbial diversity and change the structure of the tree of life. These efforts have significantly advanced our understanding of the diversity, ecology and evolution of microorganisms populating Earth's extreme environments, and have facilitated the exploration of microbiota and processes in more complex ecosystems.

16S ribosomal RNA (rRNA) gene clone library analysis The cloning and Sanger sequencing of 16S ribosomal RNA (rRNA) genes PCR-amplified from total DNA extracted from an environmental sample, resolving the composition and diversity of the microbial community.

¹School of Life Sciences, South China Normal University, Guangzhou, People's Republic of China. ²School of Life Sciences, Sun Yat-sen University, Guangzhou, People's Republic of China.

³These authors contributed equally: Wen-Sheng Shu and Li-Nan Huang.

See-mail: shuwensheng@ m.scnu.edu.cn; eseshln@ mail.sysu.edu.cn https://doi.org/10.1038/ s41579-021-00648-y Extreme environments, such as terrestrial hot springs and deep sea hydrothermal vents, glaciers and permafrost, hypersaline habitats, acid mine drainage (AMD) and the subsurface, are widely distributed around the globe^{1,2} (FIG. 1). Although harsh in physiochemical conditions (FIG. 1), these habitats provide diverse ecological niches for a wide range of microorganisms from all three domains of life. These organisms have evolved to have different strategies to cope with the extreme environmental stresses (BOX 1), define the physical and geochemical limits of life and have attracted extensive research interest from either a fundamental or a practical perspective. Moreover, owing to their reduced biological complexity, overall tractability for cultivation-independent molecular analyses and tight coupling between geochemical and biological processes, some extreme environments are ideal targets for the study of microbial ecology, evolution and environmental adaptation³⁻⁵ (BOX 2).

Historically, microbiologists mostly relied on cultivation-based methods. Thus, early ecophysiological studies focused largely on a few 'model' microorganisms that were relatively readily isolated from various extreme habitats. Later cultivationindependent surveys (for example, 16S ribosomal RNA (rRNA) gene clone library analysis) uncovered a vast microbial diversity that was not represented in the set of organisms available in pure culture⁶⁷. Furthermore, high-throughput marker gene sequencing, which allows analyses of microbial diversity more deeply and widely than ever before⁸, was applied to comprehensively examine broader patterns of microbial distribution and

the factors shaping the large-scale ecological ranges of these microorganisms9,10. Overall, these molecular surveys have substantially expanded our knowledge of the microbial diversity in Earth's major extreme environments, typically revealing relatively low to intermediate complexity microbial communities and the high abundance of a few dominant species with widespread geographical distributions. The development and application of new genomic sequencing and computational approaches have provided comprehensive insights into community metabolic capacity and dynamics at different spatio-temporal scales and along specific geochemical gradients^{4,11,12} and, more importantly, have allowed the recovery of genomes belonging to uncultivated phylum-level groups (candidate phyla), which collectively substantially expand our view of the tree of life¹³⁻¹⁵. These genomic data have provided a new perspective on the metabolic diversity and ecological roles of archaea and bacteria, and the evolution of early life and the origin of eukaryotes. In this Review, we summarize our current knowledge of the diversity, ecology and evolution of bacteria and archaea inhabiting Earth's major extreme environments.

Microbial diversity and major new lineages

Relatively low biodiversity extreme environments, especially the terrestrial geothermal springs in Yellowstone National Park (YNP), United States, featured prominently in early 16S rRNA gene clone library analyses of natural ecosystems⁶, which have revolutionized our understanding of microbial diversity on Earth.



Fig. 1 | **Global distribution of representative extreme environments.** Sites of each extreme environment^{2,17,42,75,117,184,185} are marked with a specific colour. Each extreme environment imposes multiple stresses (bottom left) on its microbial inhabitants.

Metagenomic analysis

The sequencing and genomic analysis of total community DNA directly recovered from the environment.

16S rRNA amplicon sequencing

High-throughput (next-generation) sequencing of 1 6S ribosomal RNA (rRNA) gene fragments PCR-amplified from total community genomic DNAs, allowing sampling and analysis of microbial diversity on an unprecedented scale.

Thermophily

The ability of a microorganism (that is, a thermophile) to grow at a high temperature.

Subsequent extensive gene surveys have provided a relatively comprehensive view of the composition and diversity of microbial assemblages populating various extreme habitats. The results are generally supported by more recent metagenomic analysis (FIG. 2). Extreme environments are also particularly suited for the study of microbial biogeography owing to their discrete distribution across multiple geographical scales, broad geochemical and physical gradients and relatively constrained microbial community structure^{9,10,16}. High-throughput 16S rRNA amplicon sequencing has resolved patterns of large-scale ecological ranges of these extremophilic organisms, although potential underlying ecological and evolutionary processes and their relative influence remain poorly understood.

Although diversity surveys have uncovered a vast array of previously unknown microorganisms in extreme environments, our understanding of their physiology and ecology is hampered by the lack of pure cultures^{17,18}. Genome-resolved metagenomics and single-cell genomics have allowed reconstruction of near-complete or even closed microbial genomes directly from the environments^{19,20}, providing initial insights into the metabolic potential and evolutionary history of these enigmatic extremophiles^{17,21}. Genomic analyses of extreme environments have also led to the discovery and genomic characterization of many major, deep-branching lineages that were overlooked by previous marker gene surveys^{22,23}. Collectively, these cultivation-independent genomics studies have substantially expanded the phylogenetic breadth and genome representation of the tree of life, particularly the domain Archaea (FIG. 3).

Terrestrial geothermal springs. A wide variety of continental geothermal springs exist on Earth (for instance, YNP hosts more than 14,000 geothermal features that have a temperature range of 40–92 °C)⁴, and research on their microbial inhabitants has pioneered the study of the microbiology of extreme environments and led to the historical discovery of thermophily²⁴. Cultivation-based studies and molecular surveys have established that Aquificae, Proteobacteria and Crenarchaeota often dominate hot spring waters sampled around the

Whole-genome shotgun sequencing

A process in which the genomic DNA from a pure culture is randomly fragmented and sequenced, and the sequences obtained are then assembled through bioinformatics to a whole microbial genome.

Genome bins

The population genomes (also called 'metagenomeassembled genomes') assembled from metagenomic sequence data based on G+C content, *k*-mer frequency and/or read coverage. globe^{10,25} (FIG. 2). Important taxa include the carbonassimilating *Hydrogenobaculum* spp., sulfur-oxidizing *Thermoproteus* spp. and *Sulfolobus* spp., and nitrogenfixing *Acidithiobacillus* spp. Significant differences have been found between aqueous and sediment microbial communities from the same springs^{25,26}, and this ecological differentiation is likely driven by the differences in resource availability (for example, dissolved or mineral substrates). Specifically, sediment communities exhibit relatively higher species evenness²⁵ and are dominated by taxa capable of mineral-based metabolism such as sulfur oxidation or reduction or iron oxide reduction²⁶. Microbial mats thriving in alkaline siliceous springs from YNP have been subjected to extensive research, demonstrating a very low diversity

Box 1 | Adaptation of microbial life to environmental extremes

Extreme environments feature extremely high or extremely low temperatures or pH, high salinity or pressure, low levels of nutrients and specific combinations of thereof, imposing substantial challenges to life¹. However, some microorganisms (extremophiles) have evolved to have unique mechanisms for survival in the stressful conditions (most extremophiles require such conditions to grow and reproduce). Extensive research has been conducted to uncover their adaptation at the molecular level.

High temperature adaptation in thermophiles

- Modifying cell membranes by increasing the ratio of saturated to unsaturated fatty acids (bacteria) or by adopting a lipid monolayer (archaea)¹⁹⁷.
- Producing heat-shock proteins and heat-stability proteins^{197,198}.
- Maintaining DNA stability by having a high G+C content or by positive supercoils introduced by the thermophile-specific enzyme reverse DNA gyrase¹⁹⁷.

Low temperature adaptation in psychrophiles

- Modifying the lipid composition of cell membranes (for example, by increasing the ratio of unsaturated to saturated fatty acids) to maintain fluidity¹⁶⁵.
- Producing specialized proteins or other molecules (for example, cold-adapted proteins, cold-shock proteins, cold-acclimation proteins, antifreeze and ice-binding proteins, and osmolytes) that enable the cell to survive under low-temperature conditions^{165,199,200}.
- Limiting metabolic activity by entering a dormant state^{165,199}.

High salt adaptation in halophiles

- Maintaining osmotic homeostasis by accumulating (via a K⁺/Na⁺ antiporter) high levels of inorganic salts (KCl) in the cytoplasm ('salt-in' strategy, found mainly in archaea)²⁰¹.
- Achieving osmotic balance by biosynthesizing and/or accumulating organic and compatible osmotic solutes and thus excluding salt from the cytoplasm ('salt-out' strategy, found mainly in bacteria and eukaryotes)²⁰¹.

Acid adaptation in acidophiles

- Restricting proton influx into the cytoplasm with reversed membrane potential or highly impermeable cell membranes, and promoting excess proton efflux with organic acid degradation or a predominance of secondary transporters²⁰².
- Maintaining intracellular pH with cytoplasmic buffering, stabilizing protein structure and functions of enzymes with 'iron rivets', and repairing DNA and protein damage caused by low pH with chaperones once protons enter the cytoplasm²⁰².

Metal adaptation in acidophiles

- Promoting efflux of the toxic metal out of the cytoplasm, sequestering metal by intracellular or extracellular binding to reduce its toxic effect, excluding metal with a permeability barrier, altering a cellular component to lower the sensitivity of cellular targets to the toxic metal and enzymatically converting the metal into a less toxic form²⁰³.
- Complexing free metals with sulfate to prevent the entry of metal ions into the cell, and establishing passive tolerance to metal influx through an internal positive cytoplasmic transmembrane potential²⁰⁴.

community with the predominance of chlorophototrophs from the Cyanobacteria (mostly *Synechococcus* spp.) and Chloroflexi (*Roseiflexus* and *Chloroflexus* spp.)^{27,28}. Microbial biogeography studies have often reported temperature as the strongest factor shaping community composition at different spatial and temporal scales^{16,25,28,29}. Recently, however, spatial-statistical analysis of a massive 16S rRNA gene dataset from 925 individual hot springs from the Taupō Volcanic Zone in New Zealand found that temperature has a significant effect on diversity only above 70 °C, below which pH is the primary driver¹⁰.

Korarchaeota were originally detected in the Obsidian Pool, YNP³⁰, but were later found to be widely distributed in other geographically separated terrestrial and marine thermal environments³¹ (FIG. 3). Whole-genome shotgun sequencing of Korarchaeota cells purified from an enrichment culture established with the Obsidian Pool sediments resulted in an intact composite genome, 'Candidatus Korarchaeum cryptofilum', which provides an initial look into the biology of these deeply branching archaea³². The archaeon may rely on peptide fermentation to obtain carbon and energy, with a notable lack of the ability to synthesize essential biomolecules such as purines, cofactors and coenzyme A, indicating a dependency on other community members. Recent metagenomic analyses of deep-sea hydrothermal sediments have expanded the genomic representation of Korarchaeota³³.

Aigarchaeota were first discovered in a subsurface, moderately acidic hot spring filamentous streamer community^{34,35} (FIG. 3). Characterization of the first genome, '*Candidatus* Caldiarchaeum subterraneum', suggested an aerobic hydrogen oxidation or carbon monoxide oxidation capability and a potential carbon fixation capability³⁵. Subsequent metagenomics and metatranscriptomics demonstrated that Aigarchaeota in the Octopus Spring (YNP) 'pink streamer' communities are aerobic chemoorganotrophs with autotrophic potential³⁶, whereas analysis of Aigarchaeota genome bins recovered from hot spring sediments from Tengchong, China, indicated a strict or facultative anaerobic lifestyle and a sulfide oxidation ability for energy conservation³⁷.

Robust genomic assemblies have also been recovered from novel bacterial lineages populating various geothermal environments; these include the candidate phyla Acetothermia³⁸, Fervidibacteria, Calescamantes, Atribacteria and Pyropristinus^{13,39}. Atribacteria are widely distributed in various anaerobic environments⁴⁰. This candidate phylum includes the OP9 lineage, which was originally discovered in Obsidian Pool, YNP⁴¹, and its sister lineage JS1, which is a characteristic group in subseafloor environments and is particularly abundant in methane-rich marine sediments⁴². Analyses of metagenome-assembled and single-cell genomes have established Atribacteria as heterotrophic anaerobes lacking respiratory capacity, with fermentation and syntrophy as a shared physiological feature indicating their important roles in carbon cycling in anoxic environments⁴⁰. Kryptonia were discovered in high-temperature circumneutral pH geothermal springs in a survey of a massive global set of metagenomic

Circumneutral Near neutral. data for novel microbial lineages⁴³. Analysis of the near-complete genomes indicated a heterotrophic lifestyle with unique metabolic pathways and noticeable nutritional deficiencies, suggesting an engagement in a metabolic partnership with other community members, such as the co-occurring Armatimonadetes lineage, which potentially complements the missing metabolic features in Kryptonia⁴³.

Deep-sea hydrothermal vents. Deep-sea hydrothermal vent fluids exhibit a broad range of temperatures that are dependent on vent type, namely diffuse flow vents (only a few degrees above ambient to less than 100 °C), white smoker vents (100-300 °C) and black smoker vents (up to 400 °C)⁴². These reduced, thermally charged fluids, when mixing with oxygenated seawater, provide abundant sources of oxidants and reductants for microbial metabolism in a thermal regime that can support life^{42,44}. Environmental surveys have documented different microbial communities both within and between vent fields^{45,46} (FIG. 2). Although temperature is an important determinant of the distribution of microorganisms, variation in fluid chemistry, which is directly influenced by the geological setting and defines the energy available to lithotrophs, has emerged as another key control⁴⁷. The low-pH, metal-rich and sulfide-rich thermal fluids in mafic (basalt)-hosted systems are typically dominated by sulfur-respiring and hydrogen-respiring Epsilonproteobacteria, methanogenic and sulfatereducing Euryarchaeota and the uncultivated SUP05/

Box 2 | Extreme environments as models for microbial ecology research

Natural microbial communities are inherently diverse and complex. Our inability to culture the majority of microorganisms in the environment precludes an in-depth understanding of the assembly, dynamics and evolution of these ecological communities. Extreme environments harbour relatively low diversity to intermediatediversity microbial assemblages. These unique ecosystems have potential as models for the study of microbial ecology and evolution owing to their biological and geochemical simplicity, self-sustainability and explicit species interactions, tight coupling between inorganic and biological processes, and strong environmental gradients²⁰⁵. However, the microbiological exploration of many extreme environments has faced major challenges, particularly difficulties associated with sampling (technical and financial aspects) and nucleic acid extraction (low biomass and interfering substances problems). Nevertheless, a few systems are especially outstanding in terms of their easy access and overall tractability for molecular analyses. Such examples include the acid mine drainage biofilms within the Richmond Mine at Iron Mountain, United States³, and the cyanobacterial mats of alkaline siliceous hot springs in Yellowstone National Park, United States²⁷ (with exceptionally low diversity communities), the Santa Pola multipond solar saltern communities, Spain⁷⁵, and hypersaline mats at Guerrero Negro, Mexico⁷⁸ (with communities exhibiting explicit compositional and functional shifts along the salinity or environmental gradients), the halophilic microbial communities inhabiting halite nodules in the Atacama Desert, South America (with each nodule representing a near-closed miniature ecosystem containing a specialized community vulnerable to environmental change)135, the Inferno Crater Lake, New Zealand (with microbial populations fluctuating with the unique, cyclical thermal profile in which the spring chemistry remains stable)²⁹ and so on. Notably, the Richmond Mine acidophilic biofilms have been successfully used for the development and refinement of various molecular and omics approaches, which have been then widely applied to intermediate-diversity and even moderately high diversity communities³. Overall, decades of extensive molecular and meta-omics studies have substantially advanced our understanding of the biodiversity, community structure and function, population dynamics, and evolution in extreme environments. Such progress has facilitated the exploration of other, more complex natural microbial ecosystems.

Arctic96BD-19 clade of Gammaproteobacteria^{42,48,49}, whereas the associated black smoker-type chimneys harbour Epsilonproteobacteria, Aquificales, and deepsea hydrothermal vent Euryarchaeota^{42,49,50}. The microorganisms populating ultramafic (peridotite)-influenced hydrothermal systems, which vent alkaline and hydrogen-rich and methane-rich fluids, have remained relatively unexplored. To date, marker gene surveys of the few known ultramafic sites have documented substantially different microbial assemblages as opposed to the mafic-hosted systems^{46,51-53}. At the Lost City Hydrothermal Field, Mid-Atlantic Ridge, hot, anoxic inner parts of the active carbonate chimneys are colonized largely by a single species of Methanosarcinales archaea, whereas cooler, oxygenated outer parts are populated by a more diverse community comprising sulfur-oxidizing and methane-oxidizing bacteria and sulfate-reducing bacteria^{51,52,54}. By contrast, vent fluids from hyperthermophilic habitats at this site host a community with taxa belonging to Thermococcales and uncultured Crenarchaeota52.

The co-culturing of the nanosized hyperthermophilic archaeon 'Candidatus Nanoarchaeum equitans' along with its crenarchaeal host Ignicoccus sp. from a marine hydrothermal vent led to the discovery of Nanoarchaeota⁵⁵ (FIG. 3). Genome sequencing of 'Ca. Nanoarchaeum equitans' demonstrated a reduced (less than 0.5 Mb), extremely compacted genome that encodes the machinery for information processing and DNA repair but lacks genes for amino acid, nucleotide, lipid, or cofactor biosyntheses, implying an obligate symbiotic way of life⁵⁶. Environmental surveys have revealed Nanoarchaeota populating other extreme environments, including marine hydrothermal sediments, terrestrial hot springs and hypersaline habitats⁵⁷. Recently, insights from single-cell genomics have enabled the development of effective cultivation conditions that allowed the isolation of a pure co-culture of the ectosymbiotic Nanoarchaeota 'Candidatus Nanopusillus acidilobi' and its Acidilobus host (a crenarchaeote) from a mildly acidic hot spring in YNP58.

The metagenomic discovery and characterization of the superphylum Asgard (FIG. 3), which includes the Lokiarchaeota originally discovered from the Loki's Castle hydrothermal vent field⁵⁹, and the Thorarchaeota⁶⁰, Odinarchaeota and Heimdallarchaeota⁶¹ subsequently discovered in sediments from multiple geographically separated sites (mostly geothermal environments), have provided new evidence for the process of eukaryogenesis. These candidate phyla comprise the closest archaeal relatives of eukaryotes, with the reconstructed genomes enriched in eukaryotic signature proteins, and some even encoding homologues potentially involved in membrane trafficking, vesicle biogenesis, and ubiquitin and cytoskeleton formation, and thus provide strong support for the emergence of the eukaryote host cell from within the Archaea. Most recently, the first isolate of Asgard archaea, 'Candidatus Prometheoarchaeum syntrophicum' strain MK-D1 related to Lokiarchaeota, was obtained from deep marine sediments, leading to unprecedented physiological and genomic insights into eukaryogenesis62.

Cryoconite holes

Small holes on the surface of a glacier. The deposited cryoconite (windblown dust comprising both mineral and biological material, including microorganisms) absorbs solar radiation, melting the snow and ice beneath.

Ectosymbiotic lifestyle

A form of symbiotic behaviour in which an ectosymbiont lives on the surface of its host.

Thalassohaline salt lake

Hypersaline lake of marine origin where salt ratios are similar to those of seawater and pH is near neutral to slightly alkaline.

Cryosphere. Cold ecosystems (temperatures below 5 °C) are highly diverse, and recent microbial ecology studies have focused primarily on the cryosphere^{63,64}, especially glacial habitats and permafrost. Supraglacial and subglacial ecosystems provide a range of habitats supporting substantial numbers of microbial cells. Algae (for example, Chlamydomonaceae and Zygnematophyceae) capable of strong secondary pigmentation gain ecological advantages in photosynthetically active supraglacial communities during the melt season⁶⁵. In cryoconite holes, which represent hotspots for microbial life in supraglacial habitats, cyanobacteria are the main primary producers⁶⁶, with Oscillatoria, Leptolyngbya, Phormidium and Nostoc being the most studied filamentous mat-forming genera⁶⁴. Environmental surveys have implied a global distribution of species of cyanobacteria⁶⁷ and significant temporal changes in bacterial communities in cryoconite holes⁶⁸. Specifically, autotrophic Cyanobacteria dominated communities after snowmelt, whereas heterotrophic Sphingobacteriales populations increased in abundance later in the season68. Substantial heterotrophic activities have also been documented in supraglacial habitats⁶⁹, and these communities are generally dominated by Proteobacteria, Actinobacteria, Bacteriodetes, Acidobacteria, Chloroflexi and Planctomycetes⁶⁴ (FIG. 2), although significant differences have been documented between different geographical regions (particularly between polar and alpine glaciers)⁶⁴. Subglacial environments, which include the basal ice, subglacial water and saturated sediment, are geochemically diverse but few have been sampled owing to challenges associated with access to these environments⁶⁴. These understudied habitats are typically dominated by diverse chemolithoautotrophic communities that are primarily composed of Proteobacteria⁷⁰.

Permafrost has received increasing attention owing to emerging concerns about activation of indigenous microorganisms and subsequent ecological consequences associated with global warming and permafrost thaw. Molecular approaches have revealed members of Chloroflexi, Proteobacteria and Actinobacteria as the most abundant bacteria in polar permafrost soils71, which generally exhibit both less species diversity and less functional diversity than the seasonally thawed active layers71,72. In comparison, permafrost in mountain regions, which features extreme spatial and geothermal variability, has been less studied. A recent diversity survey of European alpine permafrost documented a microbial community enriched with members of the superphylum Patescibacteria, which included several candidate phyla that were predicted to have reduced metabolic capacities and thus an ectosymbiotic lifestyle73.

Hypersaline environments. For hypersaline environments (for example, saline lakes and solar salterns), the Santa Pola multipond solar saltern (Alicante, Spain) is a remarkable model with a full range of salt concentrations that have been extensively studied⁷⁴, revealing a drastic shift of community structure through the gradient, with archaea predominating at the highest salinities (FIG. 2). Specifically, prokaryotic diversity is overwhelmingly dominated by the square archaeon *Haloquadratum*

walsbyi and the bacteroidete Salinibacter ruber in saturated brines, whereas more diverse bacterial and archaeal taxa are observed in intermediate-salinity conditions⁷⁵. Lake Tyrrell (Victoria, Australia) is a thalassohaline salt lake that experiences substantial seasonal changes in environmental conditions, especially salt concentration, thus providing an interesting comparison with the Santa Pola system. Community members are typically dynamic over time and space⁷⁶, and seasonal fluctuations in ion concentrations have been documented to drive microbial succession¹¹. Additionally, extensive benthic microbial mats, highly stratified owing to physiochemical gradients, especially light, oxygen, and sulfide gradients, thrive in saltern evaporation ponds⁵. Notably, extensive 16S rRNA gene sequencing of depth profiles of the hypersaline mats at Guerrero Negro, Mexico, revealed them to be among the most phylogenetically diverse and complex environments known, discovering several new phylum-level bacterial groups and numerous previously undetected lower-level taxa77,78.

De novo metagenomic assembly of multiple libraries from surface waters of Lake Tyrrell led to the discovery of Nanohaloarchaeota²³ within the superphylum DPANN (an acronym of the names of the first phyla included)¹³ (FIG. 3). These ultrasmall but metabolically versatile halophiles are broadly distributed in hypersaline habitats worldwide^{23,79}. Although they were once predicted to be free-living²³, recent co-cultivation of Antarctic Nanohaloarchaeota strains has demonstrated a host dependency of these little-known halophilic archaea⁸⁰. Meanwhile, single-cell genomics has resolved phylogeography, osmotic stress adaptations and the metabolic potential of the candidate bacterial phylum KB1 (REF.⁸¹), and extensive metagenomic sequencing of hypersaline soda lake sediments has recovered 871 genome bins, including members of the candidate phyla radiation (CPR), with several of which being among the most dominant taxa in situ⁸². These abundant CPR organisms are likely fermenters with a possible role in primary carbon degradation.

Subsurface environments. The deep subsurface is operationally defined as an environment that persists at least 1 m below the terrestrial surface or seafloor⁸³. Marine subsurface sediments constitute a major portion of Earth's biosphere^{84,85}. Although subject to severe energetic limitations, microorganisms may persist down to profound depths below the seafloor^{86,87}, and their longterm metabolic activities may play a fundamental role in global biogeochemical cycles88. Environmental surveys have revealed a wide array of uncultivated microorganisms from typical 'marine deep biosphere' groups, which constitute a substantial fraction of the community in geographically and geochemically distinct sites^{85,89} (FIG. 2). These common lineages include the Atribacteria, the Chloroflexi and numerous novel phylogenetic lineages affiliated with the dominant phyla within Archaea, such as Thaumarchaeota (specifically, Marine Group I), Bathyarchaeota, Lokiarchaeota and members of Marine Benthic Group E archaea^{84,89-91}. Vertical distribution patterns of microorganisms in the marine subsurface



 Fig. 2 | Microbial community composition of different extreme environments. Relative abundances of different bacterial and archaeal groups were calculated on the basis of the relative coverage of scaffolds with ribosomal protein S3 (rpS3) marker genes (only those with a relative abundance of 1% or greater are depicted in detail). Raw reads were downloaded from US National Center for Biotechnology Information and US Department of Energy Joint Genome Institute databases, quality trimmed with Trimmomatic and assembled with use of Megahit. The rpS3 sequences were identified from the assemblies with CheckM and the coverage information obtained with Bowtie2 and SAMtools. CPR, candidate phyla radiation; YNP, Yellowstone National Park. Data sources for each habitat are as follows: REFS^{130,186,187} for terrestrial geothermal springs (mats, sediments and water column); REFS^{163,164,188} for deep-sea hydrothermal vents (fluids and chimneys); REFS^{71,189-191} for glacial and permafrost environments; REFS^{74,134,192,193} for hypersaline environments (Santa Pola saltern, Lake Tyrrell and Deep Lake); REFS^{108,194,195} for terrestrial and marine subsurface (terrestrial groundwater and sediments, igneous oceanic crust and subseafloor sediments); and REFS^{12,115,155,196} for acid mine drainage (biofilm, water column and sediments).

> species richness with sediment depth or age, and significant stratification of predominant populations from the universally prevalent broad taxonomic groups^{84,92,93}. Such shifts in community composition have been shown to occur in parallel to drops in microbial activity and abundance caused by reduced energy availability⁹⁴, and are linked to changes in environmental conditions⁹⁵. The igneous ocean crust harbours the largest aquifer on Earth⁹⁶. However, the exploration of microbiology in this subsurface habitat is highly challenging despite technical advances in sampling and in situ experimentation. Pioneering studies have revealed indigenous microbial assemblages that are distinctly different from those populating nearby and associated environments (for example, hydrothermal fluids and plumes, and deep-sea water)96. This distinctness is driven largely by clades of bacteria, especially in the Firmicutes, that appear to be unique to this largely inaccessible system. Meanwhile, a recent study reported the first exploration of the microbiota that is harboured in the lithified lower oceanic crust (which constitutes two-thirds of the crust by volume), uncovering heterogeneously distributed, ultralow-biomass viable communities with unexpected heterotrophic processes to cope with the oligotrophic and fluctuating subsurface conditions⁹⁷.

have been well documented, revealing a decrease in

The terrestrial subsurface is also a major reservoir of Earth's prokaryotic biomass⁹⁸. However, the nature and extent of microbial diversity of this subsurface remain relatively underexplored owing to a lack of accessibility. Environmental surveys have documented diverse microbial communities, sometimes with an unusually high proportion of candidate divisions, in a variety of geological settings, including deep pristine aquifers and igneous and metamorphic rocks, despite the prevailing nutrient-poor and energy source-scarce conditions⁹⁹⁻¹⁰¹ (FIG. 2). A recent compilation of cell concentration and microbial diversity data from continental subsurface localities around the globe found that bacteria are numerically more abundant than archaea, with Proteobacteria dominating the bacterial communities, and Methanomicrobia and Thaumarchaeota dominating the archaeal communities¹⁰². Although no reliable predictors of species richness were identified, statistically significant correlations between community composition and sample lithology were found.

Previous individual studies also documented variations in community composition as a result of environmental variables, including temperature and geochemical composition¹⁰³, or along depth transects through the terrestrial subsurface¹⁰⁰.

Metagenomic studies have uncovered a vast phylogenetic diversity of previously unknown bacteria and archaea in Earth's subsurface, especially terrestrial regions¹⁰⁴⁻¹⁰⁷. Of special interest among the numerous candidate or little-studied phyla are those affiliated with CPR and DPANN, with many of the currently available genomes reconstructed from an aquifer adjacent to the Colorado River near Rifle, United States¹⁰⁸ (FIG. 3). Analyses of the first genomes from CPR bacteria^{104,109,110} have suggested ultrasmall cells, roles in carbon and hydrogen cycles, and metabolic limitations that have prevented their cultivation. Subsequent massive reconstructions of complete and draft CPR genomes have documented features that consistently distinguish these organisms from other bacteria, including small genomes (typically less than 1 Mb) that mostly lack numerous biosynthetic pathways (that is, incomplete tricarboxylic acid cycles and nucleotide and amino acid biosynthesis pathways, and a lack of electron transport chain complexes, thus indicating metabolic dependency on other community members), self-splicing introns and proteins encoded within their rRNA genes (a feature rarely occurring in bacteria), and unusual ribosome compositions (all missing ribosomal protein L30 and specific lineages missing ribosomal protein L9 and biogenesis factor GTPase Der), which collectively establish them as a unique subdivision within the domain Bacteria¹⁴. Meanwhile, genome bins recovered from this subsurface site have drastically expanded genomic sampling of the domain Archaea, revealing a major radiation of organisms (DPANN¹³) with small genomes, novel proteome composition and strong interdependence¹⁵, paralleling the discovery of CPR bacteria. The incomplete or absent pathways were predicted with confidence on the basis of the first complete genomes reconstructed for these enigmatic archaea¹⁵.

Acid mine drainage. Microbial species richness in AMD environments is restricted by the stressful conditions, especially the low pH, elevated levels of toxic metals and limited range of electron donors and acceptors³. These acidophilic assemblages have been extensively studied owing to their importance in AMD generation (a worldwide environmental problem) and potential application in industrial biomining processes. AMD communities are dominated by members of Proteobacteria, Nitrospira and Euryarchaeota9 (FIG. 2), which include the most extensively studied iron oxidizers and/or sulfur oxidizers Acidithiobacillus spp. and Leptospirillum spp., which are important in oxidative sulfide mineral dissolution, and the lesser-known iron oxidizers Ferrovum spp., which are abundant in moderately acidic conditions9,111. The biofilms growing within AMD ecosystems typically have very low diversity and thus are ideal targets for the study of ecological processes in natural environments³. The chemoautotrophic biofilms within the Richmond Mine at Iron Mountain, United States, are generally dominated

Biomining

The application of microorganisms to recover metals from ores or mining waste.



Fig. 3 | Genomic tree of the domain Archaea showing currently proposed major lineages (phyla). All archaeal genomes (metagenome-assembled, single-assembled and isolate-assembled genomes) in the US National Center for Biotechnology Information (NCBI) database GenBank (23 February 2021) were downloaded and dereplicated with use of dRep (version 3.2.0; with parameters '-comp 80 -con 5' selecting genomes with more than 80% completeness and less than 5% contamination). For Nanoarchaeota, Pacearchaeota, Huberarchaea and Hydrothermarchaeota, as none of the associated genomes meets this standard, a lower threshold ('-comp 70 -con 5') was used. To reduce computational workload and simplify the illustration, only representative genomes of major clades were used for the two well-established phyla Crenarchaeota and Euryarchaeota. A total of 553 genomes were included in the phylogenomic analysis. The tree was inferred on the basis of a concatenated alignment of 122 archaea-specific conserved marker genes with use of IQtree version 2.1.2 with 1,000 ultrafast bootstrapping iterations. The best fit LG + F + R10 model was determined by the Bayesian information criterion. The concatenated alignment was generated with GTDB-Tk. Support values of 70% or greater are shown as black circles. The scale bar indicates the number of substitutions per site. Coloured circles show the environments where the first genomic sampling (see related literature summarized in REFS^{20,159}) was done for each of the new potentially phylum level groups that have been described on the basis of genomes from metagenomes or single cells (leftmost), the habitat distribution of all genomes associated with each phylum currently available in the NCBI database (middle; the three extreme environments where most genomes were retrieved, followed by a pie chart showing the relative abundance of genomes retrieved from extreme environments versus non-extreme environments) and the current status of its cultivation (uncultured, enrichment culture, co-culture and pure culture) (rightmost). Note that (1) for some phyla such as Huberarchaeota and Hydrothermarchaeota, currently available genomes are limited to only one or two extreme environments, (2) for some phyla such as Helarchaeota and Aigarchaeota, only a few genomes (fewer than ten) are available, and (3) the hot spring metagenome-assembled genomes for the discovery of Geoarchaeota were deposited only in the US Department of Energy Joint Genome Institute database GOLD. Importantly, nearly all of the proposed archaeal phyla were originally discovered in various extreme environments by molecular approaches (16S ribosomal RNA gene surveys and especially more recent metagenomics or single-cell sequencing) or, in the case of Crenarchaeota, by cultivation. DPANN, the superphylum DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota).

> by one or two species of iron-oxidizing chemoautotrophic bacteria, for example, Leptospirillum group II (REF.¹⁹) (FIG. 2). Less dominant heterotrophic and mixotrophic populations of bacteria, archaea and eukaryotes may also be present with varying abundances according to biofilm developmental stages or sampling space^{19,112}. AMD ecosystems often experience substantial spatiotemporal variations in geochemical conditions. Previous investigations documented significant fluctuations in the relative abundance of archaea correlating to seasonal variations in pH and conductivity in the Richmond Mine¹¹³. More recently, a study examined the ecological distribution of microorganisms in geochemically diverse AMD sites across southeast China and around the globe9. It was found that pH is the primary factor shaping community structure regardless of geographical isolation. Interestingly, such pH-dependent patterns are also clearly evident at the local scale, as demonstrated in a subsequent diversity survey of a massive and highly heterogenous mine tailings site¹¹⁴.

Tailings

The finely ground residues from mined ores after extraction of valuable minerals and metals.

Evapotranspiration

The combined loss of water from the land both by evaporation from the soil surface and by transpiration from the plants growing thereon. that were first explored by meta-omics approaches¹⁹. Extensive Sanger metagenomic sequencing led to the discovery and genome reconstruction of the ultrasmall archaeal lineage ARMAN (archaeal Richmond Mine acidophilic nanoorganisms) — the Parvarchaeota and Micrarchaeota²² (FIG. 3). The composite genomes have abnormally high coding density, and contain a

The Richmond Mine low-complexity AMD bio-

films represent the natural microbial communities

high percentage of genes without known biological functions. Subsequent analyses of globally distributed metagenomic datasets that contain ARMAN-like organisms documented their ubiquity in AMD and hot spring environments and their potential involvements in carbon, nitrogen and iron cycling in situ²¹. Recently, a study¹¹⁵ reported the co-cultivation of an ARMAN-like organism, 'Candidatus Mancarchaeum acidiphilum' Mia14, with Cuniculiplasma divulgatum PM4 (Thermoplasmata) from the sulfidic ores in Parvs Mountain, United Kingdom. Deep genomic sequencing enabled assembly of the complete genome of the highly enriched Mia14 population, which suggests severe voids in central metabolic pathways (for example, lacking minimal sets of enzymes for the biosynthesis of nucleotides and many amino acids, and having incomplete pathways for synthesis of vitamins and cofactors), indicating a host-dependent lifestyle.

Other extreme environments. Soda lakes contain high concentrations of sodium carbonates and constitute the major naturally occurring alkaline environments on Earth. The haloalkaliphilic microbial assemblages populating these unique aquatic ecosystems encounter the multiple stresses associated with high pH (9.0-12.0) and salinity (up to saturation concentrations)¹¹⁶. Gene surveys have revealed salt concentrations as potentially the primary factor shaping microbial diversity and distribution¹¹⁷. Whereas low-salinity and moderatesalinity lakes contain relatively diverse microbial communities dominated by Alphaproteobacteria (mostly Rhodobacteraceae) and Gammaproteobacteria (for example, Halomonas and Thioalkalivibrio), the microbial populations in hypersaline soda lakes resemble those found in hypersaline solar saltern brines and demonstrate a low diversity and the predominance of extremely halophilic Halobacteria¹¹⁷.

Arid ecosystems encompass ~41% of the global land surface and are particularly vulnerable to climate change and land degradation¹¹⁸. Aridity, defined as the ratio of precipitation to potential evapotranspiration, plays dominant roles in shaping the diversity and composition of soil microbial communities in global drylands¹¹⁹. Specifically, increasing aridity significantly reduces microbial diversity and abundance by decreasing plant cover and soil organic carbon content¹¹⁹. Despite the polyextreme conditions, arid soils harbour a diverse group of bacterial lineages, such as Actinobacteria, Chloroflexi, Firmicutes, Gemmatimonadetes and Proteobacteria¹²⁰.

Low temperatures and high pressures dominate dark ocean environments (below the photic zone), which constitute the largest aqueous habitat for microbial life on Earth⁴². Prokaryotic cell abundance exhibits a decrease with depth in the water column^{42,121}. Marker gene surveys and metagenomic analyses have documented depth-dependent patterns in microbial communities^{122,123} and an increase in microbial diversity in the deepest waters of the ocean^{122,124}. Although the factors driving this diversity trend remain largely unknown, deep-water circulation and benthic processes have been implicated to play a role^{123,124}. Deep-ocean bacterial assemblages are typically dominated by Proteobacteria



Fig. 4 | Compositional and functional shifts of the microbial community along specific environmental gradients. Results of representative meta-omics studies for each type of extreme environment are summarized. Abundant taxa (phyla, genera or species/genotypes) in the community, together with major or over-represented functional processes and associated pathways, and prevailing adaptation strategies. are shown. Vertical small arrows (permafrost) indicate higher or lower (relative to the other two soils) representation of specific proteins in the metaproteomics data, whereas horizontal big arrows (hypersaline environment) indicate a relative predominance of a specific adaptation strategy or an increase in the number of sequences related to a specific proton pump in the metagenomes along the salinity gradient. CBB, Calvin-Benson-Bassham; CPR, candidate phyla radiation; DPANN, the superphylum DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota); 3-HP, 3-hydroxypropionate; rTCA, reverse tricarboxylic acid; rACP, reductive acetyl coenzyme A pathway; TCA, tricarboxylic acid; WL, Wood–Ljungdahl. Data sources for each environment are as follows: REF.¹³¹ for hot spring; REF.⁷¹ for permafrost; REFS^{74,134} for hypersaline environment; REF.¹³⁸ for terrestrial subsurface; and REF.¹¹² for acid mine drainage.

> (Alphaproteobacteria, Gammaproteobacteria and Deltaproteobacteria), Chloroflexi and Actinobacteria^{42,123}, and the archaeal communities are dominated by Marine Group I Crenarchaeota¹²⁵.

Community function and dynamics

The uncultured majority limits our ability to resolve physiological and ecological bases underlying the observed spatio-temporal patterns in microbial community structure. Meta-omics may provide insights into the functional significance and ecological roles of all community members by recovering genomic and transcribed and expressed gene information directly from the environment without the requirement for cultivation^{18,19}. Overall, multi-omics of extreme environment communities has revealed the predominance of gene families involved in energy conservation, carbon fixation and nitrogen metabolism, and resistance to the extreme environmental stresses^{18,19}. Meanwhile, comparative omics analyses have been applied to explore functional dynamics at different spatio-temporal scales, targeting microbial communities from geographically separated systems¹², along specific environmental gradients75, associated with local differentiated ecological niches or lifestyles (for example, free-living and sessile growth as biofilms)126,127, as well as of time series^{11,128}. These studies have identified interesting linkages between community or taxon functional profiles and environmental variables, which help interpret patterns observed in microbial distribution.

Several studies have reported metagenomic analyses of microbial communities populating geochemically distinct geothermal features in YNP4,129,130. Genomes of dominant members (including some for novel, previously undetected lineages) were reconstructed, resolving their metabolic potential and strategies with potential linkages to the environmental characteristics and geochemical processes in situ. Comparative metagenomics of thermophilic biofilms along the outflow of the alkaline spring Bison Pool demonstrated a transition from a chemotrophic community in the source pool (92 °C) to a peripheral (56 °C) phototrophic community and drastic shifts in community structure and metabolic functions associated with changes in environmental conditions and energy availability¹³¹ (FIG. 4). Two additional studies used metatranscriptomics approaches

to analyse phototrophic microbial mat time series (in one diel cycle) from the alkaline siliceous Mushroom Spring^{128,132}. In situ gene transcription patterns, especially those associated with photosynthesis, implicated different strategies adopted by the major types of phototrophs to maximize solar energy capture and utilization, and demonstrated how filamentous anoxygenic phototrophs responded at the gene transcription level to the changing environmental conditions and resource availability.

Omics studies of the cryosphere have largely focused on permafrost, especially community dynamics during artificial thaws or along naturally thawing gradients^{71,72,133}. Metagenomic analysis of an Alaskan permafrost microbial community during a 7-day thaw documented a rapid response in community composition and functions, particularly shifts in genes and pathways associated with carbon and nitrogen cycling⁷². Multi-omics has been applied to explore variations in microbial assemblages in soils representing different states of thaw, namely intact permafrost, the seasonally thawed active layer and thermokarst bog⁷¹. Substantial differences in the most active community members and significant shifts in dominant microbial proteins in the analysed soils were found (FIG. 4). Specifically, cold-shock proteins were relatively highly represented and transporters were less represented in the permafrost, whereas proteins involved in cold tolerance and other stress responses were less frequently detected in the bog soils71. Overall, these studies have demonstrated a good linkage between omics data and process measurements, revealed novel adaptation strategies for survival in frozen conditions and provided important information for improving model prediction of greenhouse gas emission from thawing permafrosts under climate change scenarios71,72,133.

Individual studies have reported metagenomic analyses of microbial communities in the Santa Pola saltern, revealing interesting ecological shifts along the salinity gradient^{74,134} (FIG. 4). Specifically, comparative metagenomics has documented relatively diverse microbial populations associated with intermediate salinity and predominance of the hyperhalophiles H. walsbyi and S. ruber in the sodium chloride-saturated crystallizer pond. Correspondingly, analyses of isoelectric point profiles of microbial proteins (indicative of halophilic adaptation) have revealed replacement of the 'salt-out' strategy by the 'salt-in' strategy with increasing salinity¹³⁴. Furthermore, examination of proton pumps involved in a photoheterotrophic lifestyle (note that light is extensively used as an energy source in extreme hypersaline environments) in the Santa Pola metagenomes has demonstrated an increase of bacteriorhodopsin, sensory rhodopsin and halorhodopsin levels and a decrease of proteorhodopsin level along the salinity gradient⁷⁴. Additional meta-omic analyses have been conducted on time-series samples of halophilic microbial communities inhabiting halite (salt rocks) in the Atacama Desert, South America^{135,136}. Notably, comparative metagenomics uncovered distinct dynamics of the community after a rare rainfall event: a rapid shift as the initial response and a more gradual adjustment during recovery¹³⁵.

Diel cycle

A 24-h period comprising a day and a consecutive night, or a regular day–night cycle of physiology or behaviour of an organism.

Multilocus sequence typing A molecular typing technique in which multiple targeted housekeeping genes (loci) are sequenced, often partially.

Although recent surveys have implied that subsurface environments are hotspots of biodiversity hosting numerous previously unknown archaeal and bacterial lineages, how community function (especially metabolic capacities of these uncultured taxa) varies with depth remains almost unexplored. To date, a marker gene transcript-based study has revealed different structures of active microbial communities in two (upper and lower) superimposed pristine limestone aquifers in central Germany¹³⁷. Notably, a large fraction of the community, including the two abundant genera Sulfuricella and Nitrosomonas, encode a capacity for carbon fixation via the Calvin-Benson-Bassham cycle, with energy likely deriving from oxidation of reduced sulfur and nitrogen compounds. Another study used genome-resolved metagenomics and single-cell genomics to demonstrate major shifts in community composition and functions, including the relative significance of different carbon fixation pathways, with relative groundwater source depths at the carbon dioxide-driven Crystal Geyser, United States¹³⁸ (FIG. 4). Importantly, CPR bacteria (which represent the most diverse organisms in the ecosystem) as well as DPANN and other deeply branched archaea were significantly enriched in deeper aquifers. Adaptation to the long-term nutrient limitation in the subsurface environment is likely enabled by interdependent lifestyles and metabolic cooperation involving these ultrasmall microorganisms with limited metabolic capacities.

Omics studies have documented spatially and temporally defined microbial assemblages in AMD and associated environments. Integrated metagenomics and metatranscriptomics of geographically separated AMD communities has revealed gene transcriptional profiles closely related to site geochemical characteristics¹². Proteomics of acidophilic biofilms from a range of AMD sites within the Richmond Mine has found that the proteome of the most abundant bacteria (Leptospirillum group II) responded more markedly to community composition, whereas protein expression in less abundant members was largely influenced by abiotic factors (for example, pH and temperature)¹³⁹. Meanwhile, comparative analyses of microbial communities associated with different acidification stages of mine tailings and ecological succession stages of AMD biofilms have documented explicit temporal shifts in functional profiles^{140,141}. Specifically, tailing communities of early and late stages were enriched with distinct genes involved in low-pH and heavy-metal adaptation, reflecting the different prevailing environmental stresses¹⁴⁰, and changes in physiology of the dominant species Leptospirillum group II were clearly evidenced during ecological succession of a Richmond Mine biofilm community, indicating responses to the distinct stresses (abiotic in biofilm initiation and biotic in biofilm maturation)141. Additional metaproteomics studies have investigated responses of acidophilic biofilm organisms to different pH conditions and temperatures^{112,142}. The results revealed pH-dependent niche differentiation of low-abundance taxa and pH-induced variation in specific functional categories of the dominant species¹¹² (FIG. 4), and influences of temperature at the expressed

protein level on individual organisms, the entire community and specific functions¹⁴².

Microbial evolution in extreme environments

A fundamental question is how microorganisms evolve and specialize to adapt to extreme and sometimes highly fluctuating environmental conditions. This issue was initially addressed by genomic and physiological studies of model organisms^{143,144}. Multilocus sequence typing and comparative analysis of strain genomes have allowed the examination of patterns of genetic variation and underlying processes, and their functional significance in natural populations. As an important parameter that defines the process of evolution, recombination has been explored extensively in halophilic and thermophilic archaea. Research revealed frequent recombination in populations of Halobacterium sp. AUS1 (REF.¹⁴³) and Sulfolobus islandicus144. High recombination frequencies of exchange of long chromosomal fragments have even been found between different species in the genus Haloferax145, indicating more permissive recombination barriers in halophilic archaea than in bacteria. Horizontal gene transfer (HGT) is the major source of gene gain in prokaryotic genomes, and variation in gene content has been hypothesized to be the primary mode of adaptive evolution in microorganisms¹⁴⁶. This process has been found to be particularly prevalent among hyperhalophilic bacteria and archaea147,148, and in Sulfolobales^{149,150}. Specifically, comparative genomics revealed convergence and gene exchange by HGT between S. ruber and haloarchaea, and those modular adaptive elements have a substantial impact on salt adaptation, bioenergetics and photobiology of S. ruber¹⁴⁷.

Direct shotgun genome sequencing of relatively low complexity assemblages from extreme environments has allowed examination of population-level, genome-wide heterogeneity in situ. Specifically, population genomics analysis has demonstrated large-scale homologous recombination between closely related natural populations of the dominant taxa Leptospirillum group II in Richmond Mine AMD biofilms and uncovered virus-mediated and plasmid-mediated high variability in gene content between individuals¹⁵¹. Strain-resolved community proteomics further revealed differential spatio-temporal distribution patterns of the resultant hybrid genotypes^{152,153} and their predominance in different stages of biofilm development¹⁵⁴, indicating that recombination is an important process leading to fine-scale environmental adaptation or distinct ecological strategies¹⁵⁴. Additionally, population metagenomic analyses of microorganisms in extreme environments have allowed the measurement of evolutionary rates for wild populations, a challenging task for individual species in natural microbial assemblages. Research revealed either an exceptionally high nucleotide substitution rate in the natural Leptospirillum group III populations in AMD biofilms¹⁵⁵ or extremely low mutation rates in populations of four dominant lineages in the microbiota inhabiting anoxic subseafloor sediments⁸⁵.

Extreme environments have provided excellent opportunities for the study of the origin and evolution of specific archaeal lineages. Thaumarchaeota are abundant and ubiquitous in nature, playing key roles in the global nitrogen cycle¹⁵⁶. The elusive evolutionary relationship between Thaumarchaeota and its sister lineage Aigarchaeota was recently resolved by comparative genomics upon retrieval of novel Aigarchaeota genome bins from hot spring sediments³⁷. It was found that both phyla likely originated from thermal environments, sharing a large number of genes with their thermophilic last common ancestor³⁷. Later invasion of a wide range of non-thermal habitats by Thaumarchaeota led to the functional differentiation between these two major archaeal groups. The archaea predominating saturated brine environments typically include the haloarchaea and the mostly hydrogen-dependent methanogens. Phylogenetic and phylogenomic analyses implied that haloarchaea may derive from methanogen ancestors¹⁵⁷. Subsequent comparative genomics suggested that such a drastic transformation in physiology (from an anaerobic chemolithoautotroph to an aerobic heterotroph) was enabled by acquisition (via HGT) of a massive amount of genes from bacteria^{148,150}. The recent enrichment and genomic characterization of 'Candidatus Methanonatronarchaeia' (SA1 archaea), a novel, extremely halophilic methanogenic group that branch basally to haloarchaea and inhabit hypersaline lakes, provided further insights into the early evolution of haloarchaea¹⁵⁸. It is notable that the ability for adaptation to high-salt conditions may have evolved before the divergence of haloarchaea and 'Ca. Methanonatronarchaeia'. Evolutionary genomics demonstrated that although both lineages adopt a 'salt-in' strategy, 'Ca. Methanonatronarchaeia' evolved mainly through gene loss, whereas haloarchaea undertook a gene gain-based path for genome evolution.

The recent recovery of genomes from the CPR bacteria and DPANN archaea raises the question of whether the apparent groupings of these major new lineages are artefacts of rapid evolution or reflections of early divergence. So far, drastically different metabolisms have been predicted for CPR and DPANN^{20,159}. Although many lack numerous biosynthetic capacities consistent with their small cell and genome sizes, some, as in the case of Altiarchaeota, have a metabolic potential sufficient to support a free-living lifestyle. Thus, gene loss and genome reduction, characteristic of bacteria switching to a host-associated lifestyle, may be an important phenomenon driving fast evolution in some CPR and DPANN lineages²⁰, resulting in their long branches in phylogenetic trees. Other features that could lead to increased evolutionary rates have also been documented in DPANN archaea; these include the prevalence of diversity-generating elements in Nanoarchaeota, Pacearchaeota and Woesearchaeota^{160,161}, and CRISPR-Cas9 systems encoded by ARMAN¹⁶². Meanwhile, there is also evidence that both CPR and DPANN arose from very early evolving, possibly small-genome organisms. For instance, CPR and DPANN organisms appear to be almost exclusively anaerobes, consistent with potential metabolisms that existed on early Earth²⁰. Additionally, type II/III-like and type III-like ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) enzymes, which are involved in a nucleotide-based pathway indicative of ancient heterotrophy, are widely distributed in some CPR and DPANN organisms²⁰. The presence of specific genes in widely divergent groups may imply features of ancestral organisms that were lost from many modern groups.

Different taxa in natural communities may be subject to different selective pressures and adopt distinct evolutionary strategies. Thus, findings from the study of a few specific taxa may not be generalized directly to whole environments. To this end, several studies have attempted examining evolutionary processes from a community perspective and provided complementary insights^{163,164}. For instance, analysis of a carbonate chimney biofilm metagenome from the Lost City Hydrothermal Field has revealed a remarkably high abundance and diversity of transposase genes, indicating rampant HGT events in a community with low species richness¹⁶⁴. Such a frequent exchange of genes may enable the generation of phenotypic diversity and adaptation to the extremely fluctuating environmental conditions. Another study used genome-resolved metagenomics to resolve patterns of fine-scale genomic variation within and between populations in the community, demonstrating that the evolutionary processes shaping microbial populations from different taxa differed between two geochemically distinct hydrothermal vent fields163.

Concluding remarks

Decades of diversity surveys have comprehensively revealed the phylogenetic breadth of microorganisms inhabiting Earth's major extreme environments. Recent extensive sampling and high-throughput 16S rRNA sequencing of readily accessible ecosystems (for example, hot springs and AMD) further resolved patterns of, and controls over, their large-scale ecological distribution^{9,10}. An interesting trend is that the primary parameter defining a specific type of extreme environment (for example, pH for AMD and salt concentration for hypersaline habitats) appears to be the major factor shaping community composition^{9,16,165}. Typically, overall microbial diversity decreases as the environment becomes more extreme, with a handful of taxa predominating in the harshest conditions^{75,117,140}. Although recent studies have attempted to predict these diversity patterns^{166,167} (BOX 3), resolving their nature and underlying ecological and evolutionary processes is a key emerging challenge in microbial ecology.

The accelerated reconstruction and characterization of microbial genomes from major new groups in extreme environments have largely changed the structure of the tree of life and revolutionized our understanding of archaea and bacteria. Notably, despite knowledge that archaea are present in most environments on Earth, recent discoveries of major new lineages in the Archaea tree have been made largely from genomic exploration of various extreme environments (FIG. 3). This highlights the importance of these habitats as a treasure trove for microbial 'dark matter' and as model systems for the understanding of microbial ecology and life of early Earth. Given the highly unbalanced research intensity between different types of extreme environment, and the huge scope for discovery of new classes and groups at

Box 3 | Modelling microbial communities in extreme environments

Quantitatively predicting microbial community composition and functional attributes at the spatio-temporal scale can reveal gaps in knowledge of unknown and future community states and build a more comprehensive understanding of the interactions between microorganisms and their biotic and abiotic environments. Major obstacles to developing and implementing predictions are associated with the limited material (that is, observational data) used for model construction and the complex underlying causality. An additional hurdle is that there are few widely accepted quantitative tools to make accurate predictions²⁰⁶. Nonetheless, recent advances in high-throughput DNA sequencing and bioinformatics tools have made it tractable to understand how microbial communities and functions are structured through space and time, and how they vary during environmental change²⁰⁷.

Prediction of community dynamics in extreme environments is facilitated by the relatively low complexity assemblages and the clearly evidenced patterns and drivers of microbial distribution⁶⁻¹¹. For instance, model-based approaches have recently been applied to predict the biogeographical origin of specific halite-associated archaeal populations¹⁶⁷ and the taxonomic and functional structure of microbial assemblages in geochemically diverse acid mine drainage environments¹⁶⁶. Continuing efforts will help advance the field of microbiome research beyond simple associative and descriptive analyses²⁰⁶.

lower taxonomic levels, continuing genomic exploration of the diversity and metabolic capacity of this uncultured majority will remain a major theme in this field. Such studies are important, and should be accompanied by targeted cultivation and functional verification efforts, as there is evidence that these uncultivated lineages, although often low in relative abundance, may play critical functional roles in communities, or gain ecological advantages under specific environmental conditions.

A major knowledge gap in microbial ecology in general is related to the eukaryotes and viruses in natural microbial assemblages. Although they perform many crucial ecosystem functions and play a key role shaping prokaryotic population dynamics, these microorganisms are particularly difficult to cultivate and challenging to study by genomic methods. The nature and uniqueness of extreme environments may offer tremendous opportunities for the ecological and evolutionary analyses of these little-known community members. Viruses play a particularly substantial part in the microbial food web in many extreme environments^{168,169}, and omics approaches and technical advances are just beginning to unravel the mystery of viromes in these habitats¹⁷⁰. Research has revealed high viral diversity¹⁷¹ and seasonal succession and geographical distribution of viral populations^{171,172}, discovered novel viruses (for example, virophages in Antarctic aquatic systems¹⁷³ and clades of huge phages from hot springs and hypersaline and

subsurface environments¹⁷⁴) and previously unrecognized ecological roles, and elucidated how viruses interact with specific hosts and subsequently drive community dynamics¹⁷⁵. Meanwhile, there have been relatively few omics analyses of microbial eukaryotes in extreme environments. Notably, metagenomics has enabled reconstruction of the near-complete genome of a filamentous fungus from an AMD streamer biofilm dominated by fungi in the Richmond Mine, allowing evaluation (using transcriptomics and proteomics) of its metabolic differentiation across habitats¹⁷⁶. Another meta-omics analysis documented a low-diversity but viable fungal community in the deep lithified crust at the Atlantis Bank Gabbro Massif, Indian Ocean¹⁷⁷. In principle, these emerging efforts and future efforts on viral and eukaryotic members of microbial communities will help resolve the inherently complex interactions across all trophic levels in extreme environments, which ultimately shape the structure and dynamics of the whole community. Such omics data could then be incorporated into predictive models for the prediction of community composition and function in natural settings.

As we are striving to enrich our knowledge of microbial processes and interactions in extreme environments, a fundamental question is whether ecological principles inferred in these compelling model systems (BOX 2) could predict patterns in other less constrained ecosystems. This is important because many of the features that endow the reduced complexity of these models separate them from most other environments³. However, no model system could capture all characteristics of other systems, and recent studies have demonstrated simplified model microbial consortia as useful tools to manipulate complexity and advance our understanding of complex natural communities¹⁷⁸. More importantly, parallel findings, such as deterministic community assembly¹⁷⁹, specific ecological drivers (for example, temperature, pH and salinity) of microbial biogeography patterns¹⁸⁰⁻¹⁸² and evolutionary processes shaping population genomic variation¹⁸³, have been reported in other, less well defined systems, thereby confirming the broader relevance of insights gained from the study of extreme environments3. Thus, an integrated view of microbial ecology and evolution in extreme environments will ultimately further our understanding of more complex natural microbial communities and the global ecosystem.

Published online 9 November 2021

- Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* 409, 1092–1101 (2001).
- Schmid, A. K., Allers, T. & DiRuggiero, J. Snapshot: microbial extremophiles. *Cell* 180, 818–818.e1 (2020).
- Denef, V. J., Mueller, R. S. & Banfield, J. F. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J.* 4, 599–610 (2010).
- Inskeep, W. P. et al. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front. Microbiol.* 4, 67 (2013).
- Oren, A. Halophilic microbial communities and their environments. *Curr. Opin. Microbiol.* 33, 119–124 (2015).
- Reysenbach, A. L., Wickham, G. S. & Pace, N. R. Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Vellowstone National Park. *Appl. Environ. Microbiol.* 60, 2113–2119 (1994).
- Bond, P. L., Smriga, S. P. & Banfield, J. F. Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. *Appl. Environ. Microbiol.* 66, 3842–3849 (2000).
- Huber, J. A. et al. Microbial population structures in the deep marine biosphere. *Science* **318**, 97–100 (2007).
- Kuang, J. L. et al. Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. *ISME J.* 7, 1038–1050 (2013).
- Power, J. F. et al. Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* 9, 2876 (2018).
 Extensive sampling and high-throughout 16S rRNA
 - Extensive sampling and high-throughput 165 rRNA gene sequencing have provided deeper insights into the patterns and ecological drivers of microbial communities inhabiting geothermal springs.
- Podell, S. et al. Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. *ISME J.* 8, 979–990 (2014).
- Chen, L. X. et al. Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. *ISME J.* 9, 1579–1592 (2015).
- Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437 (2013).

- Brown, C. T. et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523, 208–211 (2015).
- Castelle, C. J. et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701 (2015).

The cultivation-independent reconstruction of the first complete genomes for members of the DPANN archaea allowed confident prediction of incomplete

- or absent pathways for these enigmatic organisms.
 Sharp, C. E. et al. Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8, 1166–1174 (2014).
- Hedlund, B. P. et al. Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. *Curr. Opin. Microbiol.* 25, 136–145 (2015).
- Hua, Z. S. et al. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J.* 9, 1280–1294 (2015).
- Tyson, G. W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43 (2004).
 This is the first shotgun metagenomic sequencing study that enabled reconstruction of near-complete microbial genomes directly (without cultivation) from a natural community.
- Castelle, C. J. & Banfield, J. F. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197 (2018).
- Chen, L. X. et al. Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *ISME J.* 12, 756–775 (2018).
- Baker, B. J. et al. Enigmatic, ultrasmall, uncultivated Archaea. Proc. Natl Acad. Sci. USA 107, 8806–8811 (2010).
- Narasingarao, P. et al. *De novo* metagenomic assembly reveals abundant novel major lineage of *Archaea* in hypersaline microbial communities. *ISME J.* 6, 81–93 (2012).
- 24. Brock, T. D. Life at high temperatures. *Science* **158**, 1012–1019 (1967).
- Cole, J. K. et al. Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J.* 7, 718–729 (2013).
- Colman, D. R. et al. Ecological differentiation in planktonic and sediment-associated chemotrophic microbial populations in Yellowstone hot springs. *FEMS Microbiol. Ecol.* **92**, fiw137 (2016).
- Ward, D. M. et al. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345, 63–65 (1990).
- Miller, S. R. et al. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl. Environ. Microbiol.* 75, 4565–4572 (2009).
- Ward, L. et al. Microbial community dynamics in Inferno Crater Lake, a thermally fluctuating geothermal spring. *ISME J.* 11, 1158–1167 (2017).
- Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl Acad. Sci. USA* **91**, 1609–1613 (1994).
- Takai, K. & Yoshihiko, S. A molecular view of archaeal diversity in marine and terrestrial hot water environments. *FEMS Microbiol. Ecol.* 28, 177–188 (1999).
- Elkins, J. G. et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
- Dombrowski, N., Teske, A. P. & Baker, B. J. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* 9, 4999 (2018).
- Nunoura, T. et al. Genetic and functional properties of uncultivated thermophilic crenarchaeotes from a subsurface gold mine as revealed by analysis of genome fragments. *Environ. Microbiol.* 7, 1967–1984 (2005).
- Nunoura, T. et al. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223 (2011).
- Beam, J. P. et al. Ecophysiology of an uncultivated lineage of Algarchaeota from an oxic, hot spring filamentous 'streamer' community. *ISME J.* 10, 210–224 (2016).
- Hua, Z. S. et al. Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* 9, 2832 (2018).

- Takami, H. et al. A deeply branching thermophilic bacterium with an ancient acetyl-CoA pathway dominates a subsurface ecosystem. *PLoS ONE* 7, e30559 (2012).
- Colman, D. R. et al. Novel, deep-branching heterotrophic bacterial populations recovered from thermal spring metagenomes. *Front. Microbiol.* 7, 304 (2016).
- Nobu, M. et al. Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J.* 10, 273–286 (2016).
- Hugenholtz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* 180, 366–376 (1998).
- Orcutt, B. N., Sylvan, J. B., Knab, N. J. & Edwards, K. J. Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol. Mol. Biol. Rev.* 75, 361–422 (2011).
- 43. Eloe-Fadrosh, E. A. et al. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* 7, 10476 (2016). This is a good example of how analysis of the increasing wealth of metagenomic data collected from diverse environments may lead to the discovery of novel major lineages.
- Kelley, D. S., Baross, J. A. & Delaney, J. R. Volcanoes, fluids, and life at Mid-Ocean Ridge spreading centers. *Annu. Rev. Earth Planet. Sci.* 30, 385–491 (2002).
- Perner, M. et al. In situ chemistry and microbial community compositions in five deep-sea hydrothermal fluid samples from Irina II in the Logatchev field. *Environ. Microbiol.* 15, 1551–1560 (2013).
- Flores, G. E. et al. Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environ. Microbiol.* 13, 2158–2171 (2011).
- Dick, G. J. et al. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat. Rev. Microbiol.* **17**, 271–283 (2019).
- Campbell, B. J., Summers Engel, A., Porter, M. L. & Takai, K. The versatile *e*-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* 4, 458–468 (2006).
- Reysenbach, A. L., Longnecker, K. & Kirshtein, J. Novel bacterial and archaeal lineages from an in situ growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl. Environ. Microbiol.* 66, 3798–3806 (2000).
- Takai, K., Komatsu, T., Inagaki, F. & Horikoshi, K. Distribution of archaea in a black smoker chimney structure. *Appl. Environ. Microbiol.* 67, 3618–3629 (2001).
- Schrenk, M. O., Kelley, D. S., Bolton, S. A. & Baross, J. A. Low archaeal diversity linked to subseafloor geochemical processes at the Lost City Hydrothermal Field, Mid-Atlantic Ridge. *Environ. Microbiol.* 6, 1086–1095 (2004).
- Brazelton, W. J., Schrenk, M. O., Kelley, D. S. & Baross, J. A. Methane- and sulfur-metabolizing microbial communities dominate the Lost City Hydrothermal Field ecosystem. *Appl. Environ. Microbiol.* **72**, 6257–6270 (2006).
- Reveillaud, J. et al. Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ. Microbiol.* 18, 1970–1987 (2016).
- Brazelton, W. J. et al. Archaea and bacteria with surprising micro-diversity show shifts in dominance over 1000-year time scales in hydrothermal chimneys. *Proc. Natl Acad. Sci. USA* 107, 1612–1617 (2010).
- Huber, H. et al. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67 (2002).
- Waters, E. et al. The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* 100, 12984–12988 (2003).
- Casanueva, A. et al. Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments. *Extremophiles* 12, 651–656 (2008).
- Wurch, L. et al. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat. Commun.* 7, 12115 (2016).
 This is an interesting study demonstrating that insights from genomic studies may been develop.

insights from genomic studies may help develop effective cultivation strategies for the isolation of novel microbial species. Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179 (2015).

The discovery and genomic characterization of Lokiarchaeota have unveiled insights into eukaryogenesis.

- Construction of the second second
- Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358 (2017).
- Imachi, H. et al. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577, 519–525 (2020).
 This study reports the isolation of the first
 - member of the superplylum Asgard, confirming the existence of these archaea and their close phylogenetic relatedness to eukaryotes.
- Margesin, R. & Collins, T. Microbial ecology of the cryosphere (glacial and permafrost habitats): current knowledge. *Appl. Microbiol. Biotechnol.* 103, 2537–2549 (2019).
- Boetius, A., Anesio, A. M., Deming, J. W., Mikucki, J. A. & Rapp, J. Z. Microbial ecology of the cryosphere: sea ice and glacial habitats. *Nat. Rev. Microbiol.* 13, 677–690 (2015).
- Hoham, R. W. & Duval, B. in *Snow Ecology* (eds Jones, H. et al.) 168–228 (Cambridge Univ. Press, 2001).
- Edwards, A. et al. Coupled cryoconite ecosystem structure-function relationships are revealed by comparing bacterial communities in alpine and Arctic glaciers. *FEMS Microbiol. Ecol.* 89, 222–237 (2014).
- Jungblut, A. D., Lovejoy, C. & Vincent, W. F. Global distribution of cyanobacterial ecotypes in the cold biosphere. *ISME J.* 4, 191–202 (2010).
- Franzetti, A. et al. Temporal variability of bacterial communities in cryoconite on an alpine glacier. *Environ. Microbiol. Rep.* 9, 71–78 (2017).
- Anesio, A. M., Hodson, A. J., Fritz, A., Psenner, R. & Sattler, B. High microbial activity on glaciers: importance to the global carbon cycle. *Glob. Chang. Biol.* 15, 955–960 (2009).
- Christner, B. C. et al. A microbial ecosystem beneath the West Antarctic ice sheet. *Nature* 512, 310–313 (2014).
- Hultman, J. et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521, 208–212 (2015).
- Mackelprang, R. et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480, 368–371 (2011).
- Frey, B. et al. Microbial diversity in European alpine permafrost and active layers. *FEMS Microbiol. Ecol.* 92, fiw018 (2016).
- Fernández, A. B. et al. Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol. Ecol.* 88, 623–635 (2014).
- Ventosa, A. et al. Microbial diversity of hypersaline environments: a metagenomic approach. *Curr. Opin. Microbiol.* 25, 80–87 (2015).
- Emerson, J. B. et al. Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* 2013, 370871 (2013).
- 77. Ley, R. E. et al. Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat.
- Appl. Environ. Microbiol. **72**, 3685–3695 (2006).
 Harris, J. K. et al. Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J.* **7**, 50–60 (2013).
 - This study retrieves an unprecedented number of nearly full length 16S rRNA gene sequences from the microbial mats of the Guerrero Negro hypersaline environment, Mexico, demonstrating them to be among the most diverse, complex and novel microbial ecosystems known.
- Vavourakis, C. D. et al. Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Front. Microbiol.* 7, 211 (2016).
- Hamm, J. N. et al. Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl Acad. Sci.* USA. 116, 14661–14670 (2019).
- Nigro, L. M., Hyde, A. S., MacGregor, B. J. & Teske, A. Phylogeography, salinity adaptations and metabolic potential of the candidate division KB1 bacteria based on a partial single cell genome. *Front. Microbiol.* 7, 1266 (2016).

- 82. Vavourakis, C. D. et al. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* **6**, 168 (2018).
- Edwards, K. J., Becker, K. & Colwell, F. The deep, dark energy biosphere: intraterrestrial life on Earth. *Annu. Rev. Earth Planet. Sci.* 40, 551–568 (2012).
- Parkes, R. J. et al. A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: geosphere interactions. *Mar. Geol.* 352, 409–425 (2014).
- Starnawski, P. et al. Microbial community assembly and evolution in subseafloor sediment. *Proc. Natl Acad. Sci. USA* 114, 2940–2945 (2017).
- Ciobanu, M. C. et al. Microorganisms persist at record depths in the subseafloor of the Canterbury Basin. *ISME J.* 8, 1370–1380 (2014).
- Inagaki, F. et al. Exploring deep microbial life in coal-bearing sediment down to ~ 2.5 km below the ocean floor. *Science* **349**, 420–424 (2015).
- D'Hondt, S., Pockalny, R., Fulfer, V. M. & Spivack, A. J. Subseafloor life and its biogeochemical impacts. *Nat. Commun.* 10, 3519 (2019).
- Petro, C., Starnawski, P., Schramm, A. & Kjeldsen, K. U. Microbial community assembly in marine sediments. *Aquat. Microb. Ecol.* **79**, 177–195 (2017).
- Teske, A. & Sørensen, K. B. Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J.* 2. 3–18 (2008).
- Orsi, W. D. Ecology and evolution of seafloor and subseafloor microbial communities. *Nat. Rev. Microbiol.* 16, 671–683 (2018).
- Sørensen, K. B. & Teske, A. Stratified communities of active Archaea in deep marine subsurface sediments. *Appl. Environ. Microbiol.* **72**, 4596–4603 (2006).
- Walsh, E. A. et al. Relationship of bacterial richness to organic degradation rate and sediment age in subseafloor sediment. *Appl. Environ. Microbiol.* 82, 4994–4999 (2016).
- 94. Petro, C. et al. Marine deep biosphere microbial communities assemble in near-surface sediments in Aarhus Bay. *Front. Microbiol.* **10**, 758 (2019).
- Jorgensen, S. L. et al. Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc. Natl Acad. Sci. USA* 109, E2846–E2855 (2012).
- Edwards, K. J., Wheat, C. G. & Sylvan, J. B. Under the sea: microbial life in volcanic oceanic crust. *Nat. Rev. Microbiol.* 9, 703–712 (2011).
 Li, J. et al. Recycling and metabolic flexibility dictate
- Li, J. et al. Recycling and metabolic flexibility dictate life in the lower oceanic crust. *Nature* 579, 250–255 (2020).
 This is a multiple-approach exploration to provide

the first insights into the ultralow-biomass microbial assemblages inhabiting the lithified lower oceanic crust.

- Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl Acad. Sci. USA* 115, 6506–6511 (2018).
- Nyyssönen, M. et al. Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. *ISME J.* 8, 126–138 (2014).
- 100. Lin, X., Kennedy, D., Fredrickson, J., Bjornstad, B. & Konopka, A. Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environ. Microbiol.* 14, 414–425 (2012).
- Osburn, M. R. et al. Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Front. Microbiol.* 5, 610 (2014).
- Magnabosco, C. et al. The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* 11, 707–717 (2018).
- Navarro-Noya, Y. E. et al. Pyrosequencing analysis of the bacterial community in drinking water wells. *Microb. Ecol.* 66, 19–29 (2013).
- Wrighton, K. C. et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 337, 1661–1665 (2012).
- Bagnoud, A. et al. Reconstructing a hydrogen driven microbial metabolic network in Opalinus Clay rock. *Nat. Commun.* 7, 12770 (2016).
- Magnabosco, C. et al. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J.* 10, 730–741 (2016).
- Hernsdorf, A. W. et al. Potential for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J.* 11, 1915–1929 (2017).
- 108. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical

processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).

- Kantor, R. S. et al. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *mBio* 4, e00708–e00713 (2013).
- Wrighton, K. C. et al. Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. *ISME J.* 8, 1452–1463 (2014).
- 111. Hallberg, K. B., Coupland, K., Kimura, S. & Johnson, D. B. Macroscopic streamer growths in acidic, metal-rich mine waters in north Wales consist of novel and remarkably simple bacterial communities. *Appl. Environ. Microbiol.* **72**, 2022–2030 (2006).
- Belnap, C. P. et al. Quantitative proteomic analyses of the response of acidophilic microbial communities to different pH conditions. *ISME J.* 5, 1152–1161 (2011).
- 113. Edwards, K. J. et al. Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. *Appl. Environ. Microbiol.* 65, 3657–3652 (1999).
- 114. Liu, J. et al. Correlating microbial diversity patterns with geochemistry in an extreme and heterogeneous environment of mine tailings. *Appl. Environ. Microbiol.* **80**, 3677–3686 (2014).
- 115. Golyshina, O. V. et al. 'ARMAN' archaea depend on association with euryarchaeal host in culture and in situ. *Nat. Commun.* 8, 60 (2017).
- Antony, C. P. et al. Microbiology of Lonar Lake and other soda lakes. *ISME J.* 7, 468–476 (2013).
- 117. Sorokin, D. Y. et al. Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles* 18, 791–809 (2014).
- Reynolds, J. F. et al. Clobal desertification: building a science for dryland development. *Science* **316**, 847–851 (2007).
- Maestre, F. T. et al. Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proc. Natl Acad. Sci. USA.* **112**, 15684–15689 (2015).
- Makhalanyane, T. P. et al. Microbial ecology of hot desert edaphic systems. *FEMS Microbiol. Rev.* 39, 203–221 (2015).
- Reinthaler, T. et al. Prokaryotic respiration and production in the meso- and bathypelagic realm of the eastern and western North Atlantic basin. *Limnol. Oceanogr.* 51, 1262–1273 (2006).
- Limnol. Oceanogr. 51, 1262–1273 (2006).
 122. Hewson, I., Steele, J. A., Capone, D. G. & Fuhrman, J. A. Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. *Limnol. Oceanogr.* 51, 1274–1283 (2006).
- 123. DeLong, E. F. et al. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496–503 (2006).
- 124. Pham, V. D., Konstantinidis, K. T., Palden, T. & DeLong, E. F. Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ. Microbiol.* **10**, 2313–2330 (2008).
- 125. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
- 126. Ziegler, S. et al. Oxygen-dependent niche formation of a pyrite-dependent acidophilic consortium built by archaea and bacteria. *ISME J.* 7, 1725–1737 (2013).
- 127. Méndez-García, C. et al. Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. *ISME J.* 8, 1259–1274 (2014).
- Klatt, C. G. et al. Temporal metatranscriptomic patterning in phototrophic Chloroflexi inhabiting a microbial mat in a geothermal spring. *ISME J.* 7, 1775–1789 (2013).
- 129. Klatt, C. G. et al. Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol.* 4, 106 (2013).
- Front. Microbiol. 4, 106 (2013).
 130. Inskeep, W. P. et al. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS ONE* 5, e9773 (2010).
- 131. Swingley, W. D. et al. Coordinating environmental genomics and geochemistry reveals metabolic transitions in a hot spring ecosystem. *PLoS ONE* 7, e38108 (2012).
- 132. Liu, Z. et al. Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. *ISME J.* 5, 1279–1290 (2011).
- Woodcroft, B. J. et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* 560, 49–54 (2018).

- 134. Ghai, R. et al. New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1, 135 (2011).
- Uritskiy, G. et al. Halophilic microbial community compositional shift after a rare rainfall in the Atacama Desert. *ISME J.* **13**, 2737–2749 (2019).
- 136. Uritskiy, G. et al. Cellular life from the three domains and viruses are transcriptionally active in a hypersaline desert community. *Environ. Microbiol.* 23, 3401–3417 (2021).
- 137. Herrmann, M. et al. Large fractions of CO₂-fixing microorganisms in pristine limestone aquifers appear to be involved in the oxidation of reduced sulfur and nitrogen compounds. *Appl. Environ. Microbiol.* 81, 2384–2394 (2015).
- Probst, A. J. et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* 3, 328–336 (2018).
- 139. Mueller, R. S. et al. Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Mol. Syst. Biol.* 6, 374 (2010).
- 140. Čhen, L. X. et al. Shifts in microbial community composition and function in the acidification of a lead/zinc mine tailings. *Environ. Microbiol.* 15, 2431–2444 (2013).
- 141. Mueller, R. S. et al. Proteome changes in the initial bacterial colonist during ecological succession in an acid mine drainage biofilm community. *Environ. Microbiol.* **13**, 2279–2292 (2011).
- 142. Mosier, A. C. et al. Elevated temperature alters proteomic responses of individual organisms within a biofilm community. *ISME J.* 9, 180–194 (2015).
- 143. Papke, R. T., Koenig, J. E., Rodriguez-Valera, F. & Doolittle, W. F. Frequent recombination in a saltern population of *Halorubrum. Science* **306**, 1928–1929 (2004).
- 144. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus islandicus. Mol. Biol. Evol.* 22, 2354–2361 (2005).
- 145. Naor, A., Lapierre, P., Mevarech, M., Papke, R. T. & Gophna, U. Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* 22, 1444–1448 (2012).
- 144–1448 (2012).
 146. Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V. & Whitaker, R. J. Biogeography of the Sulfolobus islandicus pan-genome. *Proc. Natl Acad. Sci. USA* 106, 8605–8610 (2009).
- 147. Mongodin, E. F. et al. The genome of Salinibacter Ruber: convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc. Natl Acad. Sci. USA 102, 18147–18152 (2005).
- 148. Nelson-Sathi, S. et al. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl Acad. Sci. USA* 109, 20537–20542 (2012). Comparative genomics provides evidence that massive amounts of gene influx from bacterial sources may have led to the drastic change in lifestyle in the extremely salt tolerant Haloarchaea
- lifestyle in the extremely salt tolerant Haloarchaea.
 149. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* 7, 46 (2012).
- Nelson-Sathi, S. et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* 517, 77–80 (2015).
- 151. Simmons, S. L. et al. Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. *PLoS Biol.* 6, e177 (2008).
- 152. Lo, I. et al. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* 446, 537–541 (2007).
- 153. Denef, V. J. et al. Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environ. Microbiol.* 11, 313–325 (2009).
- 154. Denef, V. J. et al. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc. Natl Acad. Sci. USA* **107**, 2383–2390 (2010).
- 155. Denef, V. J. & Banfield, J. F. In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. *Science* **336**, 462–466 (2012).

This study provides a time-series population metagenomic analysis of microorganisms in exceptionally low diversity AMD biofilms, allowing

for the first time measurement of evolutionary rates for wild populations.

- 156. Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**, 245–252 (2008).
- Nat. Rev. Microbiol. 6, 245–252 (2008).
 157. Kelly, S., Wickstead, B. & Gull, K. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc. Biol. Sci.* 278, 1009–1018 (2011).
- Sci. 278, 1009–1018 (2011).
 158. Sorokin, D. Y. et al. Discovery of extremely halophilic, methyl-reducing euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat. Microbiol.* 2, 17081 (2017).
- 159. Baker, B. J. et al. Diversity, ecology and evolution of archaea. *Nat. Microbiol.* **5**, 887–900 (2020).
- Paul, B. G. et al. Targeted diversity generation by intraterrestrial archaea and archaeal viruses. *Nat. Commun.* 6, 6585 (2015).
- 161. Paul, B. G. et al. Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat. Microbiol.* 2, 17045 (2017).
- 162. Burstein, D. et al. New CRISPR-Cas systems from uncultivated microbes. *Nature* 542, 237–241 (2017).
- 163. Anderson, R. E. et al. Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nat. Commun.* 8, 1114 (2017).
- 164. Brazelton, W. J. & Baross, J. A. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J.* 3, 1420–1424 (2009).
- 165. Jansson, J. K. & Taş, N. The microbial ecology of permafrost. *Nat. Rev. Microbiol.* **12**, 414–425 (2014).
- Kuang, J. et al. Predicting taxonomic and functional structure of microbial communities in acid mine drainage. *ISME J.* **10**, 1527–1539 (2016).
- 167. Clark, D. R. et al. Biogeography at the limits of life: do extremophilic microbial communities show biogeographical regionalization? *Clob. Ecol. Biogeogr.* 26, 1435–1446 (2017).
- 168. Atanasova, N. S., Roine, E., Oren, A., Bamford, D. H. & Oksanen, H. M. Global network of specific virus-host interactions in hypersaline environments. *Environ. Microbiol* 14 405–440 (2012)
- Microbiol. 14, 426–440 (2012).
 169. Wilkins, D. et al. Key microbial drivers in Antarctic aquatic environments. *FEMS Microbiol. Rev.* 37, 303–335 (2013).
- 170. Cavicchioli, R. Microbial ecology of Antarctic aquatic systems. *Nat. Rev. Microbiol.* 13, 691–706 (2015).
- 171. López-Bueno, A. et al. High diversity of the viral community from an Antarctic lake. *Science* **326**, 858–861 (2009).
- 172. Aguirre de Cárcer, D., López-Bueno, A., Pearce, D. A. & Alcamí, A. Biodiversity and distribution of polar freshwater DNA viruses. *Sci. Adv.* 1, e1400127 (2015).
- 173. Yau, S. et al. Virophage control of Antarctic algal host-virus dynamics. *Proc. Natl Acad. Sci. USA* 108, 6163–6168 (2011).
 This is the first study to reveal the important

ecological roles of virophages and their regulation of host-virus interactions.

174. Al-Shayeb, B. et al. Clades of huge phages from across Earth's ecosystems. *Nature* 578, 425–431 (2020). Analysis of massive metagenomic datasets revealed clades of huge phages from diverse habitats, including extreme environments.

- 175. Tschitschko, B. et al. Antarctic archaea-virus interactions: metaproteome-led analysis of invasion, evasion and adaptation. *ISME J.* **9**, 2094–2107 (2015).
- Mosier, A. C. et al. Fungi contribute critical but spatially varying roles in nitrogen and carbon cycling in acid mine drainage. *Front. Microbiol.* 7, 238 (2016).
 Quemener, M. et al. Meta-omics highlights the diversity.
- 177. Quemener, M. et al. Meta-omics highlights the diversity, activity and adaptations of fungi in deep oceanic crust. *Environ. Microbiol.* **22**, 3950–3967 (2020).
- 178. Fredrickson, J. K. Ecological communities by design. *Science* **348**, 1425–1427 (2015).
- Fuhrman, J. A. et al. Annually recourring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* 103, 13104–13109 (2006).
- 180. Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
- Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl Acad. Sci. USA* 104, 11436–11440 (2007).
- 182. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103**, 626–631 (2006).
- 183. López-Pérez, M., Haro-Moreno, J. M., Coutinho, F. H., Martinez-Garcia, M. & Rodriguez-Valera, F. The evolutionary success of the marine bacterium SAR11 analyzed through a metagenomic perspective. *mSystems* 5, e00605-20 (2020).
- Altshuler, I., Goordial, J. & Whyte, L. G. in *Psychrophiles: From Biodiversity to Biotechnology* (ed. Margesin, R.) 153–180 (Springer International Publishing, 2017).
 Huang, L. N., Kuang, J. L. & Shu, W. S. Microbial
- Huang, L. N., Kuang, J. L. & Shu, W. S. Microbial ecology and evolution in the acid mine drainage model system. *Trends Microbiol.* 24, 581–593 (2016).
- Klatt, C. G. et al. Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J.* 5, 1262–1278 (2011).
- Menzel, P. et al. Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* **70**, 411–424 (2015).
- 188. Stokke, R. et al. Functional interactions among filamentous Epsilonproteobacteria and Bacteroidetes in a deep-sea hydrothermal vent biofilm. *Environ.*
- *Microbiol.* **17**, 4063–4077 (2015). 189. Zeng, Y. et al. Potential rhodopsin- and bacteriochlorophyll-based dual phototrophy in
- a High Arctic glacier. *mBio* **11**, e02641–20 (2020). 190. Simon, C., Wiezer, A., Strittmatter, A. W. & Daniel, R. Phylogenetic diversity and metabolic potential
- revealed in a glacier ice metagenome. *Appl. Environ. Microbiol.* **75**, 7519–7526 (2009).
 191. Lipson, D. A. et al. Metagenomic insights into anaerobic metabolism along an Arctic peat soil profile.
- PLoS ONE 8, e64659 (2013).
 Podell, S. et al. Assembly-driven community genomics of a hypersaline microbial ecosystem. *PLoS ONE* 8, e61692 (2013).
- 193. DeMaere, M. Z. et al. High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl Acad. Sci. USA.* **110**, 16939–16944 (2013).
- 194. Smith, A. R. et al. Carbon fixation and energy metabolisms of a subseafloor olivine biofilm. *ISME J.* 13, 1737–1749 (2019).
- 195. Zhao, R. et al. Geochemical transition zone powering microbial growth in subsurface sediments. *Proc. Natl Acad. Sci. USA.* **117**, 32617–32626 (2020).

- 196. Luo, Z. H. et al. Diversity and genomic characterization of a novel Parvarchaeota family in acid mine drainage sediments. *Front. Microbiol.* **11**, 612257 (2020).
- Lewin, A., Wentzel, A. & Valla, S. Metagenomics of microbial life in extreme temperature environments. *Curr. Opin. Biotechnol.* 24, 516–525 (2013).
- 198. Schlesinger, M. J. Heat-shock proteins. *J. Biol. Chem.* **265**, 12111–12114 (1990).
- D'Amico, S., Collins, T., Marx, J.-C., Feller, G. & Gerday, C. Psychrophilic microorganisms: challenges for life. *EMBO Rep.* 7, 385–389 (2006).
 Bakermans, C., Bergholz, P. W., Ayala-del-Rio, H.
- Bakermans, C., Bergholz, P. W., Ayala-del-Rio, H. & Tiedje, J. in *Permafrost Soils* (ed. Margesin, R.) 159–168 (Springer, 2009).
- 201. Gunde-Cimerman, N., Plemenitaš, A. & Oren, A. Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiol. Rev.* 42, 353–375 (2018).
- Baker-Austin, C. & Dopson, M. Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* 15, 165–171 (2007).
- Dopson, M., Baker-Austin, C., Koppineedi, P. R. & Bond, P. L. Growth in sulfidic mineral environments: metal resistance mechanisms in acidophilic microorganisms. *Microbiology* 149, 1959–1970 (2003).
- Dopson, M., Ossandon, F. J., Lövgren, L. & Holmes, D. S. Metal resistance or tolerance? Acidophiles confront high metal loads via both abiotic and biotic mechanisms. *Front. Microbiol.* 5, 157 (2014).
- Allen, E. E. & Banfield, J. F. Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* 3, 489–498 (2005).
- 206. Sakowski, E. et al. Current state of and future opportunities for prediction in microbiome research: report from the Mid-Atlantic Microbiome Meet-up in Baltimore on 9 January 2019. mSystems 4, e00392–19 (2019).
- Lima-Mendez, G. et al. Determinants of community structure in the global plankton interactome. *Science* 348, 1262073 (2015).

Acknowledgements

Support was received from the National Natural Science Foundation of China (grant no. 41830318 to W.-S.S. and grant nos 31570500 and 31870111 to L-N.H.). The contributions of Z. Luo and other students and postdoctoral scientists in the authors' team are gratefully acknowledged. The authors thank Z. Hua and Z. Luo for helpful discussions, and K. Anantharaman and J. F. Banfield for providing metagenome-derived taxon abundance data for the analysis of the microbial composition of the Colorado River aquifer communities (Fig. 2).

Author contributions

The authors contributed equally to all aspects of the article.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Microbiology thanks J. DiRuggiero and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2021