CellPress
OPEN ACCESS

Review

Modern microbiology: Embracing complexity through integration across scales

A. Murat Eren[1,2,3,4,5,*] and Jillian F. Banfield[6,7,8,9,10,*]
[1]Helmholtz Institute for Functional Marine Biodiversity, 26129 Oldenburg, Germany
[2]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[3]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany
[4]Marine Biological Laboratory, Woods Hole, MA, USA
[5]Max Planck Institute for Marine Microbiology, Bremen, Germany
[6]Department of Earth and Planetary Sciences, University of California, Berkeley, Berkeley, CA, USA
[7]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[8]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA
[9]Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia
[10]Department of Environmental Science Policy, and Management, University of California, Berkeley, Berkeley, CA, USA
*Correspondence: meren@hifmb.de (A.M.E.), jbanfield@berkeley.edu (J.F.B.)
https://doi.org/10.1016/j.cell.2024.08.028

SUMMARY

Microbes were the only form of life on Earth for most of its history, and they still account for the vast majority of life's diversity. They convert rocks to soil, produce much of the oxygen we breathe, remediate our sewage, and sustain agriculture. Microbes are vital to planetary health as they maintain biogeochemical cycles that produce and consume major greenhouse gases and support large food webs. Modern microbiologists analyze nucleic acids, proteins, and metabolites; leverage sophisticated genetic tools, software, and bioinformatic algorithms; and process and integrate complex and heterogeneous datasets so that microbial systems may be harnessed to address contemporary challenges in health, the environment, and basic science. Here, we consider an inevitably incomplete list of emergent themes in our discipline and highlight those that we recognize as the archetypes of its modern era that aim to address the most pressing problems of the 21st century.

INTRODUCTION

Our planet faces formidable challenges that span from climate change, environmental pollution, depletion of fertile soils, and loss of biodiversity to feeding and caring for the ever-growing human populations in a sustainable fashion. Microbes, whose evolution has been intimately interconnected with changes in Earth's physicochemical environments over biological time, remain an integral part of these interlinked crises. With its aim to understand microbial life and given the potential to harness microbes to address these huge problems, the field of microbiology is emerging as arguably the most important science of the mid-21st century.

Microbiology has made remarkable progress throughout its relatively brief, few hundred-year history. The advances seem far from reaching a plateau, as we have likely just begun to scratch the surface of major biotechnological and medical advances that are enabled by studies of microbial life. Today, microbiologists aim to elucidate and ultimately address a wide range of processes that include greenhouse gas emissions and soil carbon sinks to wastewater management, protection of clean water resources, detoxification of heavy-metal and organic pollutants, maintenance of soil fertility and rhizosphere nutrition, the emergence and mitigation of plant and animal pathogens, antibiotic resis-

tance, and pandemics. Investigations to address such diverse questions often rely on a broad range of data-producing and processing technologies, which demand increasingly higher levels of integration and synthesis for unified insights.

Microbiology draws its power from relatively well-established technologies such as molecular sequencers, mass spectrometers, high-throughput cultivation, genetic manipulation, and a large arsenal of 'omics strategies, statistics, and modeling, as well as technologies that are rapidly emerging, such as artificial intelligence, machine learning, microbiome-based genome editing, and more. As contemporary questions of microbiology demand the orchestration of diverse data generation and analysis strategies that yield increasingly large datasets, integrative approaches that span across scales of complexity emerge as one of the hallmarks of modern microbiology. Here, we take a look at the past and contemplate the future of our rapidly transforming discipline by starting with a discussion on some of the major themes that underpin the new data-intensive epoch of microbiology.

MODERN MICROBIOLOGY: FROM INDIVIDUAL MICROBES TO COMMUNITIES

It is a daunting task to precisely define what modern microbiology is, but a useful one to reminisce about how we got to where we are

and to recognize new motivations, opportunities, and challenges that influence today's microbiologists. If we could explain how we study microbes today and what we know about them to Antonie van Leeuwenhoek, who saw microbes the first time in the 1670s, to Ignaz Semmelweis, who realized washing hands reduces patient mortality in hospitals in the 1840s, or to Fanny Hesse, who established agar to grow microbes in laboratory settings in the 1880s, what would they find to be the most fascinating advance microbiology has achieved compared with its early days? We indeed would have had a myriad of surprises to show them proudly, as we now can isolate microbes in nanodroplets,[1] image them *in situ* with colorful probes[2] and *in aqua* in three dimensions,[3] survey metabolites they exchange,[4] visualize and model their metabolic activities both at the level of individual elements[5] or at planetary scales,[6] pull them around from their surface proteins,[7] stick them into little wells,[8] and even modify,[9] transform,[10] or synthesize[11] their genetic repertoire, or transfer their genes into the genomes of plants[12] and animals.[13] But apart from the permanently central role of technology in microbiology, there is one conceptual fracture in our otherwise continuous progress toward a better understanding of microbial life that can serve as a line between the still alive and well classical roots of microbiology and its aspiring modern times: our expanding focus from individual organisms in cultures to complex communities in natural habitats, with the admission that "a pure culture is not enough."[14]

The shift from individuals to communities in microbiology was enabled by a series of technological and conceptual advances: it was foreshadowed by the theory of evolution and guided by many breakthroughs, such as the discovery of the structure of DNA,[15] the recognition of the utility of the molecular information encoded in genomes as a means to document evolutionary history,[16] and the arrival of sequencing.[17] One of the latest of the most defining conceptual advances that culminated the thread of progress into a single evolutionary framework that connected all cellular life was the work led by Carl Woese, which offered a first glimpse of the tree of life, revealed by comparison of ribosomal RNA (rRNA) gene sequences in 1977.[18] The translation of this transformative idea to practical applications in microbiology started to emerge through the pioneering work by Norman Pace and colleagues that described rapid sequencing of rRNA gene fragments[19] and cloning methods that enabled gene recovery from environmental samples,[20] which gained further traction with simpler molecular strategies that require lower input biomass.[21] In the following years, PCR amplification of 16S and 18S rRNA gene fragments became an increasingly popular method to survey microbial diversity in naturally occurring habitats.[22,23] Coinciding with increasing availability and affordability of short read sequencing of early 2000s, a plethora of amplicon studies revealed the richness and dynamism of microbial communities from marine[24] and terrestrial systems[25] to host-associated habitats[26] and led to a broad recognition and awareness of Earth's flourishing microbiomes and their connection to the rest of life.[27] The transition from individuals to whole microbiomes motivated a societal-level adjustment of our perspective: microbes were no longer a band of bad actors in the public mind that typically caused diseases, but the architects of life, the wardens of planetary health, and the fundamental forces that make our ecosystems tick.

Surveys of microbial diversity through rRNA gene amplicons were a wildfire in microbiology and resulted in over 85,000 studies in the past 20 years. Yet, taxonomic insights offer limited utility to understand functional drivers of biological systems, a pinnacle desire that brings together many corners of microbiology. Despite the inspiration amplicon surveys elicited to better understand microbial systems, their integrative power remained severely constrained. Understanding of microbes in their environmental context required genomes, which first became available for isolates that were generated by indispensable yet labor-intensive isolation efforts.[28] However, isolation is inherently biased toward microbes that can grow alone and for which suitable growth conditions and substrates are known, and it has not delivered even a single representative of the majority of currently known lineages.[29]

The realization of the need for direct access to the genetic makeup of environmental populations intensified during the late 1990s. In pursuit of natural products potentially available from the untapped diversity of microbes in the environment, Handelsman and colleagues introduced a cloning-based approach to access genes and pathways of interest from soil "metagenome."[30] The early approaches to metagenomics, which promoted the idea of cultivation-independent surveys of environmental populations, relied upon "isolating DNA from an environmental sample, cloning the DNA into a suitable vector, transforming the clones into a host bacterium, and screening the resulting transformants."[14] The arduous beginnings of metagenomics offered unique insights into environmental microbes,[31,32] and demonstrated the feasibility of culture-independent surveys of natural populations. An alternative approach to metagenomics relied upon the sequencing of randomly chosen short DNA fragments.[33] Combined with the advances in sequencing chemistry and molecular approaches, this "shotgun sequencing" approach replaced cloning-based strategies and ultimately enabled the reconstruction of genomes directly from natural ecosystems.

The core idea behind reconstructing genomes from environmental metagenomes, which was laid out in the first example of its kind, stood on a relatively simple principle that relied on a few imperfect but computationally tractable steps[34]: assembling metagenomic short reads into longer contiguous DNA segments (contigs), estimating the abundance of contigs in metagenomes, and partitioning contigs into separate genome bins by exploiting differences in GC content and abundance of each discrete population. The power of genome-resolved metagenomics came from the fact that no aspect of it required any prior knowledge about organisms in a given environment. Its simplicity allowed this strategy to be encapsulated in automated workflows, from which emerged a plethora of genomes, not only from taxa microbiology knew about but also from those we had not recognized before, and not only from the primary domains of life but also from viruses, plasmids, and other exciting biological entities. Over time, the size of sequencing datasets increased, along with the power of bioinformatics tools, enabling the transition of genome-resolved surveys from relatively low-complexity systems to the microbiologically most complex ecosystems on Earth, such as oceans and soils. Metagenome-assembled genomes (MAGs), along with single-amplified genomes (SAGs)

and those that emerged from improved isolation efforts, opened the floodgates of genomic data pouring in from all environments. Today, the rapidly maturing evolutionary framework around genomes offers a playground for all corners of our discipline to integrate heterogeneous data types that slice through microbial systems in complementary ways through analytical, computational, and molecular tools to address fundamental questions of modern microbiology (Figure 1).

From communities back to individual microbes

Cultivation, the most fundamental methodological advance that shaped classical microbiology by enabling molecular insights with broad applications through manipulation of representative microbial organisms, continues to be the bedrock of contemporary studies of microbial physiology in the post-omics era.[35] The ability to reconstruct microbial genomes directly from the environment represents an effective means to improve cultivation efforts as it can provide clues regarding metabolic requirements of little-known organisms and guide their directed isolation.[36] The first demonstration of this idea was the isolation of *Leptospirillum ferrodiazotrophum*[37] from a relatively simple subsurface acid mine drainage biofilm, following the observation that it was the only organism in its community capable of nitrogen fixation. The first successful isolation of *Coxiella burnetii*,[38] the causative agent of human Q fever,[39] represents another example of how partial insights into genomic features of target organisms can help define complex media that support the axenic growth of even obligate intracellular organisms. Peering into the metabolic make up of environmental communities through MAGs continues to be a powerful tool to support cultivation efforts even from complex mixtures of microbes. For instance, Pope et al. were able to isolate a representative of the family Succinivibrionaceae, so-called WG-1, that dampens methane emissions from digestive activities of macropods using a composite genome reconstructed directly from a metagenome. Metabolic insights gained from this genome supported the development of a medium with specific carbohydrate and nitrogen sources as well as antibiotics to achieve an axenic culture.[40] Others applied similar principles to grow microbes from other environments, including human gut[41] and even extreme subsurface environments.[42] Prediction of the episymbiotic lifestyles of candidate phyla radiation (CPR) bacteria from environmentally derived genomes[43] led to the co-isolation of these tiny bacteria with their hosts[44] and opened the way for microscopic,[45] metabolic,[46] transcriptomic[47] definitions of the impact of these symbioses and their interactions with their hosts.[48] An exciting workflow that integrated the use of environmental genomes, animal immunity, and fluorescent labeling for targeted cultivation of microbes was demonstrated by Cross et al., who cultivated multiple CPR bacteria, including the elusive Absconditabacteria (SR1).[7] Cross et al. first surveyed SAGs and MAGs of target bacteria to identify membrane proteins that could serve as immunogens, then predicted extracellular domains in these candidate genes with homology to previously recognized immunogens, synthesized matching peptides, introduced them into rabbits, recovered antibodies generated by the rabbit immune system, used these antibodies to "label" bacterial cells in a human saliva sample, and were able to sort labeled cells into cultures.[7]

In addition to culture-independent insights that improve cultivation efforts and new approaches that aim to preserve natural growth conditions of environmental microbes[49,50] or their interactions,[51] modern approaches to cultivation include high-throughput strategies that exploit robotics and microfluidics. Recent studies demonstrate the utility of microfluidics to generate thousands of nanoliter droplets on petri dishes with chemical gradients,[52] and the application of robotics and machine learning to process thousands of colonies per hour.[53] One critical application of microfluidics is targeted cultivation of microorganisms based on any genetic feature that is learned from metagenomes. Using this strategy, Ma et al. were able to identify a microcolony that carried their target gene sequence among 500 microcolonies they grew on a multiplexed microfluidic device and recovered the first isolate of a previously unidentified genus of the *Ruminococcaceae* family[54] that has been among the most abundant and prevalent taxa in the human gut with no cultured representatives until then.[55] Another exciting application of microfluidics is the untargeted, high-throughput cultivation of microorganisms in droplets. The isolation of individual bacterial cells in droplets reduces the loss of microbial colonies in cultivation efforts due to resource competition, differences in growth rates, or dilution of quorum-sensing molecules.[56] In an application of this approach, Watterson et al. populated millions of picolitre droplets created from distinct culture media with individual microbial cells from a single sample.[1] Droplets they sorted based on colony density included many of the rare organisms conventional cultivation strategies missed in the same sample and recovered antibiotic-resistant strains that conventional plate-based assessments of antibiotic resistance were unable to detect.[1]

Dramatic expansion of the tree of life

Our newly found power to sequence genes from environmental samples had far-reaching implications on our understanding of the diversity and evolution of life through phylogenomics and resulted in remarkable changes to the structure of the tree of life. The sequencing of marker genes such as rRNAs had revealed many, but not all, branches of life, but inconsistencies in predicted physiology in major branches and unexpected gene inventories only became evident upon recovery of genomes. Notable features of the revised and genome-informed tree of life include the Asgard archaea with their inventories of eukaryote-associated genes,[57] making them of great interest from the perspective of the origin of eukaryotes.[58] Other major groups of particular interest include the possibly not monophyletic DPANN archaea and the seemingly monophyletic CPR, which display consistently small genomes and limited biosynthetic capacities. This led to the prediction that both the bacterial and archaeal domains of life feature large groups of microbes that they generally depend on host microorganisms for basic building blocks such as lipids and nucleotides,[59] and recent work has clarified the partitioning of host-derived lipids into some episymbionts.[60] Inclusion of sequences from metagenomes as well as isolates brought into focus the dramatic prevalence of branches of the tree that remain without any isolated representatives.[29] Further refinements have begun to clarify relationships between deep branches, for example, the Chloroflexi
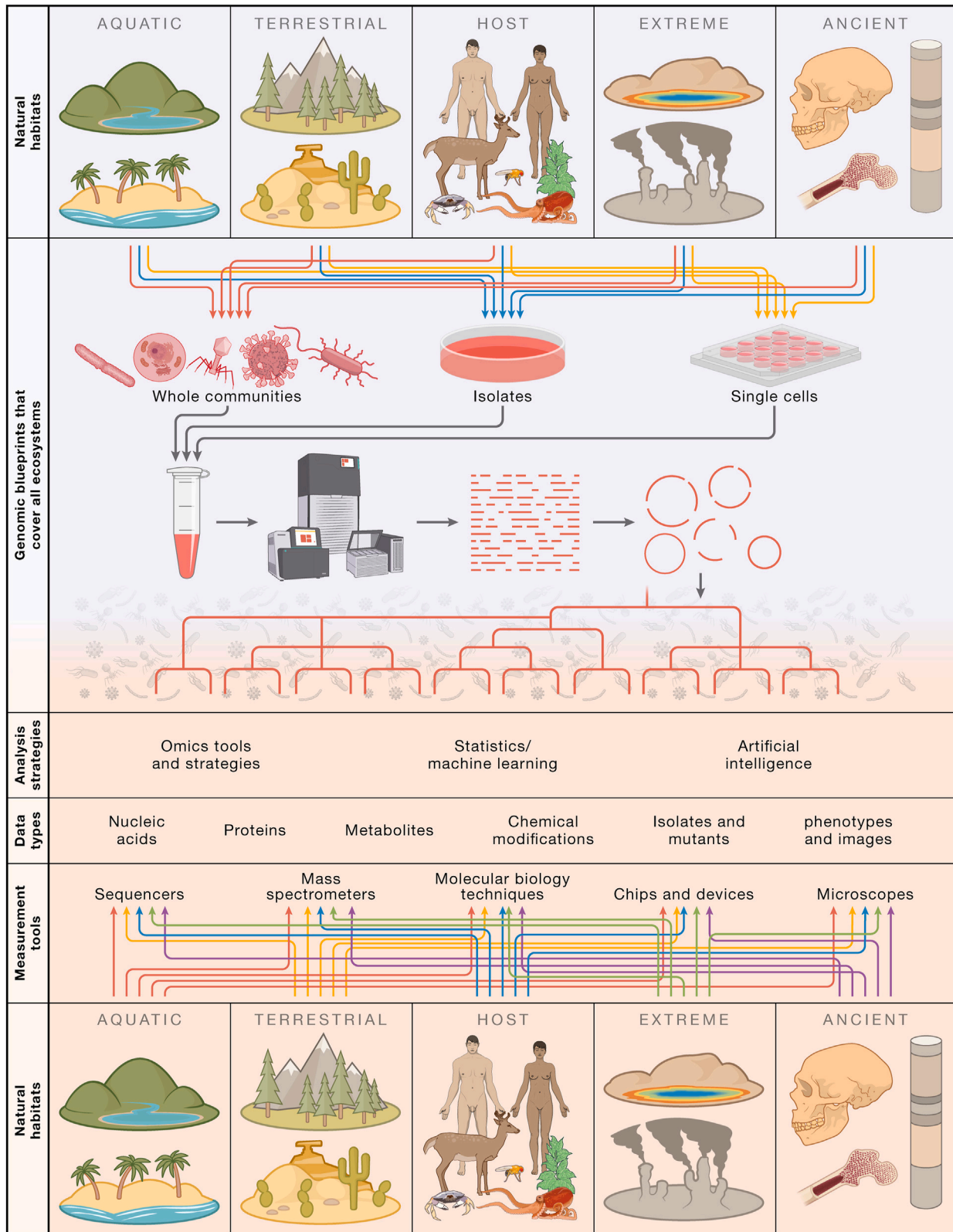
Figure 1. The measurement tools, data types, and analysis strategies commonly used in modern microbiology to study naturally occurring microbial organisms in a wide range of complex habitats through rapidly emerging genomic blueprints of the diversity of life

and CPR, and the distribution of traits.[61] The revised views of the bacteria and archaeal domains as well as the whole tree of life contrast dramatically with even contemporary views of the diversity of life that are highly eukaryote-centric and, in some cases, do not even include non-eukaryotic organisms. These renderings promote a new appreciation of our place in biological evolution and biodiversity.

The value of complete and finished genomes is hard to over emphasize. In addition to providing a quality-assured inventory of genes and metabolic potential without contamination by sequences from other sources, finished genomes ensure full genomic context for any genetic feature and define the overall genome architecture, including exact genome size and replichore structure. However, one of the hallmarks of the contemporary catalogs of MAGs is the low completion and occasional contamination issues due to the highly fragmented nature of metagenomic assemblies.[62] While some of these concerns can be addressed by manual[62,63] or automated genome curation and bin refinement approaches,[64] long-read sequencing technologies are already enabling the reconstruction of complete genomes without binning, and this will soon be possible at scale. Products of long-read sequencing can be error prone, and efforts to automate error correction and to improve the accuracy of long-reads are underway (e.g., Sacristán-Horcajada et al.,[65] Wick and Holt,[66] and Hackl et al.[67]). The number of assembly algorithms that can report circularized sequences from metagenomes is also increasing.[68–70] However, assembled long-reads are not without issues, and the resulting sequences can suffer from false circularization due to artifactual repeated sequences and misassemblies that are not supported by the original reads. High-throughput methods for detection and correction of assembly errors will ultimately bring microbiology closer to the ideal of routinely rendering accurate and complete genomes and increase the utility of genomes from environments for downstream applications.

Expanding and exploring the protein universe

Metagenomic sequencing can uncover new protein families, and whole proteome comparisons (including proteins with unknown functions) can establish relatedness among large groups of organisms,[71] including those that lack universal genes, such as extrachromosomal elements (ECEs).[72] Despite the lack of recognizable functional domains, the sequences themselves can provide information about physicochemical properties of the protein (e.g., hydrophobic domains or positively charged regions that may be involved in nucleic acid binding) and uncover metal and ligand binding sites. A new generation of computational methods such as AlphaFold,[73] RoseTTAFold,[74] and ESMFold[75] enable prediction of the three-dimensional structures of proteins and thus can provide functional clues that are not apparent based on the sequence alone. Often, predictions leverage templates in the form of the structures of biochemically validated proteins accessed via large protein databases such as the Protein Data Bank (PDB).[76] However, it is possible to improve these predictions by supplementing the underlying multi-sequence protein alignments using the ever-expanding database of amino acid sequences derived from metagenomic studies.[77] In fact, using sequences independent of templates, it is possible to discover new protein folds and vastly increase access to three-dimensional structures for proteins with no related solved structures or sequence similarity to known proteins.[78] These approaches, in combination with ecological and other information, have the potential to advance understanding the protein inventories of enigmatic entities such as viruses[79] and aid in the development of biotechnology tools.[80] Exciting new bilingual models that translate information between structural and sequence space[81] as well as computational tools that increase the accessibility of protein structure calculations[82] and searchability of existing structures[83,84] in large databases with millions of predictions[75] help locate variants of proteins of interest with potentially different size or kinetics. It may be possible to revolutionize drug development given methods that predict molecular interactions and discover small molecules or antibodies that bind to proteins identified as drug targets.[85] Indeed, computational approaches continue to improve at a rapid pace to predict a broad range of biomolecular interactions[73] and the stoichiometry of large protein complexes.[86] Altogether, the ongoing revolution in protein structure space and new opportunities to integrate structural insights with established 'omics data types and analysis strategies will undoubtedly influence every corner of microbiology.

Going beyond the tree of life—(Giant) viruses, (giant) plasmids, and other extrachromosomal entities

With their astronomical numbers and complexity, viruses represent a force of nature that is implicated in all aspects of life, starting from its very origins.[87,88] The immense impact of viruses on life continues through metabolic reprogramming of their hosts,[89,90] biogeochemical and ecological processes they influence in complex habitats,[91] and even the modifications of the reproductive behavior of animals.[92]

Contemporary virus research spans from their characterization and organization into cohesive units to linking them to particular hosts or ecosystems and investigating their biotechnological applications. The emergence of computational tools to recognize viral sequences in metagenomic assemblies[93–95] resulted in the recovery of tens of thousands of new viral genomes from the human gut,[96] oceans,[97] and soils.[98] The increasing access to metatranscriptomics further supports the recovery of RNA viruses[99] and enables spatiotemporal tracking of eukaryotic viruses of public health significance and their variants.[100] Microbiologists were surprised when "giant eukaryotic viruses," whose genome sizes surpass those of some bacteria,[101] were discovered.[102] Genome-resolved metagenomics had a remarkable impact also on our understanding of the diversity and biogeography of giant viruses,[103] their complex metabolic capabilities,[104] host reprogramming,[105] and impact on host genome evolution.[106] Despite the progress toward a global taxonomic framework for viruses,[107,108] the luxury to be represented in a universal tree of life continues to be limited to cellular organisms. However, at least at the level of individual studies, the integrated use of metagenomic assemblies or genome-resolved metagenomics, phylogenomics, and protein structure prediction offers promising insights, not only to achieve a more complete representation of viral clades but also fill the gaps between disconnected realms of viruses. For instance, in a recent study, Gaïa

et al. used phylogeny-guided genome-resolved metagenomics using DNA-dependent RNA polymerase subunits to describe a large clade of viruses, and by combining functional features and predicted protein structures, they were able to reveal a first association between giant viruses that infect unicellular marine eukaryotes and herpesviruses that infect mammals.[109] In addition to their prominent place in basic science, viruses continue to be a powerful driver of biotechnology for innumerable reasons, including therapeutic opportunities,[110] the potential of their proteins as small genome editors,[111] and their ability to deliver DNA (including editing tools) into the cells in which they replicate.[112–114] They also force us to question the classical definitions of what is a bona fide living entity[115] with opposing views,[116] yet continuously elucidate the conceptual frameworks that can offer formal organizations of the realms of biology.[117]

Plasmids, which were first defined by Joshua Lederberg in 1952[118] and have served as a workhorse of molecular biology and genetics ever since,[119] represent another exciting corner of microbiology. The significance of plasmids parallels that of viruses: plasmids also impact the ecology and evolution of their hosts,[120] often find themselves under the spotlight of public health concerns such as global antibiotic resistance[121] largely due to their contribution to the proliferation of such genetic factors,[122] encode genes that are associated with a wide range of metabolic activities,[123] and suffer from a similar exclusion from trees of life. It is conceivable that these similarities may simply be an echo of deeper evolutionary connections that are no longer present in contemporary sequence space. Indeed, recent characterizations of "virus-plasmids" that can act both as viruses and plasmids,[124] observations of plasmids that propagate through virus-like particles,[125] or phylogenetic analyses that attribute the emergence of DNA viruses to recombination events between RNA viruses and plasmids[126] reinforce the likely presence of deeper evolutionary links and hint that there is likely much more to discover at this junction. Also similar to viruses, the diversity of naturally occurring plasmids has been grossly underestimated. While metagenomic sequencing and assemblies access naturally occurring plasmids, recognizing plasmids in sequence collections remains a challenge. Some plasmids, even for some of the most rigorously studied microbial taxa such as *Wolbachia*, remained undetected for decades.[127] But our ability to recognize plasmids is improving thanks to computational strategies that exploit k-mer profiles learned from reference plasmids,[128] conserved plasmid functions,[129] circularity of sequences in metagenomes,[130] or deep learning approaches that use combinations of these features.[131] Today, we know that even for relatively well-studied ecosystems, plasmids found exclusively in metagenomes exceed the number of previously known plasmids multiple folds.[132] An interesting insight comes from a recent study that uses 68,000 non-redundant plasmids that were identified from human gut metagenomes and shows that plasmid distribution patterns are not correlated with bacterial taxonomy,[123] which reinforces the critical importance of treating them as semi-independent entities that represent an adaptive force in microbial systems,[133] similar to viruses. Yet organizing plasmids into evolutionarily cohesive units is perhaps an even bigger challenge compared with their identification. Unlike viruses, where relatively stable genes within distinct viral

realms can yield taxonomic frameworks to shed some light on evolutionary routes of individual viral genomes,[134] attempts to find cohesive units in diverse sets of plasmids have been largely limited to partitioning pairwise sequence similarity networks.[135,136] A recent attempt to identify evolutionarily cohesive "plasmid systems" in such networks through a containment-aware partitioning approach demonstrated a means to identify plasmid backbones and cargo genes and partition plasmid genes into housekeeping genes and those that respond to particular environmental conditions *de novo*.[123] Plasmids are a pillar of biopharmaceutical advances,[119] and naturally occurring plasmids represent a new asset. But the world of plasmids is chaotic: some can be so extremely essential and integrated into the host cellular physiology that they compelled researchers to call them "chromids,"[137] while others can look so small and irrelevant even though they can be some of the most numerous genetic elements in natural habitats.[138] With their unruly nature, plasmids remain an intimidating and exciting dimension of microbiology.

A significant challenge that is generally true for most extrachromosomal elements that are not typically found integrated into host genomes is the difficulty of linking them to their host organisms, which can be a significant burden in complex environments. By preserving the unity of naturally occurring genetic elements within cells that are typically lost in bulk sequencing efforts, single-cell genomics approaches offer somewhat straightforward means to capture virus-host[139] and plasmid-host[140] associations with high confidence. In some cases, it is possible to leverage the fact that bacteria and archaea incorporate snippets of their extrachromosomal element genomes into their CRISPR loci (spacers) so that invasive elements can be recognized by the protein-RNA-based surveillance system. Extrachromosomal elements whose genomes match to CRISPR spacer sequences, especially from the same or related samples, are likely the targets of that microbial immune system, thus indicating ECE-host relationships.[141] Because CRISPR loci are often very fast evolving, the effectiveness of this method can be improved when spacer sequences are recovered directly from short metagenomic reads rather than from the assembled consensus sequence.[141] Another solution to identify hosts for extrachromosomal entities in complex samples comes from Hi-C,[142] a molecular strategy that can physically link genetic entities that are adjacent to one another prior to sequencing and can reveal associations between host genomes and extrachromosomal entities, including plasmids[143] and viruses.[144,145] Finally, heterogeneity in population genomic data can identify fast evolutionary modes and uncover the exact integration sites for prophages and proviruses, simultaneously establishing host association for these elements.[146,147]

Overall, our insights into extrachromosomal elements are improving at a rapid pace and cover more than historically well-recognized ones such as viruses and plasmids. The recent description of "Borgs," which can have over 1 Mbp linear genomes, contain large inventories of metabolically relevant genes, and are not simply classifiable as plasmids or viruses,[148] "obelisks," which are ~1 kbp viroid-like extrachromosomal circular RNAs that are prevalent in gut metagenomes and metatranscriptomes,[149] the recent expansion of "viroid-like circular

RNAs,"[150] the "phage-inducible chromosomal island-like elements" that demonstrate the extent of parasitism among mobile genetic elements[151] and the overall prevalence of poorly characterized phage satellites across bacterial genomes[152] indicate that there is more to be learned about these mysterious entities.

Discovery and validation of new genetic codes

Metagenomic sequencing has brought to light unexpected phenomena that present new dimensions of modern microbiology. One of these relates to the genetic code and its variability. The genetic code was long considered immutable, and, indeed, it is widely conserved in bacteria and archaea. However, partial sequencing of a ribosomal gene cluster in the *Mycoplasma capricolum* genome revealed that, as in some mitochondrial genomes, TGA (normally a stop codon) is read as tryptophan (genetic code 4).[153] Subsequently, metagenomics-based analyses indicated that Gracilibacteria and Absconditabacteria, sibling lineages within the CPR, have reassigned the TGA codon.[43] Translation as glycine,[154] confirmed using metaproteomics,[155] ushered in a third bacterial genetic code (code 25). These code variations appear to have arisen in single, rare, and ancient evolutionary events. To date, no analogous genome-wide codon reassignments have been reported in archaea, although this domain of life remains under-explored relative to bacteria and eukaryotes. Recently, a new computational approach, "Codetta," was developed to identify new instances of codon reassignment in the now vast genomic sequence databases.[156] Analyses revealed the first evidence for changes in the amino acid encoded by sense codon changes in bacteria. In combination, these studies broaden and deepen our understanding of the modes and frequency of genetic code evolution in prokaryotes.

The origin of the relatively (but with exceptions) stable modern genetic code has been investigated recently by Žagrović and colleagues, who found that, although mRNAs and their cognate proteins are different types of polymers, mRNAs and the proteins they encode display complementarity.[157] This is particularly evident if the molecules are unstructured (disordered), as would have been the case during evolution of the primordial genetic code. In other work potentially relevant to understanding of the origin of the genetic code, analysis of large sequence databases from across the tree of life and confirmed that frame-shifted proteins have similar physicochemical properties (e.g., hydrophobicity, nucleobase affinity, and intrinsic disorder) to the original, non-frame-shifted proteins, despite essentially no primary sequence similarity.[158] It has long been known that similar codons encode amino acids with similar physicochemical properties and that the second codon position is most important for determining the properties of the amino acid encoded.[159] Thus, frame-shifted codons, which share two of three nucleotides, tend to incorporate related amino acids. Further insights into the evolution of genetic code come from integrated analyses of genomes and environmental selective pressures quantified through metagenomes. One such investigation by Shenhav and Zeevi shows that the genetic code optimizes codon assignments that minimize the impact of DNA mutations to dramatically change nutrient requirements of protein synthesis, revealing an overarching principle across the tree of life that enables relatively less risky exploration of the mutational landscape.[160]

Findings related to genetic codes of bacteriophages (phages) challenge the expectation that phages are unlikely to use a genetic code that differs from that of their hosts and the perception that genetic code changes are rare evolutionary events. In 2008, Shackelton and Holmes wrote "... it is extremely unlikely that viruses of hosts utilizing the universal genetic code would emerge, via cross-species transmission, in hosts utilizing alternative codes, and vice versa."[161] Consistent with this prediction, these authors confirmed that alternatively coded phages replicate in hosts that also use an alternative genetic code. However, Ivanova et al. reported that some phage genomes do in fact use a genetic code in which one of the three normal stop codons is read as an amino acid, yet they replicate in hosts that use the standard genetic code.[162] Recently, the reverse was suggested: standard code phages can replicate in bacteria that use a modified genetic code, specifically Gracilibacteria and Absconditabacteria, whose genomes that encode TGA as glycine.[163] Studies now show that code mismatch between phages and hosts is a relatively common phenomenon. For example, diverse alternatively coded phages replicate in Bacteroidetes and Firmicutes, bacteria that are particularly prevalent in human and animal microbiomes.[164] Combined taxonomic and genetic code analysis of a metagenomically derived database of alternatively coded phages revealed that these alternatively coded phages can be closely related to phages that use the standard code. In the most surprising case, one phage lineage features standard code as well as TGA-reassigned and TAG-reassigned genomes. Thus, contrary to evidence from prokaryote genomes, results for phages suggest that the barrier to code change is relatively low, thus codon reassignment can arise independently in different lineages and over relatively short evolutionary timescales.[164]

Indications that code shifts in phages are relatively facile raise the questions of the evolutionary path to code change and the machinery required for different codes to be interpreted by the ribosome during phage infection. Acquisition of a suppressor transfer RNA (tRNA) that translates a normal stop codon as an amino acid could occur via lateral gene transfer or, more likely, via mutation of the anticodon of a tRNA (possibly preceded by duplication of the tRNA).[165] Interestingly, to date, analyses of metagenome-derived phage genomes suggest charging of suppressor tRNAs by one of only three amino acids, suggesting convergent paths to these codon reassignments. The advantages of code mismatch remain unclear, but it may prevent premature production of structural proteins and proteins involved in host cell lysis.[164]

Another form of genetic code variation involves the reassignment of normal stop codons to non-standard amino acids. The first recognized example of this was the translation of TGA codon to selenocysteine in specific proteins, as signaled by a selenocysteine insertion sequence (SECIS) element, a specific structural mRNA feature.[166] This can substantially modify the properties (e.g., stronger nucleophilic ability and lower reduction potential) of proteins relative to the cysteine-containing homologs.[167] More recently, reassignment of the TAG stop to the 22nd amino acid, pyrrolysine, was discovered and, to date, found to be prevalent in methylamine methyltransferases and a few other proteins.[168]

Mapping biogeography to genes and genomes

Genomes have been reconstructed directly from metagenomes sampled from every major ecosystem, including microbiomes of humans, animals, insects, plants, soils, lakes, aquifers, and oceans. The large influx of genomes had an immediate impact on life sciences through broader access to predicted genes, functions, and metabolic potential of naturally occurring organisms. Surveys of traits that define the biogeographical patterns and niche partitioning of plants and animals go as far back as the eighteenth century in biology. However, a broad appreciation of the fact that microbes also display non-uniform biogeographical distribution patterns as a function of environmental gradients took time to come into focus.[169] Some of the most defining work emerged from the studies of microbial mats in Yellowstone National Park, United States. Initially using denaturing gradient gel electrophoresis to separate 16S rRNA gene fragments amplified from the environment, Ward and colleagues showed the astonishingly clear patterns of microbial population biogeography within one-millimeter slices of Mushroom Spring cyanobacterial mats.[170] These observations at the micron scale were confirmed by others at the scale of meters when Moore et al. showed that physiologically and phylogenetically distinct *Prochlorococcus* populations distribute across the water column as a function of light and temperature,[171] and across continents when Cho and Tiedje showed that genotypic differences between *Pseudomonas* populations correlated with geographic distances between soil samples from which they originated.[172] With the recognition of the impact of microbial traits, their ecology, and impact on biogeochemical systems,[173] the next challenge was to go from phylogenetic indicators of biogeography to traits that drive biogeographical partitioning of microbes.[174] The initial gene-centric analyses of short metagenomic reads[175] and longer genomic fragments recovered directly from environmental samples[176] quickly revealed non-uniform distribution patterns for genetic features that were associated with environment-specific metabolic requirements. The ability to characterize biogeography of microbial populations along with their genes was a step toward one of the holy grails of microbiology with significant biotechnological and biomedical implications: elucidating genetic determinants of fitness encoded in genomes, which benefited from ecology-enabled comparative genomics approaches that soon followed with the emergence of pangenomics.

Characterizing genetic heterogeneity in microbial populations

The first comparison of complete genomes of two closely related *Helicobacter pylori* strains revealed a large extent of gene conservancy and systematic differences[177] and the existence of "core" and "flexible" gene pools.[178] Research in the early 2000s took advantage of the increasing access to microbial genomes to reveal the utility of comparative genomics to gain insights into the likely determinants of microbial colonization and virulence,[179] or develop vaccines.[180] The idea to systematically study differentially occurring genetic features of closely related microbial populations came into fruition with the recognition of the microbial pangenomes[181] and evolutionary processes that shape them.[182,183] By partitioning the gene pool of a set of genomes, pangenomics offers insights into genes that are found

in the vast majority of genomes as well as those that differentially occur between them, which is an immense power to link gene function to population ecology.

Genomes can tell very little about their biogeography and even less about how well their gene content represents the gene content of environmental populations to which they belong. Pangenomics combined with shotgun metagenomes can uncover not only the biogeography of individual populations but also gene conservancy across habitats.[184] An application of this idea by Coleman and Chisholm to a *Prochlorococcus* pangenome and two marine metagenomes revealed that the scarcity of phosphorus in the North Atlantic Ocean resulted in a strict maintenance of accessory phosphorus acquisition genes for *Prochlorococcus* populations in the North Atlantic while their close relatives in the Pacific Ocean had largely lost them.[185] This integrated approach to pangenomics and metagenomics to identify selective forces that shape gene variation across populations and environments, also termed as metapangenomics,[186] has revealed differential enrichment of near-identical *Haemophilus parainfluenzae* populations that are millimeters apart from each other in human oral cavity, where populations enriched in tongue commonly encoded sodium-dependent oxaloacetate decarboxylase enzymes and those enriched in dental plaque completely lacked them[187] and differential distribution of transporters and enzymes within marine SAR324 populations to balance autotrophic and heterotrophic lifestyles as a function of water depth.[188] The metapangenomic approach can also be used to ensure critical genes validated experimentally in laboratory cultures were conserved in environmental populations of the same organism.[189]

The inclusion of ecological patterns into comparative genomics reduces the number of targets in our search for genes that may be particularly important under particular environmental conditions. However, difficulty with *de novo* prediction of gene function poses challenges. Efforts to shed light on the ecology of unknown genes,[190,191] profile hidden Markov model search algorithms,[192] application of deep learning to distant homology searches,[193] algorithms that enable rapid mining for protein structures,[83] and bringing in graph theory into pangenomes to infer synteny of genes[194,195] continue to improve insights into microbial gene pool.

Disentangling targets of long-term evolutionary processes

Even the genes that are shared across all members of a population can vary across individuals in subtle ways. Since the early descriptions of within-population trait variations in animals and plants by Darwin and Mendel, population genetics, the study of genetic variation among individual members of a population, has had a remarkable history in biology revealing molecular principles of evolution.[196] Even with limited access to sequencing, it was becoming clear during the 1990s and early 2000s that microbial organisms were not strictly clonal[197] and that the microbial world was a playground for evolution,[198] a powerful phenomenon that enables microbes to dynamically respond to environmental change within seasonal timescales.[199] Thus, elucidating the ways by which individuals of a single population differ from one anohter was essential to study fundamental principles of genetic innovation and divergence,[200] an ideal that

gained remarkable traction with the increasing availability of genomes and metagenomes.

Metagenomic datasets capture genomic heterogeneity of closely related populations and provide the data needed to identify genes under neutral, adaptive, or purifying forces. Indeed, the initial culture-independent observations of fine-scale population structures through microdiversity patterns found in clone libraries from marine systems[201] foreshadowed the monumental impact of shotgun metagenomics to study genome-wide spatiotemporal patterns of evolution. Identifying targets of distinct evolutionary processes in complex systems is an intimidating task, given that even in the most well-controlled laboratory systems conceivable, such as the long-term evolution experiment that exceeds 75,000 generations for 12 *Escherichia coli* populations,[202] presents us with overwhelming complexity.[203] Read recruitment from shotgun metagenomes using assembled contigs or reference genomes enables the characterization of genetic variants with access to exact allele frequencies at a single-nucleotide resolution.[204] Early studies that took advantage of shotgun metagenomics to study population genetics in relatively simple microbial systems observed the outcomes of transposase activity and purifying selection[146] and quantified genetic variability patterns throughout entire genomes of environmental populations.[205] In addition to increasing depth of metagenomic sequencing efforts, large-scale metagenomic surveys that yield public datasets with high-legacy[206,207] have been improving the application of microbial population genetics. More recent research efforts have elucidated variable rates of evolution for genes within individual microbial populations of the human gut,[208] patterns of gene- and genome-wide selective sweeps in lakes,[209] fate of subpopulations after human fecal microbiota transplantation procedures,[210] associations between microbial population structures and human biogeography,[211] location-specific purifying forces in deep-sea hydrothermal vents,[212] the impact of large-scale oceanic current temperatures on population structures of marine organisms,[213] the impact of anatomy on population structures on human skin,[214] and more.[215,216]

Many of the studies to date used subtle genetic variation in natural populations to infer to what extent divergence within microbial populations predicts or correlates with other environmental variables, such as differences in biogeography, various ecological measurements, or the health and disease states of organisms colonized by microbes. However, while such statistical inferences that treat every variant as equal are useful for broad ecological insights, they are not suitable to resolve the impact of individual variants on function or the evolutionary processes that maintain them. Increasing the utility of environmental surveys of genetic variation to advance biotechnological applications will undoubtedly benefit from consideration of protein structures,[217] given the sequence-structure-function paradigm[218] and our contemporary understanding of the direct relationships between protein sequence and function.[219] The well-understood importance of protein structures to interpret genomic variants[220] was recognized in the context of population genetics over two decades ago.[221] However, these ideals were delayed by challenges associated with predicting protein structures from amino acid sequences until recently.[222] Especially with the introduction of a new generation of deep learning frameworks that enable accurate predictions of protein structures, careful interpretations of variants observed in metagenomes are becoming more accessible through integrative 'omics platforms[223] for structure-guided mining of metagenomes.[224] A recent study by Kiefl et al. that linked protein structures and microbial population genetics demonstrated the non-uniform distribution of non-synonymous variants with respect to their relative solvent accessibility and distance to ligands and the rapidly decreasing rates of non-synonymous variants around ligand-binding sites of key metabolic genes as a function of bioavailable nitrogen concentrations in marine systems.[225] In another study, Han and Peng et al. observed variable selective pressures across the genomic features of organohalide-reducing taxa that occupy deep-sea cold seeps and demonstrated the presence of stronger purifying selective forces acting on buried residues of metabolically critical reductive dehalogenases in the environment.[226]

Elucidating how microbes respond to rapid environmental change

Environments change constantly, and survival depends on the ability to keep up. Of all organisms, microbes are unusual in that their evolutionary rates may be fast enough to respond to steady and relatively slow environmental shifts that take place in timescales of decades or longer. However, the conventional means of genome evolution is not suitable to respond to rapid changes that may occur within hours, days, or weeks. Thus, life has evolved many means to dynamically respond to ever shifting fitness landscapes, some of which are just coming to light.

Optimization of transcription and mechanisms behind transcriptional regulation is arguably one of the most well-understood means to respond to rapidly changing environmental conditions. Technological advances brought transcriptomics to fruition starting around the 1970s and continued at a rapid pace with the increasing access to sequencing technologies.[227] The first whole-genome characterization of the transcriptomic landscape of a microbe demonstrated that *Listeria monocytogenes* had remarkable transcriptional changes as a function of the environment in which they occurred.[228] With massive advances in single-cell transcriptomics, recent studies demonstrate responses such as spatial expression profiles in biofilms,[229] heterogeneous responses to antibiotics exposure,[230] and transcriptional responses to phage infection[231] at the level of individual cells across isogenic populations. Metatranscriptomics, the study of RNA molecules directly recovered from complex samples, was applied first to marine microbial populations[232] and had challenging beginnings due to the extremely low numbers of messenger RNAs per cell, short half-life of RNA molecules, lack of poly(A) tails that increase the signal-to-noise ratio in RNA libraries, and challenges associated with the quantification of RNA sequencing results.[233] Nevertheless, contemporary large-scale metatranscriptomics surveys have reached a level of maturity to track microbial responses to warming oceans[234] and elucidate disease-specific microbial responses in the human gut.[235]

Changing patterns of expression via transcriptional regulation is only one aspect of the genetic toolkit that rapidly optimizes microbial responses. Analyses of genomes and metagenomes revealed the pervasiveness of "phase variation" in bacteria

through inversion of regions in intergenic DNA to prevent the expression of costly genes in environments that do not require them.[236] Jiang and Hall et al. demonstrated that the same microbial populations that were engrafted to multiple individuals through fecal microbiota transplantation can acquire rapid and heritable changes in their promoter orientations within days, highlighting the speed by which microbes can optimize their gene use to respond to individual-specific selective forces without changing their gene content.[236] Another intriguing mechanism that enables microbes to change the efficiency of their existing genes is implemented by diversity-generating retroelements (DGRs).[237] DGRs are a family of genetic elements that generate hypervariability in target genes based on a template sequence where every adenosine is replaced with a random base prior to mutagenic homing thanks to a highly error-prone retron-type reverse transcriptase.[238] DGRs, which can generate a remarkable number of sequence variants in a single population within a few generations, were first discovered in a phage that infects *Bordetella*,[239] yet they are widespread in bacteria and archaea,[240] and the analyses of metagenomes show that they target a wide variety of functions, including those that are involved in protein binding, carbohydrate binding, and cell adhesion.[241] Within-population variants of DGR targets recovered from metagenomes can yield strong biogeographical partitioning,[242] suggesting that selective forces optimize the pool of random variants to match the fitness requirements imposed by environmental conditions.

Translational regulation is yet another layer of biology that enables cells from all domains of life to mount rapid responses to environmental change, this time without changing their genomes or patterns of transcription but by changing what is prioritized by the translational machinery. One of the essential components of the protein synthesis, tRNAs, play a significant role in this process through their relative abundances in the cell and chemical modifications that influence their interactions with the ribosome.[243,244] Cells can actively change their tRNA relative abundances to switch their metabolic state and prioritize the translation of immediate needs to respond to environmental conditions.[245] Phages and plasmids can also influence the tRNA pool as they carry large inventories of tRNAs[246] and tRNA modification enzymes[123] that are most likely driven by adaptive processes[247] that enable direct influence on translation.[248] Characterization of tRNA transcripts and their chemical modifications to study the dynamics of translational regulation is well established for cell populations that belong to the same organism, yet expanding these surveys into naturally occurring complex habitats is complicated by immense molecular and analytical challenges. For instance, due to their rigid secondary structures and chemical modifications, high-throughput sequencing of tRNA transcripts has been notoriously difficult.[249] But, improving molecular protocols promises broader access to tRNA sequencing[250] and new opportunities to study targets and implications of translational regulation also in complex systems. For instance, an application of tRNA sequencing to animal gut microbial populations revealed differences in taxon-specific tRNA chemical modification profiles as a function of diet, where highly modified tRNAs in high-fat diet condition resolved to codons that were highly en-riched in highly expressed proteins compared with low-fat diet condition,[251] and demonstrated the feasibility of this approach to track rapid changes in translational priorities and their genetic targets. Another promising approach to study translational responses of environmental microbial populations is ribosome profiling, a strategy that gives access to ribosome-protected RNA fragments that are being translated at a given time point.[252] An application of ribosome profiling to human gut microbial populations identified significant differences in transcripts and what was actually translated,[253] highlighting the need for highly resolved strategies that target critical components of translation to gain more complete insights into microbial responses to environmental conditions.

Insights into *in situ* microbial activity

From the perspective of ecological studies that seek to understand *in situ* processes, it is important to distinguish active versus inactive microbial community members, an ideal that has become more accessible in recent years. For example, bio-orthogonal non-canonical amino acid tagging (BONCAT) can visualize translationally active cells within complex environmental samples.[254] Another innovative approach uses fluorescence *in situ* hybridization (FISH) and nanoscale secondary ion mass spectrometry (FISH-nanoSIMS) in combination with ^{15}N stable isotope probing to link metabolic activity of cells to their location in complex consortia.[255,256] From the perspective of metagenomic studies, an important motivation for use of methods that can identify organisms that are active *in situ* is the recognition that a substantial fraction of DNA (e.g., in soil) is from dead organisms or is extracellular.[257] This may, in part, be a reflection of the fact that soils are incredibly biodiverse and display rapid shifts in community composition over the seasons. A key approach to identifying the subset of organisms that are active at any time is to leverage isotopes, which are selectively incorporated into proteins[258] or DNA[259] of actively growing organisms. Typically, stable isotope amendments are applied to samples of interest shortly before extraction of proteins or DNA. In the case of proteins, labeled versus unlabeled proteins (e.g., containing ^{15}N amended in the form of ammonia) are distinguished using highly accurate mass spectrometry measurements of peptides and their fragmentation productions.[260] In the case of DNA, fractions are separated by weight, and different fractions are sequenced independently. Implementation examples include the growth of plants in a $^{13}CO_2$ atmosphere and tracking of ^{13}C into the rhizosphere microbiome or addition of isotopically heavy water that is tracked into organisms that incorporate oxygen into *de novo* synthesized DNA. In the original implementation, referred to as quantitative stable isotope probing (qSIP), the heavy fraction (after consideration of the effect of genome GC content) is linked to actively growing organisms.[259] For example, one study used $H_2^{18}O$ addition and quantitative PCR of 16S rRNA genes to track growth and turnover of bacteria and fungi and thus quantified these contributions to the CO_2 flux from soil following the water addition to dry grassland soil.[261] Subsequently, DNA labeling was tracked into whole metagenomes, enabling stable isotope-informed genome-resolved metagenomics.[262] Despite the potential for artifacts when, for example, field-collected soil samples are necessarily incubated

with heavy water in the laboratory, stable isotope labeling is emerging as an important method for the study of microbial ecosystem dynamics.

Emerging opportunities enabled by genome editing of communities

Most current biological knowledge was generated via experiments in which a specific microbe was grown as an isolate (i.e., in a pure culture), and a certain gene was knocked out or added experimentally to determine its function, and much of it involved just one organism, *E. coli*. However, the tree of life is vast, and the suite of isolates available for traditional genetic manipulation is extremely limited. Even after functional assignments that leverage structural similarity, there are vast numbers of proteins for which functions cannot be predicted.

Why do microbes grow as members of communities rather than evolve into multicellular organisms such as trees or insects? The benefits may relate to adaptability to changing environmental conditions and likely come from flexibility in community membership and efficiencies gained via cooperative activities. As a consequence, microbes often depend on other organisms for key resources and devote genetic potential to communication and competition. These facets of microbial activity are understandably understudied because the majority of genetic and biochemical research conducted to date has been performed on organisms growing alone. In the absence of new methods that enable genetic manipulations of organisms without isolation, this is an essentially insurmountable problem. Understanding of the genes and pathways via which microbes interact with each other (and macroorganisms) requires the ability to perform "gold standard" genetic manipulation without the requirement for isolation. In other words, the genome editing of microbes while they exist in the context of their communities. With this approach, not only genes involved in interaction can be interrogated but also organisms that cannot be grown in isolation can become tractable for experimental manipulation.

If one is to edit a microbiome using CRISPR-Cas methods, there are several key requirements. First, stable, reproducible laboratory-based microbiomes are needed for method development and testing. Next, it is vital that the sequence that is targeted for modification is known. For this reason, genome-resolved metagenomics is foundational. The third challenge is to find a means by which the genome editing tools can be delivered into the targeted cells, and this is likely to be microbe-specific (e.g., conjugation and natural competence). To address this challenge, Rubin and others developed environmental transformation sequencing (ET-seq) and a DNA-editing all-in-one RNA-guided CRISPR-Cas transposase (DART), to enable organism- and locus-specific genetic manipulation.[263] Although direct editing of a human infant microbiome member was achieved, the editing efficiency was low and occurred in well-studied bacteria. Future delivery improvements are needed to target diverse organisms from little-studied branches of the tree of life before this potentially revolutionary approach can reach its potential.

Genetic manipulation in community context has been explored in the mouse microbiome, ultimately, with the goal of engineering communities *in situ*. The first step recovered engineered strains for reintroduction into the gut to alter overall community function. To do this, Ronda et al. developed metagenomic alteration of gut microbiome by *in situ* conjugation (MAGIC).[264] The method harnesses horizontal gene transfer to genetically modify gut bacteria. A vector introduces green fluorescent protein labeling to recipient cells that can then be cell sorted or introduces antibiotic resistance that can be selected for. Libraries included vectors bearing replication elements suitable for different gut bacteria. Among the challenges identified was long-term stability of the engineered genetic constructs, an issue that was somewhat addressed by isolation and engineering of host-derived strains that could mediate further transfer of engineered functions in the microbiome.

A recent study reported the important accomplishment of genetic editing of a population of episymbiotic Saccharibacteria in the context of a co-culture with host Actinobacteria.[265] The experiments harnessed the natural competence of the Saccharibacteria to insert heterologous sequences and targeted gene deletions. This enabled identification of novel Saccharibacterial genes as important for the growth of these cell surface-attached symbionts on their Actinobacteria host cells.

Ultimately, the ability to genetically modify strains of interest in microbiome context will make it possible to assign functions to a vast spectrum of novel genes and enable precise tuning of microbiome composition and function. By microbiome manipulation, a vast range of desirable outcomes of societal and environmental importance may be achieved.

A SHIFTING MINDSET TO ADDRESS THE CHALLENGES OF OUR MODERN ERA

A hallmark of modern microbiology is the plethora of new kinds of data generated from inconceivably intricate microbial systems as a whole with brand-new molecular and computational strategies. Taking full advantage of our observational powers demands us to focus on how to improve many of the well-established practices in microbiology, from how we responsibly disseminate data for transparency and reproducibility, or how we train life scientists to better exploit new data streams and computational resources, to how we encourage, support, and promote team science efforts to solve our immense need for integration when we mostly credit the first name and/or the last name that appear in final publications. Apart from such tangible issues that can be solved through tangible actions, some aspects of modern microbiology can also benefit from discussions regarding how the classical determinants of good scientific practice apply to this new era in which our observational powers far exceed our ability to interpret the data they yield. For instance, how critical is it to lead contemporary research in microbiology with the expectation of testable hypotheses as the first and foremost requirement to benchmark or implement new ideas? And how should we recognize, prioritize, or promote the synthesis of research products from laboratory experiments or data-driven explorations?

We believe it is important to recognize that hypothesis-driven research is inevitably limited by the ability of the human mind to imagine facets of microbiology that may be far from what we know today. Indeed, from the discovery of penicillin to insulin, from radioactivity to X-rays, history of science has a long list of

breakthroughs that emerged as unintended consequences of exploratory research and serendipity. For example, although the CRISPR-Cas system could theoretically be conceived of *de novo*, its accidental discovery through microbial genome sequencing provided a new window on virus-host interactions and ultimately opened the path for a revolution in biotechnology. Sequences from bacterial host CRISPR loci and coexisting viruses revealed infecting bacteriophage genomes as the source of the CRISPR spacer sequences,[266–268] but biochemical experiments were essential for uncovering mechanisms and thus enabling adaptation for targeted editing of eukaryotes,[269] with almost unimaginable impacts in medicine, agriculture, and basic science. This and other examples underline the opportunities that arise from exploratory research and its confluence with laboratory experimentation once the hypotheses reveal themselves from exploration.

In a world of unknowns, rigorous exploratory research provides the basis for hypothesis-driven research. Thus, given the huge scale and the microbial underpinnings of many of the problems of the modern era and the hope and potential for microbiome-based solutions, we advocate increasingly deploying the methods of microbiology in the context of ecosystems, whether they be the ocean, agricultural land, wetlands, municipal landfills, or others, and explore open-mindedly, and dare we say it, without asking for or testing any strict hypotheses. Only by developing a concrete understanding of how these ecosystems seem to work and only after conducting targeted, field-scale experiments within them will it be possible to tackle questions of utmost importance for the wellbeing of us and our planet, such as "how do we store more carbon in soil?," "how do we cut trace gas emissions from rice paddies?," "what can we do to reduce methane releases from cattle?," "how do we improve the quality of marginal drinking water, accelerate the breakdown of ocean plastic, reduce salinization of agricultural soils, and use the Earth's subsurface to store CO_2?," "how can we improve food biotechnologies to produce, for example, more palatable vegan meat alternatives?," "how do manipulation of diet, probiotics, or environmental exposures alter development of human microbiomes?," and "can changes in these factors address the ever-growing list of microbiome-associated diseases?" For example, given that around 40% of the land surface is already highly manipulated in support of food production,[270] agricultural systems are a logical target to learn how microbiome functionality can be modulated to reduce greenhouse gas emissions.[271] Yet our ability to implement investigations of such magnitude in terrestrial, marine, or host-associated habitats requires a mindset that embraces natural complexity and remains open to experiments in such systems at large scale, with integrated analyses of readouts ranging from microbiome composition to biogeochemical fluxes to crop productivity or health status.

These considerations are particularly relevant to highly critical ecosystem-level questions such as "how will future climate scenarios alter labile and recalcitrant dissolved organic carbon pools in warming oceans and impact marine food webs and the atmosphere?," "how nitrous oxide emissions from grassland soil will be altered by shifts in rainfall amount and timing or temperature driven by climate change?," or "how does soil recover after severe wildfires and how the patchiness of burned areas

impacts the reestablishment of healthy soil microbiomes?" For questions of this magnitude, harnessing natural events and variation over temporal and geographical gradients may have bigger potential to improve our understanding of key processes and guide modeling efforts to predict ecosystem behavior compared with artificial or small-scale controlled disruptions to test hypotheses. In other words, sometimes there are reasons to let nature do the experiment.

Our review highlights the motivation, progress, and potential of the data-driven epoch of microbiology to tackle systems of great complexity with vast scales of unknowns. A critical admission in our search for solutions to the most pressing needs of our planet is that the complexity of microbial systems that modern microbiology aims to elucidate may not always be compatible with the desire for hypothesis-driven research. Thus, from the level of PhD committees to funding agencies, a strict requirement for "hypotheses" should not be a critical determinant of the fate of research proposals. While we should not hold data products to a lesser degree of scrutiny, we also should not trivialize data-driven investigations at the level of grant panels or editorial decisions. In the quest to understand how nature works, it is important to embrace complexity while looking for means to reveal its driving principles.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Watterson, W.J., Tanyeri, M., Watson, A.R., Cham, C.M., Shan, Y., Chang, E.B., Eren, A.M., and Tay, S. (2020). Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes. eLife *9*, e56998.

2. Welch, J.L.M., Rossetti, B.J., Rieken, C.W., Dewhirst, F.E., and Borisy, G.G. (2016). Biogeography of a human oral microbiome at the micron scale. Proc. Natl. Acad. Sci. USA *113*, E791–E800.

3. Carroll, B.L., Nishikino, T., Guo, W., Zhu, S., Kojima, S., Homma, M., and Liu, J. (2020). The flagellar motor of Vibrio alginolyticus undergoes major structural remodeling during rotational switching. eLife *9*, e61446.

4. Quinn, A., El Chazli, Y., Escrig, S., Daraspe, J., Neuschwander, N., McNally, A., Genoud, C., Meibom, A., and Engel, P. (2024). Host-derived organic acids enable gut colonization of the honey bee symbiont Snodgrassella alvi. Nat. Microbiol. *9*, 477–489.

5. Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., Horreard, F., Amann, R., Jørgensen, B.B., and Kuypers, M.M.M. (2008). A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. Proc. Natl. Acad. Sci. USA *105*, 17861–17866.

6. Moran, M.A., Kujawinski, E.B., Schroer, W.F., Amin, S.A., Bates, N.R., Bertrand, E.M., Braakman, R., Brown, C.T., Covert, M.W., Doney, S.C.,

et al. (2022). Microbial metabolites in the marine carbon cycle. Nat. Microbiol. *7*, 508–523.

7. Cross, K.L., Campbell, J.H., Balachandran, M., Campbell, A.G., Cooper, C.J., Griffen, A., Heaton, M., Joshi, S., Klingeman, D., Leys, E., et al. (2019). Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. Nat. Biotechnol. *37*, 1314–1321.

8. Stepanauskas, R., Fergusson, E.A., Brown, J., Poulton, N.J., Tupper, B., Labonté, J.M., Becraft, E.D., Brown, J.M., Pachiadaki, M.G., Povilaitis, T., et al. (2017). Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat. Commun. *8*, 84.

9. Jin, W.-B., Li, T.-T., Huo, D., Qu, S., Li, X.V., Arifuzzaman, M., Lima, S.F., Shi, H.-Q., Wang, A., Putzel, G.G., et al. (2022). Genetic manipulation of gut microbes enables single-gene interrogation in a complex microbiome. Cell *185*, 547–562.e22.

10. Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. Nat. Rev. Microbiol. *12*, 181–196.

11. Hutchison, C.A., 3rd, Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., et al. (2016). Design and synthesis of a minimal bacterial genome. Science *351*, aad6253.

12. Lu, Y., Wu, K., Jiang, Y., Guo, Y., and Desneux, N. (2012). Widespread adoption of Bt cotton and insecticide decrease promotes biocontrol services. Nature *487*, 362–365.

13. Kaur, R., Leigh, B.A., Ritchie, I.T., and Bordenstein, S.R. (2022). The Cif proteins from Wolbachia prophage WO modify sperm genome integrity to establish cytoplasmic incompatibility. PLoS Biol. *20*, e3001584.

14. Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. *68*, 669–685.

15. Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738.

16. Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. J. Theor. Biol. *8*, 357–366.

17. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

18. Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA *74*, 5088–5090.

19. Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA *82*, 6955–6959.

20. Pace, N.R., Stahl, D.A., Lane, D.J., and Olsen, G.J. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. In Advances in Microbial Ecology, K.C. Marshall, ed. (Springer), pp. 1–55.

21. Böttger, E.C. (1989). Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. FEMS Microbiol. Lett. *53*, 171–176.

22. DeLong, E.F., and Pace, N.R. (2001). Environmental diversity of bacteria and archaea. Syst. Biol. *50*, 470–478.

23. López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., and Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature *409*, 603–607.

24. Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere.". Proc. Natl. Acad. Sci. USA *103*, 12115–12120.

25. Fierer, N., and Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. Proc. Natl. Acad. Sci. USA *103*, 626–631.

26. Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. USA *102*, 11070–11075.

27. Blaser Martin, J., Cardon Zoe, G., Cho Mildred, K., Dangl Jeffrey, L., Donohue Timothy, J., Green Jessica, L., Knight, R., Maxon Mary, E., Northen Trent, R., Pollard Katherine, S., et al. (2016). Toward a Predictive Understanding of Earth's microbiomes to Address 21st century Challenges. mBio *7*, e00714–e00716.

28. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., and Merrick, J.M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science *269*, 496–512.

29. Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., et al. (2016). A new view of the tree of life. Nat. Microbiol. *1*, 16048.

30. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. *5*, R245–R249.

31. Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J. Bacteriol. *178*, 591–599.

32. Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., et al. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science *289*, 1902–1906.

33. Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., and Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA *99*, 14250–14255.

34. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature *428*, 37–43.

35. Tamaki, H. (2019). Cultivation renaissance in the post-metagenomics era: combining the new and old. Microbes Environ. *34*, 117–120.

36. Tyson, G.W., and Banfield, J.F. (2005). Cultivating the uncultivated: a community genomics perspective. Trends Microbiol. *13*, 411–415.

37. Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. (2005). Genome-directed isolation of the key nitrogen fixer Leptospirillum ferrodiazotrophum sp. nov. from an acidophilic microbial community. Appl. Environ. Microbiol. *71*, 6319–6324.

38. Omsland, A., Cockrell, D.C., Howe, D., Fischer, E.R., Virtaneva, K., Sturdevant, D.E., Porcella, S.F., and Heinzen, R.A. (2009). Host cell-free growth of the Q fever bacterium Coxiella burnetii. Proc. Natl. Acad. Sci. USA *106*, 4430–4434.

39. Maurin, M., and Raoult, D. (1999). Q fever. Clin. Microbiol. Rev. *12*, 518–553.

40. Pope, P.B., Smith, W., Denman, S.E., Tringe, S.G., Barry, K., Hugenholtz, P., McSweeney, C.S., McHardy, A.C., and Morrison, M. (2011). Isolation of Succinivibrionaceae implicated in low methane emissions from Tammar wallabies. Science *333*, 646–648.

41. Lugli, G.A., Milani, C., Duranti, S., Alessandri, G., Turroni, F., Mancabelli, L., Tatoni, D., Ossiprandi, M.C., van Sinderen, D., and Ventura, M. (2019). Isolation of novel gut bifidobacteria using a combination of metagenomic and cultivation approaches. Genome Biol. *20*, 96.

42. Karnachuk, O.V., Lukina, A.P., Kadnikov, V.V., Sherbakova, V.A., Beletsky, A.V., Mardanov, A.V., and Ravin, N.V. (2021). Targeted isolation based on metagenome-assembled genomes reveals a phylogenetically distinct group of thermophilic spirochetes from deep biosphere. Environ. Microbiol. *23*, 3585–3598.

43. Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science *337*, 1661–1665.

44. He, X., McLean, J.S., Edlund, A., Yooseph, S., Hall, A.P., Liu, S.-Y., Dorrestein, P.C., Esquenazi, E., Hunter, R.C., Cheng, G., et al. (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc. Natl. Acad. Sci. USA *112*, 244–249.

45. Luef, B., Frischkorn, K.R., Wrighton, K.C., Holman, H.-Y.N., Birarda, G., Thomas, B.C., Singh, A., Williams, K.H., Siegerist, C.E., Tringe, S.G., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. Nat. Commun. *6*, 6372.

46. Moreira, D., Zivanovic, Y., López-Archilla, A.I., Iniesto, M., and López-García, P. (2021). Reductive evolution and unique predatory mode in the CPR bacterium Vampirococcus lugosii. Nat. Commun. *12*, 2454.

47. Kuroda, K., Nakajima, M., Nakai, R., Hirakata, Y., Kagemasa, S., Kubota, K., Noguchi, T.Q.P., Yamamoto, K., Satoh, H., Nobu, M.K., et al. (2024). Microscopic and metatranscriptomic analyses revealed unique cross-domain parasitism between phylum Candidatus Patescibacteria/candidate phyla radiation and methanogenic archaea in anaerobic ecosystems. mBio *15*, e0310223.

48. Bouderka, F., López-García, P., Deschamps, P., Krupovic, M., Gutiérrez-Preciado, A., Ciobanu, M., Bertolino, P., David, G., Moreira, D., and Jardillier, L. (2024). Culture- and genome-based characterization of a tripartite interaction between patescibacterial epibionts, methylotrophic proteobacteria, and a jumbo phage in freshwater ecosystems. Preprint at bioRxiv. https://doi.org/10.1101/2024.03.08.584096.

49. Rappé, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature *418*, 630–633.

50. Kaeberlein, T., Lewis, K., and Epstein, S.S. (2002). Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science *296*, 1127–1129.

51. Garcia, S.L. (2016). Mixed cultures as model communities: hunting for ubiquitous microorganisms, their partners, and interactions. Aquat. Microb. Ecol. *77*, 79–85.

52. Jiang, C.-Y., Dong, L., Zhao, J.-K., Hu, X., Shen, C., Qiao, Y., Zhang, X., Wang, Y., Ismagilov, R.F., Liu, S.-J., et al. (2016). High-throughput single-cell cultivation on microfluidic streak plates. Appl. Environ. Microbiol. *82*, 2210–2218.

53. Huang, Y., Sheth, R.U., Zhao, S., Cohen, L.A., Dabaghi, K., Moody, T., Sun, Y., Ricaurte, D., Richardson, M., Velez-Cortes, F., et al. (2023). High-throughput microbial culturomics using automation and machine learning. Nat. Biotechnol. *41*, 1424–1433.

54. Ma, L., Kim, J., Hatzenpichler, R., Karymov, M.A., Hubert, N., Hanan, I.M., Chang, E.B., and Ismagilov, R.F. (2014). Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in Human Microbiome Project's Most Wanted taxa. Proc. Natl. Acad. Sci. USA *111*, 9768–9773.

55. Fodor, A.A., DeSantis, T.Z., Wylie, K.M., Badger, J.H., Ye, Y., Hepburn, T., Hu, P., Sodergren, E., Liolios, K., Huot-Creasy, H., et al. (2012). The "most wanted" taxa from the human microbiome for whole genome sequencing. PLoS One *7*, e41294.

56. Boedicker, J.Q., Vincent, M.E., and Ismagilov, R.F. (2009). Microfluidic confinement of single cells of bacteria in small volumes initiates high-density behavior of quorum sensing and growth and reveals its variability. Angew. Chem. Int. Ed. Engl. *48*, 5908–5911.

57. Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature *521*, 173–179.

58. Eme, L., Spang, A., Lombard, J., Stairs, C.W., and Ettema, T.J.G. (2017). Archaea and the origin of eukaryotes. Nat. Rev. Microbiol. *15*, 711–723.

59. Castelle, C.J., and Banfield, J.F. (2018). Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell *172*, 1181–1197.

60. Ding, S., Hamm, J.N., Bale, N.J., Sinninghe Damsté, J.S., and Spang, A. (2024). Selective lipid recruitment by an archaeal DPANN symbiont from its host. Nat. Commun. *15*, 3405.

61. Witwinowski, J., Sartori-Rupp, A., Taib, N., Pende, N., Tham, T.N., Poppleton, D., Ghigo, J.-M., Beloin, C., and Gribaldo, S. (2022). An ancient divide in outer membrane tethering systems in bacteria suggests a mechanism for the diderm-to-monoderm transition. Nat. Microbiol. *7*, 411–422.

62. Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A.M., and Banfield, J.F. (2020). Accurate and complete genomes from metagenomes. Genome Res. *30*, 315–333.

63. Shaiber, A., and Eren, A.M. (2019). Composite metagenome-assembled genomes reduce the quality of public genome repositories. mBio *10*, e00725-19.

64. Orakov, A., Fullam, A., Coelho, L.P., Khedkar, S., Szklarczyk, D., Mende, D.R., Schmidt, T.S.B., and Bork, P. (2021). GUNC: detection of chimerism and contamination in prokaryotic genomes. Genome Biol. *22*, 178.

65. Sacristán-Horcajada, E., González-de la Fuente, S., Peiró-Pastor, R., Carrasco-Ramiro, F., Amils, R., Requena, J.M., Berenguer, J., and Aguado, B. (2021). ARAMIS: from systematic errors of NGS long reads to accurate assemblies. Brief. Bioinform. *22*, bbab170.

66. Wick, R.R., and Holt, K.E. (2022). Polypolish: short-read polishing of long-read bacterial genome assemblies. PLoS Comput. Biol. *18*, e1009802.

67. Hackl, T., Trigodet, F., Eren, A.M., Biller, S.J., Eppley, J.M., Luo, E., Burger, A., DeLong, E.F., and Fischer, M.G. (2021). proovframe: frame-shift-correction for long-read (meta)genomics. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.23.457338.

68. Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., et al. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat. Methods *17*, 1103–1110.

69. Feng, X., Cheng, H., Portik, D., and Li, H. (2022). Metagenome assembly of high-fidelity long reads with hifiasm-meta. Nat. Methods *19*, 671–674.

70. Benoit, G., Raguideau, S., James, R., Phillippy, A.M., Chikhi, R., and Quince, C. (2024). High-quality metagenome assembly from long accurate reads with metaMDBG. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-01983-6.

71. Méheust, R., Castelle, C.J., Jaffe, A.L., and Banfield, J.F. (2022). Conserved and lineage-specific hypothetical proteins may have played a central role in the rise and diversification of major archaeal groups. BMC Biol. *20*, 154.

72. Rohwer, F., and Edwards, R. (2002). The phage proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. *184*, 4529–4535.

73. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500.

74. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876.

75. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science *379*, 1123–1130.

76. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

77. Munsamy, G., Bohnuud, T., and Lorenz, P. (2024). Improving ALPHAFOLD2 performance with a global metagenomic & Biological Data supply chain. Preprint at bioRxiv. https://doi.org/10.1101/2024.03.06.583325.

78. Ovchinnikov, S., Park, H., Varghese, N., Huang, P.-S., Pavlopoulos, G.A., Kim, D.E., Kamisetty, H., Kyrpides, N.C., and Baker, D. (2017). Protein structure determination using metagenome sequence data. Science *355*, 294–298.

79. Flamholz, Z.N., Biller, S.J., and Kelly, L. (2024). Large language models improve annotation of prokaryotic viral proteins. Nat. Microbiol. *9*, 537–549.

80. Xiang, G., Li, Y., Sun, J., Huo, Y., Cao, S., Cao, Y., Guo, Y., Yang, L., Cai, Y., Zhang, Y.E., et al. (2024). Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. Nat. Biotechnol. *42*, 745–757.

81. Heinzinger, M., Weissenow, K., Sanchez, J.G., Henkel, A., Steinegger, M., and Rost, B. (2023). ProstT5: bilingual language model for protein sequence and structure. bioRxiv. https://doi.org/10.1101/2023.07.23.550085.

82. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682.

83. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. Nat. Biotechnol. *42*, 243–246.

84. Prochazka, D., Slaninakova, T., Olha, J., Rosinec, A., Gresova, K., Janosova, M., Cillik, J., Porubska, J., Svobodova, R., Dohnal, V., et al. (2024). AlphaFind: discover structure similarity across the entire known proteome. Preprint at bioRxiv. https://doi.org/10.1101/2024.02.15.580465.

85. Borkakoti, N., and Thornton, J.M. (2023). AlphaFold2 protein structure prediction: implications for drug discovery. Curr. Opin. Struct. Biol. *78*, 102526.

86. Shor, B., and Schneidman-Duhovny, D. (2024). CombFold: predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2. Nat. Methods *21*, 477–487.

87. Krupovic, M., Dolja, V.V., and Koonin, E.V. (2020). The LUCA and its complex virome. Nat. Rev. Microbiol. *18*, 661–670.

88. Koonin, E.V., and Martin, W. (2005). On the origin of genomes and cells within inorganic compartments. Trends Genet. *21*, 647–654.

89. Forterre, P. (2013). The virocell concept and environmental microbiology. ISME J. *7*, 233–236.

90. Rosenwasser, S., Ziv, C., van Creveld, S.G.V., and Vardi, A. (2016). Virocell metabolism: metabolic innovations during host-virus interactions in the ocean. Trends Microbiol. *24*, 821–832.

91. Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. Nature *399*, 541–548.

92. LePage, D.P., Metcalf, J.A., Bordenstein, S.R., On, J., Perlmutter, J.I., Shropshire, J.D., Layton, E.M., Funkhouser-Jones, L.J., Beckmann, J.F., and Bordenstein, S.R. (2017). Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. Nature *543*, 243–247.

93. Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ *3*, e985.

94. Kieft, K., Zhou, Z., and Anantharaman, K. (2020). VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome *8*, 90.

95. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P.S.G., Nayfach, S., and Kyrpides, N.C. (2024). Identification of mobile genetic elements with geNomad. Nat. Biotechnol. *42*, 1303–1312.

96. Camarillo-Guerrero, L.F., Almeida, A., Rangel-Pineros, G., Finn, R.D., and Lawley, T.D. (2021). Massive expansion of human gut bacteriophage diversity. Cell *184*, 1098–1109.e9.

97. Gregory, A.C., Zayed, A.A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., et al. (2019). Marine DNA viral macro- and microdiversity from pole to pole. Cell *177*, 1109–1123.e14.

98. Ma, B., Wang, Y., Zhao, K., Stirling, E., Lv, X., Yu, Y., Hu, L., Tang, C., Wu, C., Dong, B., et al. (2024). Biogeographic patterns and drivers of soil viromes. Nat. Ecol. Evol. *8*, 717–728.

99. Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., and Hill, C. (2020). Expansion of known ssRNA phage genomes: from tens to over a thousand. Sci. Adv. *6*, eaay5981.

100. Crits-Christoph, A., Kantor, R.S., Olm, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C., Flamholz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., et al. (2021). Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. mBio *12*, e02703–e02720. https://doi.org/10.1128/mBio.02703-20.

101. Abrahão, J., Silva, L., Silva, L.S., Khalil, J.Y.B., Rodrigues, R., Arantes, T., Assis, F., Boratto, P., Andrade, M., Kroon, E.G., et al. (2018). Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. Nat. Commun. *9*, 749.

102. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J.-M. (2004). The 1.2-megabase genome sequence of Mimivirus. Science *306*, 1344–1350.

103. Schulz, F., Abergel, C., and Woyke, T. (2022). Giant virus biology and diversity in the era of genome-resolved metagenomics. Nat. Rev. Microbiol. *20*, 721–736.

104. Moniruzzaman, M., Martinez-Gutierrez, C.A., Weinheimer, A.R., and Aylward, F.O. (2020). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat. Commun. *11*, 1710.

105. Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D.A., Denef, V.J., McMahon, K.D., Konstantinidis, K.T., Eloe-Fadrosh, E.A., Kyrpides, N.C., et al. (2020). Giant virus diversity and host interactions through global metagenomics. Nature *578*, 432–436.

106. Moniruzzaman, M., Weinheimer, A.R., Martinez-Gutierrez, C.A., and Aylward, F.O. (2020). Widespread endogenization of giant viruses shapes genomes of green algae. Nature *588*, 141–145.

107. Koonin, E.V., Dolja, V.V., Krupovic, M., Varsani, A., Wolf, Y.I., Yutin, N., Zerbini, F.M., and Kuhn, J.H. (2020). Global organization and proposed megataxonomy of the virus world. Microbiol. Mol. Biol. Rev. *84*, e00061-19. https://doi.org/10.1128/MMBR.00061-19.

108. Siddell, S.G., Smith, D.B., Adriaenssens, E., Alfenas-Zerbini, P., Dutilh, B.E., Garcia, M.L., Junglen, S., Krupovic, M., Kuhn, J.H., Lambert, A.J., et al. (2023). Virus taxonomy and the role of the International Committee on Taxonomy of Viruses (ICTV). J. Gen. Virol. *104*, 001840. https://doi.org/10.1099/jgv.0.001840.

109. Gaïa, M., Meng, L., Pelletier, E., Forterre, P., Vanni, C., Fernandez-Guerra, A., Jaillon, O., Wincker, P., Ogata, H., Krupovic, M., et al. (2023). Mirusviruses link herpesviruses to giant viruses. Nature *616*, 783–789.

110. Strathdee, S.A., Hatfull, G.F., Mutalik, V.K., and Schooley, R.T. (2023). Phage therapy: from biological mechanisms to future directions. Cell *186*, 17–31.

111. Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C.A., Li, Z., Cress, B.F., Knott, G.J., Jacobsen, S.E., Banfield, J.F., and Doudna, J.A. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. Science *369*, 333–337.

112. Selle, K., Fletcher, J.R., Tuson, H., Schmitt, D.S., McMillan, L., Vridhambal, G.S., Rivera, A.J., Montgomery, S.A., Fortier, L.-C., Barrangou, R., et al. (2020). In vivo Targeting of Clostridioides difficile Using Phage-Delivered CRISPR-Cas3 antimicrobials. mBio *11*, e00019–e00020. https://doi.org/10.1128/mBio.00019-20.

113. Nethery, M.A., Hidalgo-Cantabrana, C., Roberts, A., and Barrangou, R. (2022). CRISPR-based engineering of phages for in situ bacterial base editing. Proc. Natl. Acad. Sci. USA *119*, e2206744119.

114. Gencay, Y.E., Jasinskytė, D., Robert, C., Semsey, S., Martínez, V., Petersen, A.Ø., Brunner, K., de Santiago Torio, A., Salazar, A., Turcu, I.C., et al. (2024). Engineered phage with antibacterial CRISPR-Cas selectively reduce E. coli burden in mice. Nat. Biotechnol. *42*, 265–274.

115. Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from Mimivirus. Nat. Rev. Microbiol. *6*, 315–319.

116. Moreira, D., and López-García, P. (2009). Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol. *7*, 306–311.

117. Koonin, E.V., and Starokadomskyy, P. (2016). Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. Stud. Hist. Philos. Biol. Biomed. Sci. *59*, 125–134.

118. Lederberg, J. (1998). Plasmid (1952–1997). Plasmid *39*, 1–9.

119. Prazeres, D.M.F., and Monteiro, G.A. (2014). Plasmid biopharmaceuticals. Microbiol. Spectr. *2*. https://doi.org/10.1128/microbiolspec.PLAS-0022-2014.

120. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R.C., and San Millán, Á. (2021). Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. Nat. Rev. Microbiol. *19*, 347–359.

121. Prestinaci, F., Pezzotti, P., and Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. Pathog. Glob. Health *109*, 309–318.

122. Alawi, M., Torrijos, T.V., and Walsh, F. (2022). Plasmid-mediated antimicrobial resistance in drinking water. Environ. Adv. *8*, 100191.

123. Yu, M.K., Fogarty, E.C., and Eren, A.M. (2024). Diverse plasmid systems and their ecology across human gut metagenomes revealed by PlasX and MobMess. Nat. Microbiol. *9*, 830–847.

124. Pfeifer, E., and Rocha, E.P.C. (2024). Phage-plasmids promote recombination and emergence of phages and plasmids. Nat. Commun. *15*, 1545.

125. Erdmann, S., Tschitschko, B., Zhong, L., Raftery, M.J., and Cavicchioli, R. (2017). A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. Nat. Microbiol. *2*, 1446–1455.

126. Krupovic, M., Ravantti, J.J., and Bamford, D.H. (2009). Geminiviruses: a tale of a plasmid becoming a virus. BMC Evol. Biol. *9*, 112.

127. Reveillaud, J., Bordenstein, S.R., Cruaud, C., Shaiber, A., Esen, Ö.C., Weill, M., Makoundou, P., Lolans, K., Watson, A.R., Rakotoarivony, I., et al. (2019). The Wolbachia mobilome in Culex pipiens includes a putative plasmid. Nat. Commun. *10*, 1051.

128. Pellow, D., Mizrahi, I., and Shamir, R. (2020). PlasClass improves plasmid sequence classification. PLoS Comput. Biol. *16*, e1007781.

129. Robertson, J., and Nash, J.H.E. (2018). MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb. Genom. *4*, e000206.

130. Rozov, R., Brown Kav, A., Bogumil, D., Shterzer, N., Halperin, E., Mizrahi, I., and Shamir, R. (2017). Recycler: an algorithm for detecting plasmids from de novo assembly graphs. Bioinformatics *33*, 475–482.

131. Andreopoulos, W.B., Geller, A.M., Lucke, M., Balewski, J., Clum, A., Ivanova, N.N., and Levy, A. (2022). Deeplasmid: deep learning accurately separates plasmids from bacterial chromosomes. Nucleic Acids Res. *50*, e17.

132. Camargo, A.P., Call, L., Roux, S., Nayfach, S., Huntemann, M., Palaniappan, K., Ratner, A., Chu, K., Mukherjeep, S., Reddy, T.B.K., et al. (2024). IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. Nucleic Acids Res. *52*, D164–D173.

133. Pilosof, S. (2023). Conceptualizing microbe-plasmid communities as complex adaptive systems. Trends Microbiol. *31*, 672–680.

134. Simmonds, P., Adriaenssens, E.M., Zerbini, F.M., Abrescia, N.G.A., Aiewsakun, P., Alfenas-Zerbini, P., Bao, Y., Barylski, J., Drosten, C., Duffy, S., et al. (2023). Four principles to establish a universal virus taxonomy. PLoS Biol. *21*, e3001922.

135. Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E.P.C., Garcillán-Barcia, M.P., and de la Cruz, F. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. Nat. Commun. *11*, 3602.

136. Acman, M., van Dorp, L., Santini, J.M., and Balloux, F. (2020). Large-scale network analysis captures biological features of bacterial plasmids. Nat. Commun. *11*, 2452.

137. Harrison, P.W., Lower, R.P.J., Kim, N.K.D., and Young, J.P.W. (2010). Introducing the bacterial "chromid": not a chromosome, not a plasmid. Trends Microbiol. *18*, 141–148.

138. Fogarty, E.C., Schechter, M.S., Lolans, K., Sheahan, M.L., Veseli, I., Moore, R.M., Kiefl, E., Moody, T., Rice, P.A., Yu, M.K., et al. (2024). A cryptic plasmid is among the most numerous genetic elements in the human gut. Cell *187*, 1206–1222.e16.

139. Jarett, J.K., Džunková, M., Schulz, F., Roux, S., Paez-Espino, D., Eloe-Fadrosh, E., Jungbluth, S.P., Ivanova, N., Spear, J.R., Carr, S.A., et al. (2020). Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. ISME J. *14*, 2527–2541.

140. Diebold, P.J., New, F.N., Hovan, M., Satlin, M.J., and Brito, I.L. (2021). Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR. eLife *10*, e66834.

141. Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047–1050.

142. Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods *58*, 268–276.

143. Stalder, T., Press, M.O., Sullivan, S., Liachko, I., and Top, E.M. (2019). Linking the resistome and plasmidome to the microbiome. ISME J. *13*, 2437–2446.

144. Marbouty, M., Thierry, A., Millot, G.A., and Koszul, R. (2021). MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. eLife *10*, e60608.

145. Hwang, Y., Roux, S., Coclet, C., Krause, S.J.E., and Girguis, P.R. (2023). Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. Nat. Microbiol. *8*, 946–957.

146. Allen, E.E., Tyson, G.W., Whitaker, R.J., Detter, J.C., Richardson, P.M., and Banfield, J.F. (2007). Genome dynamics in a natural archaeal population. Proc. Natl. Acad. Sci. USA *104*, 1883–1888.

147. Chen, L.-X., Zhao, Y., McMahon, K.D., Mori, J.F., Jessen, G.L., Nelson, T.C., Warren, L.A., and Banfield, J.F. (2019). Wide distribution of phage that infect freshwater SAR11 bacteria. mSystems *4*, e00410–e00419. https://doi.org/10.1128/mSystems.00410-19.

148. Al-Shayeb, B., Schoelmerich, M.C., West-Roberts, J., Valentin-Alvarado, L.E., Sachdeva, R., Mullen, S., Crits-Christoph, A., Wilkins, M.J., Williams, K.H., Doudna, J.A., et al. (2022). Borgs are giant genetic elements with potential to expand metabolic capacity. Nature *610*, 731–736.

149. Zheludev, I.N., Edgar, R.C., Lopez-Galiano, M.J., de la Peña, M., Babaian, A., Bhatt, A.S., and Fire, A.Z. (2024). Viroid-like colonists of human microbiomes. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.20.576352.

150. Lee, B.D., Neri, U., Roux, S., Wolf, Y.I., Camargo, A.P., Krupovic, M., RNA Virus Discovery Consortium, Simmonds, P., Kyrpides, N., Gophna, U., et al. (2023). Mining metatranscriptomes reveals a vast world of viroid-like circular RNAs. Cell *186*, 646–661.e4.

151. Boyd, C.M., Subramanian, S., Dunham, D.T., Parent, K.N., and Seed, K.D. (2024). A Vibrio cholerae viral satellite maximizes its spread and inhibits phage by remodeling hijacked phage coat proteins into small capsids. eLife *12*, RP87611. https://doi.org/10.7554/eLife.87611.

152. de Sousa, J.A.M., Fillol-Salom, A., Penadés, J.R., and Rocha, E.P.C. (2023). Identification and characterization of thousands of bacteriophage satellites across bacteria. Nucleic Acids Res. *51*, 2759–2777.

153. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., and Osawa, S. (1985). UGA is read as tryptophan in Mycoplasma capricolum. Proc. Natl. Acad. Sci. USA *82*, 2306–2309.

154. Campbell, J.H., O'Donoghue, P., Campbell, A.G., Schwientek, P., Sczyrba, A., Woyke, T., Söll, D., and Podar, M. (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc. Natl. Acad. Sci. USA *110*, 5540–5545.

155. Hanke, A., Hamann, E., Sharma, R., Geelhoed, J.S., Hargesheimer, T., Kraft, B., Meyer, V., Lenk, S., Osmers, H., Wu, R., et al. (2014). Recoding of the stop codon UGA to glycine by a BD1-5/SN-2 bacterium and niche partitioning between Alpha- and Gammaproteobacteria in a tidal sediment microbial community naturally selected in a laboratory chemostat. Front. Microbiol. *5*, 231.

156. Shulgina, Y., and Eddy, S.R. (2021). A computational screen for alternative genetic codes in over 250,000 genomes. eLife *10*, e71402.

157. Zagrovic, B., Bartonek, L., and Polyansky, A.A. (2018). RNA-protein interactions in an unstructured context. FEBS Lett. *592*, 2901–2916.

158. Bartonek, L., Braun, D., and Zagrovic, B. (2020). Frameshifting preserves key physicochemical properties of proteins. Proc. Natl. Acad. Sci. USA *117*, 5907–5912.

159. Alff-Steinberger, C. (1969). The genetic code and error transmission. Proc. Natl. Acad. Sci. USA *64*, 584–591.

160. Shenhav, L., and Zeevi, D. (2020). Resource conservation manifests in the genetic code. Science *370*, 683–687.

161. Shackelton, L.A., and Holmes, E.C. (2008). The role of alternative genetic codes in viral evolution and emergence. J. Theor. Biol. *254*, 128–134.

162. Ivanova, N.N., Schwientek, P., Tripp, H.J., Rinke, C., Pati, A., Huntemann, M., Visel, A., Woyke, T., Kyrpides, N.C., and Rubin, E.M. (2014). Stop codon reassignments in the wild. Science *344*, 909–913.

163. Liu, J., Jaffe, A.L., Chen, L., Bor, B., and Banfield, J.F. (2023). Host translation machinery is not a barrier to phages that interact with both CPR and non-CPR bacteria. mBio *14*, e0176623.

164. Borges, A.L., Lou, Y.C., Sachdeva, R., Al-Shayeb, B., Penev, P.I., Jaffe, A.L., Lei, S., Santini, J.M., and Banfield, J.F. (2022). Widespread stop-codon recoding in bacteriophages may regulate translation of lytic genes. Nat. Microbiol. *7*, 918–927.

165. Massey, S.E. (2017). The identities of stop codon reassignments support ancestral tRNA stop codon decoding activity as a facilitator of gene duplication and evolution of novel function. Gene *619*, 37–43.

166. Fischer, N., Paleskava, A., Gromadski, K.B., Konevega, A.L., Wahl, M.C., Stark, H., and Rodnina, M.V. (2007). Towards understanding selenocysteine incorporation into bacterial proteins. Biol. Chem. *388*, 1061–1067.

167. Chung, C.Z., Miller, C., Söll, D., and Krahn, N. (2021). Introducing selenocysteine into recombinant proteins in Escherichia coli. Curr. Protoc. *1*, e54.

168. Li, J., Kang, P.T., Jiang, R., Lee, J.Y., Soares, J.A., Krzycki, J.A., and Chan, M.K. (2023). Insights into pyrrolysine function from structures of a trimethylamine methyltransferase and its corrinoid protein complex. Commun. Biol. *6*, 54.

169. Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., et al. (2006). Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. *4*, 102–112.

170. Ward, D.M., Ferris, M.J., Nold, S.C., and Bateson, M.M. (1998). A natural view of microbial biodiversity within Hot Spring cyanobacterial mat communities. Microbiol. Mol. Biol. Rev. *62*, 1353–1370.

171. Moore, L.R., Rocap, G., and Chisholm, S.W. (1998). Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. Nature *393*, 464–467.

172. Cho, J.C., and Tiedje, J.M. (2000). Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. Appl. Environ. Microbiol. *66*, 5448–5456.

173. Newman, D.K., and Banfield, J.F. (2002). Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. Science *296*, 1071–1077.

174. Green, J.L., Bohannan, B.J.M., and Whitaker, R.J. (2008). Microbial biogeography: from taxonomy to traits. Science *320*, 1039–1043.

175. Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., et al. (2005). Comparative metagenomics of microbial communities. Science *308*, 554–557.

176. DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R., et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. Science *311*, 496–503.

177. Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., et al. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. Nature *397*, 176–180.

178. Hacker, J., and Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep. *2*, 376–381.

179. Tettelin, H., Masignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., et al. (2002). Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. USA *99*, 12391–12396.

180. Maione, D., Margarit, I., Rinaudo, C.D., Masignani, V., Mora, M., Scarselli, M., Tettelin, H., Brettoni, C., Iacobini, E.T., Rosini, R., et al. (2005). Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science *309*, 148–150.

181. Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome.". Proc. Natl. Acad. Sci. USA *102*, 13950–13955.

182. Domingo-Sananes, M.R., and McInerney, J.O. (2021). Mechanisms that shape microbial pangenomes. Trends Microbiol. *29*, 493–503.

183. McInerney, J.O. (2023). Prokaryotic pangenomes act as evolving ecosystems. Mol. Biol. Evol. *40*, msac232.

184. Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science *344*, 416–420.

185. Coleman, M.L., and Chisholm, S.W. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. Proc. Natl. Acad. Sci. USA *107*, 18634–18639.

186. Delmont, T.O., and Eren, A.M. (2018). Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. PeerJ *6*, e4320.

187. Utter, D.R., Borisy, G.G., Eren, A.M., Cavanaugh, C.M., and Mark Welch, J.L. (2020). Metapangenomics of the oral microbiome provides insights into habitat adaptation and cultivar diversity. Genome Biol. *21*, 293.

188. Boeuf, D., Eppley, J.M., Mende, D.R., Malmstrom, R.R., Woyke, T., and DeLong, E.F. (2021). Metapangenomics reveals depth-dependent shifts in metabolic potential for the ubiquitous marine bacterial SAR324 lineage. Microbiome *9*, 172.

189. Bunker, J.J., Drees, C., Watson, A.R., Plunkett, C.H., Nagler, C.R., Schneewind, O., Eren, A.M., and Bendelac, A. (2019). B cell superantigens in the human intestinal microbiota. Sci. Transl. Med. *11*, eaau9356. https://doi.org/10.1126/scitranslmed.aau9356.

190. Vanni, C., Schechter, M.S., Acinas, S.G., Barberán, A., Buttigieg, P.L., Casamayor, E.O., Delmont, T.O., Duarte, C.M., Eren, A.M., Finn, R.D., et al. (2022). Unifying the known and unknown microbial coding sequence space. eLife *11*, e67667. https://doi.org/10.7554/eLife.67667.

191. Rodríguez Del Río, Á., Giner-Lamia, J., Cantalapiedra, C.P., Botas, J., Deng, Z., Hernández-Plaza, A., Munar-Palmer, M., Santamaría-Hernando, S., Rodríguez-Herva, J.J., Ruscheweyh, H.-J., et al. (2024). Functional and evolutionary significance of unknown genes from uncultivated taxa. Nature *626*, 377–384.

192. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods *9*, 173–175.

193. Hamamsy, T., Morton, J.T., Blackwell, R., Berenberg, D., Carriero, N., Gligorijevic, V., Strauss, C.E.M., Leman, J.K., Cho, K., and Bonneau, R. (2024). Protein remote homology detection and structural alignment using deep learning. Nat. Biotechnol. *42*, 975–985.

194. Gautreau, G., Bazin, A., Gachet, M., Planel, R., Burlot, L., Dubois, M., Perrin, A., Médigue, C., Calteau, A., Cruveiller, S., et al. (2020). PPanGGOLiN: depicting microbial diversity via a partitioned pangenome graph. PLoS Comput. Biol. *16*, e1007732.

195. Colquhoun, R.M., Hall, M.B., Lima, L., Roberts, L.W., Malone, K.M., Hunt, M., Letcher, B., Hawkey, J., George, S., Pankhurst, L., et al. (2021). Pandora: nucleotide-resolution bacterial pan-genomics with reference graphs. Genome Biol. *22*, 267.

196. Okazaki, A., Yamazaki, S., Inoue, I., and Ott, J. (2021). Population genetics: past, present, and future. Hum. Genet. *140*, 231–240.

197. Smith, J.M., Smith, N.H., O'Rourke, M., and Spratt, B.G. (1993). How clonal are bacteria? Proc. Natl. Acad. Sci. USA *90*, 4384–4388.

198. Cohan, F.M. (2001). Bacterial species and speciation. Syst. Biol. *50*, 513–524.

199. Rohwer, R.R., Kirkpatrick, M., Garcia, S.L., Kellom, M., McMahon, K.D., and Baker, B.J. (2024). Bacterial ecology and evolution converge on seasonal and decadal scales. Preprint at bioRxiv. https://doi.org/10.1101/2024.02.06.579087.

200. Cohan, F.M. (1994). Genetic exchange and evolutionary divergence in prokaryotes. Trends Ecol. Evol. *9*, 175–180.

201. Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. Nature *430*, 551–554.

202. Lenski, R.E. (2023). Revisiting the design of the long-term evolution experiment with Escherichia coli. J. Mol. Evol. *91*, 241–253.

203. Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., and Desai, M.M. (2017). The dynamics of molecular evolution over 60,000 generations. Nature *551*, 45–50.

204. Denef, V.J. (2018). Peering into the genetic makeup of natural microbial populations using metagenomics. In Population Genomics: Microorganisms, M.F. Polz and O.P. Rajora, eds. (Springer International Publishing), pp. 49–75.

205. Simmons, S.L., Dibartolo, G., Denef, V.J., Goltsman, D.S.A., Thelen, M.P., and Banfield, J.F. (2008). Population genomic analysis of strain variation in Leptospirillum group II bacteria involved in acid mine drainage formation. PLoS Biol. *6*, e177.

206. Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

207. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., et al. (2015). Ocean plankton. Structure and function of the global ocean microbiome. Science *348*, 1261359.

208. Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D.R., Kultima, J.R., Martin, J., et al. (2013). Genomic variation landscape of the human gut microbiome. Nature *493*, 45–50.

209. Bendall, M.L., Stevens, S.L., Chan, L.-K., Malfatti, S., Schwientek, P., Tremblay, J., Schackwitz, W., Martin, J., Pati, A., Bushnell, B., et al. (2016). Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. *10*, 1589–1601.

210. Li, S.S., Zhu, A., Benes, V., Costea, P.I., Hercog, R., Hildebrand, F., Huerta-Cepas, J., Nieuwdorp, M., Salojärvi, J., Voigt, A.Y., et al. (2016). Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science *352*, 586–589.

211. Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. *27*, 626–638.

212. Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M., Seewald, J.S., and Huber, J.A. (2017). Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. Nat. Commun. *8*, 1114.

213. Delmont, T.O., Kiefl, E., Kilinc, O., Esen, O.C., Uysal, I., Rappé, M.S., Giovannoni, S., and Eren, A.M. (2019). Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. eLife *8*, e46497. https://doi.org/10.7554/eLife.46497.

214. Conwill, A., Kuan, A.C., Damerla, R., Poret, A.J., Baker, J.S., Tripp, A.D., Alm, E.J., and Lieberman, T.D. (2022). Anatomy promotes neutral coexistence of strains in the human skin microbiome. Cell Host Microbe *30*, 171–182.e7.

215. Garud, N.R., and Pollard, K.S. (2020). Population genetics in the human microbiome. Trends Genet. *36*, 53–67.

216. Logares, R. (2024). Decoding populations in the ocean microbiome. Microbiome *12*, 67.

217. Wilke, C.O. (2012). Bringing molecules back into molecular evolution. PLoS Comput. Biol. *8*, e1002572.

218. Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223–230.

219. Park, Y., Metzger, B.P.H., and Thornton, J.W. (2024). The simplicity of protein sequence-function relationships. Preprint at bioRxiv. https://doi.org/10.1101/2023.09.02.556057.

220. Dean, A.M., and Thornton, J.W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. Nat. Rev. Genet. *8*, 675–688.

221. Sunyaev, S., Lathe, W., 3rd, and Bork, P. (2001). Integration of genome data and protein structures: prediction of protein folds, protein interactions and "molecular phenotypes" of single nucleotide polymorphisms. Curr. Opin. Struct. Biol. *11*, 125–130.

222. Kuhlman, B., and Bradley, P. (2019). Advances in protein structure prediction and design. Nat. Rev. Mol. Cell Biol. *20*, 681–697.

223. Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. Nat. Microbiol. *6*, 3–6.

224. Robinson, S.L. (2023). Structure-guided metagenome mining to tap microbial functional diversity. Curr. Opin. Microbiol. *76*, 102382.

225. Kiefl, E., Esen, O.C., Miller, S.E., Kroll, K.L., Willis, A.D., Rappé, M.S., Pan, T., and Eren, A.M. (2023). Structure-informed microbial population genetics elucidate selective pressures that shape protein evolution. Sci. Adv. *9*, eabq4632.

226. Han, Y., Peng, Y., Peng, J., Cao, L., Xu, Y., Yang, Y., Wu, M., Zhou, H., Zhang, C., Zhang, D., et al. (2024). Phylogenetically and structurally diverse reductive dehalogenases link biogeochemical cycles in deep-sea cold seeps. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.23.576788.

227. Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. Nat. Methods *19*, 534–546.

228. Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., et al. (2009). The Listeria transcriptional landscape from saprophytism to virulence. Nature *459*, 950–956.

229. Dar, D., Dar, N., Cai, L., and Newman, D.K. (2021). Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. Science *373*, eabi4882. https://doi.org/10.1126/science.abi4882.

230. Ma, P., Amemiya, H.M., He, L.L., Gandhi, S.J., Nicol, R., Bhattacharyya, R.P., Smillie, C.S., and Hung, D.T. (2023). Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. Cell *186*, 877–891.e14.

231. Wang, B., Lin, A.E., Yuan, J., Novak, K.E., Koch, M.D., Wingreen, N.S., Adamson, B., and Gitai, Z. (2023). Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. Nat. Microbiol. *8*, 1846–1862.

232. Poretsky, R.S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A., and Hollibaugh, J.T. (2005). Analysis of microbial gene transcripts in environmental samples. Appl. Environ. Microbiol. *71*, 4121–4126.

233. Gifford, S.M., Sharma, S., Rinta-Kanto, J.M., and Moran, M.A. (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. ISME J. *5*, 461–472.

234. Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., Field, C.M., Coelho, L.P., Cruaud, C., Engelen, S., et al. (2019). Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell *179*, 1068–1083.e21.

235. Schirmer, M., Franzosa, E.A., Lloyd-Price, J., McIver, L.J., Schwager, R., Poon, T.W., Ananthakrishnan, A.N., Andrews, E., Barron, G., Lake, K., et al. (2018). Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat. Microbiol. *3*, 337–346.

236. Jiang, X., Hall, A.B., Arthur, T.D., Plichta, D.R., Covington, C.T., Poyet, M., Crothers, J., Moses, P.L., Tolonen, A.C., Vlamakis, H., et al. (2019). Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. Science *363*, 181–187.

237. Medhekar, B., and Miller, J.F. (2007). Diversity-generating retroelements. Curr. Opin. Microbiol. *10*, 388–395.

238. Naorem, S.S., Han, J., Wang, S., Lee, W.R., Heng, X., Miller, J.F., and Guo, H. (2017). DGR mutagenic transposition occurs via hypermutagenic reverse transcription primed by nicked template RNA. Proc. Natl. Acad. Sci. USA *114*, E10187–E10195.

239. Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J., et al. (2002). Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science *295*, 2091–2094.

240. Paul, B.G., Burstein, D., Castelle, C.J., Handa, S., Arambula, D., Czornyj, E., Thomas, B.C., Ghosh, P., Miller, J.F., Banfield, J.F., et al. (2017). Retroelement-guided protein diversification abounds in vast lineages of Bacteria and archaea. Nat. Microbiol. *2*, 17045.

241. Roux, S., Paul, B.G., Bagby, S.C., Nayfach, S., Allen, M.A., Attwood, G., Cavicchioli, R., Chistoserdova, L., Gruninger, R.J., Hallam, S.J., et al. (2021). Ecology and molecular targets of hypermutation in the global microbiome. Nat. Commun. *12*, 3076.

242. Paul, B.G., and Eren, A.M. (2022). Eco-evolutionary significance of domesticated retroelements in microbial genomes. Mobile DNA *13*, 6.

243. Mohler, K., and Ibba, M. (2017). Translational fidelity and mistranslation in the cellular response to stress. Nat. Microbiol. *2*, 17117.

244. Samatova, E., Daberger, J., Liutkute, M., and Rodnina, M.V. (2020). Translational control by ribosome pausing in bacteria: how a non-uniform pace of translation affects protein production and folding. Front. Microbiol. *11*, 619430.

245. Torrent, M., Chalancon, G., de Groot, N.S., Wuster, A., and Madan Babu, M. (2018). Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. Sci. Signal. *11*, eaat6409.

246. Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y., et al. (2020). Clades of huge phages from across Earth's ecosystems. Nature *578*, 425–431.

247. Yang, J.Y., Fang, W., Miranda-Sanchez, F., Brown, J.M., Kauffman, K.M., Acevero, C.M., Bartel, D.P., Polz, M.F., and Kelly, L. (2021). Degradation of host translational machinery drives tRNA acquisition in viruses. Cell Syst. *12*, 771–779.e5.

248. Bailly-Bechet, M., Vergassola, M., and Rocha, E. (2007). Causes for the intriguing presence of tRNAs in phages. Genome Res. *17*, 1486–1495.

249. Zhang, W., Foo, M., Eren, A.M., and Pan, T. (2022). tRNA modification dynamics from individual organisms to metaepitranscriptomics of microbiomes. Mol. Cell *82*, 891–906.

250. Padhiar, N.H., Katneni, U., Komar, A.A., Motorin, Y., and Kimchi-Sarfaty, C. (2024). Advances in methods for tRNA sequencing and quantification. Trends Genet. *40*, 276–290.

251. Schwartz, M.H., Wang, H., Pan, J.N., Clark, W.C., Cui, S., Eckwahl, M.J., Pan, D.W., Parisien, M., Owens, S.M., Cheng, B.L., et al. (2018). Microbiome characterization by high-throughput transfer RNA sequencing and modification analysis. Nat. Commun. *9*, 5353.

252. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

253. Fremin, B.J., Sberro, H., and Bhatt, A.S. (2020). MetaRibo-Seq measures translation in microbiomes. Nat. Commun. *11*, 3268.

254. Hatzenpichler, R., Scheller, S., Tavormina, P.L., Babin, B.M., Tirrell, D.A., and Orphan, V.J. (2014). In situ visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. Environ. Microbiol. *16*, 2568–2590.

255. Musat, N., Foster, R., Vagner, T., Adam, B., and Kuypers, M.M.M. (2012). Detecting metabolic activities in single cells, with emphasis on NanoSIMS. FEMS Microbiol. Rev. *36*, 486–511.

256. McGlynn, S.E., Chadwick, G.L., Kempes, C.P., and Orphan, V.J. (2015). Single cell activity reveals direct electron transfer in methanotrophic consortia. Nature *526*, 531–535.

257. Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S., and Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat. Microbiol. *2*, 16242.

258. VerBerkmoes, N.C., Denef, V.J., Hettich, R.L., and Banfield, J.F. (2009). Systems biology: functional analysis of natural microbial consortia using community proteomics. Nat. Rev. Microbiol. *7*, 196–205.

259. Hungate, B.A., Mau, R.L., Schwartz, E., Caporaso, J.G., Dijkstra, P., van Gestel, N., Koch, B.J., Liu, C.M., McHugh, T.A., Marks, J.C., et al. (2015). Quantitative microbial ecology through stable isotope probing. Appl. Environ. Microbiol. *81*, 7570–7581.

260. Pan, C., Fischer, C.R., Hyatt, D., Bowen, B.P., Hettich, R.L., and Banfield, J.F. (2011). Quantitative tracking of isotope flows in proteomes of microbial communities. Mol. Cell. Proteomics *10*, M110.006049.

261. Blazewicz, S.J., Hungate, B.A., Koch, B.J., Nuccio, E.E., Morrissey, E., Brodie, E.L., Schwartz, E., Pett-Ridge, J., and Firestone, M.K. (2020). Taxon-specific microbial growth and mortality patterns reveal distinct temporal population responses to rewetting in a California grassland soil. ISME J. *14*, 1520–1532.

262. Vyshenska, D., Sampara, P., Singh, K., Tomatsu, A., Kauffman, W.B., Nuccio, E.E., Blazewicz, S.J., Pett-Ridge, J., Louie, K.B., Varghese, N., et al. (2023). A standardized quantitative analysis strategy for stable isotope probing metagenomics. mSystems *8*, e0128022.

263. Rubin, B.E., Diamond, S., Cress, B.F., Crits-Christoph, A., Lou, Y.C., Borges, A.L., Shivram, H., He, C., Xu, M., Zhou, Z., et al. (2022). Species- and site-specific genome editing in complex bacterial communities. Nat. Microbiol. *7*, 34–47.

264. Ronda, C., Chen, S.P., Cabral, V., Yaung, S.J., and Wang, H.H. (2019). Metagenomic engineering of the mammalian gut microbiome in situ. Nat. Methods *16*, 167–170.

265. Wang, Y., Gallagher, L.A., Andrade, P.A., Liu, A., Humphreys, I.R., Turkarslan, S., Cutler, K.J., Arrieta-Ortiz, M.L., Li, Y., Radey, M.C., et al. (2023). Genetic manipulation of candidate phyla radiation bacteria provides functional insights into microbial dark matter. Preprint at bioRxiv. https://doi.org/10.1101/2023.05.02.539146.

266. Mojica, F.J., Juez, G., and Rodríguez-Valera, F. (1993). Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol. Microbiol. *9*, 613–621.

267. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551–2561.

268. Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653–663.

269. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

270. Ritchie, H., and Roser, M. (2024). Half of the world's habitable land is used for agriculture. Our World in Data. https://www.futureoffood.ox.ac.uk/article/half-of-the-worlds-habitable-land-is-used-for-agriculture.

271. Foley, J.A., Defries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., et al. (2005). Global consequences of land use. Science *309*, 570–574.