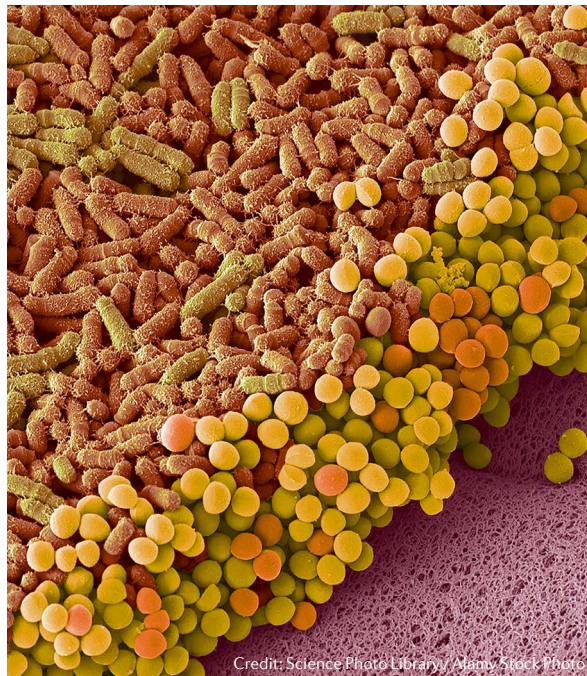


Sequencing the unculturable majority

Before the twenty-first century, the study of microorganisms typically required the ability to cultivate them in isolation. The advent of metagenomics — exemplified by two key studies published in 2004 — provided approaches that enabled unbiased and culture-independent analysis of DNA directly from communities in the environment, revolutionizing the study of complex microbial communities.

Metagenomics was a term first used in 1998 by Handelsman et al. in describing the study of the “collective genomes of soil microflora”. Yet, over the 100 years before this point, many microbiologists had noted that the number of microbial cells they could count microscopically was not aligned with the number of colonies that they could grow on plates. This phenomenon, termed the ‘great plate count anomaly’ by Staley and Konopka in 1985, led to the realization that only ~1% of microbial diversity could be accessed through standard cultivation approaches. This prompted the question: how do we study the remaining 99%?

Pioneering work starting in the late 1970s on the 16S ribosomal RNA (rRNA) gene led to various PCR-based and other molecular tools to enable quantitative and qualitative analyses of microbial identity and diversity from environmental DNA. These advances facilitated early cloning approaches that sought to access new genes from unculturable microorganisms. For example, in the first targeted attempt at metagenomic sequencing led by Edward DeLong in 1996, Stein et al. recovered a 40 kb genome fragment from an uncultivable archaeon from a marine picoplankton assemblage using PCR amplification of the 16S rRNA gene to identify clones containing archaeal DNA. However, PCR-based studies are inherently biased, and the ultimate goal was to access the full genetic potential of microbial genomes from the environment — a daunting task considering that most microbial



Credit: Science Photo Library, Alamy, Stock Photo

communities comprise hundreds to thousands of species.

The first genome-resolved metagenomics study came in 2004 in a study by Jill Banfield and colleagues. Tyson et al. reported the successful reconstruction of multiple genomes from a DNA sample taken from a biofilm in an acid mine drainage system. The near-complete genomes of a bacterial and an archaeon species, plus the partial genomes from a further three microorganisms, were recovered from environmental DNA using random shotgun sequencing, in which total DNA is fragmented, cloned and sequenced. This study was aided by the low species richness of the sampled biofilm and the low intraspecies genomic variation of the microorganisms, which facilitated assembly of the sequencing reads. Analysis of the recovered genomes from the unculturable iron-oxidizing microorganisms in the acid mine enabled characterization of metabolic pathways and provided insights into survival strategies used by extremophiles.

A second large-scale metagenomics project in 2004 provided the first whole-genome shotgun

sequencing study of oceanic microbial populations. Craig Venter and colleagues examined the Sargasso Sea using pooled environmental DNA that was filtered and extracted from seawater samples. From 1.66 million sequencing reads, 265 Mb of sequence data were generated, which led to the identification of >1.2 million previously unknown microbial genes. Data binning and phylogenetic analyses to predict the origin of the sequences led to estimates that the data derived from 1,800 genomic species, including 148 previously unknown ‘species’ (phylotypes). This study demonstrated that metagenomic sequencing could assess the taxonomical composition of complex microbial communities in an unbiased manner and confirmed previous vast underestimates of microbial biodiversity.

These two studies revealed exciting new opportunities for metagenomics, with the only real limitation being the cost of sequencing. Since 2004, the dramatic drop in the cost of sequencing following the emergence of next-generation sequencing has led to widespread adoption of metagenomics approaches at a scale that was previously thought unachievable.

Mining the huge wealth of metagenomic data has improved our understanding of microbial biodiversity, ecology and evolution and led to valuable gains in biotechnology and medicine. Many discoveries of new antibiotics, anti-cancer drugs and biosynthetic pathways of biomedical, agricultural and industrial importance have their origins in our ability to access the genomes of the unculturable majority.

Iain Dickson, *Nature Reviews Gastroenterology & Hepatology*

“ Many discoveries... have their origins in our ability to access the genomes of the unculturable majority ”

ORIGINAL ARTICLES Tyson, G. W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004) | Venter, J. C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004)

FURTHER READING Handelsman, J. et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–R249 (1998) | Staley, J. T. & Konopka, A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* **39**, 321–346 (1985) | Stein, J. L. et al. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**, 591–599 (1996)