

MICROBIOLOGY

Microbial planet

Researchers sequence single-cell genomes to target uncharted microbial diversity.

Deep in the boiling sea, trapped under permafrost and seemingly everywhere in between, researchers are revealing a rich world of microbes. The discoveries come from sequencing DNA fragments from the environment, a practice known as metagenomics. But to best understand how microbes function requires whole genome sequences, which were confined to a few well-studied groups until recently.

The Genomic Encyclopedia of Bacteria and Archaea (GEBA) project, spearheaded by scientists at the US Department of Energy's Joint Genome Institute (DOE-JGI), targeted undersampled areas of the microbiome for whole-genome sequencing of cultivated isolates. It filled gaps in the evolutionary tree, but as JGI researcher Tanja Woyke notes, "there are at least as many phyla that have no cultivated representatives as phyla that have cultivated representatives."

As the GEBA project tapered down around 2010, Woyke and a cadre of collaborators began a new effort to sequence uncultivated bacterial and archaeal genomes in unsequenced regions of the evolutionary tree, the so-called 'microbial dark matter' (they named the project GEBA-MDM).

They first needed to decide where to tap unexplored diversity. The JGI is a sequencing facility that handles metagenomics projects around the globe; amplified sequence from a ribosomal RNA gene is used in these projects to assign microbes to taxonomic clusters. The researchers took advantage of this existing diversity data to focus their sequencing.

Woyke recalls that forming collaborations was easy. Investigators from all nine selected sites, which included a gold mine runoff site, bioreactors, and fresh and brackish water sites, agreed to collect new samples.

Before the project, Woyke sequenced her first single cell by carefully plucking it with a micromanipulator and amplifying its genome. Scaling up for a large project required changes such as switching to a flow cytometer and also provided many chances to learn by troubleshooting.

What prevents some genomes from amplifying well is still a mystery. Cells vary in their propensity to break open and reveal accessible DNA. Soil and biofilms are notoriously tricky to work with because they include sub-



Christian Rinke, DOE-JGI

Examples of environments sampled for the GEBA-MDM study.

stances that may inhibit amplification, and disaggregating cells is a challenge. One hard-won lesson is to monitor amplification in real time and only sequence if there are signs of a robust reaction. "Generally, the early amplifiers get better genome coverage," says Woyke.

They assembled selected genomes from fragments of amplified single-cell DNA, about 40% complete on average. "They're definitely fragmented draft assemblies, [covering] anything from a few percent to over 90%," says Woyke. The group sequenced 201 genomes, a sliver of her estimate that 20,000 genome sequences would cover just half of known microbial diversity. Yet the work represents a considerable expansion of our knowledge of microbial life.

The researchers constructed an evolutionary tree that included the sequenced genomes, allowing them to discover new groupings such as a superphylum that includes bacteria with lean metabolic capabilities. They also discovered important proteins required for gene expression in an archaeal lineage that were thought to be specific to bacteria, crossing entirely distinct domains of life. The new genomes will serve as references that help anchor DNA fragments to known taxonomic positions in metagenomics studies.

Woyke has been adding new sample-collection sites and is optimistic about the pace of improvements in whole-genome sequencing. JGI scientists are now looking into sequencing mixtures of bacterial cells to assemble a few genomes simultaneously. "These 201 genomes, it's the beginning of something that's much greater," she says.

Tal Nawy

RESEARCH PAPERS

Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).