

MICROBIOLOGY

Assembling genomes in microbial communities

Binning based on differential read coverage, rather than sequence composition, allows separation of metagenomic sequence reads into species-level clusters that can be assembled into single chromosomes.

We can't live without bacteria, and sometimes we can't live with them. Unfortunately, there is still a large discrepancy between the known impact of the microbial world and our understanding of exactly what bacteria do that makes them so beneficial or so destructive. Bacteria that can be cultured are relatively easy to characterize; regrettably, only a small fraction of the microbial world grows in a dish. Recent sequencing technology and computational tools have made large inroads into profiling the bacterial communities in different natural habitats. "Often different metagenomic communities are characterized without looking specifically at individual genomes," says Per Nielsen of Aalborg University in Denmark. "People looked for genes to see which bacteria are present and to get an overview of ecosystem function, but to really learn about the communities you need the genomes of the individual species that are present in an ecosystem."

In recent years researchers have adopted several strategies to get at this genomic information. One is to focus on communities with lower complexities, containing between five and ten abundant species. Although it is possible to assemble individual genomes from metagenomes in these communities, such assemblies do not represent the diversity seen in most natural ecosystems. The other strategy is to perform single-cell genomics, in which the DNA of a single bacterium is amplified and sequenced, but it is difficult to obtain complete genomes with this method: rarely can more than 50% be recovered.

Nielsen and his graduate student Mads Albertsen wanted to characterize complex communities at the population level to investigate contributions of individual species to an ecosystem. A serendipitous observation provided a way to do this.

When Albertsen extracted DNA from bioreactor biomass, he used two different extraction methods: one with and one with-

out hot phenol. He assembled the paired-end reads of each data set into scaffolds and then mapped the reads to the assembled scaffold to get an idea of coverage depth along the scaffold. In doing this, he realized that the extraction efficiencies of the two methods were different for the same species, leading to different relative abundances. Plotting the abundances along the scaffold against each other allowed Albertsen to see clusters of species. Using the two coverage measures to zoom in on a specific population in the complex community led to a dramatic reduction in complexity so that the researchers could extract individual species by traditional metrics such as tetranucleotide frequencies. Albertsen was able to assemble 31 population genomes and 12 complete genomes at the species level, among them species with a relative abundance of less than 1%.

Four of these rare genomes belong to the candidate phylum TM7, of special interest to Nielsen and his coauthors Philip Hugenholtz and Gene Tyson from the University of Queensland, Australia, for its presence in wastewater and role in human gut and oral inflammation. The researchers validated the presence of the TM7 populations with specific fluorescence *in situ* hybridization probes and analyzed the metabolism of the species in greater detail. On the basis of their metabolic reconstruction of the near-complete TM7 genomes, they proposed the name *Saccharibacteria* because of the capacity of these bacteria to use sugars.

Nielsen is convinced that this approach, based on binning different relative abundances rather than depending on knowledge about sequence composition, will aid in genome extraction from metagenomic samples. But he is also realistic about the current limitations of the method. "It is a lot of work," he says. "If you just want to know which bacteria are present, it is not necessary; but if you want to know their function, you need the genomes."

Nicole Rusk

RESEARCH PAPERS

Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**, 533–538 (2013).