

# Epigenetic profiles to classify bacterial sequences

**Methylation patterns can help determine the origin of metagenomic sequences.**

An important aim of metagenomics is to piece together genomes from the ‘gemisch’ of sequence fragments found in a microbial mixture. One strategy uses overlaps to assemble long contiguous sequences, or contigs, from sequence reads. Collecting all contigs derived from the same taxon into one bin is the next step, and it is challenging.

Researchers have used DNA base composition and covariation in sequence coverage across samples to group contigs by their taxonomic origin, but these approaches falter when related strains are present, or when species are evenly represented across samples.

Gang Fang and his team at the Icahn School of Medicine at Mount Sinai, New York, have found that epigenetic data can help with metagenomic binning. About five years ago, they identified methylation patterns in a pathogenic strain of *Escherichia*

*coli*. They realized that the patterns differed between closely related strains, meaning that they could be used as signatures to bin metagenomic sequences.

Different bacterial strains have distinct combinations of methylases that each act at a set of sequence motifs, forming what Fang refers to as an epigenetic barcode. The methylases also mark resident plasmids and viruses with the same barcode as their host. Fang’s approach relies on long-read sequence data that report methylation status, which can be generated by the Pacific Biosciences and Oxford Nanopore Technologies platforms. But after his original insight, it took a few years for the available technology to effectively capture methylation in a metagenomics setting and for enough data to support the point that strains differ by methylation profile.

Fang’s team developed the mBin software, which profiles methylation motifs for each

contig. Most motifs are not methylated in a sample and are removed as noise, while the remaining motifs provide enough discriminative power to robustly cluster profiles into bins. The software was successfully used to bin contigs from nine poorly characterized genomes from a mouse gut microbiome and to assign mobile genetic elements.

Epigenetic binning is most advantageous for microbiome samples with little species divergence and similar abundance across samples, and it can be combined with other binning approaches to improve resolving power. Fang’s team is also working on ways to lower the main hurdle associated with long-read sequencing—to broaden use, “we want to bring down the cost,” he says.

## Tal Nawy

### RESEARCH ARTICLES

Beaulaurier, J. *et al.* Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nat Biotechnol.* <http://dx.doi.org/10.1038/nbt.4037> (2017).