

February 16, 2016 | By Ioana Patringenaru

## A New Method to Dramatically Improve the Sequencing of Metagenomes

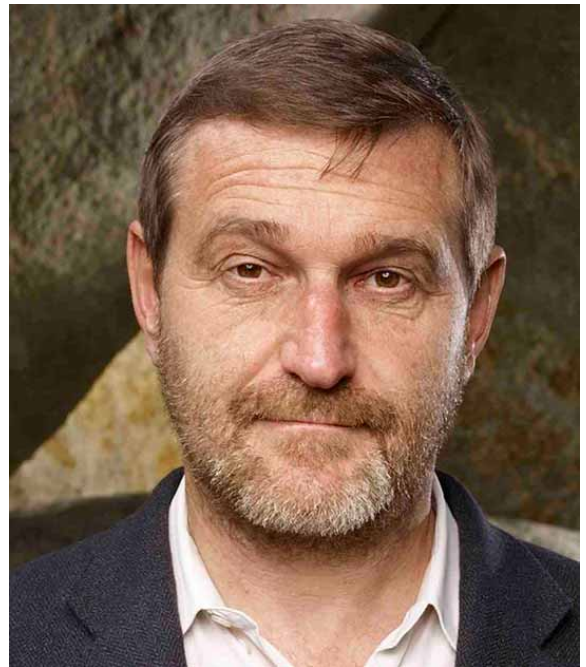
An international team of computer scientists developed a method that greatly improves researchers' ability to sequence the DNA of organisms that can't be cultured in the lab, such as microbes living in the human gut or bacteria living in the depths of the ocean. They published their work in the Feb. 1 issue of *Nature Methods*.

The method, called TruSPADES, generates via computer so-called Synthetic Long Reads, segments that are about 10,000 base pairs of the genome, from the commonly used short reads of just 300 base pairs produced by machines from San Diego-based Illumina.

Using Synthetic Long Reads instead of short reads to assemble a genome is like using entire chapters rather than single sentences to assemble a book, researchers said. So there is a strong incentive to improve sequencing with long reads.

"This is the next generation of sequencing technologies," said Pavel Pevzner, a professor of computer science at the University of California, and the lead author on the study. "It will make a significant impact on the practice of metagenomics sequencing."

Currently, the leaders in the long-read sequencing market, Pacific Biosciences and Oxford Nanopore, generate long reads that can be inaccurate and difficult to use in complex sequencing problems, such as assembling metagenomes—whole colonies of microbes



*Computer scientist Pavel Pevzner is the lead author on the Nature Methods paper.*

sampled from their natural environment. By contrast, the Synthetic Long Reads are 100 times more accurate and can be rapidly generated on a massive scale to cover a large fraction of bacteria in metagenomes.

To develop their new method, researchers took the shorter reads, 100 to 300 base pairs, equipped with barcodes. They then assembled the short reads together into Synthetic Long Reads by representing them using a de Bruijn graph, a method often used in short read sequencing. The graph allows researchers to determine which reads are connected together, resulting in the longer and more accurate Synthetic Long Reads.

The next step is to apply this method to the study of various microbial communities ranging from human to marine microbiomes. Pevzner and co-author Anton Bankevich from St. Petersburg State University, are working with Christopher Dupont, a researcher at the J. Craig Venter Institute, to do just that.

Metagenomics is especially challenging because researchers do not study a single species of bacteria but hundreds of them that live together in a community. When they extract a sample from the community and sequence it, they end up with bits of bacterial genomes from all the organisms in the community. It's very much like trying to solve hundreds of puzzles without knowing which pieces belong to which puzzle. TruSPADES and Synthetic Long Reads will help researchers solve these puzzles.

“This method gives us better results at a much smaller cost,” said Dupont. “We are now assembling genomes for organisms we didn't even know existed.”

---

## MEDIA CONTACT

**Ioana Patringenaru**, 858-822-0899, [ipatrin@ucsd.edu](mailto:ipatrin@ucsd.edu)

UC San Diego's [Studio Ten 300](#) offers radio and television connections for media interviews with our faculty, which can be coordinated via [studio@ucsd.edu](mailto:studio@ucsd.edu). To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.