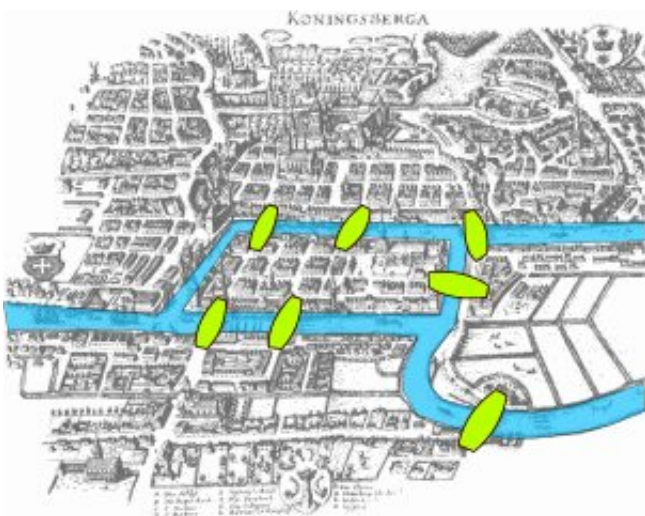# Long reads and powerful algorithms identify 'invisible' microbes

January 4 2022, by Katie E. Ismael

To solve the metagenome assembly, the team of UC San Diego's Pavel Pevzner used the algorithmic approach that is not unlike solving the "Seven Bridges of Konigsberg" puzzle, which asks participants to find a path through the middle age city of Konigsberg while walking across each bridge only once. Pevzner modeled genome assembly as a giant city with millions of bridges in which each read represents a bridge and a genome represents a path visiting each bridge. Credit: University of California - San Diego

Microbes are everywhere—in our guts, on our skin, permeating the environments around us. Studying these microbial communities has delivered tremendous insights into disease and good health, but identifying all the distinct species in a sample can be challenging.

Now, a study by an international research team has shown that highly accurate, long-read genomic sequencing technology (HiFi) can shine a light on this previously hidden biology.

Researchers at the University of California San Diego Department of Computer Science and Engineering, the U.S. Department of Agriculture, the biotechnology company Pacific Biosciences and labs in Russia, Israel and the Netherlands have shown that HiFi, combined with advanced algorithms, can differentiate between nearly identical organisms, allowing researchers to more completely catalog microbial communities. The study was published today in *Nature Biotechnology.*

"This HiFi technology, developed by Pacific Biosciences, is revolutionizing the field," said Pavel Pevzner, the Ronald R. Taylor Distinguished Professor of Computer Science at UC San Diego and co-senior author on the paper. "Here, we provide complete or nearly complete bacterial genomes, distinguishing very similar bacterial strains from a single sample. This is no small thing: Some E. coli strains are harmless, others are deadly."

HiFi recently helped sequence a [complete human genome](link), a feat that had evaded scientists since the Human Genome Project produced an incomplete version 20 years ago.

This paper builds on those findings, showing that long reads can shine a light on previously invisible organisms. Short-reads, the most common genomic sequencing technique, analyze brief DNA fragments (100 to 300 base pairs) and have trouble assembling complete genomes and differentiating between genomically-similar microbes.

Long-read technologies, such as HiFi, generate much larger DNA fragments (greater than 15,000 base pairs) and have emerged as a potential solution. As long-read accuracy has increased, the technology

has revealed hidden genomic features in amazing detail. In this case, HiFi easily differentiated microbes with only minor genomic variations.

"We can now sequence complete genomes of nearly all abundant bacteria in a microbiome," said first author Mikhail Kolmogorov, a former UC San Diego postdoctoral fellow and now a Stadtman Investigator at the National Cancer Institute. "Short-read studies rarely provided complete sequences of even a single microbe."

Long reads from sheep guts

In this study, the research team used HiFi long reads to sequence the microbial metagenome in sheep guts. Their goal was to create complete reference genomes for unique microbial species (metagenome-assembled genomes or MAGs).

They found HiFi and associated algorithms identified the genomes from 428 species with greater than 90 percent completeness. Many of these had been invisible to short-read technologies. These findings could be a tremendous boon for microbiome researchers and other scientists, providing new and powerful tools to fully delineate a sample's microbial complement.

"Characterizing microbiomes of ruminant livestock, like sheep, can be used to develop methods to reduce disease, environmental impact and greenhouse gas emissions, while improving productivity," said Timothy Smith, a research chemist at the USDA's Meat Animal Research Center and co-senior author on the paper. "Strain-level genome resolution will help track genes related to antimicrobial resistance and determine the extent animal husbandry might be contributing to the rise of antibiotic resistance in human and animal diseases."

The applications for this work are quite broad, as the ability to precisely

delineate specific microbial species in complex samples could inform many scientific endeavors.

"Although this study focuses on microbes in the sheep gut, the potential is tremendous because microbes inhabit so many places on human, animal and plant bodies and throughout the environment," said Rob Knight, director of UC San Diego's Center for Microbiome Innovation and a professor in the departments of Computer Science and Engineering, Bioengineering and Pediatrics and a co-founder of the American Gut Project.

"Having better ways to read genomes in complex environments sets the stage for improved efforts to write the genomes we need to solve many of society's most pressing problems," said Knight, who is not an author on this study.

These advances should be quite useful in medicine, including UC San Diego's groundbreaking phage therapy (IPATH) and antimicrobial (CHARM) programs. Microbes may also help diagnose cancer, decipher red tides, study plastic biodegradation in the ocean and measure carbon release and capture.

"Rather than combining similar organisms into one bucket, we can now differentiate them and get a true metagenomic picture of complex bacterial communities," said Pevzner. "Like complete genomics, which is already being applied to rare disease diagnostics, complete metagenomics may soon make its way into medicine and many other disciplines."

Provided by University of California - San Diego