Tech News

Single-cell Genomics: Defining Microbiology’s Dark Matter

We all remember this experiment from freshman biology: Take a sterile swab, rub it along the inside of your cheek, and transfer the collected microbes to a sterile agar plate.

After incubating the plate overnight at 37 °C, the agar will be awash in bacterial colonies, a glossy and somewhat smelly testament to the fact that humans, fundamentally, are giant microbial buses.

Yet that picture, it turns out, is woefully incomplete. The human microbial complement, or “microbiome,” comprises “at least 10 times more bacteria than the number of human cells in the body” (1)—trillions upon trillions of cells representing thousands of microbial strains. And yet this is merely the snowflake atop the iceberg that is the global microbial population. Many strains have never even been seen before, and the vast majority, millions perhaps, simply will not grow in the lab. They are, in the parlance of the microbiologists who would study them, “uncultured.”

Uncultured, but not impenetrable.

A decade ago those microbes might have been known only by the arrangement of their 16S ribosomal RNA signatures on a phylogenetic tree. Today those same organisms are being probed in astonishing detail thanks in large part to next-generation DNA sequencing and cell separation technologies—not to mention an ever-expanding database of annotated genes against which to compare them.

It is, says J. Craig Venter Institute professor Roger Lasken, “a fantastic time” to be a microbiologist.

The “new” microbiology

Single-cell genomics is the latest in a string of techniques microbiologists have used to delve into the worlds of unculturable microbes. The first approach, “molecular phylogenomics,” was spearheaded by Carl Woese and others in the late 1970s. Relying on PCR-amplified, Sanger-sequenced 16S ribosomal RNA genes to estimate evolutionary relationships among different organisms, the end result was a phylogenetic “tree of life.” Yet this was a tree where most leaves are mere placeholders, stubs that offer no insight into the biology of the organisms they represent, says Paul Blainey, a core faculty member at the Broad Institute in Cambridge, Massachusetts who develops technology to study unculturables.

“We know there’s some organism there that we haven’t sequenced, but we don’t really get much useful information in terms of understanding that organism from these [16S] sequences,” Blainey explains. That’s because researchers know what 16S rRNA does—it’s involved in protein translation—and that isn’t what makes most microbes interesting.

The next step in unculturable microbiology was metagenomics. In metagenomics, environmental samples are shotgun sequenced en masse, detailing the genetic and metabolic potential of samples as a whole—for instance, that something in the population can fix carbon, while another can extract energy from sulfur metabolism. The approach has been the subject of some seriously big science in recent years, from J. Craig Venter’s Sorcerer II Global Ocean Sampling Expedition (GOS) to the $157-million NIH funded Human Microbiome Project.

Such studies produce voluminous data, to be sure. Yet metagenomics researchers often struggle to assign specific sequences to particular organisms. The result is a massive collection of DNA fragments, which may or may not belong to any particular microbe—the only way to know for certain is if both an interesting function and an identifying marker, such as a 16S rRNA, exist on the same fragment. Reassembling these sequences is like an archeologist trying to piece together thousands of different pottery shards, without knowing how many objects they represent or even what they looked like whole.

This is where single-cell genomics enters the picture. In single-cell genomics, researchers physically separate cells prior to sequencing. The resulting genomes may not be intact, says Tanja Woyke, microbial genomics program lead at the DOE Joint Genome Institute, but at least all the pieces come from one organism.

The process is simple is theory, complex in practice. Cells are separated and placed in individual reaction chambers, either by flow cytometry, micromanipulation, or microfluidics. They are then lysed to release nucleic acids, which are amplified to produce single amplified genomes, or SAGs. The SAGs are typically screened to identify samples for further analysis by sequencing 16S rRNA or some other gene of interest.

Finally, selected SAGs are more fully sequenced. JCVI barcodes and pools up to 90 single cell genomes samples per Illumina lane, says Karen Nelson, director of the JCVI’s Rockville campus. At JGI, researchers pool 12 cells per lane in an Illumina HiSeq run, producing more than 4 gigabases of sequence per cell.

One recent study highlights the power of this approach. Ramunas Stepanauskas,
Lysis conditions need to be gentle enough to maintain DNA integrity, but harsh enough to be effective. It’s a difficult balance to strike: At JGI lysis (using potassium hydroxide) typically only works for about 20% of the cells.

DNA yield is another issue. Most microbial cells contain only a single chromosomal copy and femtograms of DNA, far too little to be directly sequenced, so it must be amplified. The traditional whole-genome amplification method, PCR using random primers, introduces too much sequence bias and results in short fragments, says Lasken. But in 2002, Lasken, then at biotech firm Molecular Staging, and colleagues introduced a new method called multiple displacement amplification (MDA).

MDA is a random-primed, isothermal amplification process that employs the highly processive phi29 polymerase to produce fragments on the order of tens of kilobases in length. According to Woyke, the key to phi29 is that it doesn’t stop upon hitting new double-stranded DNA, but rather keeps moving. Newly synthesized fragments thus become templates themselves, producing a hyperbranched structure sprouting multiple replication forks.

The process is highly efficient: From femtograms of starting material MDA can produce micrograms of long fragments, typically 20–40 kb in length. But MDA has issues, too. Many reagents, it turns out, contain minute quantities of contaminating DNA that can significantly impact the process, Woyke says; after all, it takes only the nucleic acid equivalent of a few bacterial cells to swamp one cell’s worth of DNA. Woyke’s team has devised several techniques, including UV irradiating MDA reagents and performing cellular sorts twice, to minimize contamination issues.

Blainey took another approach. Prior to joining the Broad Institute in early 2012, Blainey was a postdoc with Steven Quake at Stanford University. There his job was to develop methods to sequence the genomes of unculturable “dark matter” microbes “that were from the deepest, darkest regions of the phylogenetic tree.”

To do that, Blainey devised a microfluidics device that could separate up to 48 individual cells using optical tweezers, lyse them, and produce SAGs on-chip. Blainey developed his chip specifically to minimize contamination, he says. By reducing reaction volumes down to the nanoliter level, Blainey figured he could shift the balance of DNA to contaminants in favor of the former. “We’ve still got one cell, but we’ve got a lot less of that other junk that comes with the reagents,” he explains.

The team used the device to capture a single cell of a “candidate division” called OP11 from an anaerobic sulfide and sulfur-rich spring in southwestern Oklahoma called Zodletone Spring. There was nothing particularly special about OP11, Blainey says, because nothing was known about it; the organism was sequenced to fill in a phylogenetic gap.

Yet the resulting sequence provided more than just As, Cs, Gs, and Ts, according to Blainey; it provided glimpses of how OP11 lives—oxidizing organic molecules for energy, just as humans do, and breaking down complex polymers, which humans cannot. “We are actually quite excited about the data we got back—not in the sense that it encodes genes for turning lead in to gold, but that we got information about enough of the right genes to start outlining the nature of the OP11 cell we isolated.”

**Single cell SNPs**

The other major issue with MDA is amplification bias; researchers using MDA consistently observe uneven amplification across the genome. According to Blainey, genome completeness in single-cell genome projects can run the gamut from 0 to 95%; Woyke says her team typically recovers between 20% and 80%.

Such uneven reads tend to confound genome assemblers, which are designed to anticipate relatively even sequence coverage. “This is absolutely not true at all for single cell data and really fouls up the assemblies,” notes Blainey.

Patrick Chain, team lead for metagenomics and next-gen sequencing applications at the Los Alamos National Laboratory, along with LANL colleague Cliff Han, has developed a process to...
normalize genome coverage by inducing artificial polyploidy—
that is, multiple genome copies — by using cell division inhibitors
to produce cells with 2, 4, 8, or more genomes per cell. The result,
he says, is “better genome coverage and a more normalized distri-
bution of genome coverage after sequencing.”

Meanwhile a group led by Sunney Xie, professor of chemistry
and chemical biology at Harvard University, has developed an
alternative approach to genome amplification. Called MALBAC,
the technique employs a linear thermal cycling amplification step
prior to PCR amplification, rather than MDA.

“That solves the bias problem,” Xie says — so much so, in fact,
that his team has amplified the DNA from a single human cell with
93% coverage at 30x sequencing depth. The team could even call
a single SNP variant based on their data. “If one base is different,
we can call it.”

Metagenomics meets single cell genomics

Single-cell genomics and metagenomics represent two sides of the same
coin — understanding the biology of organisms that cannot be grown
in the lab. Metagenomics techniques can reveal the genetic potential of
a community, but not the players. Single-cell approaches can close that
gap, for instance by providing scaffolds for assembling metagenomics
data or reference genomes for variation studies. As a result, for many
researchers the two technologies are complementary.

According to Woyke, JGI received 12 single cell genomics
proposals during a recent call for proposals, 10 of which combine
metagenomics and single cell approaches and account for more than
400 single-cell genomes in total.

“It’s not like you should only do one or the other, they inform each
other,” says Philip Hugenholtz, director of the Australian Center for
Ecogenomics in Queensland.

In one 2009 study (3), for instance, Stepanauskas and his team
(including Woyke) isolated two individual “uncultured, proteorhodopsin-containing marine flavobacteria” from the Gulf of Maine,
collecting 1.9 Mb and 1.5 Mb of genomic DNA representing an
estimated 91% and 78% genome recovery, respectively.

The team used those genome assemblies as scaffolds to “recruit”
individual reads from the Venter’s GOS to map where in the world’s
oceans those organisms reside. “In theory we could do the same with
single-cell sequencing alone,” Stepanauskas concedes. “But it would
be much more expensive.”

Single cell viromics

Of course, to get a really complete picture of a microbial community,
researchers must study more than just its microbes; there also are the
viruses that prey upon them.

It has been estimated there are 10 viruses for every bacterial species,
and as Lasken points out, “we don’t even know how many bacterial
species there are.”

Recently Lasken, along with former JCVI colleague Shannon
Williamson, published a proof-of-principle study addressing “single
virus genomics” (4). The team mixed two known bacteriophages (T4
and lambda), sorted them into individual “cooled agarose beads” on
a microscope slide, and used those as templates for MDA and sequencing.

The team showed they could read a single lambda phage at 437x coverage,
capturing all but the first five bases of the virus’ genome.

Still, says Lasken, “It is fair to say that single viral particle
sequencing would solve a very difficult problem of how to get access
to this enormous number of viruses in the environment.”

As with most everything in biology, such sequences will be
only the beginning. After all, even well-studied of organisms have
their secrets. Says Hugenholtz, “There are still genes in E. coli that
they haven’t worked out what the function is, despite having the
genome for almost 20 years.”

1. The NIH HMP Working Group, “The NIH Human Microbiome Project,”
2. B.K. Swan et al., “Potential for chemolithoautotrophy among ubiquitous
3. T. Woyke et al., “Assembling the marine metagenome, one cell at a time,”

Written by Jeffrey M. Perkel, Ph.D.


To purchase reprints of this article, contact: biotechniques@fosterprinting.com