

Microbiology: making the best of PCR bias

Michael Eisenstein

Many factors can skew the results of a widely used amplification technique for microbiome analysis, but researchers are finding strategies for getting at the truth.

Imagine you're holding on to a vial of water that's crawling with microbes—and that's a good thing, because you're conducting a census of the myriad microscopic denizens of the lake from which you obtained that water. But how should you perform the survey? The answer depends on who you ask.

Traditionally, microbiome researchers have used PCR-based strategies to amplify the gene encoding the small ribosomal RNA subunit, known as 16S in prokaryotes. This essential gene is extremely variable between species, which offers a useful 'fingerprint' for distinguishing the microorganisms that compose a given community, whether from soil, sea or human gut. "I think it's still pretty much a cornerstone," says Jed Fuhrman, a marine microbiologist at the University of Southern California. "It gets such high coverage of a particular gene that gives you a lot of information."

Some labs are turning to a newer technique known as 'shotgun metagenomics', in which the total DNA content of a microbiome sample is fragmented, sequenced and reassembled to generate a more comprehensive picture of its contents. "You get so much more mileage from metagenomics, and from my perspective, there's absolutely no reason to do 16S," says Nikos Kyrpides, a molecular biologist at the Joint Genome Institute. As sequencing costs continue to fall, even former 16S stalwarts such as Rob Knight at the University of California San Diego School of Medicine are embracing the newer method. "We're switching our entire pipeline over to shotgun, for the sample types where it can be done," says Knight.

But 16S has vigorous defenders who, like Fuhrman, think it offers the best bang for microbiologists' buck. "It's so much faster and you can run a lot more samples," says Per Nielsen, head of Denmark's Center for



The Great Boiling Spring in Nevada contains numerous microbes that would have gone undetected in conventional 16S amplicon analysis.

Microbial Communities. "It's really easy, quite cheap and quite reliable." It performs especially well in samples where biological material is limited or that are heavily contaminated with nonmicrobial cells. However, even its champions agree that 16S profiling is vulnerable to bias from diverse sources, and considerable effort has been invested to get a handle on factors that can confound this tool.

Making a match

The presence of ribosomes is a constant across the entire tree of life. And critically, the sequence of the 16S rRNA gene is both highly conserved across species and highly distinctive, comprising nine hypervariable regions that are flanked by relatively conserved sequences. This makes it an excellent biological marker gene for species characterization.

Most labs rely on so-called short-read sequencing platforms, such as those made

by Illumina, which cannot cover all 1,500+ nucleotides of the 16S gene. As a result, 16S profiling typically relies on PCR amplification with pairs of primers that recognize conserved domains flanking informative variable regions, which can then be sequenced in vast numbers. After the initial demonstration of 16S-based microbial profiling by Carl Woese¹ in 1977, Sanger-sequencing-based methods for 16S analysis gave rise to more efficient and sensitive PCR-based methods in the 1990s—and much of this early molecular toolbox is still in use. "Almost all the really good primers were discovered some years ago," says Fuhrman.

There is still variation in the conserved domains, albeit to a lower degree than in the hypervariable domains. The primers are designed for compatibility with minor sequence variants through 'degenerate' binding that doesn't rely on strict Watson-Crick base-pairing. However, these are not



UC San Diego Health

Rob Knight, cofounder of the Earth Microbiome Project.

a perfect skeleton key for unlocking 16S data. “When you talk about universal primers, ‘universal’ has to go in air quotes,” says Pat Schloss, a microbiologist at the University of Michigan. Certain primer sets underperform when

they encounter particular mismatches that undermine hybridization to their target sequence, which results in underamplification of certain organisms. Knight cautions against simply thinking of primers and templates as abstract sequences with highly predictable behaviors. “They’re actually molecules that behave in very different ways in chemical reactions, especially with respect to hybridization under a particular set of conditions,” he says.

A 2016 study from Fuhrman and colleagues² showed how unpredictable the effects of disagreement between primer and template can be. “One mismatch in the middle can cause something to be underestimated by tenfold in a complex mixture,” says Fuhrman, “but in a simple mixture of things, it will amplify beautifully.” Certain reaction conditions may render PCR more error-tolerant. After surveying a wide range of reaction scenarios for 16S amplification, Daryl Gohl and colleagues at the University of Minnesota determined that certain proofreading polymerase enzymes are capable of ‘primer editing’, overcoming sequence incompatibilities that would otherwise cause certain microbes to disappear entirely³. “We could now detect some of these taxa with mismatches in the primer regions,” says Gohl. “The proofreading polymerases were chewing back the primers and editing them to match the templates.”

It also helps to know what one is looking for, as some primer sets are inherently ill-suited for particular taxa. “You’ve got to pick your primers for the type of organisms you expect to find in your environment,” says Schloss. “I feel pretty good about the primers we use for the gut, but I know they miss bacteria on the skin ... you’d want to modify those to amplify things like *Propionibacterium*.” For very diverse microbial communities, this can be particularly problematic, and a recent study from Kyrpides’ team⁴ used shotgun

metagenomic data to show that certain phyla may be consistently overlooked by conventional 16S amplicon surveys. The problem was particularly stark for Archaea, a microbial domain that remains notably undercharacterized in general. “The best primers were missing as few as 10% of lineages overall,” says Kyrpides. “But in particular samples, those lineages may represent up to 90% of the community.” This effort led to the discovery of a ‘new’ bacterial phylum, Kryptonia, which had been hiding in plain sight in existing metagenome data but was undetectable with existing 16S primers.

In principle, one could design additional 16S primers that capture the missing microbes, but Nielsen and his colleague Mads Albertsen have devised an alternative approach⁵. Their method goes after the rRNA itself, rather than the gene encoding it, and uses a sequence-independent molecular-tagging approach that allows them to amplify and reconstruct entire rRNA sequences instead of targeting individual variable domains. As an initial demonstration, the team derived a million small-subunit sequences. “In just one study we got similar numbers of full-length sequences as everyone else has done in the past 20 years,” says Nielsen. “In principle, we can sequence all life on the planet and get the complete ‘tree of life.’” However, he also notes that their method is labor intensive and technically complex; rather than a replacement for conventional amplicon analysis, he sees the method as a powerful tool for populating the reference databases that researchers consult to interpret 16S data.

Community standards

Although the PCR reaction is the centerpiece of 16S analysis, there are other opportunities for bias or error to skew the census. These include the initial DNA preparation steps, which must generally be customized to the particular environmental sample, and the sequencing procedures used to analyze the PCR amplicons. For example, Christopher Quince, a bioinformatician at the University of Warwick, notes that although contemporary sequencing systems such as the widely used Illumina MiSeq might be expected to deliver more reliable data than the older, more error-prone technologies used in the past, they will not necessarily do so. “The error rates are lower, but there are some other, new sources of error that are harder to deal with,” says Quince. “One of them is sample

switching, where reads from one sample ‘bleed’ into another sample.”

As a safeguard against these aberrations, many microbiome researchers strongly advocate the use of robust controls. These can include either defined ‘spike-in’ sequences added directly to the experimental sample as a barometer of assay performance or, more commonly, ‘mock communities’ of selected microbial cells or genomes that can be analyzed alongside samples of interest. “It’s very important for measuring things like error rates in terms of sequencing, and perhaps thinking about your primer bias,” says Schloss. Ideally, these should reflect the specimen of interest, with a mixture of various species found in that particular environment, but Schloss has more modest expectations. “We’re lucky if people use a mock community at all,” he says. “Even if you’re just sequencing *E. coli*, that’s a start.”

Indeed, many studies fail to include such controls, and this remains a source of frustration for some. “It has not been routine, and I’m certainly hoping it does become routine,” says Fuhrman. “If we were measuring a chemical, we would have to put a standard in and run a standard curve.” Reasons for skipping this experimental step include habits established in the early days of the field, when each additional sample incurred a considerable cost, and an emphasis on churning out maximal amounts of data rather than focusing on the quality of the results. Some researchers also think such controls are most valuable when a protocol is first being developed, as a means to characterize sources of bias, and that less may be gained by the use of mock communities in subsequent experiments. Knight also notes that over-reliance on mock communities can potentially reinforce bias rather than eliminate it. “It’s very easy to over-tune your technique to a single



UC San Diego Health

Before analyzing a microbiome sample, such as this oral swab for the American Gut Project, one must select the PCR primers most likely to amplify taxa of interest.

mock community or biological specimen,” he says.

Nevertheless, Knight still recommends using such controls to catch experimental errors, and Fuhrman notes that mock communities recently helped his team to explain the mysterious disappearance of entire microbial taxa from seawater samples in an otherwise routine study.⁶ “We got this strange sequencing result, and we could only tell by using the mock communities,” he says.

Lumpers and splitters

Once the data have been harvested, the final task is to profile the various gene sequences and make a determination about which organisms are present. But as with other aspects of the 16S procedure, there is some debate regarding how this should be done. Traditionally, amplicons are computationally clustered into what are known as operational taxonomic units, or OTUs, based on the similarity between pairs of sequences—typically, a cutoff is set at 97% identity. This gives researchers the ability to efficiently organize ribosomal sequences in a way that allows queries against existing reference databases, or comparisons to determine which organisms are present or absent in different specimens. However, it’s also something of a blunt instrument from a taxonomic perspective. “You have multiple organisms, possibly going up even to the genus or family level or above, that all basically get collapsed into one OTU and you’re no longer able to distinguish them,” says Gohl.

An alternative method now gaining traction entails the use of ‘amplicon sequence variants’ (ASVs), which essentially allow scientists to look at the entirety of the amplified sequence without any computational clustering. From Fuhrman’s perspective, this captures much richer information about 16S content and allows greater comparability across data sets than OTUs, which are more subjective and can vary from lab to lab depending on the clustering strategy used. “To me, you can’t possibly go wrong other than having too much data,” he says. “You can always aggregate your highly resolved data, but you can never take data that were reported at low resolution and then resolve it.” On the other hand, this approach can also create a lot of work for researchers, who must grapple with distinguishing meaningful variants from errors introduced during sequencing, as well as quirks of the micro-

bial genome, in the process of interpreting their data. “For example, *E. coli* has seven copies of its 16S gene, and those copies are not identical,” says Schloss. “So you could potentially split *E. coli* into multiple ASVs.”

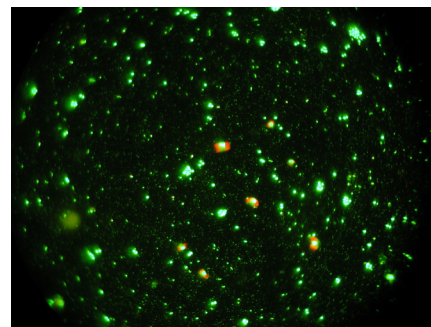
This factor can also impede quantification—Fuhrman notes that in bacteria, the number of 16S gene copies can range from a couple to a few dozen, and some eukaryotic microbes have copy numbers ranging into the thousands. Software tools can help correct for this, but more generally, a PCR-based method such as 16S amplicon analysis will never be entirely quantitative. Still, it can deliver extremely valuable and trustworthy relative measurements. “Based on what we’ve seen in terms of people who have healthy colons versus those who have colon cancer, we can take a new gut sample and classify it based on their microbiome,” says Schloss. “It’s quantitative enough to do that.” From Knight’s point of view, once the sources and manifestations of bias are known, consistency is what really determines whether an experiment will be scientifically informative. “If you’re using the same method across different samples, it doesn’t really matter that there are biases because you’re just trying to compare and find out whether the communities are identical or not,” he says.

The biased truth

Unfortunately, there is no such consistency across laboratories, and comparison and integration of data from different groups



Per Nielsen collects soil and water samples on the Danish coast for microbiome analysis.



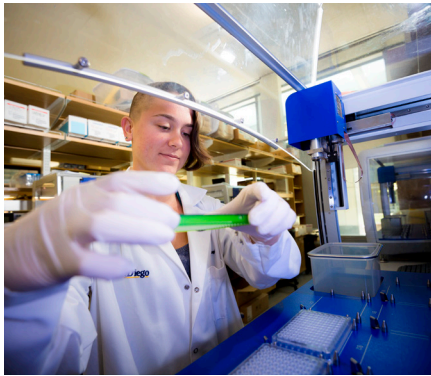
J. Fuhrman, USC

16S-based clustering can help researchers characterize the diverse constellation of bacteria, archaea and protists found in a sample of seawater.

remains a formidable task. “There really is a huge amount of variation between different labs and their data-generation processes,” says Gohl, “and the technical noise that can be introduced into these microbiome measurements is on the scale of the biological variability that people are trying to measure.”

Several multi-institutional initiatives have been launched to assess sources of variability and establish best practices for reproducibility, including the Earth Microbiome Project (EMP) and the Microbiome Quality Control (MBQC) project. “The beauty of EMP was to make the data as comparable as possible,” says Fuhrman, who was on the project’s advisory board. “We had 300 or so authors with all this incredible data, analyzing it all the same way.” Over the course of eight years, the EMP has helped establish robust protocols for conducting 16S rRNA analysis, but this is not a one-size-fits-all experimental approach. Indeed, the EMP received considerable pushback early on for compelling participants to use a common DNA preparation method regardless of the sample type being studied. This kind of uniformity is helpful for standard-setting, but real-world research often requires tweaking and adaptation of protocols to get the most out of a particular sample.

Thus, there is a parallel push for microbiome researchers to share their experimental and analytical procedures as structured metadata. Most investigators are dutiful about uploading 16S data into public repositories such as the Sequence Read Archive, but this alone is insufficient. “Uploading raw sequences is quite useless in terms of actually sharing what you’ve done in a paper,” says Quince. “You can only do meta-analyses if you’ve actually put the metadata on the server as well—which hardly anybody does.”



UC San Diego Health

A student at UCSD's Center for Microbiome Innovation, which is coordinated by Knight.

It remains unclear what combination of carrots and sticks might incentivize the community to make this extra effort, but Knight is hopeful that the ideas of standardization and transparency will gain momentum. “The really good news is that a lot of people have used a relatively small number of protocols,” he says. “Making it easy to integrate your data if you use those protocols will play

a very important role in making it possible to reuse a lot of data.”

Even though the methodology is becoming more robust and reliable, there are scientific questions for which 16S analysis remains ill-suited. For example, it is generally inadequate for species-level identification of organisms, and it offers no insight into the functional properties of a microbial community—for example, which metabolic processes are active. For such experiments, shotgun is a better way forward, although even that method may be supplanted as ‘long-read’ sequencing technologies mature. “With time, certainly you will not only get the full-length 16S but probably the full genome,” says Nielsen. “That’s just a matter of a few years.”

But there’s still a lot of life left in 16S, and labs should think carefully before making the investment to switch. “The process of getting a protocol that you have set up nicely on amplicon working for shotgun is not trivial,” says Knight. “It’s taken us several years and several million

dollars.” And as always, the method one chooses needs to match the question one intends to ask—and sometimes 16S is the best tool in the box. “Every method has biases, even shotgun metagenomics,” says Schloss. “And if you’re looking for a community-level appreciation of what’s going on in a microbiome sample, then even though there’s biases, we don’t have a better alternative.”

Michael Eisenstein is a freelance science writer based in Philadelphia (michael@eisensteinium.com).

1. Woese, C.R. & Fox, G.E. *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090 (1977).
2. Parada, A.E., Needham, D.M. & Fuhrman, J.A. *Environ. Microbiol.* **18**, 1403–1414 (2016).
3. Gohl, D.M. *et al. Nat. Biotechnol.* **34**, 942–949 (2016).
4. Eloë-Fadrosh, E.A., Ivanova, N.N., Woyke, T. & Kyrpides, N.C. *Nat. Microbiol.* **1**, 15032 (2016).
5. Karst, S.M. *et al. Nat. Biotechnol.* **36**, 190–195 (2018).
6. Yeh, Y.-C., Needham, D.M., Sieradzki, E.T. & Fuhrman, J.A. *mSystems* **3**, e00023-18 (2018).