

# With metagenomic sequencing, no pathogen can hide

## William Check, PhD

**January 2017**—Detecting pathogenic organisms with PCR has become a staple of the clinical microbiology laboratory, so much so that it seems like it has always been there. A more advanced molecular technique—unbiased metagenomic next-generation sequencing—will increasingly become a part of infectious disease diagnosis because it has several advantages over PCR. While it will be demanding to perform at first, it, too, may become a standard method in the clinical microbiology laboratory.

Platform	Throughput	Speed	Per run costs	Per base costs
MISeq	+	+++	\$	\$\$\$
NextSeq	++	++	\$\$	\$\$
HiSeq	+++	+ / ++	\$\$\$	\$
Ion Proton	++	+++	\$\$	\$\$

Whether samples need to be run individually or can be batched influences costs and workflow. Compiled by Robert Schlaberg, MD, MPH.

“In contrast to most of our current tests, you can use a metagenomic approach to find any and all potential pathogens in a patient sample without having to know what you’re looking for,” says Robert Schlaberg, MD, MPH, an assistant professor of pathology at the University of Utah and medical director of infectious diseases at ARUP Laboratories. While PCR can be fast and effective, it needs to be targeted. “So it is difficult to use when a condition can be caused by many microbes and pathogens. In that situation, you need to use a battery of tests, which can become lengthy and expensive and often doesn’t yield a result.” Such is the case for pneumonia, sepsis, encephalitis, meningitis, and diarrhea.

“The advantage of a metagenomic approach is that you can start without having a hypothesis,” says Dr. Schlaberg, who spoke with CAP TODAY and presented in November at the Association for Molecular Pathology meeting on universal pathogen detection directly from specimens in the diagnostic laboratory. Another advantage of metagenomics is that it can detect bacteria, viruses, fungi, and parasites in one assay.

“In our lab, the process from sample to report takes about 1.5 days,” Dr. Schlaberg says of the complex workflow. “It takes a full shift at least to process the sample; sequencing takes overnight.” Still, he can envision a simpler workflow and a future in which metagenomic sequencing replaces many current tests. “Right now to get started is challenging,” he says, but it’s easy to forget that PCR was also once labor-intensive and demanding to run. “It will be the same as with PCR,” Dr. Schlaberg says. “Technology will evolve, and we will get to a point where metagenomic sequencing is done in most large labs.”

Dr. Schlaberg and colleague Mark Yandell, PhD, a professor of human genetics at the University of Utah School of Medicine and co-director of the USTAR Center for Genetic Discovery, have made a start toward the future by co-developing a semiautomated informatics package for analysis of metagenomic sequence data as a clinical microbiology detection tool. Their program, called TaxonomerDx, will be deployed in a few months at ARUP Laboratories. Its initial indication will be for pneumonia.

Unexplained infectious illness ranges from 40 percent upward for sepsis, infectious diarrhea, pneumonia, and febrile neutropenia. CNS infections top the list with an 80 percent “unknown” rate.

“Even for [CNS] cases where we know what causes an infection, we need a very sensitive test to detect an organism,” Dr. Schlaberg says. “Often with viral infections they are there for only a few days, so we could be collecting a sample after the organism has gone.” Then, too, noninfectious conditions can look similar to a CNS

infection. Dr. Schlaberg cites post-infectious encephalitis and autoimmune conditions as examples of conditions that may clinically look like an infection. Often when a patient presents to the hospital with a suspected infection, he or she is treated quickly, even before optimal specimens are collected. "So the antibiotics can be in the specimen and inhibit growth of the organisms in the lab." Many of these obstacles can be overcome by a metagenomic technique, like the Taxonomer-based technology platform the University of Utah has developed, Dr. Schlaberg says.

Unbiased metagenomic sequencing means all the nucleic acid in the sample is submitted for sequencing and analysis. "Unbiased" means all the nucleic acid is analyzed just as it is in the specimen, without being amplified.

In the metagenomic technique, as in all next-generation sequencing, sensitivity depends on the number of times each nucleic acid segment is sequenced, or read ("read depth"). External and internal controls are essential. "Many routine kits and reagents used in NGS processes are contaminated with microbial DNA and RNA," Dr. Schlaberg warns. "With unbiased metagenomics sequencing, we routinely detect these microbial nucleic acids. Using the right negative controls helps identify these as contaminants. This sounds trivial, but the importance is hard to overstate." He shares a case in which a novel virus found in clinical samples turned out to originate from contaminated silica-binding spin columns used for nucleic acid extraction (Naccache SN, et al. *J Virol.* 2013;87[22]:11966-11977).

Selecting an instrument for metagenomic sequencing involves tradeoffs among throughput, turnaround time, and cost. Dr. Schlaberg has tabulated relative throughput and speed for several instruments and both per run and per base costs (see box, this page). "In infectious disease diagnosis you need to be fast," he notes. "We use the Illumina NextSeq platform. Of the available instruments, it gives us the best combination of rapid turnaround time, per base sequencing cost, and per run cost."

This is just a start, though, he says, adding, "We are always looking for ways to speed up laboratory protocols and sequencing times."

Still needed, he says, are user-friendly data analysis solutions for the diagnostic setting. Ideally they should require minimal bioinformatics expertise and be easily updatable and expandable. All current databases are imperfect and incomplete and contain errors, he cautions, and they require curation. "Informatics for metagenomic application used to be very slow."

"Until a couple of years ago, alignment-based methods were used for data analysis," Dr. Schlaberg explains. Because every one of millions of sequences from the patient sample needs to be compared to millions of reference sequences, these are computationally intense analyses and took days or weeks to complete. Advances in computer science made possible alignment-free analysis methods, which can be completed in minutes.



"We can't say for sure that these

organisms were responsible. But in about 30 percent of cases where many tests were all negative, we could find putative pathogens.”

**Taxonomer is the approach Dr. Schlaberg** and Dr. Yandell developed to solve these issues. It is publicly available through a Web-based user interface ([www.taxonomer.com](http://www.taxonomer.com); Flygare S, et al. *Genome Biol.* 2016;17[1]:111). “Taxonomer is based on alignment-free algorithms to analyze DNA and protein sequences by comparison to very large, curated databases,” Dr. Schlaberg says. Taxonomer analyzes more than 1 million sequences per minute. “Therefore, completing a standard analysis of metagenomic data takes just minutes, which brings this method into the realm of diagnostic application,” he says.

Through a startup company called IDbyDNA that Dr. Schlaberg and Dr. Yandell helped co-found, a diagnostic version of Taxonomer—TaxonomerDx—was developed for routine use in diagnostic laboratories. “TaxonomerDx contains an extensive suite of tools that turn this DNA search engine into a semiautomated diagnostic tool,” Dr. Schlaberg explains. This includes software that helps with sample management, laboratory processing, curated and validated reference sequence databases, interpretive algorithms and thresholds, comprehensive result review interfaces, and automatic result reporting features. “It provides a complete solution for all steps from sample to report,” he says.

They are using IDbyDNA’s TaxonomerDx platform now at ARUP Laboratories and say the first CLIA-validated test will be available for clinical use in the first half of this year. The first indication will be testing of lower respiratory tract specimens from patients with pneumonia. Even though the method itself is fairly universal, “diagnostic protocols need to be geared to the intended use of the test,” he says.

The pneumonia version of TaxonomerDx was validated in accordance with the recommendations of a CAP committee (Schrijver I, et al. *J Mol Diagn.* 2014;16[3]:283–287). In silico validation consisted of more than 50,000 virtual samples, while wet bench validation was done on more than 400 samples. The test was fully validated for about 200 viral, bacterial, and fungal pathogens. Data have been submitted for publication.

Dr. Schlaberg explains one way in which the test has been customized for pneumonia diagnosis. “We can compare challenges to those in constitutional genetics. People used to sequence one or a number of genes in patients with suspected inherited diseases. Then they started sequencing whole exomes and genomes. When you do that, you find things you were not looking for that you have to interpret in light of the patient’s presentation.”

An analogous problem arises with metagenomic sequencing. “In doing metagenomics, you can find microbes that are not pathogens but part of the normal flora,” Dr. Schlaberg says. “What we decided to do to simplify the results was to select what we think are the relevant pathogenic organisms that we want to report in this context and to select out those that we wouldn’t want to report. So we only look at a limited range of organisms.”

Another criterion for identifying a causative organism is to evaluate its relative abundance in the specimen. “The basic problem is to prioritize. Our approach,” Dr. Schlaberg says, “is to select the most relevant organisms we can reliably detect with relevance defined based on an extensive literature review.”

In some cases, manual review will still be necessary. “There is going to be some learning,” he says, “and the technology will become more automated based on results of previously analyzed samples.” This is why having analyzed hundreds of samples as part of research projects and an extensive validation is helpful, he adds. “But there is going to continue to be a need for expert review of the data, at least in some situations.”

In the validation study, accuracy of TaxonomerDx as measured by agreement with results of antigen detection tests, culture, and PCR (performed previously on the specimens) and after PCR confirmation of discrepant results ranged from 80 percent to 96 percent for different groups of viruses and bacteria. (Accuracy for yeast was 67

percent.) “Most of the time it does work across a wide range of organisms,” he says, noting there are limitations to a study like this, which uses previously tested samples that have been stored for a time.

In addition to accuracy, one strength of the method, Dr. Schlager says, is that it detected many organisms not previously reported or suspected.

To underscore this value, he shares examples in which IDbyDNA’s metagenomics platform identified previously unrecognized viral, bacterial, and fungal pathogens in bronchoalveolar lavage samples from immunocompromised children. Pathogens included human parainfluenza virus type 4, *Streptococcus pneumoniae*, *Klebsiella pneumoniae*, *Pneumocystis jirovecii*, and *Fusarium* and *Mucor* species.

Comparison studies with conventional tests were also informative. One involved 109 nasal swabs, 42 positive by a respiratory virus panel and 67 unselected specimens. Taxonomer showed high agreement with the commercial RVP and higher diagnostic yield (Graf EH, et al. *J Clin Microbiol.* 2016;54[4]: 1000-1007).

Taxonomer was also evaluated in a subset of samples from a large Centers for Disease Control and Prevention study. The study aimed at identifying causes of pneumonia in children after implementation of vaccination programs for several common respiratory pathogens. In about 20 percent of the children all tests were negative. Dr. Schlager and colleagues analyzed some of those negative samples by their method. In about 30 percent of the children in whom no pathogen had been identified by conventional methods, Taxonomer found a possible pathogen, mostly viruses with one bacterium (*Chlamydia trachomatis*). “We can’t say for sure that these organisms were responsible. But in about 30 percent of cases where many tests were all negative, we could find putative pathogens,” Dr. Schlager says.

What will be the context for ordering TaxonomerDx-based tests? While Dr. Schlager notes that it will be clinicians who choose when and for which patients to order the test, his own recommendation would be to use it in conjunction with rapid tests such as PCR initially. While it provides a much broader scope, metagenomic testing is not as rapid right now, and a better understanding of its performance in clinical use would be desirable. “Probably the responsible thing at this point would be to use it in conjunction with conventional tests,” he says. “I expect that to change as we get more experience.” Current tests are negative in 20 percent of children and 60 percent of adults with pneumonia, he adds.

“We think that at first TaxonomerDx-based tests will be most useful in critically ill ICU patients. Then it will move toward first-line use.”

[hr]

*William Check is a writer in Ft. Lauderdale, Fla.*