Technology comes to typing

As mass spectrometry makes inroads into pathogen identification in the clinical laboratory, deep sequencing—even nanopore sequencing—is waiting in the wings. Jeffrey L. Fox investigates.

In September, a team of investigators from Warwick Medical Center in the UK and the Medical Research Council's unit in Gambia, reported using shotgun metagenomic sequencing to characterize a tuberculosis strain directly from sputum samples[1]. This followed an earlier report using next-generation sequencing (NGS) to solve the riddle of the microbe responsible for a case of intractable encephalitis in a young man[2]. Over the past 12 months, however, it has been another high-tech approach to microbial typing—mass spectrometry (MS)—that has changed practice in clinical microbiology. Since its regulatory approval, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS has been transforming how pathogens are identified in the clinical laboratory.

MS and NGS are recent arrivals on the clinical laboratory scene, and both are creating excitement among researchers, particularly the possibilities of nanopore sequencing (**Box 1**). But integrating these technologies into existing microbiology workflows poses daunting challenges, particularly at a time when clinical medicine is adapting to healthcare reform and tightening budgets.

MS takes flight

Although various molecular technologies for pathogen identification have infiltrated clinical laboratories—PCR, for example, or immunoassays—these techniques are specific for particular pathogens, and thus not every species can be identified in a single assay. Educated guesses must be made in choosing which assay to use. In addition, the assortment of available assays is limited and costly.

MALDI-TOF MS provides something closer to a universal pathogen detection device; it identifies crucial proteins with specific signature patterns and compares them to a collection of patterns that have been deposited in a database. Each time a clinical specimen is analyzed, its protein ion pattern, or signature, is automatically and nearly instantaneously compared to standard patterns within the database.

The major corporate players developing MALDI-TOF MS for the clinical microbiology sector include bioMérieux of Durham, North Carolina (with international headquarters in Paris) and Bruker of Billerica, Massachusetts. The bioMérieux system, called Vitek MS, was approved for use as a clinical diagnostic system in the EU in 2012 and then in 2013 by the US Food and Drug Administration (FDA). The system is capable of detecting Gram-positive and Gram-negative bacteria as well as fungal pathogens, according to Bert Top, senior marketing manager of US clinical marketing at bioMérieux. In fewer than 12 months since FDA approval, the company has placed more than 100 of its systems in clinical laboratories throughout the US. The early users are mainly high-profile laboratories at academic hospitals or at major metropolitan hospitals, he says. The initial investment for the system is about $300,000, whereas the reagent costs for testing a patient specimen come to a mere $1 or so. Such testing requires growth of samples on plates, before they are picked, sometimes acid-treated, and then placed on a matrix before going into the instrument for automated reading and analysis.

The Bruker system, MALDI Biotyper, was approved in 2009 in Europe, and by the FDA in 2013 for Gram-negative bacteria only, and is available for clinical microbiology in many other countries. "For the past seven years, Bruker has been working on continuous innovation in the field of MALDI-TOF–based microbial identification, bringing the MALDI Biotyper platform into clinical routine laboratories," says Frank Laukien, president and CEO of Bruker.

According to the company, the overall accuracy of the MALDI Biotyper CA System "is comparable to" that of 16S ribosomal RNA gene sequencing, based on multisite evaluations of its performance with clinical samples. Midway through 2014, nearly 1,200 such systems were either owned or are being leased worldwide. To further expand the system's clinical applicability, Bruker is sponsoring a clinical trial to evaluate the system's capabilities for identifying rare Gram-negative bacteria, as well as aerobic Gram-positive bacteria, yeasts and anaerobic microorganisms cultured from human specimens, according to George Goedesky, Bruker Daltonics vice president for microbiology.

Both of these systems are capable of identifying additional microorganisms through specialized, user-constructed databases, and both manufacturers are continuing to expand their proprietary databases to meet regulatory requirements and expand the scope of their respective analytic systems. Although neither system is adept at assigning antimicrobial drug susceptibilities, typically the next crucial step once an infection-causing microbe is identified, both companies are working on this challenge and already provide alternative approaches to solving this part of the diagnostic puzzle. Clinically minded researchers outside the companies also are developing approaches to analyzing samples for drug-resistance markers.

For instance, bioMérieux has done feasibility studies to explore how its system might be adapted for antibiotic drug-susceptibility testing, according to Christine Ginocchio, vice president for global microbiology affairs at the company. Nonetheless, using MALDI-TOF MS to identify the pathogen, when used with a more conventional approach to determine antibiotic susceptibility, speeds the overall process from over 18 hours to about 6 to 8, she says. For some patients, particularly those with

Oxford Nanopore


Pacific Biosciences

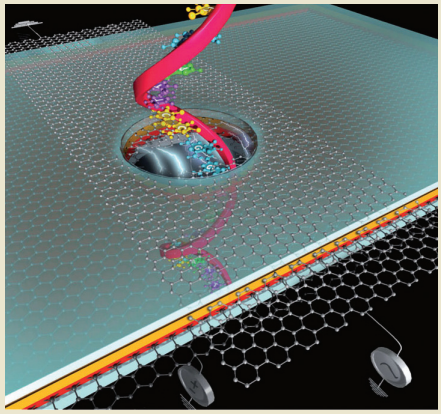
Single-molecule sequencing is making a play for microbe identification and detection.

Box 1 Going long

Conventional wisdom and the marketplace make Illumina the undisputed leader in the sequencing arena, particularly outside of microbiology applications. Yet short reads generated in the Illumina suite of sequencers can pose challenges in terms of chromosome assembly and amplification bias. Single-molecule, long-read platforms, such as PacBio's Single Molecule, Real Time (SMRT) instrument and nanopore-based platforms in development by UK's Oxford Nanopore Technologies and Genia of Mountain View, California, address some of these shortcomings and complement short-read Illumina sequencing.

According to Mark Pallen at the University of Warwick and Warwick Medical School, the field is focused on "squeezing the most out of short reads." In part, this is due to existing investments in the technology at many sequencing centers, but according to Pallan, "We all got suckered into thinking that [short reads are] all there is. We used to get longer reads with Sanger. If we could have reads of tens of kilobases or even longer, it would be wonderful."

Nick Loman, of the Institute of Microbiology and Infection at the UK's University of Birmingham, is one of hundreds of researchers participating in a research program for Oxford Nanopore's sequencer, termed the MinION Access Programme (MAP). According to Loman, long reads for the system make it easier to put together the pieces of complex genomes. Consider, for example, the work done by NIH's Segre on *Klebsiella*. Short-read technology couldn't resolve plasmid from chromosomal genes, whereas fully contiguous genome sequence data "make the analysis much more straightforward," Loman says.



Solid state nanopore architecture for DNA sequencing. Reprinted by permission from Macmillan Publishers Ltd: Venkatsean, B.M. & Bashir, R. Nanopore sensors for nucleic acid analysis. *Nat. Nanotechnol.* **6**, 615–624 (2011).

In addition, with single-molecule sequencing, you don't have to wait for a run to finish before analyzing it. With clinical samples from a collaborating hospital in Birmingham, Loman can say within 20 minutes what species he is looking at; within an hour, he gets a rough serotype; and within 2 hours, he can say whether it's part of an outbreak.

And useful information can be obtained with single reads of more than a few hundred bases. "Even if you have a quite inaccurate read, if it's very long, it's giving you a lot of information that is completely unambiguous. It'll tell you that this gene follows this gene follows that gene. And it doesn't matter if it's 80% accurate or 90% accurate," says Loman.

PacBio's SMRT technology and Oxford Nanopore's MinION each sequence single molecules, but the former measures base incorporation into a DNA strand by fluorescence, whereas the latter passively 'reads' the bases that comprise a single molecule of DNA through changes in current through a pore. SMRT technology reads fluorescence as tagged bases are incorporated into a DNA template, aided by a proprietary technology called zero mode waveguides, detecting single fluorescent molecules against a backdrop of unincorporated nucleotides. It's been offered for sale since 2011 and is installed in over 100 labs worldwide, according to Korlach. But unlike the portable MinION, it requires a one-ton box to house the detector.

The MinION system is so small (~4 inches in length) it fits into a USB port on a laptop. The solid-state pore senses changes in current as a DNA strand passes through, and the fluctuation is interpreted and/or resolved into base information. Although the MinION is not yet for sale, Oxford Nanopore has invited hundreds of laboratories to try, use, debug and develop their novel applications of NGS in their open access MAP, according to the company's chief technology officer Clive Brown. Although some see this approach as a risky proposition, Brown counters, "Our philosophy was to trust the participants, believing that they are well intentioned and motivated, and that they would also like to see the technology become successful and thus help themselves in their own research objectives."

Laura DeFrancesco

blood infections leading to sepsis, this faster turnaround makes for a big difference in terms of clinical outcomes, she says. And Bruker says that research-use-only versions of the MALDI Biotyper allow "selected high-value antimicrobial resistance tests."

New dimensions from sequencing

Although MALDI-TOF MS is already being adopted by clinical microbiologists, commercial applications of NGS in the field are also gathering pace. Two major players seeking to adapt NGS for microbial pathogen identification are Pacific Biosciences in Menlo Park, California, and Illumina in San Diego. No NGS system is yet approved by FDA for routine use in diagnosing infectious diseases; however, last November, the agency approved the Illumina MiSeqDx as a device to detect DNA changes in the cystic fibrosis transmembrane conductance regulator gene (*CFTR*) for use in diagnosing cystic fibrosis. This approval opened the door for using NGS in clinical settings.

Meanwhile, the Illumina and Pacific Biosciences NGS platforms are being used extensively in microbiology research to determine the genomic sequences of microorganisms and to identify emerging pathogens in infectious disease outbreaks. And the possibility of doing patient diagnoses was realized when a team of researchers at the University of California, San Francisco (UCSF), employed the Illumina system to diagnose the encephalitis patient mentioned above. In that case, Charles Chiu, Joseph Derisi and their collaborators analyzed a cerebrospinal fluid specimen from the patient, who was being cared for at the University of Wisconsin, Madison. Notably, the specimen had tested negative for pathogenic agents by conventional methods, but analysis using the Illumina MiSeq DNA sequencer indicated that the sample contained DNA sequences from the relatively exotic bacterial pathogen *Leptospira santarosai*, making it the likely culprit responsible for the patient's encephalitis.

"This is one test to rule them all," Derisi exclaims, praising the power of DNA sequencing for identifying microbial pathogens. Even so, Chiu is quick to point out that the testing

approach that they followed is neither validated nor approved by regulatory agencies for routine use in diagnosing infections in patients. "This case was a test run for how you would validate and deploy a clinically approved test," he notes. Chiu is taking steps to gain approval for this approach, with plans to offer such tests at UCSF for diagnosing infectious diseases. But gaining approval for clinical testing at UCSF is several steps away from the Illumina DNA sequencing platform itself receiving FDA approval for general use in clinical laboratories.

The same status holds true for the DNA sequencing system being developed by Pacific Biosciences for identifying microbial pathogens, according to CSO Jonas Korlach. "We are serving the research-only-use environment, so I can't make predictions regarding FDA," he says; however, there is little doubt that the company is moving in that direction, signaled in large part by its distribution partnership with Roche Diagnostics of Pleasanton, California, which was announced in October 2013. "We are interested in this area and hope to leverage Roche's expertise in clinical diagnostics to bring this technology into that space," he says.

Without need for regulatory approvals, NGS is being used increasingly to analyze infectious disease outbreaks and for surveillance, according to Susan Knowles, who is senior market manager at Illumina. "It has the highest resolution to trace back outbreaks," she says. "Traditional methods and even MALDI-TOF MS don't have that ability." Sequencing gives you every gene, every conceivable marker, whereas MALDI-TOF MS relies on a handful of signature proteins. In some outbreaks, knowledge of a handful of particular genes can prove dispositive, but in other cases, microbiologists and physicians may need to look broadly before determining which virulence genes in a particular strain might account for an outbreak. A notable example was a foodborne outbreak in Germany several years ago where the culprit proved to be a novel strain of *Escherichia coli* O157, and its novelty turned on the presence of several specific genes[3]. MALDI-TOF MS and multilocus typing would provide some information, but sequencing of the genome to find those particular genes proved highly informative. Public health investigators at FDA and at the Centers for Disease Control and Prevention (CDC) in Atlanta are using genomic sequencing for these purposes, as are investigators at universities and in hospital settings. In 2012, for example, FDA awarded the company a five-year, $17-million contract to provide sequencing systems and reagents for analyzing samples from produce for *Salmonella* and other foodborne pathogens.

CDC is using genomic sequencing instruments to track foodborne pathogens and is also using them to analyze other kinds of infectious disease outbreaks, according to Peter Gerner-Smidt, a branch chief in the CDC Division of Foodborne, Bacterial and Mycotic Diseases. Genomic sequencing is "more efficient and precise" than multilocus typing during outbreaks and provides a better means for determining what pathogen might be causing sporadic infections as well, he says. Genomic sequencing also provides "better resolution" than does pulsed field gel electrophoresis, the current standard tool at CDC for analyzing bacterial pathogens responsible for foodborne disease outbreaks, he says. "We've also used genomic sequencing to exclude food sources as not being responsible for disease." Embrace of this technology for public health surveillance duties "is going to happen," he adds.

For infection control and surveillance in healthcare settings, genomic sequencing "is available now, and it's superior to other technologies," says Lynn Bry, director of the Center for Clinical and Translational Metagenomics at Brigham & Women's Hospital of Harvard Medical School in Boston. Genomic sequencing is also valuable in clinical settings when dealing with difficult-to-culture or notoriously slow-growing pathogens such as *Mycobacterium tuberculosis*, or when specimens test negative even though a microbial pathogen is likely responsible for a patient's illness, she adds.

In a step closer to patient care, investigators working in the Clinical Center at the NIH are using sequencing to help in analyzing whether patients being treated for other illnesses acquired their infections while on the hospital wards or whether they brought particular pathogens into the center with them, according to Julie Segre, a senior investigator with the National Human Genome Research Institute. During an outbreak involving antibiotic-resistant *Klebsiella pneumonia* several years ago, she and her collaborators were able to track the movement of antibiotic resistance–carrying plasmids among patients, pathogens and in the hospital environment using long-read genome sequencing. The approach not only uncovered surprises about the epidemiology of carbapenem resistance, but also led to the identification of a new carbapenemase-encoding plasmid of "potentially high clinical impact"[4] (**Box 1**). "We used the sequencing data to overrule some scenarios and to see how the pathogen was being transmitted," she says. Such information is critical for "seeing what practices needed to be changed to prevent such transmissions. We needed to understand what was happening in the hospital." In this case, knowing that the pathogen was introduced separately spared the infection control staff the time and costs of revising practices that were working as they should.

Similarly, whole genome sequencing is being used to determine whether plasmids or other mobile genetic elements carrying antimicrobial drug-resistance genes are being transferred among the bacterial pathogens infecting patients within the NIH Clinical Center, according to Segre. "The plasmid sequences are the hardest data to assemble," she says. "Without sequencing, we have no idea how drug-resistance factors are trafficking through the hospital." Recasting this research effort into a reliable diagnostic procedure will depend in part on development of suitable DNA databases to pinpoint what is really happening, she adds. "It's still in an experimental phase."

Whereas, increasingly, investigators are relying on genomic sequencing to track pathogens in healthcare settings or among the general population, the technology is not yet suited for routine diagnostic use on individual patients, according to Bry, Segre and others. For one thing, the cost for the equipment, reagents and supporting software requires a substantial capital investment—something shy of $1 million plus a comparable investment in bioinformatics. That, along with sample preparation costs, comes to plenty more than current guidelines will allow laboratories to be reimbursed per patient, which amounts to a mere $20 per test, Bry says.

"We see high-throughput sequencing as something we'll want to incorporate when it's 'do-able,'" says Joe Campos, interim chief of the Division of Laboratory Medicine at Children's National Health System in Washington, DC. But its place in clinical laboratories faces "limits now," particularly because of the current emphasis on "cost containment in the real world of the lab and hospital," he adds. "Cost effectiveness is tough to justify except for special cases." Beyond costs, this analytic method needs to be very much more automated if it is to prove useful for identifying pathogens in clinical specimens. "There's a wide spectrum of microbiology labs out there, but the easier it becomes, the more widely adopted it will be," he says.

"I'm familiar with the constraints that economics places on clinical labs," says Ginocchio of bioMérieux, who until early in 2014 ran the clinical laboratory of North

Shore-LIJ Health System, which does hundreds of thousands of infectious disease tests annually. "But you can't look only at initial costs. You need to look also at downstream savings in costs of reagents, technicians' time and the dwindling work force, as well as the effect on the hospital and the impact on clinical outcomes."

Although Ginocchio was referring specifically to MALDI-TOF MS, her comments apply to genomic sequencing in terms of initial costs and eventual returns on investment through improved health outcomes. "I understand the cost of high-throughput sequencing," Segre of NIH says. However, hospital patients who develop difficult-to-treat or life-threatening infections can prove very costly to the system, she adds. "My costs from sequencing seem orders of magnitude less than the costs with those patients." For infection control in hospitals, DNA sequencing is superior to other approaches, Bry says. Moreover, it can be very helpful in cases where other analytic methods report a culture to be negative, particularly if turnaround times could be reduced, allowing patients to be treated with appropriate drugs more quickly.

FDA comes into play

Regulatory approval stands as another major barrier to wider use of microbial genomic sequencing for diagnostic purposes. But FDA officials, who are gearing up for this technology, convened a public workshop in April 2014 to develop a better sense of what needs to be done before high-throughput sequencing is ready for the wide gamut of clinical laboratories[5]. Agency officials also formed a working group and are collaborating with several outside groups to collect sequencing data for hundreds of clinically relevant bacterial pathogens, according to Peyton Hobson from the FDA Center for Devices and Radiological Health. Efforts are underway to set up searchable databases within the public domain and to set standards—"suitable for a regulatory body"—for those data, he says. To achieve such goals, FDA plans to consult outside experts and to work together with industry.

Meeting the broad bioinformatics challenge at hand might not prove so easy, according to Tom Slezak of Lawrence Livermore National Laboratory in Berkeley, California, who is working with several federal agencies to analyze this and other bioinformatics issues. For one thing, he says, "It's not all that simple to screen out human reads from samples. Even a tiny amount of contamination can cause real problems." Further, distinguishing pathogens from closely related microorganisms that humans may carry as part of the human microbiota also poses technical problems. For example, the bacterial pathogen *Staphylococcus aureus* is difficult to tell apart from the more benign *Staphylococcus epidermidis*, a very similar species also found on skin. Another problem is "sample carry-over contamination," he adds, when going from the analysis of one specimen to another. "There are a lot of challenges, but I think we can pull this off. It is a good time for precompetitive cooperation with companies in the sequencing industry."

The ever-expanding abundance of sequencing data is part of the problem, suggests David Lipman, director of the NIH National Center for Biotechnology Information. Data describing many different isolates of the same microbial species can be counted on to contain discrepancies, no matter whether the analyses come from many different laboratories or all from one. Such data sets contain mistakes as well as genuine differences between microbial isolates such as single-nucleotide polymorphisms (SNPs). "We want clarity but also want the data to be comprehensive," he says. "How do we get both?" Finding fast and efficient ways to compare isolates with SNPs to determine if they derive from the same clone or are different will be one way to reduce such analytic "noise."

The National Institute of Allergy and Infectious Diseases (NIAID) at NIH has its own mandate to develop accessible databases based on the microbial genomics research that it funds, according to Vivien Dugan of the NIAID Office of Genomics and Advanced Technologies. One of the big challenges is integrating different data types, a challenge that includes coping with quality differences and a lack of standardization, she says. "Clinical data is a work in progress." Nonetheless, NIAID and the microbial genome research centers it is supporting are moving in that direction while also trying to incorporate data pertaining to antibiotic resistance, another important component of this effort if genomic sequencing is to prove effective in clinical laboratories.

A future for both platforms?

MALDI-TOF MS can be expected to gain broader acceptance in clinical laboratories, particularly as research outside and within companies continues to expand the range of microbial pathogens that will be subject to such analysis. Whether the technology can also be extended to reveal drug-resistance traits is another matter. In such cases, genomic sequencing could well have the upper hand, at least for detecting the genes encoding well-established antibiotic resistance traits.

Although the capital costs for genomic sequencing remain high, the costs for determining individual microbial genomes continue to fall, and now cost as little as $100 per sequence, according to Alexander McAdam, director of infectious diseases at Boston Children's Hospital. "Sequencing is 'slow' in batches but it's getting cheaper, faster and easier," he says. Although "unlikely" to replace MALDI-TOF MS for routine analyses, it may be superior already for identifying "novel" pathogens, detecting virulence genes, and not only for identifying drug-resistance genes but also for providing insights into drug-resistance mechanisms and maybe new targets at which to direct novel therapies. "There are big hurdles to overcome as we try to understand how to use these data, but we need to get ready now," he says.

In the face of such optimism, it is important not to forget that both public health laboratories and hospital laboratories are faced with budget cuts, and they are likely to face further decreases because of a widely held "perception that labs are being overpaid," says clinical microbiologist Vickie Baselski from the University of Tennessee Health Science Center in Memphis. However, such laboratories "don't make money but save money," she adds. "Things are not good and, hence, new technologies face delays. It always comes down to money."

Jeffrey L Fox, Washington, DC

1. Doughty, E.L. *et al. PeerJ* **2**, e585 (2014).
2. Wilson, M.R. *et al. N. Engl. J. Med.* **370**, 2408–2417 (2014).
3. Loman, N.J. *et al. J. Am. Med. Assoc.* **309**, 1502–1510 (2013).
4. Conlan, S. *et al. Sci. Transl. Med.* **6**, 254ra126 (2014).
5. FDA Public Workshop. Advancing Regulatory Science for High Throughput Sequencing Devices for Microbial Identification and Detection of Antimicrobial Resistance Markers. Silver Spring, Maryland, April 1, 2014.