Next-Generation Sequencing of Infectious Pathogens

Marta Gwinn, MD, MPH; Duncan MacCannell, PhD; Gregory L. Armstrong, MD

Next-generation sequencing holds potential for improving clinical and public health microbiology. In addition to identifying pathogens more rapidly and precisely than traditional methods, high-throughput technologies and bioinformatics can provide new insights into disease transmission, virulence, and antimicrobial resistance. The US public health system is transitioning to next-generation sequencing from an older molecular biology processes with a single workflow that accommodates a wide array of pathogens.

How Next-Generation Sequencing of Pathogens Works

Next-generation sequencing is a versatile technology, broadly applicable to viruses, bacteria, fungi, parasites, animal vectors, and human hosts. Choosing among available methods depends on sequencing objectives and involves tradeoffs in accuracy, efficiency, and cost. For routine sequencing, most US clinical and public health microbiology laboratories use short-read sequencing platforms, which produce sequence fragments up to 1000 base-pairs long. Although microbial genomes are generally smaller and less complex than human genomes, long-read sequencing technologies (such as single-molecule real-time sequencing) are useful for constructing complete, highly accurate genomes and sorting out plasmids, repeats, and other complex regions.

A different approach, nanopore sequencing, relies on threading individual DNA or RNA molecules through engineered protein nanopores and monitoring the electric current across each pore. The first such commercially available instrument offers relatively long sequence reads and allows data analysis to begin while sequencing is still in progress. Early limitations in throughput and accuracy have been mitigated by continued improvements in hardware and reagents. Because of device portability, fast sample preparation, flexibility, and relatively low cost, nanopore sequencing is becoming a feasible first-line strategy for pathogen sequencing in clinical and public health settings.

The transformation of raw sequence data into actionable information is complex and computationally intensive (Figure). The first step is typically to assemble shorter fragments into a complete sequence, either by mapping against a known reference genome or by assembling the sequence de novo using overlapping reads. Comparing the assembled genome with reference strains facilitates many different inferences, such as pathogen identification, high-resolution strain typing, and prediction of important phenotypic characteristics (eg, virulence, antimicrobial resistance). Well-curated and up-to-date reference databases are crucially important because microbial pathogens evolve rapidly and bacteria can exchange plasmids—often encoding virulence and antimicrobial resistance traits—across strains and species. Assembled genomes can be compared with others to look for phylogenetic clustering as evidence of transmission. Each step—assembly, strain typing, phenotyping, and clustering—requires different bioinformatics tools that must be harmonized into a consistent workflow.

Important Practice Considerations

In public health, next-generation sequencing offers crucial advantages for surveillance and outbreak investigation in terms of speed and resolution of sequence differences. For example, the transition to next-generation sequencing from an older molecular method (pulsed-field gel electrophoresis) is well underway in PulseNet, the foodborne disease surveillance system maintained by CDC and its public health partners. PulseNet is now able to detect outbreaks earlier, to distinguish clusters of related cases more accurately, and to link illnesses to potential contaminated food sources more quickly.

Integrating pathogen genomics with epidemiology is enhancing public health efforts to prevent transmission of communicable diseases, such as tuberculosis. Genotyping tuberculosis isolates can corroborate transmission inferred from contact investigations or suggest connections among apparently unrelated cases, helping health departments to better focus their resources. Next-generation sequencing has the potential to yield information on
likely anti-mycobacterial drug susceptibility more quickly than conventional testing, enabling more specific and timely treatment.8

Next-generation sequencing data are amenable to standardization and sharing, important advantages for global health partnerships like the World Health Organization’s influenza surveillance system. An open, “sequencing first” approach can help produce timely and accurate data for selection of candidate influenza vaccines, quickly identifying prevalent variants while monitoring the dynamics of co-circulating viral populations.1

Next-generation sequencing also offers advantages for challenging field investigations. For example, a research team from the United Kingdom packed a nanopore sequencing laboratory into standard luggage for transport to Guinea during the 2015 Ebola outbreak.9 During an 8-month period, they sequenced 142 Ebola virus genomes on site, usually within 1 working day; data were transmitted to the cloud for analysis and results returned the next day. Despite significant logistical challenges, including unreliable electrical power and internet service, the team provided actionable information for epidemic response without exporting samples from the country.

US public health laboratories, with support from the AMD program, are rapidly adopting next-generation sequencing for surveillance and investigation of foodborne disease, tuberculosis, hepatitis C, Legionella, and other pathogens.2 Nevertheless, the transition from research to routine public health and clinical use will have to overcome substantial challenges.3 At the laboratory level, these include infrastructure, workforce development, efficiency, and cost. At a broader, systemic level, substantial efforts are needed to develop standard protocols, proficiency-testing programs, professional guidelines, and regulatory requirements.3

Value

Compared with conventional methods, next-generation sequencing increases speed, accuracy, and detail but also increases cost. For example, a CDC analysis estimated that next-generation sequencing cost approximately $150 to $250 per bacterial isolate, compared with $94 for pulsed-field gel electrophoresis. Consolidating workflows for multiple pathogens may improve laboratory efficiency and help offset this cost; however, the transition to next-generation sequencing also entails significant up-front investment in laboratory equipment, computer resources, and training. Much more information will be needed to evaluate the value of next-generation sequencing technologies for microbiology at patient, programmatic, and societal levels.

Evidence

Evidence-based guidelines exist for only a few specific, clinical uses of pathogen sequence data, for example, in selecting antiretroviral treatment for HIV infection. Informatative sequences from bacterial, viral, fungal, and parasite genomes are the basis for many new, nucleic-acid-based diagnostic tests, including “point-of-care” tests that bypass the microbiology laboratory completely. As multiplex panels for syndromic diagnosis (eg, for diarrhea) become more widely used, systematic efforts will be needed to assess their clinical validity and utility, as well as their effect on laboratory-based public health surveillance.

Bottom Line

Next-generation sequencing is transforming the public health approach to infectious diseases, as well as the treatment of individual patients. Better coordination in establishing quality standards, reporting, and interpretation of next-generation sequencing data could make these efforts synergistic.

ARTICLE INFORMATION

Author Affiliations: CFOL International, Centers for Disease Control and Prevention, Atlanta, Georgia (Gwinn); Office of Advanced Molecular Detection, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia (MacCannell, Armstrong).

Corresponding Author: Marta Gwinn, MD, MPH, CFOL International, Centers for Disease Control and Prevention, 1600 Clifton Rd, Mailstop E-61, Atlanta, GA 30333 (Mgwinn@cdc.gov).

Section Editor: W. Gregory Feero, MD, PhD, Associate Editor, JAMA.

Published Online: February 14, 2019. doi:10.1001/jama.2018.21669

Conflict of Interest Disclosures: Dr Armstrong reported that the Bill and Melinda Gates Foundation awarded the CDC Foundation a grant to do a landscape analysis of software to support pathogen genomics in public health. Dr Armstrong participated in the application and is helping to administer the grant. The funding goes entirely to the Africa CDC and to the University of Washington; no entity at CDC receives any of the funding. No other disclosures were reported.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the CDC/the Agency for Toxic Substances and Disease Registry. Use of trade names and commercial sources is for identification only and does not imply endorsement of the Office of Advanced Molecular Detection, National Center for Emerging and Zoonotic Diseases, CDC, the Public Health Service, or the US Department of Health and Human Services.

REFERENCES


