Editorials represent the opinions
of the authors and *JAMA* and
not those of the American Medical Association.

Metagenomics, Infectious Disease Diagnostics, and Outbreak Investigations

Sequence First, Ask Questions Later?

David A. Relman, MD

Revolutionary advances in DNA sequencing technology are radically changing the approach for studying and characterizing the microbial world. The sequencing of a microbial genome, which can now be achieved in preliminary form in hours, provides a wealth of information about the functional potential of the organism, its evolutionary relationships with other organisms, and clues about niche adaptation, and without the need for microbial cultivation or isolation.[1] Likewise, so-called shotgun or metagenomic sequencing, which is the high-throughput simultaneous sequencing of random fragments from complex mixes of different genomes, provides insights into the potential activities of a microbial community, possible interactions between microbial community members, and the nature of their relationships with their environment (eg, a human host).

However, the use of this technology to solve the day-to-day needs of clinicians has been more slowly realized. Targeted sequencing of phylogenetically informative microbial genes, such as the bacterial *16S rRNA* gene, directly from clinical specimens, has had a modest but increasing influence on the practice of clinical microbiology. Whole-genome sequencing of viruses and bacteria has substantively enhanced outbreak investigations and strain tracking.[2,3] Yet, the rapid identification and characterization of microbial agents in routine cases of infectious disease is a major and increasing unmet need, made even more pressing by the alarming increase in antimicrobial resistance, and the dwindling availability of effective antimicrobial agents.[4]

In this issue of *JAMA*, Loman et al[5] describe an application of metagenomic sequencing that has yielded promising results, but more needs to be accomplished in this area. They selected fecal specimens from 34 patients with diarrhea who had tested positive in Germany for the Shiga-toxigenic *Escherichia coli* (STEC) O104:H4 2011 outbreak strain as well as fecal specimens from 5 patients with diarrhea who tested negative for STEC but positive for another bacterial pathogen. The authors evaluated whether they could detect and characterize these bacterial disease agents by preparing a pool of DNA from the complex mix of organisms in feces (ie, thousands of different organisms in each sample, and abundant host DNA); sequence millions of short, random DNA fragments from this pool; and then attempt to reassemble as many of the genomes in the original specimen as possible, without using any information learned through other means about the microbiological diagnosis. The investigators filtered their sequences by selecting those that were shared by 20 or more STEC-positive samples, and by ignoring sequences that matched those found in a previous study of the fecal microbial communities of healthy study participants.[6] About two-thirds of the resulting genome assemblies were recognized as belonging to the family Enterobacteriaceae, which is typically present in feces at less than 0.1% relative abundance. Further efforts led to a near-complete assembly of the genome of the known *E coli* outbreak strain from some specimens.

From the perspective of an outbreak investigator, this is an important achievement. However, from the perspective of a clinician caring for individual patients, this is only a mixed success story. Even though the causative agent could be recognized in the majority of case specimens—once its genome sequence had been assembled, roughly one-third of the STEC-positive diarrheal specimens failed to yield an STEC genome sequence. From this study, it is unclear why the sensitivity was this low; the sensitivity must be improved for the technology to be used in a diagnostic setting. The results from the 5 STEC-negative cases of diarrhea were also promising but mixed. A *Clostridium difficile* genome sequence, or fragments thereof, was recovered from the 2 diagnosed cases of *C difficile*–associated diarrhea. From 1 of these cases, many more genome fragments belonged to *Campylobacter concisus* than to *C difficile*. *Campylobacter concisus* is a known but rare intestinal pathogen, and one that was unexpected in this case. A *Campylobacter jejuni* genome sequence was assembled from a case of *C jejuni* culture-positive diarrhea; and *Salmonella*-specific genome sequences were identified in 1 of the 2 culture-positive cases of *Salmonella enterica* diarrhea.

While the shotgun metagenomics approach of Loman et al[5] was not as sensitive as traditional microbiological test-

See also p 1502.

Author Affiliations: Departments of Medicine and Microbiology & Immunology, Stanford University School of Medicine, Stanford, California; and Veterans Affairs Palo Alto Health Care System, Palo Alto, California.
Corresponding Author: David A. Relman, MD, Veterans Affairs Palo Alto Health Care System, 3801 Miranda Ave, Palo Alto, CA 94304 (relman@stanford.edu).

©2013 American Medical Association. All rights reserved.

Downloaded From: http://jama.jamanetwork.com/ by a University of Hawaii at Manoa User on 09/07/2013

ing in providing a diagnosis in STEC and non-STEC cases, it was successful in reconstructing an outbreak strain genome sequence, and in revealing an unanticipated, potential pathogen. As sequencing costs and turnaround times continue to decrease, deeper sequencing surveys of fecal specimens likely will provide ever-improving assessments of the complex microbial community.

Several issues deserve mention. First, the assembly of a genome sequence from a less abundant member of a microbial community as complex as those typically found in the human adult distal gut, starting with a pool of bulk DNA, is still a major technical challenge. As a result and for example, important antibiotic resistance genes might be detected, but without knowing to which strain or species they belong. Yet rapid advances are being made; currently, genomes from organisms with relative abundance levels of 0.1% to 1.0% have been assembled from such disparate settings as feces of a newborn premature infant and groundwater from deep subsurface aquifers, and have provided unexpected insights into the biology of these important ecosystems.[7,8]

Second, despite improved capabilities, the goal of understanding microbial behavior and lifestyle from a genome sequence in isolation remains problematic. For example, the presence of virulence-associated genes does not identify an organism as a pathogen. Rather, a set of co-regulated genes and choreographed cellular activities distinguishes an organism with the capacity for causing disease, given the right host and ecological context. Furthermore, even the identification of a known organism that typically causes disease in an immunocompetent host does not necessarily indicate that the organism is the cause of disease at that particular time and place. This problem is especially clear with advancing capabilities for detecting increasingly rare genes and genomes from heavily colonized sites within the human body, in which it is common to find low levels of a pathogen not to be related to pathology. Partial solutions will come from assessments of the broader ecological and clinical contexts within which these genes and genomes are found (eg, structure and function of the local microbiota as well as host factors[9,10]), more extensive assessments of states of health as well as disease (eg, healthy control participants in studies like those of Loman et al), and continued efforts to measure and understand the biological activity of microbes in relevant model systems. Proof of causation in this increasingly molecular age will require data on abundance of the candidate disease marker, spatial and temporal linkage to pathology, and the host.[11]

Finally, microbial cultivation and purified isolates of specific organisms remain necessary for clinical diagnosis as well as research at the present time. In particular, purified isolates provide an ongoing resource with which to study the critical behaviors of microbes that are not so easily predicted from genomic sequences. Microbial genome and community sequence data are destined to affect clinical and public health decision making in a profound manner. However, clinician-investigators know that there remain many critical, clinically relevant questions that demand more than genome sequence data, requiring biological measurements and a deeper understanding of the ecological and clinical setting.

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Relman reported serving as a consultant to Cepheid Corporation, NanoBio Corporation, Novartis Vaccines, and Procter & Gamble; and owning stock in Cepheid Corporation.

REFERENCES

1. Relman DA. Microbial genomics and infectious diseases. *N Engl J Med*. 2011; 365(4):347-357.
2. Snitkin ES, Zelazny AM, Thomas PJ, et al; NISC Comparative Sequencing Program Group. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci Transl Med*. 2012;4(148):ra116.
3. Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med*. 2008;358(10):991-998.
4. Davies J, Davies D. Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev*. 2010;74(3):417-433.
5. Loman NJ, Constantinidou C, Christner M, et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA*. 2013;309(14):1502-1510.
6. Qin J, Li R, Raes J, et al; MetaHIT Consortium. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285): 59-65.
7. Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res*. 2013;23(1):111-120.
8. Wrighton KC, Thomas BC, Sharon I, et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science*. 2012;337(6102): 1661-1665.
9. Lemon KP, Armitage GC, Relman DA, Fischbach MA. Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med*. 2012;4(137):rv5.
10. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. *Science*. 2012;336(6086):1255-1262.
11. Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev*. 1996;9(1): 18-33.

 ©2013 American Medical Association. All rights reserved.

Downloaded From: http://jama.jamanetwork.com/ by a University of Hawaii at Manoa User on 09/07/2013