AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

COMMENTARY

Societal Implications of the Internet of Pathogens

Alexander L. Greninger[a]

[a]Department of Laboratory Medicine, University of Washington, Seattle, Washington, USA

ABSTRACT The growth of pathogen genomics shows no signs of abating. Whole-genome sequencing of clinical viral and bacterial isolates continues to grow in nearly exponential bounds. Reductions in cost driven by new technology have created a seamless environment for generating, sharing, and analyzing pathogen genomes. The high-resolution view of infectious disease transmission dynamics offered by analyzing whole genomes from pathogens, coupled with the genomicist ethic of widespread data sharing, has created a veritable Internet of pathogens, which inadvertently produces new threats to patient privacy and protected heath information. The health care system, and society more generally, have yet to explore the far-reaching privacy concerns raised by readily accessible pathogen genomic data. The recent use of human genomic databases, the existence of freely available alternative data and metadata sources, and lax regulation of collecting publicly available genomes to identify individuals in a criminal context raise concerning parallels about what is possible with pathogen genomics. The growing ability to ascertain culpability for infectious disease transmission at a nearly individual level could change our perspective on disease outbreaks from one based on public health to one based on individual liability. These technological breakthroughs in the absence of an understanding of potential privacy and liability issues lead to questions about the dominant paradigm of better living through pathogen genomics.

KEYWORDS Cassandra, genomics, Internet of pathogens, liability, metagenomics, pathogen sequencing, privacy

AN INTERNET OF PATHOGENS

Slightly more than 20 years after its advent, pathogen genome sequencing is still increasing like a one-step viral growth curve and spreading throughout the world like a newly plated *Proteus*. Although routine genome sequencing was unthinkable until recently, the Centers for Disease Control (CDC), Food and Drug Administration (FDA), state, and global public health laboratories now routinely sequence more than 200 foodborne bacterial isolates a day and more than 6,000 influenza virus genomes a year (1). The rise of pathogen genomics is credited by the CDC for the record number of foodborne outbreak investigations conducted in 2018. The cumulative nature of the accelerating data collection creates increasing returns to scale and forces us to consider the future today. Isolates collected during surveillance and investigations may not shed light on an outbreak, but the data are kept for future use, populating the database for the next query. Combined with new scalable analysis and visualization tools, these pathogen sequences create a rapidly growing Internet of pathogens (2–5).

The current state of microbiological sequencing is not unlike that of the Internet of 1997. At slightly more than a million webpages, the 1997 Internet had demonstrated its utility, disrupting classified advertisements. But it was still adolescent in its development. While many people could foresee continued growth at the time, few could have imagined the novel search algorithms, social media, scale of data collection, and the widespread generational and sociological changes surrounding the use of these technologies with their downstream implications for privacy.

Genome sequences of almost all human pathogens are now publicly available. This

Citation Greninger AL. 2019. Societal implications of the Internet of pathogens. J Clin Microbiol 57:e01914-18. https://doi.org/10.1128/JCM.01914-18.

Editor Daniel J. Diekema, University of Iowa College of Medicine

Address correspondence to agrening@uw.edu.

The views expressed in this article do not necessarily reflect the views of the journal or of ASM.

Accepted manuscript posted online 10 April 2019
Published 24 May 2019

allows us to take full advantage of sequencing technologies to use sequences both to arbitrate pathogen identification and to suggest potential treatments (6). The move to metagenomics for primary microbiological diagnosis may create a surfeit of clinical pathogen sequencing data (7). And if there is a consequence to the growing amount of clinical microbiology sequence, it is that such data are likely to be used in full to infer everything from antimicrobial resistance to phylogenetic relationships.

The scale of sequencing for common pathogens is remarkable. Sequencing projects in microbiology and clinical metagenomics are now measured in the tens of thousands to hundreds of thousands of isolates and are likely to follow human genomics into the millions of samples (6, 8–10). All of pus are being sequenced (11–13). As sequencing becomes less expensive and the methods more readily available, the resolution of sampling of individuals increases in both time and space, from decades to day-by-day, from countries to block-by-block maps (14, 15).

Building the required databases and sequencing more pathogens are not inexpensive. Academics have initiated the analytical framework and seeded the sequencing databases. Subsequently, genomic-based epidemiology has been sustained by public health organizations such as the FDA and CDC at the federal level and privately by denizens of data such as Bill Gates and Mark Zuckerberg (https://www.npr.org/2011/01/31/133377748/bill-gates-goal-get-rid-of-polio-forever and https://www.czbiohub.org/projects/infectious-disease). That may be enough to continue the growth. Medicare and other health insurers are likely only to reimburse pathogen sequencing as it relates to influencing direct care for a tested beneficiary (e.g., antimicrobial resistance and, potentially, metagenomics) and will not reimburse hospital infection control or other epidemiological uses since those are not generally considered a payable benefit for Medicare beneficiaries. However, if costs continue to plunge and the informational utility increases from the scale of relational sequence data and metadata available, pathogen sequencing may yet transition into routine health care or at least clinical microbiology laboratory practice, fueling even more waves of sequencing growth.

ALL OUR DATA BELONG TO US

Recently, familial human genomic data have been used to (putatively) solve high-profile cold cases such as that of the Golden State killer. This has taken place in the context of lax or absent regulations of ownership of human genetic data. Parabon Nanolabs, a new company that investigates such cases, says that they have used human genomics to close more than a dozen cases in the last half year alone and forecasts that there will be hundreds more (https://www.wkyc.com/article/news/investigations/investigator-dna-experts-helping-to-crack-cold-cases-everywhere-including-northeast-ohio/95-612308915). The concomitant growth and use of human genomics in this fashion have intriguing parallels with pathogen genomics.

Since its inception, genomics has overwhelmingly been governed by an ethic of rapid and open sharing of data. The Bermuda Principles codified the release of human genome contigs in less than 24 h of assembly while the Fort Lauderdale Agreement promoted continued rapid public release of genomic information (16, 17). This ethic has also been adopted in the pathogen genomics community (18). While the privacy issues associated with broad sharing of human genomics data have been debated extensively (17, 19, 20), comparatively few researchers have taken on this issue in pathogen genomics (21–23). However, in the same manner as depositing one's genome in GEDmatch, an open personal genomics database, can allow inferences about your cousins without their consent, sequencing my enterovirus may have implicit implications for you (24).

Pathogen genome sequences have not been seen as protected health information or as potentially identifying information in the same way as human genomes are. After all, pathogen genomes change, and most infections are transient. Rather, it is the combination of pathogen sequence and associated metadata, such as collection date/time or location, that come with the pathogen sequence that might be considered problematic. Dates associated with patient care that are more precise than the year and

locations that are smaller than the first three numbers of a zip code are expressly named as protected health information under the "safe harbor" method of deidentification in U.S. Department of Health and Human Services guidance. Combined with phylogenetic inferences, the metadata and sequence from non-deidentified isolates can be used to impute metadata and protected health information from other individuals with sequenced pathogens to levels of accuracy greater than current law would otherwise allow (25). In this way, a pathogen's sequence may allow reidentification of a deidentified sample and may need to be considered potentially identifiable information. Deidentified HIV sequences obtained for antiviral resistance testing are already being used by the CDC in 27 states to monitor local transmission networks and, combined with additional metadata, to identify individuals in the transmission networks for enhanced public health interventions (26). Going forward, it will likely become nearly impossible to meet the deidentification standard for the whole-genome sequence of any pathogen under the safe harbor method of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

Many people could argue that the ends justify the means. Stopping a killer such as *Escherichia coli* O157 or Ebola virus has mobilized a diverse community to organize a broad data-sharing mandate (18). To date, many of these efforts have been focused on newsworthy outbreak scenarios, best exemplified by the 2014 Ebola virus crisis. But focusing on discrete outbreak situations may not be helpful since every infectious disease exists in some form of an outbreak. The same logic holds for seemingly innocuous agents such as rhinovirus (27).

LIABLE FOR THE FUTURE

A perverse corollary to the growth in pathogen genomics to determine transmission epidemiology is that it may drive increasing interest in infectious disease liability. It is already easy to prove damage in these cases. With the increasing ability to demonstrate or refute links in transmission, there will now be recourse for those affected. The United Nations dedicated a self-imposed dollar amount of $400 million for its role, elucidated and confirmed by pathogen genomics, in introducing cholera into Haiti and thereby causing an outbreak that has resulted in 800,000 infections and 9,000 deaths (28). Viral sequencing may have exonerated "Patient 0" for introducing HIV into the United States, but it is now equally used to adjudicate hepatitis C virus (HCV) transmission in health care settings and among injected drug users (29–31). In infection prevention parlance, everyone from your coworker to your bus copassenger who possibly gave you the flu this year is guilty of some form of "presenteeism," or working while sick (32). Personal injury attorneys focused on *Legionella* sp. infections are increasingly aware of the growing use of sequencing to determine transmission sources by public health authorities (https://www.pritzkerlaw.com/personal-injury/2019/can-dna-tests-solve-albany-legionnaires-disease-outbreak/). Discovering the source of a decade-long outbreak through genomic sleuthing also equates to discovery for potential litigants (33). Filing subpoenas for pathogen sequence data may become the clinical microbiology paternity suit (34). Even without litigation, pathogen survivors may have to live with the knowledge of the damage their strain caused in the community.

Personal knowledge of one's role in disease transmission may not be a bad thing. Private awareness of public health is a cornerstone of disease prevention and control. The public health system does not wash your hands for you. But a liability- and culpability-based ethic to deter pathogen transmission is certainly a different manifestation than the one currently envisioned for pathogen sequencing (35).

Of course, caveats abound, and the evolutionary models are still being built. Not all pathogens have sufficient evolutionary rates to directly equate transmission with consensus genome identity (36). Intrahost single nucleotide variants present in heterogeneous RNA viruses can both aid and complicate the inference of transmission (37–39). Clinical and research laboratories are not often set up for chain of custody. The resolution of imputation of transmission will likely be dependent on the richness of the social networks involved: mapping transmission in an isolated hospital ward is easier

than in a subway system. And while it will always be difficult to prove that an unknown third person was not the basis of transmission, additional ubiquitous digital data sources may fill in the missing information (40, 41).

INVOLVING CLINICAL MICROBIOLOGY BEYOND THE ISOLATES

The growing use of sequencing in academic and public health laboratories will eventually make its way into the clinical microbiology laboratory. The evolutionary relationship of your pathogen to the greater community is not yet a clinically reportable result. But it may be some day. Sequencing pathogens brings new importance to detailed epidemiological questions that have not traditionally been under the purview of clinical microbiologists. These conversations will soon become the responsibility of the 21st century clinical microbiologist.

ACKNOWLEDGMENTS

I thank Alexander McAdam, Amin Addetia, Samia Naccache, Keith Jerome, and Ryan Shean for helpful comments.

I have received consultant fees from Abbott Molecular.

REFERENCES

1. Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R. 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. J Clin Microbiol 54:1975–1983. https://doi.org/10.1128/JCM.00081-16.
2. Greninger AL. 2018. A decade of RNA virus metagenomics is (not) enough. Virus Res 244:218–229. https://doi.org/10.1016/j.virusres.2017.10.014.
3. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34:4121–4123. https://doi.org/10.1093/bioinformatics/bty407.
4. Burall LS, Grim CJ, Mammel MK, Datta AR. 2016. Whole genome sequence analysis using JSpecies tool establishes clonal relationships between Listeria monocytogenes strains from epidemiologically unrelated listeriosis outbreaks. PLoS One 11:e0150797. https://doi.org/10.1371/journal.pone.0150797.
5. Shu Y, McCauley J. 2017. GISAID: global initiative on sharing all influenza data—from vision to reality. Euro Surveill 22:30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.
6. CRyPTIC Consortium and the 100,000 Genomes Project. 2018. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. N Engl J Med 379:1403–1415. https://doi.org/10.1056/NEJMoa1800474.
7. Greninger AL. 2018. The challenge of diagnostic metagenomics. Expert Rev Mol Diagn 18:605–615. https://doi.org/10.1080/14737159.2018.1487292.
8. Weimer BC. 2017. 100K Pathogen Genome Project. Genome Announc 5:e00594-17. https://doi.org/10.1128/genomeA.00594-17.
9. Liu S, Huang S, Chen F, Zhao L, Yuan Y, Francis SS, Fang L, Li Z, Lin L, Liu R, Zhang Y, Xu H, Li S, Zhou Y, Davies RW, Liu Q, Walters RG, Lin K, Ju J, Korneliussen T, Yang MA, Fu Q, Wang J, Zhou L, Krogh A, Zhang H, Wang W, Chen Z, Cai Z, Yin Y, Yang H, Mao M, Shendure J, Wang J, Albrechtsen A, Jin X, Nielsen R, Xu X. 2018. Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history. Cell 175:347–359.e14. https://doi.org/10.1016/j.cell.2018.08.016.
10. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–209. https://doi.org/10.1038/s41586-018-0579-z.
11. Guet-Revillet H, Jais J-P, Ungeheuer M-N, Coignard-Biehler H, Duchatelet S, Delage M, Lam T, Hovnanian A, Lortholary O, Nassif X, Nassif A, Join-Lambert O. 2017. The microbiological landscape of anaerobic infections in hidradenitis suppurativa: a prospective metagenomic study. Clin Infect Dis 65:282–291. https://doi.org/10.1093/cid/cix285.
12. Yu H-J, Deng H, Ma J, Huang S-J, Yang J-M, Huang Y-F, Mu X-P, Zhang L, Wang Q. 2016. Clinical metagenomic analysis of bacterial communities in breast abscesses of granulomatous mastitis. Int J Infect Dis 53:30–33. https://doi.org/10.1016/j.ijid.2016.10.015.
13. Drancourt M, Nkamga VD, Lakhe NA, Régis J-M, Dufour H, Fournier P-E, Bechah Y, Scheld WM, Raoult D. 2017. Evidence of archaeal methanogens in brain abscess. Clin Infect Dis 65:1–5. https://doi.org/10.1093/cid/cix286.
14. Pybus OG, Rambaut A. 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet 10:540–550. https://doi.org/10.1038/nrg2583.
15. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, Wiley MR, White S, Thézé J, Magnani DM, Prieto K, Reyes D, Bingham AM, Paul LM, Robles-Sikisaka R, Oliveira G, Pronty D, Barcellona CM, Metsky HC, Baniecki ML, Barnes KG, Chak B, Freije CA, Gladden-Young A, Gnirke A, Luo C, MacInnis B, Matranga CB, Park DJ, Qu J, Schaffner SF, Tomkins-Tinch C, West KL, Winnicki SM, Wohl S, Yozwiak NL, Quick J, Fauver JR, Khan K, Brent SE, Reiner RC, Lichtenberger PN, Ricciardi MJ, Bailey VK, Watkins DI, Cone MR, Kopp EW, Hogan KN, Cannons AC, Jean R, et al. 2017. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature 546:401–405. https://doi.org/10.1038/nature22400.
16. Guyer M. 1998. Statement on the rapid release of genomic DNA sequence. Genome Res 8:413.
17. Reardon J, Ankeny RA, Bangham J, W Darling K, Hilgartner S, Jones KM, Shapiro B, Stevens H. 2016. Bermuda 2.0: reflections from Santa Cruz. GigaScience 5:1–4. https://doi.org/10.1093/gigascience/giw003.
18. Yozwiak NL, Schaffner SF, Sabeti PC. 2015. Data sharing: make outbreak research open access. Nature 518:477–479. https://doi.org/10.1038/518477a.
19. McCombie WR, McPherson JD. 26 November 2018. Future promises and concerns of ubiquitous next-generation sequencing. Cold Spring Harb Perspect Med. https://doi.org/10.1101/cshperspect.a025783.
20. Goodman KW. 1996. Ethics, genomics, and information retrieval. Comput Biol Med 26:223–229. https://doi.org/10.1016/0010-4825(95)00059-3.
21. Mehta SR, Vinterbo SA, Little SJ. 2014. Ensuring privacy in the study of pathogen genetics. Lancet Infect Dis 14:773–777. https://doi.org/10.1016/S1473-3099(14)70016-7.
22. Geller G, Dvoskin R, Thio CL, Duggal P, Lewis MH, Bailey TC, Sutherland A, Salmon DA, Kahn JP. 2014. Genomics and infectious disease: a call to identify the ethical, legal and social implications for public health and clinical practice. Genome Med 6:106. https://doi.org/10.1186/s13073-014-0106-2.
23. Fairchild AL, Gable L, Gostin LO, Bayer R, Sweeney P, Janssen RS. 2007. Public goods, private data: HIV and the history, ethics, and uses of identifiable public health information. Public Health Rep 122:7–15. https://doi.org/10.1177/00333549071220S103.
24. Erlich Y, Shor T, Pe'er I, Carmi S. 2018. Identity inference of genomic data using long-range familial searches. Science 362:690–694. https://doi.org/10.1126/science.aau4832.
25. Shean RC, Greninger AL. 2018. Private collection: high correlation of sample collection and patient admission date in clinical microbiological

testing complicates sharing of phylodynamic metadata. Virus Evol 4:vey005. https://doi.org/10.1093/ve/vey005.
26. McClelland A, Guta A, Gagnon M. 10 February 2019. The rise of molecular HIV surveillance: implications on consent and criminalization. Crit Public Health. https://doi.org/10.1080/09581596.2019.1582755.
27. Seo S, Waghmare A, Scott EM, Xie H, Kuypers JM, Hackman RC, Campbell AP, Choi S-M, Leisenring WM, Jerome KR, Englund JA, Boeckh M. 2017. Human rhinovirus detection in the lower respiratory tract of hematopoietic cell transplant recipients: association with mortality. Haematologica 102:1120–1130. https://doi.org/10.3324/haematol.2016.153767.
28. Gladstone R. 27 June 2017. U.N. brought cholera to Haiti. Now it is fumbling its effort to atone, p A4. New York Times, New York, NY.
29. Worobey M, Watts TD, McKay RA, Suchard MA, Granade T, Teuwen DE, Koblin BA, Heneine W, Lemey P, Jaffe HW. 2016. 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature 539:98–101. https://doi.org/10.1038/nature19827.
30. Escobar-Gutiérrez A, Vazquez-Pichardo M, Cruz-Rivera M, Rivera-Osorio P, Carpio-Pedroza JC, Ruíz-Pacheco JA, Ruiz-Tovar K, Vaughan G. 2012. Identification of hepatitis C virus transmission using a next-generation sequencing approach. J Clin Microbiol 50:1461–1463. https://doi.org/10.1128/JCM.00005-12.
31. Apostolou A, Bartholomew ML, Greeley R, Guilfoyle SM, Gordon M, Genese C, Davis JP, Montana B, Borlaug G, Centers for Disease Control and Prevention (CDC). 2015. Transmission of hepatitis C virus associated with surgical procedures—New Jersey 2010 and Wisconsin 2011. MMWR Morb Mortal Wkly Rep 64:165–170.
32. Hansen CD, Andersen JH. 2008. Going ill to work—what personal circumstances, attitudes and work-related factors are associated with sickness presenteeism? Soc Sci Med 67:956–964. https://doi.org/10.1016/j.socscimed.2008.05.022.
33. Johnson RC, Deming C, Conlan S, Zellmer CJ, Michelin AV, Lee-Lin S, Thomas PJ, Park M, Weingarten RA, Less J, Dekker JP, Frank KM, Musser KA, McQuiston JR, Henderson DK, Lau AF, Palmore TN, Segre JA. 2018. Investigation of a cluster of Sphingomonas koreensis infections. N Engl J Med 379:2529–2539. https://doi.org/10.1056/NEJMoa1803238.
34. Miller JM. 2004. Liability relating to contracting infectious diseases in hospitals. J Leg Med 25:211–227. https://doi.org/10.1080/01947640490457497.
35. Gardy J, Loman NJ, Rambaut A. 2015. Real-time digital pathogen surveillance—the time is now. Genome Biol 16:155. https://doi.org/10.1186/s13059-015-0726-x.
36. Campbell F, Strang C, Ferguson N, Cori A, Jombart T. 2018. When are pathogen genome sequences informative of transmission events? PLoS Pathog 14:e1006885. https://doi.org/10.1371/journal.ppat.1006885.
37. Campo DS, Zhang J, Ramachandran S, Khudyakov Y. 2017. Transmissibility of intra-host hepatitis C virus variants. BMC Genomics 18:881. https://doi.org/10.1186/s12864-017-4267-4.
38. Ramachandran S, Thai H, Forbi JC, Galang RR, Dimitrova Z, Xia G-L, Lin Y, Punkova LT, Pontones PR, Gentry J, Blosser SJ, Lovchik J, Switzer WM, Teshale E, Peters P, Ward J, Khudyakov Y, Hepatitis C Investigation Team. 2018. A large HCV transmission network enabled a fast-growing HIV outbreak in rural Indiana, 2015. EBioMedicine 37:374–381. https://doi.org/10.1016/j.ebiom.2018.10.007.
39. Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin Microbiol Infect 24:335–341. https://doi.org/10.1016/j.cmi.2017.10.013.
40. Carreras I, Matic A, Saar P, Osmani V. 2012. Comm2Sense: detecting proximity through smartphones, p 253–258. *In* 2012 IEEE International Conference on Pervasive Computing and Communications Workshops. Institute of Electrical and Electronics Engineers, Piscataway, NJ.
41. Sapiezynski P, Stopczynski A, Wind DK, Leskovec J, Lehmann S. 2017. Inferring person-to-person proximity using WiFi signals. Proc ACM Interact Mob Wearable Ubiquitous Technol 1:24. https://doi.org/10.1145/3090089.