Deciphering the Origins and Tracking the Evolution of Cholera Epidemics with Whole-Genome-Based Molecular Epidemiology

Yonatan H. Grad,[a,b] Matthew K. Waldor[a,c,d]

Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts, USA[a]; Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard School of Public Health, Boston, Massachusetts, USA[b]; Department of Microbiology & Immunobiology, Harvard Medical School, Boston, Massachusetts, USA[c]; HHMI, Boston, Massachusetts, USA[d]

ABSTRACT The devastating Haitian cholera outbreak that began in October 2010 is the first known cholera epidemic in this island nation. Epidemiological and genomic data have provided strong evidence that United Nations security forces from Nepal introduced toxigenic *Vibrio cholerae* O1, the cause of epidemic cholera, to Haiti shortly before the outbreak arose. However, some have contended that indigenous *V. cholerae* contributed to the outbreak. In a recent paper (*mBio* 4:e00398-13, 2013), L. S. Katz et al. explored the nature and rate of changes in this ancient pathogen's genome during an outbreak, based on whole-genome sequencing of 23 Haitian *V. cholerae* clinical isolates obtained over a 20-month period. Notably, they detected point mutations, deletions, and inversions but found no insertion of horizontally transmitted DNA, arguing strongly against the idea that autochthonous *V. cholerae* donated DNA to the outbreak strain. Furthermore, they found that Haitian epidemic *V. cholerae* isolates were virtually untransformable. Comparative genomic analyses revealed that the Haitian isolates were nearly identical to isolates from Nepal and that the Nepalese-Haitian isolates were distinguishable from isolates circulating elsewhere in the world. Reconstruction of the phylogeny of the Haitian isolates was consistent with a single introduction of *V. cholerae* to Haiti sometime between late July and late October 2010, dates remarkably concordant with epidemiological observations. In aggregate, this paper provides additional compelling evidence that the *V. cholerae* strain responsible for the Haitian cholera epidemic came from Nepal and illustrates the power of whole-genome-based analyses for epidemiology, pathogen evolution, and forensics.

During the last 2 centuries, cholera has spread from the Indian subcontinent to cause seven global pandemics, during which this rapidly dehydrating diarrheal disease has killed millions in Asia, Africa, Europe, and the Americas. How has cholera toxin-producing *Vibrio cholerae* serogroup O1, the cause of the sixth and ongoing seventh cholera pandemics, spread from continent to continent? Remarkably, John Snow, who founded modern epidemiology with his groundbreaking analyses of the spread of cholera in London in the late 1840s and 1850s (3 decades before Koch's discovery of *V. cholerae* in 1883), provided the answer. He noticed that the first case of cholera in 1848 in London, which had been free of the disease for several years, occurred in a seaman from Hamburg, a city that had recently experienced a large cholera outbreak. To Snow, the conclusion was clear: the causative agent of cholera was transported from Germany to England by the infected seaman (1). His hypothesis that contamination of the local environment by this sick seaman caused the outbreak provided a controversial but ultimately compelling alternative to the miasma theory of cholera transmission, the predominant model at the time, which linked disease to the noxious air of unsanitary spaces.

In subsequent years, the sixth and ongoing seventh cholera pandemics have wreaked havoc in the Americas, but there is no reported history of cholera in Haiti until 2010 (2). However, an earthquake in January 2010 severely damaged the already marginal Haitian public sanitation system, creating ideal conditions for the spread of the disease following its introduction. The extent of the outbreak has been devastating. Since the first cholera cases were recognized in October 2010, nearly 8,200 cholera deaths and more than 668,000 cholera cases (almost 7% of the population) have been reported by the Haitian Ministry of Public Health (http://www.mspp.gouv.ht/site/index.php). Despite compelling epidemiological and genomic evidence linking United Nations security forces from Nepal to the introduction of cholera to Haiti (3–5), questions have been raised about the role of autochthonous *V. cholerae* strains in triggering the Haitian cholera outbreak and serving as a source of horizontally transmitted genetic material that was incorporated into the initial outbreak strain (6, 7). Given questions about accountability for the outbreak and the possible associated legal ramifications (see http://www.law.yale.edu/news/17237.htm), new insights into this controversy will be welcomed by scientists and lay persons alike.

In a recent *mBio* paper, Katz et al. used whole-genome-based analyses to investigate the nature and pace of genomic changes in a series of *V. cholerae* O1 isolates from the ongoing Haitian epidemic (8). As the authors noted, the apparent single-source introduction of *V. cholerae* O1 to Hispaniola provided "an unprecedented natural experiment for characterizing in detail the intrinsic tempo and mode of genome evolution in this deadly pathogen." Their findings reinforce the evidence that the Haitian outbreak began after the introduction of a *V. cholerae* O1 strain from Nepal, argue convincingly against the idea that indigenous *V. cholerae* contributed to the genome of the outbreak strain, provide insights into the evolution of this ancient pathogen, and demonstrate the enormous power and utility that whole-genome-based investigations can bring to epidemiology and forensics.

The authors sequenced 23 *V. cholerae* O1 genomes drawn from over 20 months, 13 locations dispersed across Haiti, and multiple

Published 10 September 2013

Citation Grad YH, Waldor MK. 2013. Deciphering the origins and tracking the evolution of cholera epidemics with whole-genome-based molecular epidemiology. mBio 4(5): e00670-13. doi:10.1128/mBio.00670-13.

Address correspondence to Matthew K. Waldor, mwaldor@research.bwh.harvard.edu.

pulsed-field gel electrophoresis (PFGE) patterns. Comparisons of these genomes with 108 *V. cholerae* genomes from other global sources, including contemporaneous isolates from Thailand, Bangladesh, Nepal, Cameroon, India, Pakistan, and Benin, demonstrated not only that the Haitian isolates were nearly identical to isolates from Nepal but also that the Nepal-Haiti isolates were clearly distinct from isolates circulating elsewhere in the world. Notably, reconstruction of the phylogeny of the Haitian isolates with a Bayesian coalescent model was consistent with a single introduction of *V. cholerae* to Haiti sometime between 23 July and 17 October 2010, dates remarkably in keeping with epidemiological findings of the beginnings of the cholera outbreak in Haiti (4). Although the authors did not extend their model to include the three closely related isolates from Nepal, the very few single-nucleotide polymorphisms (SNPs) separating the Haiti and Nepal isolates (as shown in Fig. 1 [and Fig. S4 in the supplemental material] of reference 8) suggest that the two sets of isolates diverged very shortly before the estimated date of introduction into Haiti, a narrow window of time that further fastens the link of transmission of cholera from Nepal to Haiti.

The SNP-based phylogenetic reconstruction offers a framework for understanding *V. cholerae* evolution over this short time span. This reconstruction is built on SNPs determined by mapping reads to a reference genome and therefore is based on regions of the genome that reflect shared descent. To characterize larger-scale genome variation, the authors compare gene content and detect rearrangements by using *de novo* assemblies. These analyses revealed several independent deletions and inversions within SXT, a mobile genetic element that contains most of the genes that confer antibiotic resistance on *V. cholerae* (9). The observations of deletions and inversions may also explain the variations among the PFGE patterns of these strains, as well as phenotypic variations in antibiotic susceptibility. The notable absence of insertions from the sequenced genomes firmly establishes that horizontal gene transfer from indigenous *V. cholerae* did not introduce novel genetic elements into the genome of the outbreak strain, contrary to a recent hypothesis (6). Furthermore, the authors found that outbreak isolates were virtually untransformable, perhaps explaining the absence of DNA from environmental organisms.

The work by Katz and coauthors adds to a growing body of work that explores the power and utility of pathogen population genomics in elucidating the origins and dynamics of outbreaks (10). A key issue facing investigators who use this approach to study outbreaks is determining the ideal sampling strategy that can provide statistically robust answers while also being cost-effective. While the continuing decline in sequencing costs is making monetary considerations less important, the collection of clinical microbiological isolates and their storage and processing for sequencing remain nonstandard and still require considerable financial support. What prior knowledge is needed to devise rational, evidence-based sampling strategies that provide sufficient power to test hypotheses about the origins of an outbreak? Answers may be outbreak or pathogen specific and may require integration of transmission modeling, considering, for example, the impact of the mode of transmission on the ideal sampling method. One important aid to devising sampling strategies may come from published estimates of molecular clock rates (see reference 11 for an estimate for *V. cholerae*). We anticipate that these rates may establish a starting point for making inferences about the history of a pathogen population from its diversity. It will be important to test the assumption that the molecular clock for a given species remains relatively stable across niches and lineages. Additionally, public databases with pathogen genome sequences and associated metadata (epidemiological and demographic information associated with the sequences) broaden the range of questions and provide a historical set of samples with which to compare data or integrate them into an analysis, as done by Katz and colleagues.

The rapid expansion of microbial genome sequences in public databases raises another critical point—the importance of standardization of genome sequence data, annotation, and analysis. Quality metrics are needed so that results can be compared and data readily imported from one study into another. For example, Katz et al. refer to "high-quality SNPs," presumably to distinguish them from those derived from less stringent SNP-calling thresholds and to emphasize the low likelihood of false-positive SNPs. Ideally, study-specific methods of SNP calling and quality assessment should be replaced with rigorous methods yielding reproducible results that account for varying sequencing and analytic platforms and study quality (12).

The ongoing revolution in DNA sequencing technology is ushering in a new era of whole-genome-based molecular epidemiology. Coupled with the adoption of standards that permit rigorous cross-study analyses and the development and implementation of sophisticated statistical methods, these new technologies will transform our ability to understand microbial evolution and epidemics and ultimately to create improved public health interventions. For cholera in particular, we can take advantage of the power of these new methods to address a fundamental question: are humans always involved in the spread of this ancient disease over great distances (such as between continents)? If so, such a conclusion has profound implications for our ability to eliminate this ancient scourge from our planet with appropriate interventions.

ACKNOWLEDGMENTS

We thank Brigid Davis, John Mekalanos, and Nicholas J. Croucher for their insightful comments on the manuscript.

Y.H.G. is supported by an ASTDA developmental award and K08 AI-104767, and M.K.W. is supported by R37 AI-42347 and HHMI.

REFERENCES

1. **Johnson S.** 2006. The ghost map. Riverhead, New York, NY.
2. **Jenson D, Szabo V, Duke FHI Haiti Humanities Laboratory Student Research Team.** 2011. Cholera in Haiti and other Caribbean regions, 19th century. Emerg. Infect. Dis. **17:**2130–2135.
3. **Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK.** 2011. The origin of the Haitian cholera outbreak strain. N. Engl. J. Med. **364:**33–42.
4. **Frerichs RR, Keim PS, Barrais R, Piarroux R.** 2012. Nepalese origin of cholera epidemic in Haiti. Clin. Microbiol. Infect. **18:**E158–E163.
5. **Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, Bortolaia V, Pearson T, Waters AE, Upadhyay BP, Shrestha SD, Adhikari S, Shakya G, Keim PS, Aarestrup FM.** 2011. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. mBio **2**(4):e00157-11. doi:10.1128/mBio.00157-11.
6. **Hasan NA, Choi SY, Eppinger M, Clark PW, Chen A, Alam M, Haley BJ, Taviani E, Hine E, Su Q, Tallon LJ, Prosper JB, Furth K, Hoq MM, Li H, Fraser-Liggett CM, Cravioto A, Huq A, Ravel J, Cebula TA, Colwell RR.** 2012. Genomic diversity of 2010 Haitian cholera outbreak strains. Proc. Natl. Acad. Sci. U. S. A. **109:**E2010–E2017.
7. **Jutla A, Whitcombe E, Hasan N, Haley B, Akanda A, Huq A, Alam M, Sack RB, Colwell R. 29** July 2013, posting date. Environmental factors influencing epidemic cholera. Am. J. Trop. Med. Hyg. (Epub ahead of print.) doi:10.4269/ajtmh.12-0721.

8. **Katz LS, Petkau A, Beaulaurier J, Tyler S, Antonova ES, Turnsek MA, Guo Y, Wang S, Paxinos EE, Orata F, Gladney LM, Stroika S, Folster JP, Rowe L, Freeman MM, Knox N, Frace M, Boncy J, Graham M, Hammer BK, Boucher Y, Bashir A, Hanage WP, Van Domselaar G, Tarr CL.** 2013. Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. mBio **4**(4):e00398-13. doi:10.1128/mBio.00398-13.
9. **Wozniak RA, Waldor MK.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. **8:**552–563.
10. **Pybus OG, Fraser C, Rambaut A.** 2013. Evolutionary epidemiology: preparing for an age of genomic plenty. Philos. Trans. R. Soc. Lond. B Biol. Sci. **368:**20120193. doi:10.1098/rstb.2012.0193.
11. **Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M, Niyogi SK, Kim EJ, Ramamurthy T, Chun J, Wood JL, Clemens JD, Czerkinsky C, Nair GB, Holmgren J, Parkhill J, Dougan G.** 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature **477:**462–465.
12. **Croucher NJ, Harris SR, Grad YH, Hanage WP.** 2013. Bacterial genomes in epidemiology—present and future. Philos. Trans. R. Soc. Lond. B Biol. Sci. **368:**20120202. doi: 10.1098/rstb.2012.0202.

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