

Next-Generation Epidemiology: Using Real-Time Core Genome Multilocus Sequence Typing To Support Infection Control Policy

John P. Dekker, Karen M. Frank

Department of Laboratory Medicine, National Institutes of Health Clinical Center, Bethesda, Maryland, USA

Multidrug-resistant bacteria are responsible for substantial morbidity and mortality worldwide. Tracking the nosocomial spread of resistant bacteria is critical to infection control. Mellmann et al. (J. Clin. Microbiol. 54:2874–2881, 2016, <http://dx.doi.org/10.1128/JCM.00790-16>) have described prospective whole-genome sequencing with core genome multilocus sequencing typing (cgMLST) analysis for real-time surveillance and have addressed the practical aspects of implementing this type of operation in the hospital setting.

Multilocus sequencing typing (MLST), introduced by Maiden and colleagues in 1998, has come to represent one of the most widely used and successful molecular methods for strain-level classification of bacterial isolates (1). MLST schemes work by indexing sequence variation in a small set of housekeeping genes to a numerical classification system. Each novel sequence occurring at any of the selected loci is named as a unique numbered allele, and the profile of a given isolate is specified by the numbers representing the allelic composition. Each combination of alleles is assigned a unique sequence type (ST), and STs, in turn, can be grouped into higher levels of classification (2, 3). The ability of MLST to resolve closely related strains is limited in principle by the absolute sequence diversity present in the loci covered by the scheme and by the manner in which this sequence diversity distributes across the population to be classified. As MLST schemes require conserved PCR primer binding sites, the targets must also be sufficiently conserved so that the method works robustly, and this puts limits on the resolution of classical MLST (3). MLST schemes have now been developed for >100 genera and species, and continuously updated databases are maintained by the University of Oxford, United Kingdom (pubmlst.org), Imperial College London, London, United Kingdom (<http://www.mlst.net/>), and others.

Classical MLST analysis, developed to be compatible with Sanger sequencing methods, is based on 400-to-500-nucleotide segments of (usually) seven genes, encompassing <0.2% of the bacterial genome in many cases (3). Advances in whole-genome sequencing (WGS) and the availability of a large number of assembled bacterial genomes have made possible new approaches to strain- and clone-level epidemiologic tracking of isolates, including the ability to apply MLST schemes on a genome-wide scale. Inference of clonal relationships from the comparison of whole-genome sequences for the purpose of epidemiologic investigations can involve a bit more subtlety than classical MLST analysis, however, and there are abundant opportunities for drawing incorrect conclusions. Perhaps the simplest of the many available approaches is that of direct-alignment-based (single nucleotide polymorphism [SNP]-based) phylogenetic analysis, often built on the underlying statistical assumptions of independent individual substitution events. Direct-alignment-based phylogenetic analyses have been used to study the population structures of bacterial outbreaks and to construct transmission maps in a variety of epidemiologic investigations (4–6). However, single recombination

events, mobile genetic element insertions, or other horizontal genetic transfer events can result in multiple SNPs or more-complex rearrangements over a contiguously exchanged or interrupted region. Consequently, genome-wide direct-alignment methods, in addition to being computationally expensive, may result in the generation of incorrect tree topologies due to such rearrangements (7). Alternatively, MLST-derived approaches applied to whole-genome sequences offer computationally simplified methods for dealing with these issues, as the unit of analysis is the allele as opposed to the individual SNP (2, 3). These methods include ribosomal MLST (rMLST), whole-genome MLST (wgMLST), and core genome MLST (cgMLST).

The rMLST scheme is based on analysis of (usually) 52 to 53 genes encoding ribosomal protein subunits (*rps*) present in almost all bacteria, conferring the advantage of universality to this method (2). The wgMLST strategies, in contrast, analyze all loci in a given isolate, including intergenic regions if they are defined. A challenge of wgMLST is that comparisons between loci shared only by subsets, and not all, of the members of the group to be analyzed can present statistical difficulties. The cgMLST methods address this problem by comparing only genes common to a set of bacterial isolates (2). These methods are now mature and have been applied successfully to the analysis of a number of bacterial outbreaks (8–21).

Given the developments in WGS-based strain typing described above, there is much excitement about applying these methods to real-time hospital epidemiology, but the actual cost-benefit equation, the feasibility of achieving results on an actionable time scale, and the technical requirements for implementation in a “real world” clinical microbiology laboratory are difficult to study. In work published in this issue of the *Journal of Clinical Microbiology*,

Accepted manuscript posted online 14 September 2016

Citation Dekker JP, Frank KM. 2016. Next-generation epidemiology: using real-time core genome multilocus sequence typing to support infection control policy. *J Clin Microbiol* 54:2850–2853. doi:10.1128/JCM.01714-16.

Editor: P. Bourbeau

Address correspondence to Karen M. Frank, Karen.frank@nih.gov.

For the article discussed, see doi:10.1128/JCM.00790-16.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

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

Mellmann and colleagues attempt to address these issues through the implementation of a prospective hospital multidrug resistance (MDR) epidemiologic surveillance program based on real-time WGS and cgMLST at the University Hospital Muenster, Muenster, Germany, a 1,450-bed tertiary care center (22).

In their study, WGS-based strain typing was used to complement classical epidemiologic investigations of transmission. The study was divided into two 6-month intervals, during which all methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococcus faecium* (VRE), MDR *Escherichia coli*, and MDR *Pseudomonas aeruginosa* isolates underwent whole-genome sequencing with a bench-top Illumina MiSeq system. Note that swab-based surveillance screening was performed only for MRSA in this hospital during the study period; all other isolates were derived from clinical cultures, precluding assessment of baseline carriage rates. During interval I of the study, 645 isolates underwent cgMLST analysis with SeqSphere+ (Ridom GmbH), a popular proprietary software package developed in part by one of the study's coauthors, and baseline transmission rates of the two most common species were calculated: MRSA at 5.8% and MDR *E. coli* at 2.3%. Based on these transmission rate levels, which the authors deemed to be low, the decision was made to discontinue certain isolation protocols for interval II of the study. Patients infected with Gram-negative bacteria resistant to representatives of three defined antimicrobial classes but susceptible to imipenem/meropenem (3MDR-GN) were isolated only in specified high-risk areas, but isolation precautions were discontinued for other areas of the hospital. It should be noted that although the authors refer to the baseline transmission rates as "low," there are no benchmarks for what rate would be considered acceptably low. Transmission from a single index patient can lead to an outbreak with associated deaths for some resistant organisms of high virulence, so the tolerance for any transmission would be quite small in those cases (6, 23, 24).

During interval II, 550 isolates were sequenced, and genotyping data suggesting transmission events were relayed immediately to infection control personnel. Analysis of these isolates by cgMLST did not reveal a statistically significant change in inferred transmission rates between the two intervals. However, the overall number of 3MDR *E. coli* patient cases did increase from 86 in interval I to 120 in interval II (~40% increase). The authors concluded that the bulk of this increase was associated with genetically distinct isolates based on cgMLST analysis, arguing against transmission. As noted above, a limitation in the study design was the lack of rectal surveillance cultures to measure asymptomatic colonization with MDR *Enterobacteriaceae*. Without such measurements, one cannot rule out a significant increase in "silent" transmission events that were not detected. The authors included epidemiological data in the algorithm to define transmission, after an initial exclusion of transmission based on genomic analysis. There are published examples of studies in which WGS analysis identified presumed transmissions that were not identified by classical epidemiological approaches (6, 25). There have also been studies demonstrating the importance of considering environmental samples in defining nosocomial transmission for some species. Environmental sampling, which can be quite labor-intensive and may be most appropriate in selected circumstances (26–28), was not included in the study by Mellmann et al.

To assess the turnaround time and technical feasibility of the approach for other institutions, it is important to consider both

the initial microbiology testing and the computational workflow. Sequencing was performed in 2 to 3 batches per week, and genomes were assembled with either proprietary CLC Genomic Workbench software (Qiagen) or SeqSphere+ and then analyzed with SeqSphere+ for cgMLST computation. For MRSA analyses, the authors used the public cgMLST scheme (11). For VRE, MDR *E. coli*, and MDR *P. aeruginosa* analyses, the authors created *ad hoc* schemes using the SeqSphere+ target definer (14). Assessment of sequencing quality was based on the number of successfully extracted cgMLST targets. A measure of genetic distance was calculated based on a count of the number of loci where alleles differed between pairs of isolates. Minimum spanning trees were generated from these difference counts, and isolates that differed by greater than a defined number of alleles were considered to be ruled out for transmission. For this purpose, a threshold of >6 was used for MRSA, a threshold of >8 for VRE, a threshold of >10 for MDR *E. coli*, and a threshold of >14 for MDR *P. aeruginosa*. These thresholds represent critical parameters in defining transmission rates in this type of analysis and were selected as species specific. After an initial exclusion procedure based on genomic difference criteria, epidemiologic data were incorporated to define likely transmission events.

Mean sequencing turnaround time ranged from 4.4 to 5.3 days after initial culture, identification, and susceptibility testing in the clinical microbiology laboratory. Implementation of this study represents a technical *tour de force*, and only a small number of hospital laboratories would have the capability to implement such an operation at this time. The delay in possible recognition of an outbreak required by the sequencing and analysis steps of this workflow (which, it should be noted, is not necessarily any longer than that involved in pulsed-field gel electrophoresis [PFGE] or classical MSLT analysis) could be mitigated by other measures, such as the use of parallel, rapid PCR for specific examples of worrisome plasmid-borne resistance such as that mediated by KPC/NDM/OXA family genes. Also, identification of 4MDR-GN isolates, pan-resistant isolates, or isolates in high-risk wards could prompt action before WGS analysis was completed. Additionally, the authors make the point that genomic data could be used for rapid development of outbreak-specific PCR for rapid screening.

The authors calculate the cost of their study, including an instrument depreciation model and assuming the sequencing of 1,500 isolates per year, at ~€200 per isolate. Though a modest technical labor cost is included in the above calculation, many clinical microbiology laboratories without existing bioinformatics expertise would require a full-time, embedded bioinformaticist, or tight collaboration with a core sequencing and bioinformatics facility, at significant expense to implement such a program. The authors then compare the total cost of their study with an estimate of the amount saved by discontinuation of isolation procedures. The logic of this cost calculation is a bit tricky. The calculated savings are derived from the reduction in the number of incompletely occupied double rooms resulting from the decrease in the isolation precautions between intervals I and II. The generalizability of this savings would depend on the bed configuration for a given hospital and might not apply to a hospital with only single rooms, as with many newly constructed hospitals in the United States.

Mellmann and colleagues are to be commended on their large and complicated real-world study, which contributes substantially to the field. As in any study of that size and complexity, there

are perhaps as many interesting questions raised as answered. One might wonder how the conclusions might have been different had a large, hospital-wide outbreak occurred during interval II. Such real-world epidemiologic events will demonstrate how effectively cgMLST and related methods can be used in real time to recognize and break transmission chains. One must be cautious in generalizing conclusions based on this single hospital experience to other hospitals with different policies, practices, bed configurations, MDR bacterial species prevalences, and baseline transmission rates. Each institution must define isolation policies with careful consideration of their specific data and resources. Additional laboratories are striving to develop sequencing and analysis capabilities such as those described in the study by Mellmann et al., and these issues and others await further large-scale implementation studies in diverse hospital contexts. One fact that is beyond argument is that WGS has fundamentally changed the study of bacterial outbreaks and will play an increasingly central role in hospital epidemiology in the future.

ACKNOWLEDGMENT

This work was supported by the Intramural Research Program of the National Institutes of Health Clinical Center.

REFERENCES

- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95:3140–3145. <http://dx.doi.org/10.1073/pnas.95.6.3140>.
- Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 11:728–736. <http://dx.doi.org/10.1038/nrmicro3093>.
- Jolley KA, Maiden MC. 2014. Using multilocus sequence typing to study bacterial variation: prospects in the genomic era. *Future Microbiol* 9:623–630. <http://dx.doi.org/10.2217/fmb.14.24>.
- Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamai-chi 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. <http://dx.doi.org/10.1056/NEJMoa1012928>.
- Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, Chen W, Pu F, Peng Y, Li J, Xi F, Li S, Li Y, Zhang Z, Yang X, Zhao M, Wang P, Guan Y, Cen Z, Zhao X, Christner M, Kobbe R, Loos S, Oh J, Yang L, Danchin A, Gao GF, Song Y, Li Y, Yang H, Wang J, Xu J, Pallen MJ, Wang J, Aepfelbacher M, Yang R; E. coli O104 H4 Genome Analysis Crowd-Sourcing Consortium. 2011. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. *N Engl J Med* 365:718–724. <http://dx.doi.org/10.1056/NEJMoa1107643>.
- Snitkin ES, Zelazny AM, Thomas PJ, Stock F; NISC Comparative Sequencing Program Group, Henderson DK, Palmore TN, Segre JA. 2012. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci Transl Med* 4:148ra116.
- Posada D, Crandall KA, Holmes EC. 2002. Recombination in evolutionary genomics. *Annu Rev Genet* 36:75–97. <http://dx.doi.org/10.1146/annurev.genet.36.040202.111115>.
- Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* 6:e22751. <http://dx.doi.org/10.1371/journal.pone.0022751>.
- Hauser E, Mellmann A, Semmler T, Stoeber H, Wieler LH, Karch H, Kuebler N, Fruth A, Harmsen D, Weniger T, Tietze E, Schmidt H. 2013. Phylogenetic and molecular analysis of food-borne Shiga toxin-producing *Escherichia coli*. *Appl Environ Microbiol* 79:2731–2740. <http://dx.doi.org/10.1128/AEM.03552-12>.
- Kovanen SM, Kivisto RI, Rossi M, Schott T, Karkkainen UM, Tuuminen T, Uksila J, Rautelin H, Hanninen ML. 2014. Multilocus sequence typing (MLST) and whole-genome MLST of *Campylobacter jejuni* isolates from human infections in three districts during a seasonal peak in Finland. *J Clin Microbiol* 52:4147–4154. <http://dx.doi.org/10.1128/JCM.01959-14>.
- Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. 2014. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J Clin Microbiol* 52:2365–2370. <http://dx.doi.org/10.1128/JCM.00262-14>.
- Revez J, Zhang J, Schott T, Kivisto R, Rossi M, Hanninen ML. 2014. Genomic variation between *Campylobacter jejuni* isolates associated with milk-borne-disease outbreaks. *J Clin Microbiol* 52:2782–2786. <http://dx.doi.org/10.1128/JCM.00931-14>.
- Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, Weniger T, Niemann S. 2014. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J Clin Microbiol* 52:2479–2486. <http://dx.doi.org/10.1128/JCM.00567-14>.
- Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, Harmsen D, Mellmann A. 2015. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. *J Clin Microbiol* 53:2869–2876. <http://dx.doi.org/10.1128/JCM.01193-15>.
- de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, Brouwer E, Rogers M, Kraat Y, Bonten M, Corander J, Westh H, Harmsen D, Willems RJ. 2015. Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. *J Clin Microbiol* 53:3788–3797. <http://dx.doi.org/10.1128/JCM.01946-15>.
- Chen Y, Gonzalez-Escalona N, Hammack TS, Allard M, Strain EA, Brown EW. 12 August 2016. Core genome multilocus sequence typing for the identification of globally distributed clonal groups and differentiation of outbreak strains of *Listeria monocytogenes*. *Appl Environ Microbiol* <http://dx.doi.org/10.1128/AEM.01532-16>.
- Moran-Gilad J, Prior K, Yakunin E, Harrison TG, Underwood A, Lazarovitch T, Valinsky L, Luck C, Krux F, Agmon V, Grotto I, Harmsen D. 2015. Design and application of a core genome multilocus sequence typing scheme for investigation of Legionnaires' disease incidents. *Euro Surveill* 20(28):pii=21186. <http://dx.doi.org/10.2807/1560-7917.ES2015.20.28.21186>.
- Haendiges J, Jones J, Myers RA, Mitchell CS, Butler E, Toro M, Gonzalez-Escalona N. 2016. A nonautochthonous U.S. strain of *Vibrio parahaemolyticus* isolated from Chesapeake Bay oysters caused the outbreak in Maryland in 2010. *Appl Environ Microbiol* 82:3208–3216.
- Kingry LC, Rowe LA, Respicio-Kingry LB, Beard CB, Schriefer ME, Petersen JM. 2016. Whole genome multilocus sequence typing as an epidemiologic tool for *Yersinia pestis*. *Diagn Microbiol Infect Dis* 84:275–280. <http://dx.doi.org/10.1016/j.diagmicrobio.2015.12.003>.
- Toro M, Retamal P, Ayers S, Barreto M, Allard M, Brown EW, Gonzalez-Escalona N. 12 August 2016. Whole genome sequencing analysis of *Salmonella* Enteritidis isolated in Chile provides insights about possible transmission between gulls, poultry and humans. *Appl Environ Microbiol* <http://dx.doi.org/10.1128/AEM.01760-16>.
- Raphael BH, Baker DJ, Nazarian E, Lapierre P, Bopp D, Kozak-Muiznieks NA, Morrison SS, Lucas CE, Mercante JW, Musser KA, Winchell JM. 2016. Genomic resolution of outbreak-associated *Legionella pneumophila* serogroup 1 isolates from New York State. *Appl Environ Microbiol* 82:3582–3590. <http://dx.doi.org/10.1128/AEM.00362-16>.
- Mellmann A, Bletz S, Boking T, Kipp F, Becker K, Schultes A, Prior K, Harmsen D. 2016. Mellmann A, Bletz S, Boking T, Kipp F, Becker K, Schultes A, Prior K, Harmsen D. 2016. Real-time genome sequencing of resistant bacteria provides precision infection control in an institutional setting. *J Clin Microbiol* 54:2874–2881. <http://dx.doi.org/10.1128/JCM.00790-16>.
- Ross AS, Baliga C, Verma P, Duchin J, Gluck M. 2015. A quarantine process for the resolution of duodenoscope-associated transmission of multidrug-resistant *Escherichia coli*. *Gastrointest Endosc* 82:477–483. <http://dx.doi.org/10.1016/j.gie.2015.04.036>.
- Munoz-Price LS, Zembower T, Penugonda S, Schreckenberger P, Lavin

- MA, Welbel S, Vais D, Baig M, Mohapatra S, Quinn JP, Weinstein RA. 2010. Clinical outcomes of carbapenem-resistant *Acinetobacter baumannii* bloodstream infections: study of a 2-state monoclonal outbreak. *Infect Control Hosp Epidemiol* 31:1057–1062. <http://dx.doi.org/10.1086/656247>.
25. Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, Petrovska L, Ellis RJ, Elson R, Underwood A, Green J, Hanage WP, Jenkins C, Grant K, Wain J. 2015. Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clin Infect Dis* 61:305–312. <http://dx.doi.org/10.1093/cid/civ318>.
26. Bartley PB, Ben Zakour NL, Stanton-Cook M, Muguli R, Prado L, Garnys V, Taylor K, Barnett TC, Pinna G, Robson J, Paterson DL, Walker MJ, Schembri MA, Beatson SA. 2016. Hospital-wide eradication of a nosocomial *Legionella pneumophila* serogroup 1 outbreak. *Clin Infect Dis* 62:273–279. <http://dx.doi.org/10.1093/cid/civ870>.
27. Garvey MI, Bradley CW, Tracey J, Oppenheim B. 2016. Continued transmission of *Pseudomonas aeruginosa* from a wash hand basin tap in a critical care unit. *J Hosp Infect* 94:8–12. <http://dx.doi.org/10.1016/j.jhin.2016.05.004>.
28. Gherardi G, Creti R, Pompilio A, Di Bonaventura G. 2015. An overview of various typing methods for clinical epidemiology of the emerging pathogen *Stenotrophomonas maltophilia*. *Diagn Microbiol Infect Dis* 81: 219–226. <http://dx.doi.org/10.1016/j.diagmicrobio.2014.11.005>.