Whole Genome Sequencing Applications in Hospital Epidemiology and Infection Prevention

Michelle Doll[1] · Alexandra L. Bryson[1] · Tara N. Palmore[2]

Abstract

Purpose of Review Whole genome sequencing is increasingly used in epidemiologic surveillance in healthcare centers, shedding new light on the transmission of healthcare-associated infections. As processing times for these technologies shorten, the ability to use sequencing data for targeted infection prevention is seemingly attainable and of great interest to infection prevention practitioners in an era of limited resources.

Recent Findings Common healthcare-associated infections such as *C. difficile* and methicillin-resistant *Staphylococcus aureus* may be acquired in diverse settings including the community, rather than direct patient-to-patient transmission as previously thought. This along with the emergence of new organisms such as *Candida auris* may indicate that infection prevention interventions should be all-encompassing rather than organism-specific in focus.

Summary Whole genome sequencing technologies are providing a level of detail in assessing organism relatedness that is changing our understanding of the transmission of infections in healthcare settings and may impact infection prevention strategies in the future.

Introduction

Over the last 2 decades, whole genome sequencing (WGS) has emerged from the research realm and is increasingly applied to facilitate infection prevention efforts in healthcare facilities [1]. The ability to process whole genomes quickly and inexpensively is due to high throughput or "next generation" sequencing methods in which DNA sequences are fragmented, sequenced in pieces simultaneously, and then reconstructed in the correct order using areas of overlap to inform orientation [2, 3]. The resulting sequence can then be compared against other sample sequences or a reference sequence. The number of base pair differences (single nucleotide variants or SNVs) between 2 samples in coding regions of the genome can be used to assess relatedness [2, 3]. Depending on the organism, the differences can also inform how long ago the isolates may have diverged from the evolutionary predecessor, allowing for the construction of detailed phylogenic trees [3].

Not only have the technologies supporting WGS improved in terms of cost, usability, and ease of analysis, but the turnaround times for the data produced are reaching to point of supporting "real-time" infection prevention analytics [4]. When potential pathogens from patient isolates are prospectively sequenced and compared across a healthcare setting, these data can alert IP teams of a potential transmission event that they can then act upon in real time. Using traditional methods, there is often a delay in the recognition of clusters or outbreaks, owing to the dependence on a clear epidemiologic link. This complicates control measures as the issues leading to transmission may have been occurring unchecked for long periods of time or extend beyond the identified unit to other areas of the facility.

✉ Michelle Doll
Michelle.doll@vcuhealth.org

1 Virginia Commonwealth University School of Medicine, Richmond, 1000 E. Marshall Street, Richmond, VA 23298, USA

2 George Washington School of Medicine, Washington, D.C, USA

Whole genome sequencing data continue to challenge long-standing dogma around the transmission of pathogens in the hospital. For the first time, we are able to appreciate the complexity of transmission dynamics, unique to each pathogen, which prior to WGS were largely invisible. Some instances of healthcare-associated transmission are shown to be unrelated, while other transmission events previously missed are now identified. Still, uncertainty in the interpretation of data exists, and epidemiologic data remain integral to understanding the webs of transmission that exist in healthcare settings as well as in the larger community. Some of the questions raised are more philosophical in nature, for example, how much transmission necessitates intensive infection prevention responses and resources? In an era of zero infection targets, some may argue that even one instance of spread is unacceptable, while others may argue that in the setting of predominantly non-healthcare-associated transmission, resources should be focused elsewhere in the absence of an outbreak.

As WGS is increasingly available to IP programs, we open the door to a new era of patient safety, in which threats could theoretically be identified and eliminated prior to causing widespread infections in vulnerable patients. Further experience in the best ways to apply these methods to infection prevention is an area of intense interest and investigation.

New Insights into the Etiology of Purported Hospital-Acquired Infections

Methicillin-resistant *Staphylococcus aureus* (MRSA) infections in hospitalized patients were traditionally attributed to breakdowns of basic infection control practices. As such, it has been a common target for quality programs. Whole genome sequencing as applied to outbreaks of MRSA in healthcare settings has generally helped to identify a common source, at least in the published literature [5]. However, when Price et al. [6•] investigated the epidemiology of *Staphylococcus* colonization in patients and staff under non-outbreak conditions, some important insights arise. The authors performed 14 months of serial screening of all patients in an adult ICU, at admission and weekly thereafter, collecting over 275 isolates. Both full genome sequencing and *spa*-typing were performed, and definitions for transmission were < 40 SNV difference for WGS and matching *spa*-type with overlapping ICU stay for *spa*-typed comparisons. Spa-typing is a focused molecular method for assessing the relatedness of *S. aureus* species, based on sequencing a specific polymorphic X region of the organism's *spa* gene [7]. New acquisition of *S. aureus* occurred in 44 patients, and *spa*-typing identified 5 possible transmission events. However, WGS of the same isolates discounted 3 of those transmissions, confirmed 2, and identified

another 5 transmission events. The authors conclude that traditional typing methods (e.g., *spa*-typing) do not provide the necessary resolution to drive infection prevention interventions. Furthermore, most acquisitions of *S. aureus* colonization in their study did not have evidence of transmission from patient to patient by genetic analysis (18.9% of acquisitions), and some baseline or admission strains of MRSA and methicillin-susceptible *S. aureus* (MSSA) were highly genetically related, raising the possibility of common community sources [6•]. While in the current era, 7 transmission events in the space of 14 months would be considered unacceptable, the fact that the vast majority of acquisitions did not stem from a clear nosocomial source challenged the dogma of MRSA as an organism transmitted from patient to patient in the hospital.

Acknowledging the missing links potentially influencing the results of their first study, Price et al. went on to perform another 14-month survey of *S. aureus* in an ICU and a high-dependency unit, this time including healthcare personnel (HCP) and environmental surfaces [8]. In addition to sampling patients as previously described [6•], they sampled HCP weekly and a range of environmental surfaces monthly; air samples were also collected monthly. Samples were grouped into subtypes after full genome sequencing, defined as those isolates sharing a similar genome with < 40 SNV differences; isolates with > 40 SNV differences would be considered unrelated and assigned a different subtype. When applied to their dataset, authors found that the majority of *S. aureus* subtypes collected during the study were unique: 380/416 subtypes from patient samples, 131/159 HCP samples, and 37/78 environmental samples. However, 11 subtypes were found in patients, HCP, and the environment. Patients and HCP shared another 6 subtypes, patients and the environment shared 19 subtypes, and HCP and the environment shared 21 subtypes. Some of these shared subtypes were classified as transmissions, as the new subtype was acquired at some timepoint after first appearing in another source (i.e., patient, environment, and HCP) [8]. While *S. aureus* transmission in hospitals does occur in a minority of cases, the vast majority of newly acquired *S. aureus* has no identified common source within the healthcare setting.

Today, new *S. aureus* infections in healthcare are thought to derive primarily from a patient's own endogenous flora. While new strains could be acquired from breakdowns in infection prevention, strategies to decrease the risk of *S. aureus* infection while in the hospital often aim to decolonize patients of this organism. Whole genome sequencing helped make the case for MRSA as an endemic and not epidemic organism in the population-level study by Ulhemann et al. [9]. The group compared 348 *spa*-type 8 MRSA isolates from community dwellers in the New York City region with all published reference genomes of the same type that included sources from geographically distant regions of

California and Texas. The authors found expected low variability within household samples and higher variability in the community. Unrelated isolates, including those from geographically distinct regions, were interspersed among their New York City samples, suggesting multiple introductions of these strains into communities over prolonged periods of time, rather than the epidemic spread of a single strain [9]. Some healthcare systems have abandoned the use of isolation precautions for MRSA given the apparent endemicity of the organism and the observation that standard infection prevention practices such as hand hygiene are fundamentally important to break the chains of transmission for this organism [10].

Whole genome sequencing studies have challenged our epidemiologic understanding of other organisms as well. Prior to 2013, *Clostridioides difficile* was the prototypical hospital-onset pathogen, with widespread healthcare-associated outbreaks associated with a new, aggressive NAP-1/072 strain terrorizing the National Health Service in the UK, as well as sites in the USA and Canada [11]. However, whole genome sequencing data from a 5-year, regional study in Oxfordshire, UK, unearthed an astonishing genetic diversity among *C. difficile* isolates with 45% of the > 1200 specimens distinct from all others [12•]. Only a minority (19%) appeared linked to another isolate to suggest transmission in this study. The authors suggest that additional reservoirs, possibly within the community, drive acquisition of *C. difficile* and continually introduce new strains into healthcare settings [12•].

Where could those other reservoirs be? The observation that virtually all healthy infants become colonized with *C. difficile* in the first year of life points to a widespread environmental presence of this organism [13]. In addition, a One Health link between humans, the environment, and animals has been demonstrated in other studies [14, 15]. Whole genome sequencing performed on 248 globally collected isolates of *C. difficle* ribotype 078 from both human and animal sources demonstrated significant genomic overlap among the human and animal isolates as well as among isolates from disparate regions [16]. The authors conclude that their data suggest "a highly-linked, inter-continental transmission network between humans and animals," [16] in which new strains continually circulate between community-derived reservoirs and healthcare settings.

Understanding Emergence of Novel Organisms and Antimicrobial Resistance

Emergence or re-emergence of important pathogens is occurring with escalating frequency in our world, thought to be due to multiple factors such as increased crowding and environmental strains. Whole genome sequencing studies

attempting to elucidate the epidemiology of emerging and re-emerging infections have allowed us to observe the manner in which new infections are transmitted and spread at the global level. In addition, study of novel antimicrobial resistance mechanisms can help inform our understanding of the epidemiology of multidrug-resistant organisms.

Carbapenem-resistant Enterobacterales (CRE) have been involved in high-profile outbreaks, particularly those bacteria that have the ability to produce carbapenemases (i.e., Carbapenemase-producing CRE or CP-CRE) that often confer high levels of antimicrobial resistance making infections very difficult to treat and thus potentially highly morbid. In one example, the National Institutes of Health Clinical Center suffered a nosocomial outbreak in 2011 involving an index patient transferred from another facility with known CP-CRE colonization. The isolate spread to 18 patients, 7 of whom died of CP-CRE infection [17, 18]. Whole genome sequencing used to support that outbreak investigation revealed some important insights. First, the index patient was only present in the ICU for two separate 24-h periods, demonstrating that prolonged admissions are not required for widespread transmission in a facility, as unseen breaches of infection control can occur at any time. In fact, WGS demonstrated some heterogeneity in the index patients' CP-CRE isolates taken from different body sources and traced these isolates to 3 separate transmission events driving the overall outbreak [17]. Second, there was a 3-week interval between the index patient's first ICU stay and the identification of CP-CRE in the next patient [17]. Lastly, evolution of colistin resistance and other genetic diversification occurred within the isolates just in the 8 months comprising the outbreak [17], a stark reminder that genetic heterogeneity does not ensure the absence of an outbreak situation given the rapidity with which recombination events can occur for some pathogens.

Outside of an outbreak setting, researchers performed WGS (full genome sequencing including plasmids) on carbapenem-resistant isolates from clinical cultures in 3 Boston area hospitals and 1 hospital in California in an effort to detect unrecognized outbreaks, transmission between facilities, and/or transmission of resistance genes between different bacterial species [19••]. Similar to prior studies, an astounding degree of diversity was present among the samples, and only 2 instances of potential relatedness were found, in samples differing by 17 and 19 SNV. For context, the authors reported *K. pneumoniae* ST258 isolates from the same patient differing by 29 SNV and noted the difficulty in establishing SNP cut-offs for these gram-negative enteric organisms based on the mobility of their genomes [19••]. However, in the absence of clear-cut transmission events, they found identical portions of genomes encoding resistance mechanisms [19••], raising concern that clinical cultures alone do not provide enough data to link transmission

pathways together within healthcare systems, and our larger community clinical cultures may represent the tip of the iceberg, missing human and environmental colonization. Early identification and isolation of these organisms may not be sufficient to halt transmission while their genetic material is shuffled potentially in environmental and other background reservoirs.

A more extensive bank of isolates encompassing 5 years of a government surveillance program in Singapore allowed researchers there to locate additional evidence of nosocomial CP-CRE transmission [20]. In this study, clonal transmission was classified if two isolates had the same core genome and carbapenemase gene allele and differed by less than a SNV threshold based on mutation rates and Bayesian probabilities [20]. The study also looked at the relatedness of plasmids alone, defining plasmid transmission as the acquisition of a matching plasmid containing carbapenemase genes [20]. From 901 patient isolates, 779 acquisitions were identified: of these, 327 were related by the core genome, and 349 represented shared plasmid sequences. The remaining acquisitions had no related isolates within the dataset [20]. It is not surprising that many more presumptive transmission events can be identified with a larger and more comprehensive dataset. This study demonstrates that transmission of not only multidrug-resistant organisms but also their mobile genetic elements is underappreciated based on the level of detection and analysis available in most settings. Containment via isolation of known clinical cases only scratches the surface; real control efforts will require comprehensive infection prevention strategies applied universally in addition to persistent disinfection of the healthcare environment.

Global emergence of novel organisms causes great public health concern, such as the case of *Candida auris.* First identified and reported in the scientific literature in 2009 [21], *C. auris* may have emerged as early as 1996 but has only been identified as such in retrospect [22]. *C. auris* has distinct biological characteristics such as tolerance for desiccation, resistance to antifungal drugs and disinfectants, and colonization of skin, rendering it well adapted for nosocomial spread. Globally, *C. auris* exists in distinct geographical clades, but within-clade isolates are difficult to distinguish using molecular methods analyzing only parts or subsections of the genome [23]. Whole genome sequencing capabilities for *C. auris* were supported by the Centers for Disease Control and Prevention (CDC) as part of a global collaboration to analyze 54 patient samples from Pakistan, India, Venezuela, and South Africa [24]. Existing "draft" genomes for *C. auris* were used as well as a reference genome constructed from one of the study samples [24]. The analysis of 47 of the samples suitable for WGS revealed very little heterogeneity between samples from the same region and more variation between regions, suggesting that *C. auris* arose independently and simultaneously in these 4 regions rather than from a single epidemic strain [24]. Thus, the selective pressure favoring the rise of these public health threats may be present in multiple geographic regions with the ability to amplify in each area once it emerged [24]. Since that analysis, *C. auris* has spread globally to become a major international nosocomial threat.

Prospective Infection Prevention with Incorporation of Routine WGS

Clinical recognition of an outbreak is often delayed, resulting in further transmission within the healthcare facility before the problem is identified. With routine, real-time use of WGS for epidemiologically important organisms, however, the window between outbreak occurrence and recognition is shrinking. Such was the experience of centers using large-scale WGS on SARS-CoV-2 isolates during the COVID-19 pandemic [4, 25, 26]. The UK was able to leverage their WGS data to inform facilities of evidence of possible transmission events, leading to rapid deployment of infection prevention resources [4]. This targeted approach would have been particularly important in the context of the overall strain COVID-19 put on healthcare systems; it was important for infection prevention teams to understand where their efforts were most needed.

Routine sequencing of bacterial pathogens may be the logical next step. Berbel Caban et al. [27] describe the development of a novel data integration program in which WGS data could be overlaid with epidemiologic data from the patient record in order to better detect outbreaks. By applying this program to archived MRSA strains, they were able to detect outbreaks previously unrecognized, including a long-term outbreak spanning 21 months [27]. At the University of Pittsburgh Medical Center-Presbyterian Hospital (UPMC), a machine learning program was put into place to facilitate a truly prospective monitoring system that combined routinely collected WGS data from clinically obtained targeted pathogens with EMR data [28••]. The EMR data included procedural charge codes, and recognizing transmission in hospital facilities determined by WGS is not always able to be linked by overlapping patient locations [28••]. During implementation, the prospective WGS-EMR alert system was compared to infection prevention-driven requests for WGS as a result of epidemiologic concerns for an outbreak. The WGS-EMR alert system performed better, identifying 65 potential clusters over the 2-year period (another 33 clusters identified by WGS could not be tied by the EMR algorithms or manual chart reviews). Simultaneous "reactive" WGS requested by infection prevention occurred 15 times during the 2 years, of which only 2 clusters were found to represent likely transmission events. Not only was the WGS-EMR system able to identify more events, but also it

uncovered transmission occurring in procedural areas which otherwise might have gone unrecognized [28••]. These data suggest that most transmission events and undetected breaches in infection prevention are underappreciated in healthcare facilities. At the same time, supposed transmission events are often debunked by WGS, similar to what we have seen presented in the literature when WGS is applied to understand the epidemiology of "nosocomial" organisms. A prospective approach to WGS could allow over-tasked infection preventionists to focus efforts on those areas that truly need interventions.

WGS has also been used to assess infection control practices prospectively to assess a major change in practice. Mellmann et al. [29] applied real-time WGS to all MRSA, VRE, MDR *E. coli*, and MDR *Pseudomonas* species collected at their 1450-bed tertiary medical center in two 6-month intervals. In the first interval, all patients were isolated on contact precautions if known to be colonized or infected with these organisms. In the subsequent period, contact precautions were discontinued for patients infected or colonized with the MDR gram-negative organisms. Sequences of isolates collected in these time periods were compared along with available epidemiologic data to detect clusters. Clusters were detected for both MRSA and MDR *E. coli* throughout the study; in the units that discontinued contact precautions, no major transmission events were detected for MDR *E. coli*, and no difference was found in the number of MRSA transmission events during the second 6-month interval [29]. Authors conclude that WGS allowed infection preventionists to focus efforts on the areas and types of interventions that were most relevant to WGS-documented risks in their facility [29]. The study also raises the possibility that WGS could support interventional infection prevention studies in a more robust way than traditional microbiologic methods, by offering unequivocal confirmation rather than just species-level suggestion of transmission.

Logistical Challenges for Clinical Settings

Despite the promise of WGS to enhance infection prevention, these technologies remain out of reach for most facilities. While the material costs of sequencing equipment and reagents have fallen, upfront investments in equipment are substantial. Significant costs for human resources and expert interpretation of results have kept these capabilities in the research realm and out of most infection control programs, even those within academic medical centers. Laboratory standards for the performance of WGS in clinical labs are being developed, but the reliability of results obtained by different labs or even different operators within the same lab may be lacking [30]. Once sequences are produced, analysis of WGS data is complex, relying on automated computer

programs of which there are numerous options. For infectious diseases, reference databases for unusual organisms and in particular parasitic and fungal pathogens are notoriously lacking [30]. Thus, there are multiple opportunities for the introduction of subjectivity in the process at present in navigating imperfect sequence runs, determination of cut-offs for related versus not, and mutation rates and other assumptions about specific infectious organisms. Calls for standardized protocol for WGS that are species-specific are attempting to address the lack of uniformity in these processes [30].

Cost-effectiveness studies suggest that despite these challenges, WGS would remain cost-effective in preventing HAIs [29, 31, 32••]. Dymond et al. [31] published modeling data to support WGS effectiveness in reducing costs of acquisition and treatment of MRSA in hospitals in the UK over a 1-year period. Their model assumed WGS would be 90% effective in this prevention, but sensitivity analyses suggested that the cost-effectiveness would be durable over a range of effectiveness estimates as well as MRSA prevalence and volume of isolates undergoing WGS [31].

Attempting to strengthen assumptions in these models, Kumar et al. [32••] used past outbreak experience in their facility over a 6-year period to improve their estimates regarding the preventability of subsequent acquisitions. The authors developed a mathematical model encompassing a range of carefully established estimates for variables such as the effectiveness of IP interventions targeting specific transmission routes, time to WGS results, attributable mortality, and costs of treating specific infection types. Based on knowledge about prior outbreaks and published IP intervention effectiveness, an overall effectiveness of IP interventions in this study was estimated to be 30%. They were able to conclude that WGS would be cost-effective so long as healthcare facilities were willing to pay over $2400 to prevent each transmission event [32••]. This analysis underscores that the various factors represented in these cost-effectiveness estimations are not living within the same budgets. Whether facilities are willing to make the initial investments in this technology for infection prevention remains to be seen. In consideration of costs, it is unlikely that WGS could be applied to every potential pathogen or outbreak, and facilities with experience in using WGS for infection prevention have targeted specific organisms or populations that they consider to be the highest risk [1, 26, 28••, 29].

Conclusions

Despite the infrastructural costs of building local WGS capacity, centers that have succeeded in incorporating real-time WGS in routine infection prevention report ongoing

value added to their programs. This value is realized in targeting limited infection prevention resources to the areas in which it makes the most impact for patients. The value can be counted in actual infections prevented, often involving organisms of high epidemiologic significance, going beyond individual patient benefits to serving a public health mission. In the absence of sequencing data, the scientific literature shows that infection preventionists are spending time mitigating outbreaks that do not actually exist, while other transmission events fly under our radar unrecognized. It is important for infection prevention professionals to recognize both the opportunities and limitations inherent in available WGS technologies and to begin incorporating genomic epidemiology into infection prevention education. Infection prevention professionals will increasingly use WGS data to understand the transmission dynamics of organisms within healthcare settings and beyond.

Author Contributions MD, TP, and AB shared in literature review, drafting of the manuscript, and manuscript editing.

Compliance with Ethical Standards

Conflict of Interest The authors declare no competing interests.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
● Of importance
●● Of major importance

1. Parcell BJ, Gillespie SH, Pettigrew KA, Holden MTG. Clinical perspectives in integrating whole-genome sequencing into the investigation of healthcare and public health outbreaks – hype or help? J Hosp Infect. 2021;109:1–9. https://doi.org/10.1016/j.jhin.2020.11.001.

2. Price JR, Didelot X, Crook DW, Llewelyn MJ, Paul J. Whole genome sequencing in the prevention and control of Staphylococcus aureus infection. J of Hosp Infect. 2013;83:14–21.

3. Quainoo S, Coolen JPM, Sacha AFT van Hijum C, Martijn A, Huynen CWJGM, Willem van Schaik E, Wertheim HFL. Whole-genome sequencing of bacterial pathogens : the future of nosocomial outbreak analysis. Clin Microbiol Rev. 2017;30(4):1015–1064.

4. Francis RV, Billam H, Clarke M, et al. The impact of real-time whole-genome sequencing in controlling healthcare-associated SARS-CoV-2 outbreaks. J Infect Dis. 2022;225(1):10–8. https://doi.org/10.1093/infdis/jiab483.

5. Popovich KJ, Snitkin ES. Whole genome sequencing—implications for infection prevention and outbreak investigations. Curr Infect Dis Rep. 2017;19(4). https://doi.org/10.1007/s11908-017-0570-0.

6.● Price JR, Golubchik T, Cole K, et al. Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of staphylococcus aureus in an intensive care unit. Clin Infect Dis. 2014;58(5):609–18. https://doi.org/10.1093/cid/cit807. **One of the first studies to employ WGS to understand transmission dynamics on a unit outside of an outbreak situation**.

7. Hallin M, Friedrich AW, Struelens MJ. spa typing for epidemiological surveillance of Staphylococcus aureus. Methods Mol Biol. 2009;551:189–202. https://doi.org/10.1007/978-1-60327-999-4_15.

8. Price JR, Cole K, Bexley A, et al. Transmission of Staphylococcus aureus between health-care workers, the environment, and patients in an intensive care unit: a longitudinal cohort study based on whole-genome sequencing. Lancet Infect Dis. 2017;17(2):207–14. https://doi.org/10.1016/S1473-3099(16)30413-3.

9. Uhlemann AC, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MT, Peacock SJ, Lowy FD. Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proc Natl Acad Sci U S A. 2014;111(18):6738–43. https://doi.org/10.1073/pnas.1401006111. Epub 2014 Apr 21. PMID: 24753569; PMCID: PMC4020051.

10. Morgan DJ, Murthy R, Munoz-Price LS, Barnden M, Camins BC, Johnston BL, Rubin Z, Sullivan KV, Shane AL, Dellinger EP, Rupp ME, Bearman G. Reconsidering contact precautions for endemic methicillin-resistant Staphylococcus aureus and vancomycin-resistant Enterococcus. Infect Control Hosp Epidemiol. 2015;36(10):1163–72. https://doi.org/10.1017/ice.2015.156. Epub 2015 Jul 3.

11. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, Frost E, McDonald LC. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. Lancet. 2005;366(9491):1079–84. https://doi.org/10.1016/S0140-6736(05)67420-X.

12.● Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CLC, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TEA, Walker AS. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med. 2013;369(13):1195–205. https://doi.org/10.1056/NEJMoa1216064. **WGS applied to a *C. difficile* in a health system and community challenged the paradigm of *C. difficile* infection as a hospital-acquired conditions**.

13. Rousseau C, Poilane I, De Pontual L, Maherault AC, Le Monnier A, Collignon A. *Clostridium difficile* carriage in healthy infants in the community: a potential reservoir for pathogenic strains. Clin Infect Dis. 2012;55(9):1209–15.

14. Keel K, Brazier JS, Post KW, Weese S, Songer JG. Prevalence of PCR ribotypes among *Clostridium difficile* isolates from pigs, calves, and other species. J Clin Microbiol. 2007;45:1963–4.

15. Stone NE, Sidak-Loftis LC, Sahl JW, Vazquez AJ, Wiggins KB, Gillece JD, et al. More than 50% of *Clostridium difficile* isolates from pet dogs in Flagstaff, USA, carry toxigenic genotypes. PLoS ONE. 2016;11(10):e0164504. https://doi.org/10.1371/journal.pone.0164504.

16. Knetsch CW, Kumar N, Forster SC, Connor TR, Browne HP, Harmanus C, Sanders IM, Harris SR, Turner L, Morris T, Perry M, Miyajima F, Roberts P, Pirmohamed M, Songer JG, Weese JS, Indra A, Corver J, Rupnik M, Wren BW, Riley TV, Kuijper EJ, Lawley TD. Zoonotic transfer of *Clostridium difficile* harboring antimicrobial resistance between farm animals and humans. J Clin Microbiol. 2018;56(3):e01384-e1417. https://doi.org/10.1128/JCM.01384-17.

17. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program Group, Henderson DK, Palmore TN, Segre JA. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med. 2012;4(148):148ra116. https://doi.org/10.1126/scitranslmed.3004129.

18. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song Y, Tsai YC, Boitano M, Dayal J, Brooks SY, Schmidt B, Young AC, Thomas JW, Bouffard GG, Blakesley RW, NISC Comparative Sequencing Program, Mullikin JC, Korlach J, Henderson DK, Frank KM, Palmore TN, Segre JA. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med. 2014;6(254):254ra126. https://doi.org/10.1126/scitranslmed.3009845.

19.•• Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M, Chapman SB, Reis-Cunha JL, Shea TP, Young S, Zeng Q, Delaney ML, Kim D, Peterson EM, O'Brien TF, Ferraro MJ, Hooper DC, Huang SS, Kirby JE, Onderdonk AB, Birren BW, Hung DT, Cosimi LA, Wortman JR, Murphy CI, Hanage WP. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A. 2017;114(5):1135–40. https://doi.org/10.1073/pnas.1616248114. **WGS of gram negative organisms across 4 hospitals revealed remarkable diversity and suggests transmission of resistance mechanisms between bacterial species, and unrecognized outbreaks of these organisms in hospitals**.

20. Marimuthu K, Venkatachalam I, Koh V, et al. Whole genome sequencing reveals hidden transmission of carbapenemase-producing Enterobacterales. Nat Commun. 2022;13(1):1–11. https://doi.org/10.1038/s41467-022-30637-5.

21. Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. Candida auris sp. nov., a novel ascomycetous yeast isolated from the external ear canal of an inpatient in a Japanese hospital. Microbiol Immunol. 2009;53:41–4.

22. Lee WG, Shin JH, Uh Y, Kang MG, Kim SH, Park KH, Jang HC. First three reported cases of nosocomial fungemia caused by Candida auris. J Clin Microbiol. 2011;49(9):3139–42. https://doi.org/10.1128/JCM.00319-11.

23. Jeffery-Smith A, Taori SK, Schelenz S, Jeffery K, Johnson EM, Borman A, Candida auris Incident Management Team, Manuel R, Brown CS. Candida auris: a review of the literature. Clin Microbiol Rev. 2017;31(1):e00029-17. https://doi.org/10.1128/CMR.00029-17.

24. Lockhart SR, Etienne KA, Vallabhaneni S, et al. Simultaneous emergence of multidrug-resistant Candida auris on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. Clin Infect Dis. 2017;64(2):134–40. https://doi.org/10.1093/cid/ciw691.

25. Berggreen H, Løvestad AH, Helmersen K, Jørgensen SB, Aamot HV. Lessons learned: use of WGS in real-time investigation of suspected intrahospital SARS-CoV-2 outbreaks. J Hosp Infect. 2023;131:81–8. https://doi.org/10.1016/j.jhin.2022.10.003.

26. Marinelli TM, Dolan L, Jenkins F, Lee A, Davis RJ, Crawford S, Nield B, Ronnachit A, Van Hal SJ. The role of real-time, on-site, whole-genome sequencing of severe acute respiratory coronavirus virus 2 (SARS-CoV-2) in guiding the management of hospital outbreaks of coronavirus disease 2019 (COVID-19). Infect Control Hosp Epidemiol. 2022;9:1–5. https://doi.org/10.1017/ice.2022.220.

27. Berbel Caban A, Pak TR, Obla A, Dupper AC, Chacko KI, Fox L, Mills A, Ciferri B, Oussenko I, Beckford C, Chung M, Sebra R, Smith M, Conolly S, Patel G, Kasarskis A, Sullivan MJ, Altman DR, van Bakel H. PathoSPOT genomic epidemiology reveals under-the-radar nosocomial outbreaks. Genome Med. 2020;12(1):96. https://doi.org/10.1186/s13073-020-00798-3.

28.•• Sundermann AJ, Chen J, Kumar P, Ayres AM, Cho ST, Ezeonwuka C, Griffith MP, Miller JK, Mustapha MM, Pasculle AW, Saul MI, Shutt KA, Srinivasa V, Waggle K, Snyder DJ, Cooper VS, Van Tyne D, Snyder GM, Marsh JW, Dubrawski A, Roberts MS, Harrison LH. Whole-genome sequencing surveillance and machine learning of the electronic health record for enhanced healthcare outbreak detection. Clin Infect Dis. 2022;75(3):476–82. https://doi.org/10.1093/cid/ciab946. **Description of a novel implementation of a WGS strategy to identify possible outbreaks of high-risk pathogens in hospitals and comparison to standard surveillance practices**.

29. Mellmann A, Bletz S, Böking T, Kipp F, Becker K, Schultes A, Prior K, Harmsen D. Real-time genome sequencing of resistant bacteria provides precision infection control in an institutional setting. J Clin Microbiol. 2016;54(12):2874–81. https://doi.org/10.1128/JCM.00790-16.

30. Gargis AS, Kalman L, Lubin IM. Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. J Clin Microbiol. 2016;54(12):2857–65. https://doi.org/10.1128/JCM.00949-16.

31. Dymond A, Davies H, Mealing S, et al. Genomic surveillance of methicillin-resistant Staphylococcus aureus: a mathematical early modeling study of cost-effectiveness. Clin Infect Dis. 2020;70(8):1613–9. https://doi.org/10.1093/cid/ciz480.

32.•• Kumar P, Sundermann AJ, Martin EM, Snyder GM, Marsh JW, Harrison LH, Roberts MS. Method for economic evaluation of bacterial whole genome sequencing surveillance compared to standard of care in detecting hospital outbreaks. Clin Infect Dis. 2021;73(1):e9–18. https://doi.org/10.1093/cid/ciaa512. **Cost effectiveness analysis of WGS for detecting/responding to outbreaks compared to standard surveillance and response practices**.