Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Hospital Infection 122 (2022) 180–186

Available online at www.sciencedirect.com

Journal of Hospital Infection

journal homepage: www.elsevier.com/locate/jhin

Review

Infection prevention and control insights from a decade of pathogen whole-genome sequencing

D.W. Eyre [a,b,c,*]

[a] *Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford, UK*
[b] *National Institiute for Health Research, Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, UK*
[c] *Oxford University Hospitals, Oxford, UK*

ARTICLE INFO

Article history:
Received 25 January 2022
Accepted 31 January 2022
Available online 12 February 2022

Keywords:
Whole-genome sequencing
Epidemiology

SUMMARY

Pathogen whole-genome sequencing has become an important tool for understanding the transmission and epidemiology of infectious diseases. It has improved our understanding of sources of infection and transmission routes for important healthcare-associated pathogens, including *Clostridioides difficile* and *Staphylococcus aureus*. Transmission from known infected or colonized patients in hospitals may explain fewer cases than previously thought and multiple introductions of these pathogens from the community may play a greater a role. The findings have had important implications for infection prevention and control. Sequencing has identified heterogeneity within pathogen species, with some subtypes transmitting and persisting in hospitals better than others. It has identified sources of infection in healthcare-associated outbreaks of food-borne pathogens, *Candida auris* and *Mycobacterium chimera*, as well as individuals or groups involved in transmission and historical sources of infection. SARS-CoV-2 sequencing has been central to tracking variants during the COVID-19 pandemic and has helped understand transmission to and from patients and healthcare workers despite prevention efforts. Metagenomic sequencing is an emerging technology for culture-independent diagnosis of infection and antimicrobial resistance. In future, sequencing is likely to become more accessible and widely available. Real-time use in hospitals may allow infection prevention and control teams to identify transmission and to target interventions. It may also provide surveillance and infection control benchmarking. Attention to ethical and wellbeing issues arising from sequencing identifying individuals involved in transmission is important. Pathogen whole-genome sequencing has provided an incredible new lens to understand the epidemiology of healthcare-associated infection and to better control and prevent these infections.

Introduction

Over the last decade pathogen whole-genome sequencing has transformed from an emerging technology to become established as an important tool for understanding pathogen transmission and the epidemiology of infectious diseases. It has

* Address: Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK.
E-mail address: david.eyre@bdi.ox.ac.uk.

https://doi.org/10.1016/j.jhin.2022.01.024

led to improved understanding of the sources of infection and routes of transmission for several important healthcare-associated pathogens. In this personal perspective, commissioned following a Healthcare Infection Society Early Career Award, I outline this progress and my own involvement, with selected other illustrative studies. I also discuss how this has been associated with changes in infection prevention and control priorities.

Large-scale sequencing has challenged infection prevention and control orthodoxies

Clostridioides difficile can spread readily in healthcare settings in the absence of appropriate infection control. *C. difficile* was previously believed to be predominantly acquired from other symptomatic cases in healthcare settings, with interventions focused on preventing this.

However, contingent on control efforts in place at the time, large-scale sequencing of more than 1200 consecutive *C. difficile* infection cases in Oxfordshire, UK, during 2007–2010, revealed only a minority of infections, 35%, were sufficiently genetically related to have been plausibly acquired from another known case [1]. Additionally, only 19% of cases overall were both genetically related and shared some form of hospital contact. Similar findings have since been reproduced in Leeds and Liverpool in the UK and in Canada [2–4].

These findings suggest that most *C. difficile* infections are acquired from sources other than symptomatic infected hospital inpatients. Recent exposure to *C. difficile* in hospitals from other sources may still be important, with the associated infection prevention and control implications. Supporting the importance of recent acquisition leading to infection, in some studies pre-existing colonization with *C. difficile* has been reported to be protective against subsequent disease [5]. However, more recent data suggest the opposite with colonization increasing subsequent risk of disease, therefore highlighting the potential role of earlier healthcare and community-based acquisition [6].

The search for other sources of *C. difficile* infection has prompted studies of the role that asymptomatically colonized hospital inpatients may play, with evidence from sequencing and other high-resolution molecular typing, that asymptomatic patients may be a source of some healthcare-associated transmission [7–10]. Asymptomatic screening for *C. difficile* has been investigated as a control strategy, with its introduction associated with reduced infection incidence in one interrupted time-series study [11]. However, the efficacy and cost-effectiveness of such an approach still needs further study, ideally using cluster-randomized designs. Patients colonized with toxigenic *C. difficile* with diarrhoea of another cause may also be a source of transmission, and may be missed by infection control teams as they may test GDH-positive, but faecal toxin-negative [12].

In part prompted by findings of limited within-hospital transmission, other investigators have focused on community-based acquisition and the role that disease-causing *C. difficile* lineages in food production and domestic animals might play. Isolates from these sources have been found to be closely genetically related to those causing human disease [13]. One specific example is *C. difficile* ribotype 078 where genetic overlap between strains in pigs, farmers and clinical isolates was seen in a sequencing study from the Netherlands [14]. Demonstrating directionality of transmission, i.e. from an animal reservoir to human disease, is challenging without temporal data showing human acquisition (*C. difficile* negative followed by positive samples) associated with an appropriate exposure. However, if genome sequences from human *C. difficile* infection isolates are nested within the genetic diversity found in an animal reservoir this supports transmission from animals to humans. A limited example of this was recently seen in a study of clinical and porcine isolates from Ireland [15].

The logistical challenges in preventing acquisition with these multiple potential sources of infection underline the importance of antimicrobial stewardship as an intervention that may prevent both acquisition and transition from colonization to infection. Combined analysis of antimicrobial usage data and antimicrobial resistance determinants in sequencing data from Oxfordshire, UK, suggest that restrictions in fluoroquinolone prescribing were responsible for the successful control of *C. difficile* in England over the last decade [16]. As a result of these measures, the reduction in the prevalence of fluoroquinolone-resistant *C. difficile* in England may mean that the risk of *C. difficile* infection following fluoroquinolone exposure is now not as high as it has been historically (although selection pressure from increased fluoroquinolone use could still potentially reverse this).

Sequencing studies of *Staphylococcus aureus* have also yielded unexpected results. In common with *C. difficile* and contingent on infection prevention and control practice, sequencing suggests that the contribution of direct healthcare-associated transmission may be smaller than previously thought and that multiple introductions of *S. aureus* into hospitals may be more important than has been realized. In a study comprehensively sampling patients, healthcare workers, and the environment in an intensive care unit in Brighton, UK, over 14 months, colonization of all three was common [17]. However, more than 600 genetically distinct subtypes were recovered, and only 25 out of 92 acquisition events in patients could be attributed to other sampled patients (16 instances), healthcare workers (seven instances), or the environment (two instances).

This study and the *C. difficile* studies above highlight a limitation of pathogen sequencing in this context; it may pose more questions than it answers. In both cases there was marked genetic diversity in the bacterial isolates obtained from a single geographic area over a relatively short time-period. This suggests that the sequenced cases are unlikely to be responsible for most transmission, but still leaves the question of what is responsible? Several explanations are possible. First, it may be that we have not sampled comprehensively enough to recover all the bacterial lineages present in the known infected sources. However, at least in the case of *C. difficile*, such mixed infections do not appear to explain transmission when a sweep of all bacterial growth is sequenced from potential sources and compared to closely epidemiologically linked cases not related on standard single isolate sequencing [18]. Another explanation is recent exposure to unsampled sources in hospital, e.g. other patients, healthcare workers, visitors, or the environment. Exhaustive contemporaneous sampling of all these potential sources is challenging or may be impossible, especially when colonization may be transient such that frequent sampling is needed. The Brighton *S. aureus* study is close to what is feasibly achievable even with highly motivated

researchers and clinical staff. A third possible explanation is that patients may be colonized at admission, and this is either not detected due to the absence of admission screening, or not detected as the admission screening is imperfectly sensitive due to the organism being present at a low level, which may subsequently be amplified, e.g. by antimicrobial exposure disrupting competing flora.

Sequencing reveals epidemiological heterogeneity within pathogens

Returning to the example of *C. difficile*, sequencing has highlighted that the extent of healthcare-associated transmission and environmental persistence may vary within a species. For example, higher proportions of ribotype 027 cases are closely genetically related to previous cases than many other ribotypes [3]. Applying Bayesian statistical approaches to sequencing and hospital data from Oxfordshire demonstrate that ribotypes 027, 001, and 106 transmit more readily between patients on the same hospital ward, and also persist for longer in the ward environment following discharge or recovery of infected cases [19]. Notably this study also showed that, by 2010, transmission of *C. difficile* from known cases had been largely stopped, with most apparently healthcare-associated *C. difficile* acquired from other sources. In a pan-European survey, healthcare-associated ribotypes such as 027 and 001 were found to cluster genomically by country and region consistent with local transmission, whereas many other ribotypes, including 078, showed no geographic structure, consistent with transmission via widely disseminated sources, such as food.

These findings have led some clinicians to implement different infection control approaches for different *C. difficile* lineages. In a Swiss hospital with robust standard precautions and predominantly one- and two-bed hospital rooms, only patients with ribotypes 027 or 078 or faecal incontinence were subject to contact precautions and all other patients with *C. difficile* infection underwent standard precautions with a dedicated toilet. During 10 years, 451 contacts were exposed to 279 index patients in two- to four-bed rooms, only six (1.3%) contacts had *C. difficile* detected with the same ribotype, and, of these six case—contact pairs, four pairs had isolates sequenced and only two found to be closely genetically related [20]. Therefore, stratification of infection control by transmission risk appears safe as implemented in this setting and has facilitated fewer barriers to patient care and conserved resources. However, this strategy has not been widely reported elsewhere.

Sequencing supports identification of specific sources of infection

Sequencing can support identifying specific sources of infection. Some of the clearest examples are for food-borne infection, e.g. *E. coli* O104:H4 and *Salmonella* outbreaks across Europe [21—23]. In a healthcare context, a country-wide outbreak of nine listeriosis cases occurred in England in 2017 associated with hospital-provided prepared sandwiches [24]. National prospective whole-genome sequencing allowed the closely related cases of a not previously seen strain to be identified, triggering epidemiological investigations and subsequent identification of the food source of the outbreak, with food isolates confirmed to be part of the same genomic cluster.

Candida auris is an emerging multidrug-resistant fungus that has caused large hospital outbreaks, particularly in high-acuity settings. Between 2015 and 2017, 70 cases of colonization or infection occurred in Oxford, UK, associated with a neurosciences intensive care unit. Epidemiological investigations revealed that *C. auris* infection or colonization was associated with use of reusable axillary temperature probes. Sequenced isolates from patients and the temperature probes formed part of the same genomic cluster. The outbreak was only successfully controlled when the probes were withdrawn despite a bundle of other infection control interventions [25]. The outbreak underlines the dynamic nature of infection prevention and control, where precautions that were previously sufficient may not adequately control a new threat. Although reusable equipment is a well-recognized potential route of transmission, it serves as a reminder that specific decontamination products and methods may be needed for different pathogens.

Mycobacterium chimera infections associated with cardiopulmonary bypass heater—cooler units are another example where sequencing has helped to confirm epidemiological findings [26]. Isolates from cardiac surgery-related infections, a specific manufacturer's heater—cooler unit and its production facility all formed a distinct genetic clade, supporting the implicated heater—cooler unit as the source of the outbreak and that contamination likely occurred at the production site.

Sequencing and the role of individuals in transmission

Sequencing can also point to specific individuals as sources of infection. This has potential personal, ethical, and legal implications [27,28]. One early example relates to potential transmission of cholera from Nepalese soldiers serving as United Nations peacekeepers following an earthquake in Haiti in 2010. Sequencing of isolates from Nepal and a global collection revealed a cluster of isolates from Nepal that were highly genetically related to those from Haiti [29,30].

Another investigation receiving public attention was a meticillin-resistant *S. aureus* (MRSA) outbreak in Cambridge, UK, associated with a special care baby unit [31]. The study was one of the first using rapid benchtop sequencing as an infection control tool, along with other similar studies [32]. A cluster of 26 related cases of MRSA carriage were identified, including spanning a 64-day period following a deep clean during which no admitted patients were colonized. A healthcare worker was shown to be colonized during the intervening period, and detailed sequencing of multiple MRSA colonies from the healthcare worker revealed that their colonization was the likely source of the reintroduction of MRSA back into the unit.

Sequencing also yields historical insights

Sequencing can be used to reconstruct the past history, or phylogenetic ancestry, of a group of pathogens. This allows sequencing of recently obtained samples to yield insights into much earlier events. When combined with geographic or host species data, sampling times and rates of evolution, this can be used to reconstruct when specific lineages emerged and how they have spread between places or species. For example, this approach has been used to reconstruct the emergence of

fluoroquinolone resistance twice in *C. difficile* ribotype 027 and its subsequent spread from North America to Europe [33]. Recently, similar approaches have been used to show that MRSA appeared in the pre-antibiotic era in European hedgehogs, with β-lactams produced by the hedgehog dermatophyte *Trichophyton erinacei* providing a selective environment for resistance to emerge [34].

Sequencing as a diagnostic tool

Whole-genome sequencing can also be used as a diagnostic tool. It has replaced culture as the first-line antimicrobial susceptibility test in England for *Mycobacterium tubuculosis* [35,36]. Resistance prediction for other pathogens, e.g. Enterobacterales or *Neisseria gonorrhoeae*, is possible, but error rates are not yet consistently low enough to meet regulatory standards across commonly used antibiotics [37,38]. Sequencing also has an increasing role in reference laboratories for confirming resistance mechanisms, e.g. as in the most resistant case of *N. gonorrhoeae* infection described to date [39]. Sequencing may also be useful to identify virulence mechanisms; genome-wide association studies can be used to search for genetic correlates of virulent phenotypes — for example, in *S. aureus*, Panton–Valentine leucocidin has been shown to be a key determinant of pyomyositis using this approach [40].

Clinical metagenomic sequencing can be used to identify the causative organism in an infection directly from a clinical sample without the need for culture. As such it potentially provides a rapid, culture-independent diagnostic and with some methods it may also identify any antimicrobial resistance determinates present. However, it remains largely at the proof-of-concept stage with sensitivity versus culture in common sample types (blood, cerebrospinal fluid, orthopaedic infections) ranging from 75% to 90% and specificity between 67% and 96% [41]. However, it may detect additional plausible pathogens, both where prior antibiotic exposure has made cultures negative or fastidious organisms including anaerobes. Clinical metagenomics may also be useful where routine diagnostic workflows fail to reach a diagnosis, e.g. in central nervous system infection [42].

SARS-CoV-2 sequencing and hospital infection control

The COVID-19 pandemic has seen pathogen sequencing conducted on an industrial scale, e.g. through the UK's COVID-19 Genomics Consortium. Sequencing-defined entities such as the alpha, delta and omicron variants have become part of routine public language. The COVID-19 pandemic has also necessitated protection of healthcare workers being a major focus for infection prevention and control teams to a much greater extent than previously, with healthcare workers at increased risk of infection [43]. Sequencing and epidemiological studies have identified healthcare workers as sources for healthcare-associated transmission, but with most patient infections attributable to transmission from other patients, and patients with hospital-onset infection in particular [44,45]. There is also variation in the extent of onward transmission, with relatively few highly infectious individuals being the source for many infections [45,46], but also instances where apparent ongoing outbreaks are the result of multiple introductions of SARS-CoV-2 into a hospital. In addition to detecting new variants associated with increased transmissibility, virulence, or immune escape, sequencing may also be used in future for surveillance for resistance to SARS-CoV-2 therapeutics and for targeting these treatments for individual patients.

Future directions

More accessible sequencing and democratization of access

To date, high-quality sequencing studies have required specialist laboratory expertise and relatively complex bioinformatic workflows. In addition, interpreting sequencing results requires appropriate context including the reproducibility of sequencing and its intrinsic error rates, and the distribution and pattern of genetic differences associated with recent transmission. In some cases, this can be identified directly, e.g. in relatively small outbreaks with clearly defined transmission events, but in many cases with endemic or widespread epidemic disease there are multiple plausible sources for each infection. In these settings genetic distances associated with transmission must be inferred from the extent of within-host diversity and rates of evolution, alongside an understanding of the background genetic diversity within the wider community [1,47]. These metrics vary across different pathogens.

Several developments promise to make sequencing more accessible and available as a tool to a much wider group of users. First, the knowledge base to interpret genetic distances is increasingly mature for the major pathogens. How to define it is also well understood for an emerging novel pathogen, albeit requiring the necessary data, samples, and analysis. Laboratory sequencing workflows are increasingly routine, and improved capacity in molecular diagnostics as a result of the COVID pandemic is likely to increase access to sequencing in hospital laboratories. Processing the resulting data will become simpler via availability of sequence data processing services from commercial, academic and public health providers. Ideally, these services will process data in automatic workflows, to predefined and regulated standards, and generate standardized and exchangeable outputs and reports.

For several pathogens, hundreds of thousands or even millions of sequenced genomes now exist. This raises major challenges when it comes to comparing each new genome with what is already sequenced. Strategies for rapidly comparing genomes and identifying closely related genomes are needed and are in development, to replace existing tools [48,49]. Once the closest 'neighbours' of a new infection are identified, existing methods can be used to reconstruct relationships with other closely related infections and likely transmission events identified. For such a system to work well, sharing of sufficient data across institutions, regions and countries will be required, in a way that also respects data sovereignty.

Smarter sampling and refined insights from sequencing

Whereas the *C. difficile* and *S. aureus* sequencing studies described above were able to show that sampled patients are not the source for many infections, quantifying the extent of transmission from other sources will require carefully designed studies that undertake longitudinal sampling of humans, hospital and community environments, and likely animals as well.

There is also a need to better understand the directionality of transmission to generate actionable information on sources of transmission. Sequencing can identify closely related or indistinguishable infections, but it may not be clear who infected whom. This is partly a limitation of the relative rates of transmission and evolution. Often multiple transmission events can occur between each observed mutation event, resulting in several individuals with genetically indistinguishable infections. Addressing this, linkage to epidemiological data — e.g. sampling times, contact events, or contact networks — may allow joint reconstruction of transmission chains, ideally within a probabilistic framework so the degree of certainty about who infected whom can be captured too. Improvements in sequencing technology may also help, as current 'whole-genome' sequencing may only reconstruct 80—95% of the genome due to divergence of samples from reference genomes and the inability of short-read sequencing platforms to resolve repetitive regions of the genome. Another approach, possible with current technology but more resource intensive, is to sequence multiple bacterial colonies from each infected or colonized individual. Where sufficient within-host diversity exists, this allows higher-resolution reconstruction of transmission [50].

Even with better sampling and these approaches, it may not be possible to exhaustively sample and sequence all possible sources of infection. Here ecological approaches that model rates of transmission between particular host types (e.g. infected patients, healthcare workers, domestic pets, etc.), reservoirs, or niches based on a representative sample of all possible infections/colonizations may be needed. There are also further challenges involved in developing methods to model the transmission of Gram-negative pathogens where transmission of a specific gene, mobile genetic element, or plasmid between host bacteria adds additional complexity [51].

Sequencing as a real-time tool for infection prevention and control

In addition to the epidemiological insights discussed, sequencing has been proposed as a real-time tool for infection control. This is possible where the necessary genomic context is well understood, such that the species-specific genomic distances between sequences that are compatible with transmission have been robustly defined, as discussed above.

Potentially real-time sequencing has advantages: transmission events and pathways supported by sequencing can be targeted for infection prevention and control efforts, and time is saved by not focusing on instances where transmission is excluded based on sequencing.

However, evidence is limited that implementation of sequencing improves outcomes, e.g. reduces healthcare-associated infection, and is cost-effective [52,53]. In part this is because the range of possible interventions triggered by sequencing — that would not otherwise be implemented as part of routine infection prevention and control efforts following identification of a case — is not well defined. Randomized trials to assess the impact of sequencing should be considered, which could include a pre-determined suite of additional measures that might follow a sequencing-confirmed transmission.

Sequencing for benchmarking and surveillance

Whereas the incidence of healthcare-associated infections can be monitored, sequencing can be used to also assess the proportion of infections that were likely acquired in hospital. Proof of concept for this has been shown for *C. difficile* where routine sequencing of all cases during a year at six English hospitals showed differing incidence and rates of transmission [54].

Sequencing also has a role in population-level surveillance where it may be used to identify emerging lineages, e.g. with enhanced virulence or antimicrobial resistance. Prospective sequencing may also help to detect clusters of infection more rapidly than traditional outbreak detection algorithms at an institutional level, as exclusion of transmission by sequencing can reduce background noise.

Ensuring consent and understanding of sequencing

Most pathogen sequencing is performed without explicit consent. This may be because it is done retrospectively, on an opt-out basis, as part of service planning and delivery or epidemiological research. In this context, the findings from sequencing are unlikely to relate back to a specific individual, although care is needed to prevent inadvertent disclosure of identities if de-identified data are made public.

However, where sequencing is performed to reconstruct individual transmission events, then it may be possible that individual patients, healthcare workers, or members of the public become aware or suspect that they are a source for someone else's infection. This may have implications for their wellbeing and for healthcare workers may also have occupational health implications. Similarly, it may also be possible that individuals become aware of who may have infected them.

Ongoing ethical research and open patient, public and healthcare worker engagement is needed to ensure that the benefits of sequencing remain well supported and that people are protected from avoidable harms. Training of healthcare professionals to understand, interpret and communicate sequencing results will also be needed.

Conclusion

Pathogen whole-genome sequencing has provided a remarkable new lens through which to understand the epidemiology of healthcare-associated infection. Insights gained have improved our understanding and ability to better control and prevent these infections. Whether real-time pathogen sequencing becomes routine for all healthcare-associated infections depends on better demonstrating its benefits, and this will likely become clearer over the next few years.

Acknowledgements

I thank Professors T. Peto and K. Jeffery for their helpful comments on this article.

HIS Early Career Award 2021

The Council of the Healthcare Infection Society (HIS) announced Dr D. Eyre as the recipient of the HIS Early Career Award 2021 in recognition of his outstanding innovative contributions to the prevention and control of healthcare-associated infections as an early career clinician. The winner of the HIS Early Career Award receives a medal and a prize of £1,000, and an invitation to submit an article to the JHI for peer review and, if appropriate, publication.

Conflict of interest statement
None declared.

Funding sources
DWE is a Robertson Foundation Fellow.

References

[1] Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med 2013;369:1195—205.
[2] Kumar N, Miyajima F, He M, Roberts P, Swale A, Ellison L, et al. Genome-based infection tracking reveals dynamics of *Clostridium difficile* transmission and disease recurrence. Clin Infect Dis 2016;62:746—52.
[3] Martin JSH, Eyre DW, Fawley WN, Griffiths D, Davies K, Mawer DPC, et al. Patient and strain characteristics associated with *Clostridium difficile* transmission and adverse outcomes. Clin Infect Dis 2018;67:1379—87.
[4] Kong LY, Eyre DW, Corbeil J, Raymond F, Walker AS, Wilcox MH, et al. *Clostridium difficile*: investigating transmission patterns between infected and colonized patients using whole genome sequencing. Clin Infect Dis 2018;68:204—9.
[5] Shim JK, Johnson S, Samore MH, Bliss DZ, Gerding DN. Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea. Lancet 1998;351:633—6.
[6] Zacharioudakis IM, Zervou FN, Pliakos EE, Ziakas PD, Mylonakis E. Colonization with toxinogenic *C. difficile* upon hospital admission, and risk of infection: a systematic review and meta-analysis. Am J Gastroenterol 2015;110:381—90.
[7] Eyre DW, Griffiths D, Vaughan A, Golubchik T, Acharya M, O'Connor L, et al. Asymptomatic *Clostridium difficile* colonisation and onward transmission. PloS One 2013;8:e78445.
[8] Curry SR, Muto CA, Schlackman JL, Pasculle AW, Shutt KA, Marsh JW, et al. Use of multilocus variable number of tandem repeats analysis genotyping to determine the role of asymptomatic carriers in *Clostridium difficile* transmission. Clin Infect Dis 2013;57:1094—102.
[9] Halstead FD, Ravi A, Thomson N, Nuur M, Hughes K, Brailey M, et al. Whole genome sequencing of toxigenic *Clostridium difficile* in asymptomatic carriers: insights into possible role in transmission. J Hosp Infect 2019;102:125—34.
[10] Sheth PM, Douchant K, Uyanwune Y, Larocque M, Anantharajah A, Borgundvaag E, et al. Evidence of transmission of *Clostridium difficile* in asymptomatic patients following admission screening in a tertiary care hospital. PloS One 2019;14:e0207138.
[11] Longtin Y, Paquet-Bolduc B, Gilca R, Garenc C, Fortin E, Longtin J, et al. Effect of detecting and isolating *Clostridium difficile* carriers at hospital admission on the incidence of *C. difficile* infections: a quasi-experimental controlled study. JAMA Intern Med 2016;176:796.
[12] Mawer D, Eyre DW, Griffiths D, Fawley WN, Martin JSH, Quan TP, et al. Contribution to *Clostridium difficile* transmission of symptomatic patients with toxigenic strains who are fecal toxin negative. Clin Infect Dis 2017;64:1163—70.
[13] Lim SC, Knight DR, Riley TV. *Clostridium difficile* and one health. Clin Microbiol Infect 2019;26:857—63.
[14] Knetsch CW, Connor TR, Mutreja A, van Dorp SM, Sanders IM, Browne HP, et al. Whole genome sequencing reveals potential spread of *Clostridium difficile* between humans and farm animals in the Netherlands, 2002 to 2011. Euro Surveill 2014;19:20954.
[15] Moloney G, Eyre DW, Aogáin MM, McElroy MC, Vaughan A, Peto TEA, et al. Human and porcine transmission of *Clostridioides difficile* ribotype 078, Europe. Emerg Infect Dis 2021;27:2294—300.
[16] Dingle KE, Didelot X, Quan TP, Eyre DW, Stoesser N, Golubchik T, et al. Effects of control interventions on *Clostridium difficile* infection in England: an observational study. Lancet Infect Dis 2017;17:411—21.
[17] Price JR, Cole K, Bexley A, Kostiou V, Eyre DW, Golubchik T, 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;17:207—14.
[18] Eyre DW, Cule ML, Griffiths D, Crook DW, Peto TE, Walker AS, et al. Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in *Clostridium difficile* transmission. PloS Comput Biol 2013;9:e1003059.
[19] Eyre DW, Laager M, Walker AS, Cooper BS, Wilson DJ. CDC Modeling Infectious Diseases in Healthcare Program (MInD-Healthcare). Probabilistic transmission models incorporating sequencing data for healthcare-associated *Clostridioides difficile* outperform heuristic rules and identify strain-specific differences in transmission. PloS Comput Biol 2021;17:e1008417.
[20] Widmer AF, Frei R, Erb S, Stranden A, Kuijper EJ, Knetsch CW, et al. Transmissibility of *Clostridium difficile* without contact isolation: results from a prospective observational study with 451 patients. Clin Infect Dis 2016;64:ciw758.
[21] Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, Fitzgerald M, et al. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci USA 2012;109:3065—70.
[22] Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PloS One 2011;6:e22751.
[23] Dallman T, Inns T, Jombart T, Ashton P, Loman N, Chatt C, et al. Phylogenetic structure of European *Salmonella enteritidis* outbreak correlates with national and international egg distribution network. Microb Genom 2016;2:e000070.
[24] Public Health England. Investigation into an outbreak of *Listeria monocytogenes* infections associated with hospital-provided pre-prepared sandwiches, UK May to July 2019. 2020. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/937907/2019-05-Listeria-CC8-Outbreak-Report.pdf last accessed January 2022.
[25] Eyre DW, Sheppard AE, Madder H, Moir I, Moroney R, Quan TP, et al. A *Candida auris* outbreak and its control in an intensive care setting. N Engl J Med 2018;379:1322—31.
[26] Ingen JV, Kohl TA, Kranzer K, Hasse B, Keller PM, Katarzyna Szafrańska A, et al. Global outbreak of severe *Mycobacterium chimaera* disease after cardiac surgery: a molecular epidemiological study. Lancet Infect Dis 2017;17:1033—41.
[27] Johnson SB, Parker M. The ethics of sequencing infectious disease pathogens for clinical and public health. Nat Rev Genet 2019;20:313—5.
[28] Johnson S, Parker M. Ethical challenges in pathogen sequencing: a systematic scoping review. Wellcome Open Res 2020;5:119.
[29] Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. Mbio 2011;2. e00157-11.
[30] Eppinger M, Pearson T, Koenig SSK, Pearson O, Hicks N, Agrawal S, et al. Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. Mbio 2014;5. e01721-14.
[31] Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. Lancet Infect Dis 2013;13:130—6.
[32] Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of

Staphylococcus aureus and *Clostridium difficile* for outbreak detection and surveillance. BMJ Open 2013;2:e001124.
[33] He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. Nat Genet 2013;45:109—13.
[34] Larsen J, Raisen CL, Ba X, Sadgrove NJ, Padilla-González GF, Simmonds MSJ, et al. Emergence of methicillin resistance predates the clinical use of antibiotics. Nature 2022;602(15):135—41.
[35] CRyPTIC Consortium and the 100,000 Genomes Project, Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, et al. DNA sequencing predicts 1st-line tuberculosis drug susceptibility profiles. N Engl J Med 2018;379:1403—15. https://doi.org/10.1056/NEJMoa1800474.
[36] Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 2015;15:1193—202.
[37] Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, et al. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. J Antimicrob Chemother 2013;68:2234—44.
[38] Eyre DW, Silva DD, Cole K, Peters J, Cole MJ, Grad YH, et al. WGS to predict antibiotic MICs for *Neisseria gonorrhoeae*. J Antimicrob Chemother 2017;72:1937—47.
[39] Eyre DW, Sanderson ND, Lord E, Regisford-Reimmer N, Chau K, Barker L, et al. Gonorrhoea treatment failure caused by a *Neisseria gonorrhoeae* strain with combined ceftriaxone and high-level azithromycin resistance, England, 2018 February. Eurosurveillance 2018;23:364.
[40] Young BC, Earle SG, Soeng S, Sar P, Kumar V, Hor S, et al. Panton—Valentine leucocidin is the key determinant of *Staphylococcus aureus* pyomyositis in a bacterial GWAS. Elife 2019;8:e42486.
[41] Govender KN, Street TL, Sanderson ND, Eyre DW. Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis of diagnostic test accuracy studies. J Clin Microbiol 2021;59:e02916—20.
[42] Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med 2019;380:2327—40.
[43] Eyre DW, Lumley SF, O'Donnell D, Campbell M, Sims E, Lawson E, et al. Differential occupational risks to healthcare workers from SARS-CoV-2 observed during a prospective observational study. Elife 2020;9:e60675.
[44] Mo Y, Eyre DW, Lumley SF, Walker TM, Shaw RH, O'Donnell D, et al. Transmission of community- and hospital-acquired SARS-CoV-2 in hospital settings in the UK: a cohort study. PloS Med 2021;18:e1003816.
[45] Lumley SF, Constantinides B, Sanderson N, Rodger G, Street TL, Swann J, et al. Epidemiological data and genome sequencing reveals that nosocomial transmission of SARS-CoV-2 is underestimated and mostly mediated by a small number of highly infectious individuals. J Infect 2021;83:473—82.
[46] Illingworth CJ, Hamilton WL, Warne B, Routledge M, Popay A, Jackson C, et al. Superspreaders drive the largest outbreaks of hospital onset COVID-19 infections. Elife 2021;10:e67308.
[47] Silva DD, Peters J, Cole K, Cole MJ, Cresswell F, Dean G, et al. Whole-genome sequencing to determine transmission of *Neisseria gonorrhoeae*: an observational study. Lancet Infect Dis 2016;16:1295—303.
[48] Mazariegos-Canellas O, Do T, Peto T, Eyre DW, Underwood A, Crook D, et al. BugMat and FindNeighbour: command line and server applications for investigating bacterial relatedness. BMC Bioinformatics 2017;18:477.
[49] Eyre DW, Peto TEA, Crook DW, Walker AS, Wilcox MH. Hash-based core genome multilocus sequence typing for *Clostridium difficile*. J Clin Microbiol 2019;58. e01037-19.
[50] Hall MD, Holden MT, Srisomang P, Mahavanakul W, Wuthiekanun V, Limmathurotsakul D, et al. Improved characterisation of MRSA transmission using within-host bacterial sequence diversity. Elife 2019;8:e46402.
[51] Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, et al. Nested Russian Doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blaKPC. Antimicrob Agents Chemother 2016;60:3767—78. https://doi.org/10.1128/AAC.00464-16.
[52] Stirrup O, Hughes J, Parker M, Sebra R, Kasarskis A, Anson LW, et al. Rapid feedback on hospital onset SARS-CoV-2 infections combining epidemiological and sequencing data. Elife 2021;10:e65828.
[53] Dymond A, Davies H, Mealing S, Pollit V, Coll F, Brown NM, et al. Genomic surveillance of methicillin-resistant *Staphylococcus aureus*: a mathematical early modeling study of cost-effectiveness. Clin Infect Dis 2020;70:1613—19.
[54] Eyre DW, Fawley WN, Rajgopal A, Settle C, Mortimer K, Goldenberg SD, et al. Comparison of control of *Clostridium difficile* infection in six English hospitals using whole-genome sequencing. Clin Infect Dis 2017;65:433—41.