MICROBIAL GENOMICS

PERSONAL VIEW

Werner *et al.*, *Microbial Genomics* 2023;9:001008

DOI 10.1099/mgen.0.001008

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Taking hospital pathogen surveillance to the next level

Guido Werner[1,*], Natacha Couto[2,3], Edward J. Feil[2], Angela Novais[4,5], Kristin Hegstad[6,7], Benjamin P. Howden[8], Alexander W. Friedrich[9] and Sandra Reuter[10]

Abstract

High-throughput bacterial genomic sequencing and subsequent analyses can produce large volumes of high-quality data rapidly. Advances in sequencing technology, with commensurate developments in bioinformatics, have increased the speed and efficiency with which it is possible to apply genomics to outbreak analysis and broader public health surveillance. This approach has been focused on targeted pathogenic taxa, such as *Mycobacteria*, and diseases corresponding to different modes of transmission, including food-and-water-borne diseases (FWDs) and sexually transmitted infections (STIs). In addition, major healthcare-associated pathogens such as methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant enterococci and carbapenemase-producing *Klebsiella pneumoniae* are the focus of research projects and initiatives to understand transmission dynamics and temporal trends on both local and global scales. Here, we discuss current and future public health priorities relating to genome-based surveillance of major healthcare-associated pathogens. We highlight the specific challenges for the surveillance of healthcare-associated infections (HAIs), and how recent technical advances might be deployed most effectively to mitigate the increasing public health burden they cause.

INTRODUCTION

Disease notifications are collected over a broad range of geographical scales, from local, regional, national and continental up to the global level, and consist of three classes – syndromic surveillance, clinical surveillance and laboratory-confirmed cases. Not all of these three pillars are required in all instances; some diseases require a diagnostic, laboratory-driven notification, whilst others require both clinical and diagnostic confirmation. For novel diseases, such as coronavirus disease 2019 (COVID-19), which emerged in late 2019, only syndromic surveillance is possible in the initial stages of an outbreak, an epidemic or a pandemic due to a lack of knowledge concerning the disease and/or pathogen. Other infections are non-notifiable for various reasons, providing an even leaner portfolio of available data for public health. Outbreak investigation has followed the same basic activities for decades: a patient presenting with a suspected infectious disease will elicit a battery of diagnostic tests combined with epidemiological investigations to identify likely sources of the disease and potential cases of onward transmission. Strain typing to elucidate putative outbreak scenarios and transmission routes has traditionally been a specialty that is not necessarily linked to the other activities and medical disciplines and is essentially performed by specialists in a small number of well-equipped and dedicated expert and reference laboratories.

The results of strain typing should confirm (or disprove) epidemiological hypotheses deduced from clinical and patient data, or alert clinicians to potential outbreaks before these are epidemiologically recognized. Strain typing comprises many different techniques and approaches, complemented by an increasing knowledge of bacterial populations as well as technical and analytical

Received 01 November 2022; Accepted 03 March 2023; Published 26 April 2023

Author affiliations: [1]Department of Infectious Diseases, Robert Koch Institute, Wernigerode Branch, Germany; [2]The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, UK; [3]Centre for Genomic Pathogen Surveillance, Big Data Institute, University of Oxford, Oxford OX3 7LF, UK; [4]UCIBIO-Applied Molecular Biosciences Unit, REQUIMTE, Laboratory of Microbiology, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [5]Associate Laboratory i4HB-Institute for Health and Bioeconomy, Faculty of Pharmacy, University of Porto, 4050-313 Porto, Portugal; [6]Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway; [7]Research Group for Host–Microbe Interactions, Department of Medical Biology, UiT the Arctic University of Norway, Tromsø, Norway; [8]Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia; [9]Medical Faculty, University Hospital, Westphalia-Wilhelms-University Muenster, Germany; [10]Institute for Infection Prevention and Hospital Hygiene, University Hospital Freiburg, Germany.
***Correspondence:** Guido Werner, wernerg@rki.de

Keywords: antimicrobial resistance; VRE; MRSA; carbapenem resistance; nosocomial pathogen.

Abbreviations: AMR, antimicrobial resistance; cgMLST, coregenome MLST; FWD, Food- and Water-borne Disease(s); HAI, Healthcare-Associated Infection(s); MLST, multi-locus sequence type/typing; STI, sexually-transmitted disease(s).

Impact Statement

- High-throughput genomic sequencing enables hypothesis-free data evaluation and prospective pathogen surveillance in large and transversal isolate collections.
- Public health-oriented surveillance of healthcare pathogens is to a large extent surveillance of their antimicrobial resistance (AMR) determinants of interest.
- AMR may spread clonally and horizontally and both ways of dissemination should be covered by genomic surveillance efforts.
- AMR determinants are often located within flexible mobile genetic structures and long-read sequencing is superior to next-generation sequencing for assembling them.
- The highly variable nature of how AMR determinants or pathogens spread requires specific and adaptable technical platforms and bioinformatic approaches to follow their dissemination.
- Genomic surveillance of some AMR determinants of public health importance requires a wider, One Health-oriented approach.

advances. Not all methods are appropriate for all pathogens, standardization is challenging or impossible for many techniques, and for gel-based methods typing results are mostly not transferable between laboratories. Certain methods, such as ribotyping, rapid PCR-based typing (RAPD), amplified-fragment length polymorphism (AFLP), multi-locus variable number of tandem repeats analysis (MLVA) and macrorestriction analyses in pulsed-field gel electrophoresis (PFGE) – just to mention the most prominent ones – have replaced or accompanied each other for some years [1–5].

The first change came with the application of multi-locus sequence typing (MLST) as an early sequence-based approach, which generated robust and digital data that were readily transferrable and comparable between laboratories [5]. However, resolution is variable according to the species/pathogen and limited by comparing only seven of the several hundred to thousands of genes. Despite these limitations, MLST typing has retained its utility as a typing tool for defining clonal lineages within species. For the most part, lineages defined as sequence types (STs), clonal groups or clusters of closely related STs have remained robust in the genomics era. This has meant that nomenclature based on ST labels are still valid and useful, and web-based MLST platforms such as Enterobase (https://enterobase.warwick.ac.uk/), PubMLST (https://pubmlst.org/) and BIGSdb-Pasteur (https://bigsdb.pasteur.fr/) have facilitated the global establishment of this nomenclature [6].

The advent of high-throughput sequencing introduced a 'one-size-fits-all' approach, which allows the highest discriminatory power by determining almost the entire genetic information of a pathogen. The introduction of accurate genomic sequencing techniques is associated with decreased turnaround times, lowered costs and multiplexing options, which has allowed outbreak analyses and phylogenetic comparisons to be accessible to a broader scientific and medical community, rather than being limited to a small number of well-equipped research groups or referral centres. This process started about 10 years ago and heralded a new era in outbreak detection and genomic pathogen surveillance for public health purposes, with the latter being in the focus of this paper.

TURNING PATHOGEN SURVEILLANCE UPSIDE DOWN

The introduction of an increasingly fast, highly reliable, accurate, more affordable and technically standardizable, genome-based typing method underpinned a revolution in pathogen surveillance. Genome sequencing and subsequent data analyses are more scalable and allow hypothesis-free data evaluation. It is no longer necessary to limit typing to a small number of isolates selected based on comprehensive and representative clinical, patient, epidemiological and diagnostic data; rather, high-throughput genomic sequencing itself provides the ability to generate hypotheses of transmission pathways and/or source attribution from large, prospective and transversal isolate collections [7].

There are numerous examples in the literature demonstrating the identification of previously unknown sources and protracted outbreaks closer to real time than ever before [8]. This has been especially noticeable in countries such as Denmark and the UK, where substantial investment was made early on to integrate this new technology into a routine public health application [9, 10].

Interdisciplinary working groups at national and continental levels attached to international agencies such as the European Centre for Disease Prevention and Control (ECDC), Centers for Disease Control and Prevention (CDC), United States Food and Drug Administration (FDA) and European Food Safety Authority (EFSA) published concept papers providing roadmaps for the scaling up of genomic pathogen and disease surveillance using high-throughput genomic sequencing technology [11–13].

Early systematic public health initiatives focused on genome-based surveillance for pathogens of food- and water-borne diseases (FWDs) such as *Listeria*, entero-haemorrhagic *Escherichia coli* (EHEC) and *Salmonella*, but also *Mycobacterium tuberculosis* and bacteria causing sexually transmitted infections (STIs) [11, 14–16]. The reasons for selecting these pathogens in particular were manifold, comprehensible and justifiable, but public health priorities vary between different nations and initiatives and not all

national and global action plans (on antimicrobial resistance and/or on genomic pathogen surveillance) were congruent when compared to each other and across the different initiatives.

The transition to whole-genome sequencing (WGS) is not, however, free from challenges, and a bottleneck remains in the technical and analytical skills required to process and evaluate the samples. Studies such as EuSCAPE and the CCRE survey of the European Antimicrobial Resistance Genes Surveillance Network (EURGen-Net: https://www.ecdc.europa.eu/en/about-us/who-we-work/disease-and-laboratory-networks/EURGen-net) and the EU-funded research project COMPARE (https://www.compare-europe.eu/) have successfully endeavoured to establish and build capacity for genomic surveillance in recent years. A similar activity was instigated by the Global Microbial Identifier (GMI) initiative (https://www.globalmicrobialidentifier.org/).

In addition, novel technological approaches and achievements such as metagenomics sequencing for pathogen diagnostic and strain typing, machine learning for antimicrobial resistance (AMR) prediction, and large-scale environmental sampling to extrapolate the influx of AMR and pathogens from 'One-Health' sources were developed and successfully applied in several of these activities [17, 18].

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic further demonstrated the public health value of genomic epidemiology and has triggered international investments in sequencing capacity and digitalization of surveillance infrastructures. Long-standing scientific collaborative networks with respect to training throughout South America and Africa have also come to fruition, as evidenced by the genomic epidemiology of SARS-CoV-2 and the submission of sequencing data from these countries during the pandemic [19].

HOSPITAL PATHOGEN SURVEILLANCE – IT IS MAINLY ABOUT RESISTANCE

For STI and FWD pathogens the elucidation of a clonal spread and/or the identification of a potential source are primary objectives, whereas AMR is analysed as a secondary insight. For healthcare-associated (HAI) pathogens, there is a primary focus on AMR determinants and their potential for transmission, as this has urgent implications for therapeutic interventions and infection control. The ability to distinguish between clonal expansion versus horizontal resistance gene spread is central. Of course, susceptible nosocomial pathogens also spread; however, from the current public health view the spread of vancomycin-susceptible *Enterococcus faecalis* or methicillin-susceptible *Staphylococcus aureus* is of minor therapeutic concern and public health importance. In essence, therefore, molecular surveillance of bacterial hospital pathogens is in fact AMR surveillance.

Resistance in healthcare pathogens spreads in narrower or wider patient and hospital networks [20, 21]. Depending on the pathogen, the extent of clonal versus horizontal spread may differ, as well as transmissibility potential. Multidrug- and methicillin-resistant staphylococci, either *S. aureus* or *S. epidermidis,* rather spread clonally, within a hospital, region, nation, or even globally [22, 23]. On the other hand, vancomycin-resistant enterococci (VRE) spread regionally, with differences at the national and global level and variable dynamics of *vanA* and *vanB* vancomycin resistance determinant transfer within *Enterococcus faecium* [20, 24]. The same holds true for multidrug-resistant Gram-negative pathogens, where ESBL- and carbapenemase-mediated resistance in the nosocomial setting show variable transmission routes – (i) clone-associated, as with KPC-2-producing ST258 *Klebsiella pneumoniae* or CTX-M-15-producing ST131 *E. coli*; (ii) plasmid-mediated, as with IncL carrying $bla_{\text{OXA-48}}$ or IncX4 carrying *mcr*-1; and (iii) IS-mediated, as with IS*CR1,* IS*26* or IS*Ecp1*, which can mobilize a diverse set of resistance genes [25–28].

To investigate clonal spread, strain typing based on WGS data often uses core genome MLST (cgMLST) approaches, allowing a common nomenclature and a standardized data format. In fact, a recent proposal for strain typing nomenclature based on cgMLST and life identification number (LIN) coding promises full stability and continuity of strain taxonomy [29]. However, cgMLST typing also has some limitations, for instance, a limited discriminatory power for all public health demands [30, 31], and is highly dependent on the quality and breadth of the scheme used. More flexible and adaptable methods will be needed to overcome this, such as working with identifying unique sequences (kmer), as has been suggested recently, as well as other promising alternatives, especially for typing multidrug-resistant HAI pathogens [31]. Whole-genome MLST (wgMLST) could be another alternative; however, it is difficult to standardize and lacks a systematic nomenclature, as the number of genes used varies according to the strains included in the analysis.

TECHNIQUES AND APPROACHES FOR GENOMIC AMR SURVEILLANCE – FROM SHORT-READ TO LONG-READ SEQUENCING

AMR determinants are usually plasmid-borne and are commonly flanked by highly repetitive sequences that are difficult to assemble using short-read sequencing data only. Antibiotic resistance plasmids and other mobile genetic elements (MGEs) such as composite transposons and integrative mobilizable or conjugative elements are complex, flexible and mosaic structures that are not easily reconstructed using short-read sequencing data [32]. Reference plasmids deposited in data archives and novel technological approaches to reconstruct MGEs ease data extraction from short-read data [33, 34]. Long-read sequencing can

circumvent this drawback, but at the expense of higher costs and higher computational demands, mainly with Pacific Bioscience's technology (https://www.pacb.com/).

Oxford Nanopore Technology (ONT; https://nanoporetech.com/) represents a huge advance in scalability, affordability and flexibility, offering possibilities from field applications (MinION) and establishment of rapid sequence-based surveillance in low-resource regions, to medium- and high-throughput options such as GridION and PromethION; however, this comes at the cost of lower sequencing accuracy.

Until recently, long-read sequencing applications in genomic AMR surveillance and public health were research-driven, but with the option of multiplexing several samples per run it is now possible to drastically reduce costs per sample and increase throughput, which now makes long-read sequencing attractive and applicable for genomic AMR surveillance and public health applications. For several population-based research applications, huge sets of short-read sequencing data were improved by long-read sequencing [35, 36]. Combining long- and short-read data in this way is the most promising approach at present for routine, public health-oriented, WGS-based hospital pathogen surveillance. New bioinformatic pipelines can utilize the advantages of both data types to generate fully closed, or nearly fully closed, assemblies for both chromosomes and plasmids. This offers exciting new opportunities for outbreak and phylogenetic analyses, AMR prediction and genomic hospital pathogen and AMR surveillance [37, 38].

BEYOND SEQUENCING – THE FUN STARTS AFTER THE SEQUENCING HAS FINISHED

As much as sequencing capacities and capabilities have increased, other sides of the genomics process need to be considered. The data produced should be scrutinized for quality before being used in any downstream analyses. The workflows that can be used to process and evaluate the data are numerous, depending on the user. Commercial solutions such as EPISEQ (https://www.biomerieux-diagnostics.com/biomerieux-episeq-cs) [39] or AREScloud (https://www.opgen.com/ares/) [40], or open-source applications such as the DTU suite of bioinformatics analyses, are available and provide easy access (http://www.genomicepidemiology.org/services/) [41]. Commercial software solutions may have some advantages in terms of a user-friendliness and workflow accessibility for non-specialists. However, they are particularly well suited to specific clinical and research applications and thus do not fulfil every requirement. Moreover, dependence on commercial software solutions comes with a risk, for instance, if the providing companies withdraw future support for their products due to competitive and internal considerations. In this way, efforts in creating an open access streamlined reporting workflows that are especially well suited for clinical and public health epidemiology can significantly boost genomic surveillance. Command line-based tools are often free to use but require a working knowledge of Unix and installation of software. Cloud infrastructure such as MRC-CLIMB (https://www.climb.ac.uk) [42] or Galaxy (https://galaxyproject.org) [43] can help with challenges related to installation and accessibility, and can provide an important entry point for more complex analyses.

In any case, the user reporting the results needs to be able to make informed inferences from the results and to identify artefacts resulting from contamination or software failures. Training of laboratory microbiologists and standardization of tools and pipelines are essential. Large research consortia such as COMPARE and European bodies such as the ECDC and societies such as the European Society of Clinical Microbiology and Infectious Diseases (ESCMID), with their several study groups (e.g. ESCMID Study Group on Epidemiological Markers, ESGEM) have undertaken and will always make great efforts in offering various educational and practical activities regarding high-throughput sequencing applications for a larger user community (https://www.escmid.org/profession_career/educational_activities/; https://www.ecdc.europa.eu/en/training). Future considerations will revolve around quality control and management requirements, external quality assessments and potential International Organization for Standardization (ISO) accreditation, which are essential but currently mainly lacking in the field of genomic sequencing for diagnostic and public health purposes [44].

INTEGRATED DATA ANALYSIS AND SHARING OF DATA – THE CHANGING POSITION OF NATIONAL REFERENCE LABORATORIES

With much more data generated at local and regional levels, the role of national reference laboratories is likely to expand to oversee long-term developments with respect to AMR and bacterial evolution. Rather than sending strains to be typed across the country, dedicated platforms are required that allow the – initially protected – upload of the raw data alongside the relevant metadata such as sampling date, place of isolation and clinical patient details. Active data and strain sharing therefore is essential and must be reinforced according to open science and FAIR principles (findable, accessible, interoperable, reproducible; https://www.openaire.eu/openaire-and-eosc; https://www.go-fair.org/fair-principles/).

The data then need to be processed to be accessible for a wider community to put local developments into context. The SARS-CoV-2 pandemic highlighted the power of this approach for addressing questions such as whether specific variants are locally restricted or form part of a larger transmission network. The answer to this question very much informs intervention strategies to manage their further spread. This will greatly enhance surveillance efforts if carried out in a sustainable manner. Well-advanced examples of

interactive platforms already exist with, for instance, Pathogenwatch [45], BIGSdb [6] and Enterobase [46] focusing mainly on FWD and healthcare pathogens, 'nextstrain' built for viral surveillance [47] and 'GenomeTrakr'/"GalaxyTrakr' [48] targeting food-borne pathogens (https://nextstrain.org/; https://pathogen.watch/; https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-network).

Genome-based surveillance and data sharing for healthcare-associated bacterial pathogens will greatly benefit from SARS-CoV-2 genomic surveillance efforts put into place in many countries and regions around the world, complemented by informative visualization and imaging tools. Lastly, long-term data storage needs to be ensured, and may in part be solved by depositing the raw reads and associated metadata in public repositories to be used for broad population-based analyses.

CONCLUSION – 'DESCRIPTIVE IS GOOD, PREDICTIVE WOULD BE BETTER'

In the last decade, financial, infrastructural and technical efforts have been made to build up genome-based, One-Health-oriented surveillance for FWD pathogens in several European and American countries and on a global scale (https://www.eurgen-reflabcap.eu) [49], whilst the SARS-CoV-2 pandemic provided the impetus for the generation, dissemination and analysis of worldwide genome-based surveillance over a time scale short enough to inform real-time political decision making.

Hospital pathogen surveillance goes far beyond simple outbreak analyses and has a much broader longitudinal dimension, building on patient networks and patient-to-patient transmission chains. Pathogen or AMR dissemination could go on for years if unnoticed or if the implemented infection prevention and control measures are less effective [50–52].

The focus of hospital pathogen surveillance from a public health perspective is to a large extent AMR surveillance, which needs to detect and distinguish both vertical and horizontal, real-time transmission of AMR determinants. This requires a broader adoption of long-read sequencing for more detailed characterization of plasmids and other MGEs and novel bioinformatics solutions to assemble genetic AMR structures of interest from WGS data.

Genes conferring resistance to last-resort antibiotics such as *mcr*-mediated colistin resistance in Enterobacterales and gene-mediated linezolid resistance show a strong link to and interconnection with sectors outside the hospital setting and, as such, their genomic surveillance requires a much wider, One-Health-oriented approach [53, 54].

SARS-CoV-2 genome-based surveillance combined with epidemiological modelling and novel visualization tools have already demonstrated the future of genomic pathogen surveillance. The challenge and benchmark will be to move from retrospective and descriptive studies to real-time and predictive pathogen surveillance that not only confirms epidemiological hypotheses, but acts as an early warning system by identifying possible hidden or unnoticed reservoirs and future, upcoming risks.

Funding information
This work received no specific grant from any funding agency.

Acknowledgements
All authors are members of the ESCMID study group for epidemiological markers, ESGEM. The article has been written on behalf of ESGEM.

Author contributions
Conceptualization: G.W., E.F., N.C. and S.R. Drafted the original manuscript: G.W. and S.R. Extensive revision and editing: all authors. Finalization: G.W. and S.R.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Deplano A, Schuermans A, Van Eldere J, Witte W, Meugnier H, *et al*. Multicenter evaluation of epidemiological typing of methicillin-resistant *Staphylococcus aureus* strains by repetitive-element PCR analysis. The European Study Group on Epidemiological Markers of the ESCMID. *J Clin Microbiol* 2000;38:3527–3533.
2. Willems RJ, Top J, van Den Braak N, van Belkum A, Endtz H, *et al*. Host specificity of vancomycin-resistant *Enterococcus faecium*. *J Infect Dis* 2000;182:816–823.
3. Murchan S, Kaufmann ME, Deplano A, de Ryck R, Struelens M, *et al*. Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J Clin Microbiol* 2003;41:1574–1585.
4. Lindstedt BA. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* 2005;26:2567–2582.
5. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, *et al*. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 1998;95:3140–3145.
6. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018;3:124.
7. Peacock SJ, Parkhill J, Brown NM. Changing the paradigm for hospital outbreak detection by leading with genomic surveillance of nosocomial pathogens. *Microbiology* 2018;164:1213–1219.
8. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, *et al*. Whole-genome sequencing for analysis of an outbreak of

meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis* 2013;13:130–136.

9. Chattaway MA, Dallman TJ, Larkin L, Nair S, McCormick J, *et al*. The transformation of reference microbiology methods and surveillance for *Salmonella* with the use of whole genome sequencing in England and wales. *Front Public Health* 2019;7:317.
10. Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, *et al*. An assessment of different genomic approaches for inferring phylogeny of *Listeria monocytogenes Front Microbiol* 2017;8:2351.
11. ECDC TECHNICAL REPORT. ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations 2019–2021. Stockholm, ECDC. 2019.
12. European Food Safety Authority (EFSA). EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. *EFSA J* 2021;19:e06506.
13. U.S. Food and Drug Administration. Whole Genome Sequencing Program. GenomeTrakr Network; (n.d.). https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-network [accessed 18 January 2023].
14. GLASS whole-genome sequencing for surveillance of antimicrobial resistance. Licence: CC BY-NC-SA 3.0 IGO. Geneva: World Health Organization. 2020.
15. ECDC. Tuberculosis surveillance and monitoring in Europe 2021 –2019 data; Report of the ECDC and WHO. ISBN 978-92-9498-534-7; 2021
16. NIHR Global Health Research Unit on Genomic Surveillance of AMR. Whole-genome sequencing as part of national and international surveillance programmes for antimicrobial resistance: a roadmap. *BMJ Glob Health* 2020;5:e002244.
17. Matamoros S, Hendriksen RS, Pataki BÁ, Pakseresht N, Rossello M, *et al*. Accelerating surveillance and research of antimicrobial resistance - an online repository for sharing of antimicrobial susceptibility data associated with whole-genome sequences. *Microb Genom* 2020;6:e000342.
18. Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, *et al*. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat Commun* 2019;10:1124.
19. Aanensen DM, Carlos CC, Donado-Godoy P, Okeke IN, Ravikumar KL, *et al*. Implementing whole-genome sequencing for ongoing surveillance of antimicrobial resistance: exemplifying insights into *Klebsiella pneumoniae*. *Clin Infect Dis* 2021;73:S255–S257.
20. Gorrie CL, Da Silva AG, Ingle DJ, Higgs C, Seemann T, *et al*. Key parameters for genomics-based real-time detection and tracking of multidrug-resistant bacteria: a systematic analysis. *Lancet Microbe* 2021;2:e575–e583.
21. Sherry NL, Gorrie CL, Kwong JC, Higgs C, Stuart RL, *et al*. Multisite implementation of whole genome sequencing for hospital infection control: a prospective genomic epidemiological analysis. *Lancet Reg Health West Pac* 2022;23:100446.
22. Lee JYH, Monk IR, Gonçalves da Silva A, Seemann T, Chua KYL, *et al*. Global spread of three multidrug-resistant lineages of *Staphylococcus epidermidis*. *Nat Microbiol* 2018;3:1175–1185.
23. Coll F, Raven KE, Knight GM, Blane B, Harrison EM, *et al*. Definition of a genetic relatedness cutoff to exclude recent transmission of meticillin-resistant *Staphylococcus aureus*: a genomic epidemiology analysis. *Lancet Microbe* 2020;1:e328–e335.
24. Arredondo-Alonso S, Top J, McNally A, Puranen S, Pesonen M, *et al*. Plasmids shaped the recent emergence of the major nosocomial pathogen *Enterococcus faecium*. *mBio* 2020;11:e03284-19.
25. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, *et al*. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. *PLoS Genet* 2019;15:e1008114.
26. David S, Cohen V, Reuter S, Sheppard AE, Giani T, *et al*. Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among *Klebsiella pneumoniae*. *Proc Natl Acad Sci U S A* 2020;117:25043–25054.
27. Weber RE, Pietsch M, Frühauf A, Pfeifer Y, Martin M, *et al*. IS26-mediated transfer of blaNDM–1 as the main route of resistance transmission during a polyclonal, multispecies outbreak in a German hospital. *Front Microbiol* 2019;10:2817.
28. León-Sampedro R, DelaFuente J, Díaz-Agero C, Crellen T, Musicha P, *et al*. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. *Nat Microbiol* 2021;6:606–616.
29. Hennart M, Guglielmini J, Bridel S, Maiden MCJ, Jolley KA, *et al*. A dual barcoding approach to bacterial strain nomenclature: genomic taxonomy of *Klebsiella pneumoniae* strains. *Mol Biol Evol* 2022;39:msac135.
30. Gona F, Comandatore F, Battaglia S, Piazza A, Trovato A, *et al*. Comparison of core-genome MLST, coreSNP and PFGE methods for *Klebsiella pneumoniae* cluster analysis. *Microb Genom* 2020;6:e000347.
31. Higgs C, Sherry NL, Seemann T, Horan K, Walpola H, *et al*. Optimising genomic approaches for identifying vancomycin-resistant *Enterococcus faecium* transmission in healthcare settings. *Nat Commun* 2022;13:509.
32. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb Genom* 2017;3:e000128.
33. Redondo-Salvo S, Bartomeus-Peñalver R, Vielva L, Tagg KA, Webb HE, *et al*. COPLA, a taxonomic classifier of plasmids. *BMC Bioinformatics* 2021;22:390.
34. Arredondo-Alonso S, Bootsma M, Hein Y, Rogers MRC, Corander J, *et al*. gplas: a comprehensive tool for plasmid analysis using short-read graphs. *Bioinformatics* 2020;36:3874–3876.
35. Top J, Arredondo-Alonso S, Schürch AC, Puranen S, Pesonen M, *et al*. Genomic rearrangements uncovered by genome-wide co-evolution analysis of a major nosocomial pathogen, *Enterococcus faecium Microb Genom* 2020;6:12.
36. Martin J, Phan HTT, Findlay J, Stoesser N, Pankhurst L, *et al*. Covert dissemination of carbapenemase-producing *Klebsiella pneumoniae* (KPC) in a successfully controlled outbreak: long- and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission. *J Antimicrob Chemother* 2017;72:3025–3034.
37. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 2021;39:1348–1365.
38. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, *et al*. Oxford Nanopore R10.4 long-read sequencing enables near-perfect bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Microbiology*. DOI: 10.1101/2021.10.27.466057
39. Durand G, Javerliat F, Bes M, Veyrieras JB, Guigon G, *et al*. Routine whole-genome sequencing for outbreak investigations of *Staphylococcus aureus* in a national reference center. *Front Microbiol* 2018;9:511.
40. Ferreira I, Beisken S, Lueftinger L, Weinmaier T, Klein M, *et al*. Species identification and antibiotic resistance prediction by analysis of whole-genome sequence data by use of ARESdb: an analysis of isolates from the Unyvero lower respiratory tract infection trial. *J Clin Microbiol* 2020;58:e00273-20.
41. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, *et al*. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;75:3491–3500.
42. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, *et al*. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2016;2:e000086.
43. Community G. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res* 2022;50:W345–51.
44. Sherry NL, Horan KA, Ballard SA, Gonçalves da Silva A, Gorrie CL, *et al*. An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. *Nat Commun* 2023;14:60.

45. Argimón S, David S, Underwood A, Abrudan M, Wheeler NE, *et al*. NIHR global health research unit on genomic surveillance of antimicrobial resistance rapid genomic characterization and global surveillance of *Klebsiella* using pathogenwatch. *Clin Infect Dis* 2021;73(Suppl_4):S325–S335.
46. Achtman M, Zhou Z, Charlesworth J, Baxter L. EnteroBase: hierarchical clustering of 100 000s of bacterial genomes into species/subspecies and populations. *Philos Trans R Soc Lond B Biol Sci* 2022;377:20210240.
47. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, *et al*. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;34:4121–4123.
48. Gangiredla J, Rand H, Benisatto D, Payne J, Strittmatter C, *et al*. GalaxyTrakr: a distributed analysis tool for public health whole genome sequence data accessible to non-bioinformaticians. *BMC Genomics* 2021;22:114.
49. Timme RE, Wolfgang WJ, Balkey M, Venkata SLG, Randolph R, *et al*. Optimizing open data to support one health: best practices to ensure interoperability of genomic data from bacterial pathogens. *One Health Outlook* 2020;2:20.
50. Lisotto P, Couto N, Rosema S, Lokate M, Zhou X, *et al*. Molecular characterisation of vancomycin-resistant *Enterococcus faecium* isolates belonging to the lineage ST117/CT24 causing hospital outbreaks. *Front Microbiol* 2021;12:728356.
51. Stoesser N, Phan HTT, Seale AC, Aiken Z, Thomas S, *et al*. Genomic epidemiology of complex, multispecies, plasmid-borne bla(KPC) carbapenemase in enterobacterales in the United Kingdom from 2009 to 2014. *Antimicrob Agents Chemother* 2020;64:e02244–19.
52. Bender JK, Haller S, Pfeifer Y, Hogardt M, Hunfeld KP, *et al*. Combined clinical, epidemiological, and genome-based analysis identified a nationwide outbreak of *Burkholderia cepacia* complex infections caused by contaminated mouthwash solutions. *Open Forum Infect Dis* 2022;9:ofac114.
53. Schwarz S, Zhang W, Du XD, Krüger H, Feßler AT, *et al*. Mobile oxazolidinone resistance genes in Gram-positive and Gram-negative bacteria. *Clin Microbiol Rev* 2021;34:e0018820.
54. Bastidas-Caldes C, de Waard JH, Salgado MS, Villacís MJ, Coral-Almeida M, *et al*. Worldwide prevalence of *mcr*-mediated colistin-resistance *Escherichia coli* in isolates of clinical samples, healthy humans, and livestock-a systematic review and meta-analysis. *Pathogens* 2022;11:659.