Workshop 4: Innovations in Genomic AMR Surveillance

Nicole E. Wheeler, PhD Biochemistry, Institute of Microbiology and Infection, University of Birmingham, Birmingham, Edgbaston, B15 2TT, UK

Vivien Price*, MSc, Department of Clinical Infection, Immunology and Microbiology, University of Liverpool, Liverpool Centre for Global Health Research, Waterhouse Building Block E, 70 Pembroke Place, Liverpool, L69 3GF, UK

Edward Cunningham-Oakes*, PhD, Department of Infection Biology and Microbiomes, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, L69 7ZB, UK

Kara K. Tsang*, PhD, Department of Infection Biology, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

Jamie G. Nunn, MSc, Infectious Disease Challenge Area, Wellcome Trust, 7 Rojack Road, London, SE23 2DF, UK

Janet T. Midega, PhD, Drug Resistant Infections, Wellcome Trust, 215 Euston Road, London, NW1 2BE, UK

Muna F. Anjum, PhD, Department of Bacteriology, Animal and Plant Health Agency, Woodham Lane, New Haw, Surrey, KT15 3NB, UK

Matthew J. Wade, PhD, Data Analytics and Surveillance Group, UK Health Security Agency, Nobel House, London, SW1P 3JR, UK; School of Engineering, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK

Nicholas A. Feasey, PhD, Clinical Sciences, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK; Malawi Liverpool Wellcome Research Programme, Malawi

Sharon J. Peacock, PhD, Department of Medicine, University of Cambridge, Box 157, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

Elita Jauneikaite, PhD, Department of Infectious Diseases, School of Public Health, Imperial College London, Praed St, London, W2 1NY, UK; NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Department of Infectious Disease, Imperial College London, Hammersmith Hospital, Du Cane Road, London, W12 0NN, UK

Kate S. Baker^, PhD, Centre for Clinical Infection, Microbiology and Immunology, University of Liverpool, Crown Street, L69 7ZB, Liverpool, UK; Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, UK

The SEDRIC Genomics Surveillance Working Group

* These authors contributed equally

^ author for correspondence: Professor Kate Baker, Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, UK; email: kb827@cam.ac.uk; telephone: +44 (0) 7595626210

Word count: 3475 words

Abstract: 149 words

References: 75

Summary

Whole genome sequencing of antimicrobial resistant (AMR) pathogens is increasingly being used for AMR surveillance, particularly in high-income countries. Innovations in genome sequencing and analysis technologies promise to revolutionise AMR surveillance and epidemiology even further. However, routine adoption of these technologies is a challenge, particularly in low- and middle-income countries. As part of a wider series of workshops and online consultations, a group of experts in AMR pathogen genomics and computational tool development conducted a situational analysis, identifying the following under-utilised innovations in genomic AMR surveillance: clinical metagenomics, environmental metagenomics, gene or plasmid tracking, and machine learning. The group recommended developing cost-effective use cases for each approach and mapping data outputs to clinical outcomes of interest to justify additional investment in capacity, training, and staff required to implement these technologies. Harmonisation and standardisation of methods, and the creation of equitable data sharing and governance frameworks, will facilitate successful implementation.

1. Background

In early 2022, the Surveillance and Epidemiology of Drug Resistant Infection Consortium (SEDRIC) held a series of workshops to map the current and future landscape of genomic antimicrobial resistance (AMR) surveillance. While genomics has already begun to complement or replace phenotypic, immunological, and molecular approaches for isolate-based surveillance, recent innovations in genomic technologies promise to improve surveillance and epidemiology of AMR further. These developments will allow for tailoring of AMR interventions in real-time and address some of the barriers that impede widespread implementation. Held on 5th May 2022, the fourth workshop focused on 'Innovations in Genomic AMR Surveillance' and brought together stakeholders to conduct a situational analysis and reach a qualified consensus on the use of several predefined innovations for the surveillance and monitoring of AMR. Innovations were identified on the basis of proven benefits in research and a lack of implementation in routine surveillance. The innovations selected were clinical metagenomics, environmental metagenomics, gene or plasmid tracking, and machine learning. Through discussions on the benefits, barriers, and potential implementation pathways for these approaches, we identified common challenges that need to be addressed to enable the integration of these innovations with genomic AMR surveillance systems.

2. Clinical metagenomics

Metagenomics is the study of genetic material directly from environmental or clinical samples, without the need for isolation or lab cultivation of individual organisms. For the purposes of the workshop, "clinical metagenomics" was used as a broad term encompassing any approach that aims to analyse all genetic material from microorganisms and their hosts in clinical samples. [1] This is achieved through next-generation sequencing (NGS) and aims not only to detect potential pathogens but also to enable an understanding of the host, the microbiome, and host-microbe interactions.

Clinical metagenomics offers multiple advantages for AMR surveillance over traditional laboratory culture and single-isolate sequencing approaches. Removing the need to culture an isolate has the potential to enable much faster diagnosis and the detection of as-yet-uncultured pathogens.[2] These approaches have the potential to generate robust aetiological and phenotypic information in clinically relevant turn-around times (hours rather than days) and could therefore increase the uptake of genomic AMR surveillance on the clinical frontline (Figure). Metagenomic data can also enable the detection of infections caused by multiple strains (polyclonal) and species/pathogens (polymicrobial),[3] and lends itself to the discovery of unexpected or previously unknown pathogens[4] and pathogens not subject to routine surveillance. Additional benefits include the ability to identify and characterize all antibiotic resistance genes in a sample,[5] determine reservoirs for resistance,[6] provide real-time reporting of resistance elements during sequencing[8], and identify at-risk individuals.[8]

Analysing metagenomic data in a clinically meaningful way can be challenging owing to complexities in accurately predicting AMR phenotypes,[9] achieving enough sequencing depth to comprehensively detect resistance determinants,[10] and correctly assigning the origin of metagenomic reads to a microorganism in the sample,[11] which is particularly challenging when a resistance mechanism is carried on a plasmid. Public datasets will be essential for clinical metagenomic analyses as they allow baseline trends and patterns to be established against which samples can be compared. However, there are currently large representation gaps in these sources of data.[12] Consequently, the implications of finding under-represented populations of bacteria present in a sample are hard to determine and require expertise to interpret. Data coordination is also inherently challenging on a global scale, with geopolitical sensitivities presenting challenges to routine data sharing.[13] These challenges mirror the representation and data governance issues identified in previous workshops (see Workshops 1-3 [ref D-23-00144, D-23-00145, D-23-00146]). Additional barriers include technical issues around methodology and reproducibility, such as standardising controls between datasets, the handling and automated processing of large datasets, as well as accessing funding, data storage, and resources for analysis, all of which vary by geographical region. As highlighted for isolate-based genomic AMR surveillance in previous workshops 1-3 (D-23-00144, D-23-00145, D-23-00146), certain barriers to implementation (e.g., training, infrastructure development and maintenance) are likely to be particularly challenging to meet in LMICs.

Improvements in frameworks, databases, and software are needed to address challenges in the biological interpretation of raw metagenomic data. Regional differences in funding and coordination create barriers that can be mitigated by developing models in which routine metagenomic sequencing might begin at regional 'hubs' before being disseminated to local and national 'spoke' laboratories. Such models enable retention of the support and benefits of large regional hubs (e.g., centralised training and individual skillset development), whilst eventually permitting a more tailored approach at the local level, similar to that outlined for isolate-based surveillance [Overview, D-23-00143]. While we should ensure that training provision is guided by geographically-tailored blueprints to ensure that individuals are trained in using the tools and approaches available to them, wherever possible it will also be important to develop standard methods such as the use of mock communities during lab pre-processing and sequencing[14] and the use of well-characterised and robust pipelines for bioinformatic analysis.[15] However, it should be acknowledged that certain aspects of standardisation are not always feasible across all locations (e.g., the type of sequencing technologies available). Additionally, limited standardisation across databases adds further complexity. For example, while the National Database of Antibiotic Resistant Organisms (NDARO) possesses a higher number of acquired resistance genes, the Comprehensive Antibiotic Resistance Database (CARD) provides better coverage of mutation-conferred resistance.[16] As such, in common with other domains covered by the working group, training for competency in varied approaches for genomic analysis and epidemiology will be crucial to success in clinical settings and clinical care of infectious disease cases. This training will also be vital in facilitating infrastructure development to enable both country-level self-sufficiency and international collaboration.

3. Environmental metagenomics

Sequencing of genetic material from environmental samples – known as environmental metagenomics[17] – creates opportunities to better understand the niches, reservoirs, and transmission routes of AMR bacteria. It also holds promise as a tool to monitor the impact of public health measures (see Figure 1). Environmental samples make it possible to survey overall levels of AMR, including carriage among the healthy, non-symptomatic population, and are comparatively accessible, which is particularly valuable in settings where access to healthcare is less equitable. Furthermore, environmental metagenomics is culture-independent, and may function as an early warning of AMR yet to be observed in clinical isolates.

However, sample sources for environmental metagenomics can be highly variable. For AMR surveillance in a One Health context, surveillance of sources such as wastewater, waterways, farms, and air can all provide valuable information on the flow of AMR within and between potential reservoirs (Workshop 3, D-23-00146).[18,19] For example, wastewater sampling has been successfully employed

in the surveillance of a number of infectious diseases and public health threats.[20,21] Experience from the COVID-19 pandemic[20] has indicated that community wastewater surveillance can be an important supplement to hospital case reporting and that, while not a perfect proxy for population incidence, corresponds well with infection trends and incidence data.[22]

Difficulty in resolving gene-pathogen relationships, determining viability, and quantifying abundance (or relative abundance) challenge the interpretation and actionability of environmental metagenomic surveillance data.[6] However, recent genomic approaches offer the promise of overcoming some of these challenges; for example, metagenomic chromosome conformation capture technologies have the potential to allow the attribution of plasmids to their hosts in metagenomic samples through physically linking plasmids to the chromosomes of their hosts.[23]

Epidemiological interpretation of a single environmental sample is fraught with difficultly.[24] For environmental metagenomics to be most effective, long-term funding (a minimum of 10 years) and implementation will be needed to establish baseline rates and diversity of AMR in a given environment (e.g. wastewater), track trends, and identify disruptions in these patterns across harmonised sampling frameworks. Analysis workflows are also required to enable comparison across settings. Understanding the uncertainties and variability in environmental samples,[25] which stem from multiple stochastic and systematic sources, will be crucial to fostering confidence in the utility of the datasets. Establishing sample archives now (e.g., biobanking), which can be analysed with future pipelines, will be crucial to understanding and interpreting environmental metagenomic data over time. Effective data management processes and systems, including data storage, transfer, and sharing, must also be developed. Of note, clarity and consistency on the ethical and regulatory implications regarding the handling of human reads found within environmental samples is vital to enable the sharing of this data.[26]

The use of environmental metagenomics for AMR surveillance has, so far, largely been confined to the research domain,[6] and a lack of political will to invest in advancing this technology into active surveillance and/or monitoring was identified as a barrier to its future use, though this picture is changing. The best ways to apply these approaches as a surveillance tool are not yet well understood, and we have not yet evaluated the additional information provided over traditional clinical surveillance measures versus the time and cost of more comprehensive environmental surveillance. Researchers must generate successful use cases that demonstrate proportionate, cost-effective, and timely actionable insights that correlate with indicators relevant to clinical and/or public health.

Defining a purpose (for example, early warning, monitoring of control measure success), harmonisation of sampling and analysis processes, and validation are all areas where consensus is needed for progress to occur. This will require long-term partnerships between researchers and policymakers to align surveillance efforts with data needed to inform concrete actions that address AMR. Establishing global leadership to maintain a long-term vision and highlight the potential importance of the environment in AMR is a key area for advocacy, which could lead to embedding environmental surveillance within AMR Action Plans and disrupting current research silos.

4. Gene/Plasmid based tracking

Tracking AMR genes or the mobile genetic elements that carry them (rather than, or as a complement to, the pathogen lineages in which they reside) offers several advantages for enhancing AMR surveillance. Targeted sequencing of AMR genes and plasmids can provide clearer insights into the presence,[27,28] emergence,[29,30] and direction[31–33] of transmission of AMR among different ecological compartments (e.g. human and animal, hospital and community), individual hosts, and microorganisms than that provided by metagenomic surveillance or isolate-based sequencing (see Figure). Gene or plasmid-based tracking can aid in assessing the risk of critically important resistance mechanisms moving between compartments, for instance from humans to animals or vice versa.[34] It can also determine the means by which AMR spreads, whether by a small mobile genetic element such as a transposon, a larger mobile genetic element such as a plasmid, or a chromosomally integrated gene, which can be used to anticipate dissemination patterns among bacterial populations.

Plasmid monitoring can function across all One Health sectors; at a local level for outbreak detection[32,35] and at a national and global level for surveillance and larger contextual understanding.[36,37] Tracking plasmid backbones of concern, even when they do not carry AMR genes, is key to understanding plasmid epidemiology and identifying high-risk settings before an AMR gene is acquired. Similarly, tracking the presence of AMR genes in non-pathogenic bacteria will reveal the flow of AMR genes through different reservoirs, including the environment.[38] Consideration must be given to how existing risk assessment frameworks based on single-isolate models can be adapted to mobile cassette or plasmid transmission. Mathematical modelling to establish the minimal sampling required[39] and cost-effectiveness studies would help justify the cost of targeted gene and plasmid monitoring.

Tracking of AMR genes and plasmids currently relies on a range of sequencing technologies and platforms that have different requirements for consumables and offer specific benefits and limitations. While short-read sequencing currently dominates AMR surveillance, long-read sequencing that produces more complete plasmid sequences has the potential to offer more detailed insights into plasmid epidemiology.[40] Lessons can be transferred from the metagenomics community on how to standardise and harmonise different methods.[41] The working group noted that from an implementation perspective, a single composite genomics platform that would deliver complete genomes sufficient for resolving AMR context with a throughput comparable to short-read technologies would be ideal.

Defining plasmid similarity and classification can be challenging, especially for plasmids with highly plastic sequences. Nomenclature standardisation and identification of plasmid characteristics, including their individual structure, would benefit these efforts. Building upon the National Collection of Type Cultures[42] and American Type Culture Collection,[43] institutional and stakeholder endorsement of a physical set of standardised plasmids from clinical, environmental, and animal samples would allow for platform and phenotype benchmarking.

Improving the interpretability of plasmid surveillance data by linking plasmid characteristics with health outcomes in human and animal hosts is a priority for future research. We must better characterise plasmid diversity,[44] transmission rates,[45] permissiveness,[46,47] stability,[46,47] and other phenotypes.[48] Epidemiological data can be used as a starting point and as a supplement to investigating the clinical and public health risk of plasmids, but a better understanding of plasmid biology will allow academic researchers and public health bodies to produce increasingly accurate databases[49,50] and tools[51,52] to better define and describe plasmids and their associated AMR risk.

5. Machine learning

Machine learning (ML) is a subfield of artificial intelligence that focuses on enabling computer systems to learn from and make decisions based on data. The goal of machine learning is to create systems that can autonomously improve their accuracy and performance over time, without being explicitly programmed to do so. ML tools for the analysis of large, diverse, and often complex data streams have improved a great deal over recent years,[53] creating an opportunity to improve genomic AMR surveillance through the integration and analysis of different data streams. Two such use cases were highlighted during this workshop.

The first application is the use of ML to predict AMR using genomic and antimicrobial susceptibility testing (AST) data (see Figure). The advantages of adopting ML approaches include identifying novel resistance mechanisms,[54] predicting AMR from incomplete data, e.g. metagenomics,[55] and modelling the interaction between resistance mechanisms.[54] However, several challenges exist in translating promising preliminary work in this area into public health benefits. Large, internationally representative datasets of high-quality phenotypic AST data are required to develop accurate ML algorithms.[56] Quantitative minimum inhibitory concentration (MIC) data are more easily combined and compared than categorical SIR (sensitive, intermediate, resistant) data owing to the breakpoints between categories and conventions changing over time,[57] but are less commonly made available to the research community. Centralised AST tends to produce more consistent phenotyping results but distributed international capacity for AST coupled with external quality assessment (EQA) would enable more sustainable and equitable data generation.[58,59] Good model training data should have an

even representation of sensitive and resistant isolates, and of different resistance mechanisms, a distribution unlikely to be captured by a standard surveillance sampling framework. In the future, algorithms would ideally be independently benchmarked against a shared set of test isolates that are changed regularly to capture new AMR mechanisms[60] and prevent over-training of algorithms to the test data. A valuable adjunct to standard surveillance sampling frameworks would be to preferentially sequence treatment failures and rare resistance phenotypes to enrich for potentially novel resistance mechanisms.[61]

The second promising application of ML to AMR surveillance is the use of forecasting tools to measure and predict changes in resistance rates over time and space.[62,63] Models can be trained to predict resistance rates based on historical prevalence data and identify unexpected fluctuations in resistance that may reflect a successful intervention or a concerning new trend. However, heterogeneous data obtained at different scales is difficult to integrate.[64] Fine-scale data such as monthly AST data broken down by postcode is ideal,[62] but increased granularity in surveillance poses greater privacy concerns. These concerns could be addressed by providing restricted access to data, similar to GISAID[65] and the United Kingdom Biobank.[66] Another limitation of current models is that there is little or no integration of the mechanisms causing AMR in the modelling process.[67] This could be achieved by integrating other data sources such as gene/plasmid tracking, antimicrobial usage, and/or human and animal movement data.[63]

Several general challenges in the application of ML to AMR surveillance were identified. ML algorithm accuracy can degrade over time following deployment.[68] Existing training datasets are biased toward a small number of high-income countries, making it likely that algorithms will perform better in these settings and may falter in different locations or over longer time periods. As such, these surveillance tools should be regularly validated against traditional surveillance data to evaluate their robustness.[69] Successful deployment and translational impact of ML for AMR surveillance will require bringing together experts and stakeholders from different disciplines from the conception of a project to its deployment. But the development of ML for health is still, to a large degree, undertaken in silos, without the integration of knowledge and expertise available from other fields of application.[70] Using community platforms such as Kaggle (www.kaggle.com) or CAMDA (http://www.camda.info/) could improve the interaction between the machine learning community and stakeholders to identify key challenges and describe robust solutions.

6. Recommendations from the working group

Several key themes emerged when discussing potential innovations in genomic AMR surveillance. First, funding and political will are key requirements for the implementation of technological innovations in AMR surveillance, across regional, national, and international scales. A proposed concrete step for engaging funders and policy makers was for researchers and public health organisations to provide clear use cases where these innovations add demonstrable value, and present a thorough economic assessment of the costs versus benefits of investing in additional capacity. Ultimately, a long-term strategy and supporting funding and infrastructure are needed to realise the potential of innovations in genomic AMR surveillance.[9] In the long-term it will be important to harmonise these strategies on multiple levels: globally, by working with key partners like the World Health Organisation; nationally, by working with governments and public health providers; and locally, working with health professionals familiar local needs and priorities.

Second, introducing innovative technologies and data streams requires training, capacity building, infrastructure, and collaboration to provide actionable insights. The successful translation of innovations therefore requires multidisciplinary stakeholder involvement from the initiation of a project (e.g., choosing a problem to address) to the deployment and maintenance of an effective surveillance system. This includes the involvement of stakeholders (from research, industry, public health, and policy) from across the One Health landscape to ensure data can be integrated and compared in a meaningful way and avoid duplication of efforts. Harmonisation and standardisation of methods and overcoming technical challenges are still required to get the best use out of these innovations. Importantly, it will be vital to map how well these data streams correspond to outcomes of interest,

such as clinical infection incidence and disease outcomes, to justify the additional investment in capacity, training, and staff.

Data governance challenges must also be addressed. For example, human reads in metagenomic sequencing pose potential privacy concerns. Consensus guidelines on addressing these concerns should be established. The availability of representative data for training machine learning algorithms is also a challenge, particularly where providers of the data may be risking potential reputational damage (e.g. in the food industry), or reduction of their research competitiveness.

Agreement on sampling frameworks and standards for data quality and sharing for denominator populations (which illustrate the baseline frequency of an AMR mechanism of interest) is needed, particularly for environmental and plasmid monitoring. These could be modelled from existing frameworks, such as the European Food Safety Authority (EFSA) harmonised monitoring of AMR[71] or the UK Veterinary Antimicrobial Resistance and Sales Surveillance frameworks.[72] Part of achieving this agreement will be resolving a core tension between standardising genomic AMR surveillance practices globally and tailoring approaches to provide cost-effective solutions to local problems. Tiered models of adopting genomic AMR surveillance at different price points are a good example of a strategy for addressing these tensions.[59]

7. Conclusions

Moving to more technically sophisticated AMR surveillance will improve human and animal health provided that the approaches focus on solving the right problem, produce actionable data, and can be integrated into existing systems. To achieve this transition, researchers must provide clear evidence of the marginal utility of these innovations. The innovations discussed here are likely to function as a complement to some degree of ongoing isolate-based surveillance, rather than a replacement. Indeed, each of these innovations relies on high-quality reference data produced from whole-genome sequencing and phenotyping of single isolates. Each innovation offers potential improvements in surveillance relative to isolate-based sequencing, either by allowing analysis of many isolates at once across multiple ecological compartments, allowing targeted tracking of AMR mechanisms of interest at a lower cost, or enabling the analysis of data at a scale and speed not achievable by other approaches.

While nations work to translate these methods from research to practice, technical innovation continues. Promising developments, such as on-site sequencing that dynamically enriches and depletes specific sequences,[73] rapid point-of-care AMR diagnostics,[74] and automated literature mining for novel AMR mechanisms,[75] may play important roles in the future. To achieve ongoing enhancements, it is essential to provide platform-agnostic support for integration as outlined in the recommendations from the working group (Overview, D-23-00143). By implementing these strategies, we can proactively enhance AMR surveillance and effectively confront the challenges presented by antimicrobial resistance.

Figure legend

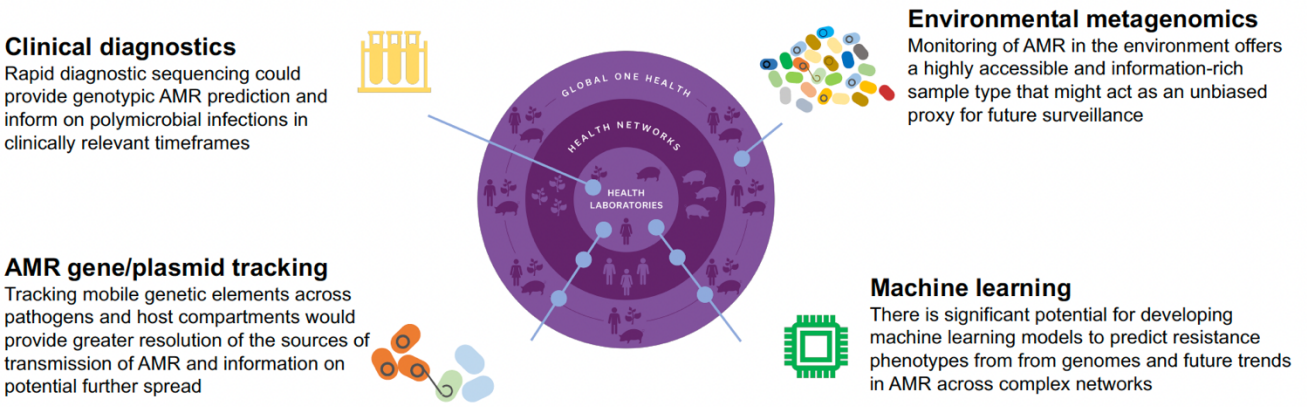

Figure 1. Four genomic innovations and AMR surveillance domains that could be positively affected by their realisation, indicated by dots overlaying the lines intersecting with the central schematic of the nested surveillance domains.

Acknowledgements

This research was funded by the Wellcome Trust. The funding source had no role in study/workshop design, data collection, analysis, interpretation, writing of the paper and, in the decision to submit the paper for publication. For the purposes of Open Access the authors have applied a CC-BY public copyright license to any author accepted manuscript version arising from this submission. Developmental editing support for this work was provided by Germinate Science Consulting ltd. KSB is affiliated with the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with the United Kingdom Health Security Agency, in collaboration with University of Warwick. EJ is an Imperial College Research Fellow, funded by Rosetrees Trust and the Stoneygate Trust; and is affiliated with the National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Imperial College London in partnership with the UK Health Security Agency (formerly Public Health England), in collaboration with, Imperial Healthcare Partners, University of Cambridge and University of Warwick. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or UKHSA.

Declaration of interests

NEW reports funding from Nuclear Threat Initiative, MRC, Open Philantropy and Shionogi as well as consulting fees from Nuclear Threat Initiative. VP reports funding from the Welldome Trust and NIHR. NAF reports funding from the BMGF, UKRI and NIHR. SJP is a member of the Scientific Advisory Board of Next Gen Diagnostics and was supported by Illumina to attend the ECCMID conference. EJ had partial salary cover from Wellcome Trust over the course of this work. KSB reports funding from the BBSRC and MRC and partial salary cover from Wellcome Trust and UKHSA over the course of this work. All other authors declare no conflicts of interest.

Author contributions

Conceptualisation – SJP, NAF, KSB, EJ, JGN, JTM, NEW all Data Curation – KSB, EJ, JGN Formal analysis – KSB, EJ, JGN Funding acquisition – SJP, NAF, JGN, JTM Investigation – SJP, NAF, KSB, EJ, JGN, JTM, NEW, all Methodology – SJP, NAF, KSB, EJ Project administration - SJP, NAF, KSB, EJ, JGN, JTM Supervision – SJP, NAF, NEW, KSB, EJ Visualisation – KSB Writing – original draft – NEW, VP, ECO, KKT, EJ, KSB Writing – reviewing and editing – all Workshop engagement and participation – all

References

1 Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20: 341–355.

2 Cheng J, Hu H, Kang Y, *et al.* Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis. *Ann Clin Microbiol Antimicrob* 2018; **17**: 43.

3 Xie F, Duan Z, Zeng W, *et al.* Clinical metagenomics assessments improve diagnosis and outcomes in community-acquired pneumonia. *BMC Infect Dis* 2021; **21**: 352.

4 Bouquet J, Melgar M, Swei A, Delwart E, Lane RS, Chiu CY. Metagenomic-based Surveillance of Pacific Coast tick Dermacentor occidentalis Identifies Two Novel Bunyaviruses and an Emerging Human Ricksettsial Pathogen. *Sci Rep* 2017; **7**: 12234.

5 De R, Mukhopadhyay AK, Dutta S. Metagenomic analysis of gut microbiome and resistome of diarrheal fecal samples from Kolkata, India, reveals the core and variable microbiota including signatures of microbial dark matter. *Gut Pathog* 2020; **12**: 32.

6 Danko D, Bezdan D, Afshin EE, *et al.* A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell* 2021; **184**: 3376–93.e17.

7 D'Souza AW, Boolchandani M, Patel S, *et al.* Destination shapes antibiotic resistance gene acquisitions, abundance increases, and diversity changes in Dutch travelers. *Genome Med* 2021; **13**: 79.

8 Whittle E, Yonkus JA, Jeraldo P, *et al.* Optimizing Nanopore Sequencing for Rapid Detection of Microbial Species and Antimicrobial Resistance in Patients at Risk of Surgical Site Infections. *mSphere* 2022; **7**: e0096421.

9 McArthur AG, Tsang KK. Antimicrobial resistance surveillance in the genomic age. *Ann N Y Acad Sci* 2017; **1388**: 78–91.

10 Gweon HS, Shaw LP, Swann J, *et al.* The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environ Microbiome* 2019; **14**: 7.

11 Liu S, Moon CD, Zheng N, Huws S, Zhao S, Wang J. Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome* 2022; **10**: 76.

12 Abdill RJ, Adamowicz EM, Blekhman R. Public human microbiome data are dominated by highly developed countries. *PLoS Biol* 2022; **20**: e3001536.

13 Maxmen A. Why some researchers oppose unrestricted sharing of coronavirus genome data. *Nature* 2021; **593**: 176–7.

14 Tourlousse DM, Narita K, Miura T, *et al.* Characterization and Demonstration of Mock Communities as Control Reagents for Accurate Human Microbiome Community Measurements.

Microbiol Spectr 2022; **10**: e0191521.

15 Kalmar L, Gupta S, Kean IRL, Ba X, Hadjirin N. HAM-ART: An optimised culture-free Hi-C metagenomics pipeline for tracking antimicrobial resistance genes in complex microbial communities. *PLoS* 2022. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009776.

16 Papp M, Solymosi N. Review and Comparison of Antimicrobial Resistance Gene Databases. *Antibiotics (Basel)* 2022; **11**. DOI:10.3390/antibiotics11030339.

17 Vuong P, Wise MJ, Whiteley AS, Kaur P. Ten simple rules for investigating (meta)genomic data from environmental ecosystems. PLoS Comput Biol. 2022;18: e1010675.

18 Hendriksen RS, Munk P, Njage P, *et al.* Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat Commun* 2019; **10**: 1124.

19 Liguori K, Keenum I, Davis BC, *et al.* Antimicrobial Resistance Monitoring of Water Environments: A Framework for Standardized Methods and Quality Control. *Environ Sci Technol* 2022; **56**: 9149–60.

20 Zhu Y, Oishi W, Maruo C, *et al.* Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. *Sci Total Environ* 2021; **767**: 145124.

21 Falman JC, Fagnant-Sperati CS, Kossik AL, Boyle DS, Meschke JS. Evaluation of Secondary Concentration Methods for Poliovirus Detection in Wastewater. *Food Environ Virol* 2019; **11**: 20–31.

22 Morvan M, Jacomo AL, Souque C, *et al.* An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. *Nat Commun* 2022; **13**: 4313.

23 Stalder T, Press MO, Sullivan S, Liachko I, Top EM. Linking the resistome and plasmidome to the microbiome. *ISME J* 2019; **13**: 2437–46.

24 O'Reilly KM, Allen DJ, Fine P, Asghar H. The challenges of informative wastewater sampling for SARS-CoV-2 must be met: lessons from polio eradication. *Lancet Microbe* 2020; **1**: e189–90.

25 Wade MJ, Lo Jacomo A, Armenise E, *et al.* Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. *J Hazard Mater* 2022; **424**: 127456.

26 Gable L, Ram N, Ram JL. Legal and ethical implications of wastewater monitoring of SARS-CoV-2 for COVID-19 surveillance. *J Law Biosci* 2020; **7**: lsaa039.

27 Khezri A, Avershina E, Ahmad R. Plasmid Identification and Plasmid-Mediated Antimicrobial Gene Detection in Norwegian Isolates. *Microorganisms* 2020; **9**. DOI:10.3390/microorganisms9010052.

28 Duggett N, AbuOun M, Randall L, *et al.* The importance of using whole genome sequencing and extended spectrum beta-lactamase selective media when monitoring antimicrobial resistance. *Sci Rep* 2020; **10**: 1–10.

29 Dunn SJ, Connor C, McNally A. The evolution and transmission of multi-drug resistant Escherichia coli and Klebsiella pneumoniae: the complexity of clones and plasmids. *Curr Opin Microbiol* 2019; **51**: 51–6.

30 Baker S, Thomson N, Weill F-X, Holt KE. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science* 2018; **360**: 733–8.

31 Redondo-Salvo S, Fernández-López R, Ruiz R, *et al.* Pathways for horizontal gene transfer in

bacteria revealed by a global map of their plasmids. *Nat Commun* 2020; **11**: 3602.

32 Peter S, Bosio M, Gross C, *et al.* Tracking of antibiotic resistance transfer and rapid plasmid evolution in a hospital setting by Nanopore sequencing. DOI:10.1101/639609.

33 Baker KS, Dallman TJ, Field N, *et al.* Horizontal antimicrobial resistance transfer drives epidemics of multiple Shigella species. *Nat Commun* 2018; **9**: 1462.

34 Duggett N, Ellington MJ, Hopkins KL, *et al.* Detection in livestock of the human pandemic Escherichia coli ST131 fimH30(R) clone carrying blaCTX-M-27. *J Antimicrob Chemother* 2021; **76**: 263–5.

35 Harris PNA, Alexander M W. Beyond the Core Genome: Tracking Plasmids in Outbreaks of Multidrug-resistant Bacteria. *Clin Infect Dis* 2020; **72**: 421–2.

36 Waddington C, Carey ME, Boinett CJ, Higginson E, Veeraraghavan B, Baker S. Exploiting genomics to mitigate the public health impact of antimicrobial resistance. *Genome Med* 2022; **14**: 1–14.

37 Portes AB, Rodrigues G, Leitão MP, Ferrari R, Conte Junior CA, Panzenhagen P. Global distribution of plasmid-mediated colistin resistance mcr gene in Salmonella: A systematic review. *J Appl Microbiol* 2022; **132**: 872–89.

38 Anjum MF, Schmitt H, Börjesson S, Berendonk TU, WAWES network. The potential of using E. coli as an indicator for the surveillance of antimicrobial resistance (AMR) in the environment. *Curr Opin Microbiol* 2021; **64**: 152–8.

39 Mishra S, Fisman DN, Boily M-C. The ABC of terms used in mathematical models of infectious diseases. *Journal of Epidemiology & Community Health* 2011; **65**: 87–94.

40 David S, Cohen V, Reuter S, *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**: 25043–54.

41 Szóstak N, Szymanek A, Havránek J, *et al.* The standardisation of the approach to metagenomic human gut analysis: from sample collection to microbiome profiling. *Sci Rep* 2022; **12**: 8470.

42 National Collection of Type Cultures (NCTC). https://www.culturecollections.org.uk/collections/nctc.aspx (accessed July 10, 2022).

43 ATCC: The Global Bioresource Center. https://www.atcc.org/ (accessed July 10, 2022).

44 de Toro M, Garcilláon-Barcia MP, De La Cruz F. Plasmid Diversity and Adaptation Analyzed by Massive Sequencing of Escherichia coli Plasmids. *Microbiol Spectr* 2014; **2**. DOI:10.1128/microbiolspec.PLAS-0031-2014.

45 San Millan A, MacLean RC. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiol Spectr* 2017; **5**. DOI:10.1128/microbiolspec.MTBP-0016-2017.

46 Li L, Dechesne A, Madsen JS, Nesme J, Sørensen SJ, Smets BF. Plasmids persist in a microbial community by providing fitness benefit to multiple phylotypes. *ISME J* 2020; **14**: 1170–81.

47 Wein T, Wang Y, Hülter NF, Hammerschmidt K, Dagan T. Antibiotics Interfere with the Evolution of Plasmid Stability. Current Biology. 2020; **30**: 3841–7.e4.

48 Malaka De Silva P, Stenhouse GE, Blackwell GA, *et al.* A tale of two plasmids: contributions of plasmid associated phenotypes to epidemiological success among Shigella. *Proc Biol Sci* 2022; **289**: 20220581.

49 Douarre P-E, Mallet L, Radomski N, Felten A, Mistou M-Y. Analysis of COMPASS, a New Comprehensive Plasmid Database Revealed Prevalence of Multireplicon and Extensive Diversity of IncF Plasmids. Frontiers in Microbiology. 2020; **11**. DOI:10.3389/fmicb.2020.00483.

50 Carattoli A, Hasman H. PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). *Methods Mol Biol* 2020; **2075**: 285–94.

51 Roosaare M, Puustusmaa M, Möls M, Vaher M, Remm M. PlasmidSeeker: identification of known plasmids from bacterial whole genome sequencing reads. *PeerJ* 2018; **6**: e4588.

52 Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 2018; **4**. DOI:10.1099/mgen.0.000206.

53 Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity* 2020; **124**: 525–34.

54 Kim JI, Maguire F, Tsang KK, *et al.* Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective. *Clin Microbiol Rev* 2022; : e0017921.

55 Nguyen M, Olson R, Shukla M, VanOeffelen M, Davis JJ. Predicting antimicrobial resistance using conserved genes. *PLoS Comput Biol* 2020; **16**: e1008319.

56 Anahtar MN, Yang JH, Kanjilal S. Applications of Machine Learning to the Problem of Antimicrobial Resistance: an Emerging Model for Translational Research. *J Clin Microbiol* 2021; **59**: e0126020.

57 van der Bij AK, van Dijk K, Muilwijk J, *et al.* Clinical breakpoint changes and their impact on surveillance of antimicrobial resistance in Escherichia coli causing bacteraemia. *Clin Microbiol Infect* 2012; **18**: E466–72.

58 Okeke IN, Feasey N, Parkhill J, *et al.* Leapfrogging laboratories: the promise and pitfalls of high-tech solutions for antimicrobial resistance surveillance in low-income settings. *BMJ Glob Health* 2020; **5**. DOI:10.1136/bmjgh-2020-003622.

59 Vegvari C, Underwood A, Kekre M, *et al.* Whole-Genome Sequencing as Part of National and International Surveillance Programmes for Antimicrobial Resistance: A Roadmap. *BMJ Global Health* 2020.

60 Nunez-Garcia J, AbuOun M, Storey N, *et al.* Harmonisation of in-silico next-generation sequencing based methods for diagnostics and surveillance. *Sci Rep* 2022; **12**: 14372.

61 Hicks AL, Kissler SM, Lipsitch M, Grad YH. Surveillance to maintain the sensitivity of genotype-based antibiotic resistance diagnostics. *PLoS Biol* 2019; **17**: e3000547.

62 Jeffrey B, Aanensen DM, Croucher NJ, Bhatt S. Predicting the future distribution of antibiotic resistance using time series forecasting and geospatial modelling. *Wellcome Open Res* 2020; **5**: 194.

63 Colson AR, Megiddo I, Alvarez-Uria G, *et al.* Quantifying uncertainty about future antimicrobial resistance: Comparing structured expert judgment and statistical forecasting methods. *PLoS One* 2019; **14**: e0219190.

64 Lee EC, Asher JM, Goldlust S, Kraemer JD, Lawson AB, Bansal S. Mind the Scales: Harnessing Spatial Big Data for Infectious Disease Surveillance and Inference. *J Infect Dis* 2016; **214**: S409–13.

65 Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro Surveill* 2017; **22**. DOI:10.2807/1560-7917.es.2017.22.13.30494.

66 Sudlow C, Gallacher J, Allen N, *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**: e1001779.

67 Birkegård AC, Halasa T, Toft N, Folkesson A, Græsbøll K. Send more data: a systematic review of mathematical models of antimicrobial resistance. *Antimicrob Resist Infect Control* 2018; **7**: 117.

68 Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014; **343**: 1203–5.

69 Simonsen L, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. *J Infect Dis* 2016; **214**: S380–5.

70 Wiens J, Saria S, Sendak M, *et al.* Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019; **25**: 1337–40.

71 European Food Safety Authority (EFSA), Aerts M, Battisti A, *et al.* Technical specifications on harmonised monitoring of antimicrobial resistance in zoonotic and indicator bacteria from food-producing animals and food. *EFSA J* 2019; **17**: e05709.

72 Veterinary Antimicrobial Resistance and Sales Surveillance 2020. https://www.gov.uk/government/publications/veterinary-antimicrobial-resistance-and-sales-surveillance-2020 (accessed Sept 26, 2022).

73 Kovaka S, Fan Y, Ni B, Timp W, Schatz MC. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat Biotechnol* 2021; **39**: 431–41.

74 Vasala A, Hytönen VP, Laitinen OH. Modern Tools for Rapid Diagnostics of Antimicrobial Resistance. *Front Cell Infect Microbiol* 2020; **10**: 308.

75 Edalatmand A, McArthur AG. CARD*Shark: automated prioritization of literature curation for the Comprehensive Antibiotic Resistance Database. *Database* 2023; **2023**.