

Original Article

Rapid whole genome characterization of antimicrobial-resistant pathogens using long-read sequencing to identify potential healthcare transmission

Chin-Ting Wu MS¹, William C. Shropshire PhD² , Micah M Bhatti MD, PhD³, Sherry Cantu MPH, MT⁴, Israel K Glover BS⁵, Selvalakshmi Selvaraj Anand MS⁶, Xiaojun Liu PhD¹, Awdhesh Kalia PhD¹, Todd J. Treangen PhD^{7,8} , Roy F Chemaly MD², Amy Spallone MD^{2,4}  and Samuel Shelburne MD, PhD^{2,5} 

¹Graduate Program in Diagnostic Genetics and Genomics, School of Health Professions, MD Anderson Cancer Center, University of Texas, Houston, TX, USA, ²Department of Infectious Diseases, Infection Control, and Employee Health, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, ³Department of Laboratory Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, ⁴Infection Control, Chief Quality Office, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, ⁵Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, ⁶PhD Program in Synthetic Biology Institute Systems, Synthetic, and Physical Biology, Rice University, Houston, TX, USA, ⁷Department of Computer Science, Rice University, Houston, TX, USA and ⁸Department of Bioengineering, Rice University, Houston, TX, USA

Abstract

Objective: Whole genome sequencing (WGS) can help identify transmission of pathogens causing healthcare-associated infections (HAIs). However, the current gold standard of short-read, Illumina-based WGS is labor and time intensive. Given recent improvements in long-read Oxford Nanopore Technologies (ONT) sequencing, we sought to establish a low resource approach providing accurate WGS-pathogen comparison within a time frame allowing for infection prevention and control (IPC) interventions.

Methods: WGS was prospectively performed on pathogens at increased risk of potential healthcare transmission using the ONT MinION sequencer with R10.4.1 flow cells and Dorado basecaller. Potential transmission was assessed via Ridom SeqSphere+ for core genome multilocus sequence typing and MINTyper for reference-based core genome single nucleotide polymorphisms using previously published cutoff values. The accuracy of our ONT pipeline was determined relative to Illumina.

Results: Over a six-month period, 242 bacterial isolates from 216 patients were sequenced by a single operator. Compared to the Illumina gold standard, our ONT pipeline achieved a mean identity score of Q60 for assembled genomes, even with a coverage rate as low as 40X. The mean time from initiating DNA extraction to complete analysis was 2 days (IQR 2–3.25 days). We identified five potential transmission clusters comprising 21 isolates (8.7% of sequenced strains). Integrating ONT with epidemiological data, >70% (15/21) of putative transmission cluster isolates originated from patients with potential healthcare transmission links.

Conclusions: Via a stand-alone ONT pipeline, we detected potentially transmitted HAI pathogens rapidly and accurately, aligning closely with epidemiological data. Our low-resource method has the potential to assist in IPC efforts.

(Received 27 August 2024; accepted 1 November 2024; electronically published 27 December 2024)

Introduction

Healthcare-associated infections (HAIs) cause tens of thousands of deaths and cost around \$3 and \$27 billion annually in England and the U.S., respectively.¹ Infection prevention and control (IPC) teams are critical to mitigating HAI pathogen transmission, but typical IPC methodologies rely on infection control professional intuition, are time consuming and can either over- or under-identify outbreaks.² As whole genome sequencing (WGS) becomes

more affordable, routine WGS has proven effective in detecting pathogen clusters not meeting typical IPC HAI transmission criteria. For example, Sundermann *et al.* found that over 10% of isolates were part of genetically related clusters, and only 44% of the isolates identified to be genetically related were classified as transmissions using National Healthcare Safety Network (NHSN) criteria.^{3,4} Similarly, Australian investigators used WGS for HAI transmission surveillance and found that over 30% of patients acquired multi-drug-resistant (MDR) pathogens from the hospital.⁵ Other research has demonstrated the effectiveness of WGS in distinguishing between methicillin-resistant *Staphylococcus aureus* (MRSA) outbreaks and pseudo-outbreaks.^{6,7} These WGS-based efforts highlight the importance of timely identification of HAI transmission in order to facilitate outbreak control by IPC teams.

Corresponding author: Samuel Shelburne; Email: SShelburne@mdanderson.org

Cite this article: Wu C-T, Shropshire WC, Bhatti MM, MD, PhD, *et al.* Rapid whole genome characterization of antimicrobial-resistant pathogens using long-read sequencing to identify potential healthcare transmission. *Infect Control Hosp Epidemiol* 2025. 46: 129–135, doi: [10.1017/ice.2024.202](https://doi.org/10.1017/ice.2024.202)

© The Author(s), 2024. Published by Cambridge University Press on behalf of The Society for Healthcare Epidemiology of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



A major barrier to real-time HAI analysis using WGS is the reliance on highly accurate Illumina short-read sequencing, which requires extensive preparation and batch processing, particularly for institutions that do not have large sequencing facilities.⁸

Oxford Nanopore Technologies (ONT) provide a long-read sequencing alternative that generally requires less sample preparation, can be adapted to sequence varying numbers of strains, and facilitates complete genome assemblies including plasmids.⁹ Whereas ONT previously had unacceptably high error rates for assessing bacterial genetic relatedness, advancements such as improved V14 chemistries, double-sensor R10 nanopores, and enhanced consensus basecalling models have significantly increased accuracy.¹⁰ Thus, ONT sequencing could be a viable standalone real-time WGS HAI analysis option. In this study, we aimed to establish an ONT-only sequencing pipeline capable of rapidly and accurately producing WGS data to classify the genetic relatedness among potentially transmitted MDR pathogens in a tertiary care cancer hospital.

Methods

Study population

The study took place between August 2023 and March 2024 at the University of Texas MD Anderson Cancer Center (MDACC), a 760-bed tertiary care facility in Houston, TX, USA, and was approved by the MDACC quality improvement institutional review board. To optimize the chances of identifying transmitted MDR pathogens, we studied organisms previously identified at high-risk of HAI transmission, namely MRSA, vancomycin-resistant *Enterococcus faecium* (VREfm), and carbapenem-resistant forms of Enterobacterales, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa*.^{4,11} Screening of the electronic health record was performed twice weekly to identify bacteria of interest with final inclusion being those isolated from patients hospitalized for ≥ 48 hours at the time of infection onset or from patients with recent contact (≤ 30 days) with the MDACC healthcare system (ie admitted to the hospital or undergoing an outpatient procedure).

Genomic DNA extraction, long-read sequencing, and data analysis

Genomic DNA (gDNA) was extracted directly from plates obtained from the MDACC clinical microbiology laboratory using the GenElute™ Bacterial Genomic DNA Kit. Long-read libraries were prepared using the ONT Rapid Barcoding Kit 96 V14 and were sequenced on the MinION device using R10.4.1 flow cells following manufacturer instructions. All pod5 reads were base-called using Dorado v0.5.1 in super high accuracy mode (dna_r10.4.1_e8.2_400bps_sup@v4.3.0) with a minimum quality score filter of 8 to produce FASTQ files. Dorado v0.5.1 was also used to demultiplex and remove adapters from the sequencing results (GitHub: <https://github.com/nanoporetech/dorado>). Long-read assemblies were generated using an in-house hybrid assembly pipeline (GitHub: <https://github.com/wshropshire/flyest>). AMR gene presence was assessed using AMRFinderPlus v3.11.14.¹²

Accuracy assessment of stand-alone ONT sequencing

Illumina sequencing was performed on 55 (22%) samples using the NextSeq500 platform at the MDACC core sequencing facility with a target sequencing depth of $\sim 100\times$. Trimmed paired-end Illumina short reads were quality controlled using FastQC. We used these short reads as input for the variant calling pipeline Snippy v4.6.0

Table 1. Cutoff values for two-step screening

| | Primary screening | | Secondary analysis |
|----------------------|--|------------------------|------------------------------|
| | Number of SeqSphere+ cgMLST target genes | cgMLST cutoff (allele) | core-genome SNP cutoff (SNP) |
| <i>S. aureus</i> | 1861 | < 24 | < 24 |
| <i>E. faecium</i> | 1423 | < 20 | < 20 |
| <i>E. coli</i> | 2513 | < 15 | < 15 |
| <i>K. pneumoniae</i> | 2358 | < 15 | < 15 |
| <i>P. aeruginosa</i> | 3867 | < 20 | < 40 |

with default variant calling parameters (<https://github.com/teemann/snippy>) to assess the accuracy of complete ONT long-read assemblies (ie self vs self). In particular, variants detected using this pipeline were considered potential ONT long-read assembly errors. Using these results, we calculated the mean identity score (ie Q score) as follows: Q score = $-10 * \log_{10}$ (total number of variants identified by Snippy/genome length assembled by Flyer). We utilized Rasusa v0.8.0 (GitHub: <https://github.com/mhall88/rasusa>) to subsample our ONT FASTQ files to coverages of 100 \times , 80 \times , 60 \times , 40 \times , and 20 \times for the 35 strains with at least 100 \times ONT coverage. The subsampled data were used to generate assembled genomes, and the Illumina data were used to identify potential errors as a function of coverage depth using the approach described above.

Measures of genetic relatedness

Sequence types (ST) and clonal complexes (CC) were determined by PubMLST databases.¹³ To identify strains with sufficient genetic similarity to indicate potential transmission, we conducted a two-step screening with cutoff values for different species derived from previously published studies (Table 1).^{14–17} First, SeqSphere+ software (Ridom SeqSphere+ version 9.0.10) was used for core genome multilocus sequence typing (cgMLST). The FASTA file from the ONT assembly was utilized, requiring at least 95% of cgMLST target genes for subsequent analysis. For strains that met the cgMLST cutoff, reference-based single nucleotide polymorphism (SNP) calling was performed using MINTyper version 1.1.0 directly on the FASTQ files generated from Dorado demultiplexing¹⁸ For strains where cgMLST schema are currently not available in SeqSphere+ (eg *Enterobacter cloacae*), genetic relatedness was exclusively assessed using MINTyper with cutoffs derived from previously published data.¹⁹

Integration of genetic and epidemiologic data to identify potential transmission

For strains that met our potential transmission genetic cutoff values, the IPC team was notified and potential epidemiologic links were evaluated using standard procedures such as physical location during each MDA admission including service line at time of isolate identification as well as common procedures. Given the observational nature of the study, no IPC interventions were made based on these data. Ultimately, transmission likelihood was classified based on these previously published definitions: (1) probable transmission assigned to patients who stayed on the same ward with at least 24 hours of overlap; (2) possible

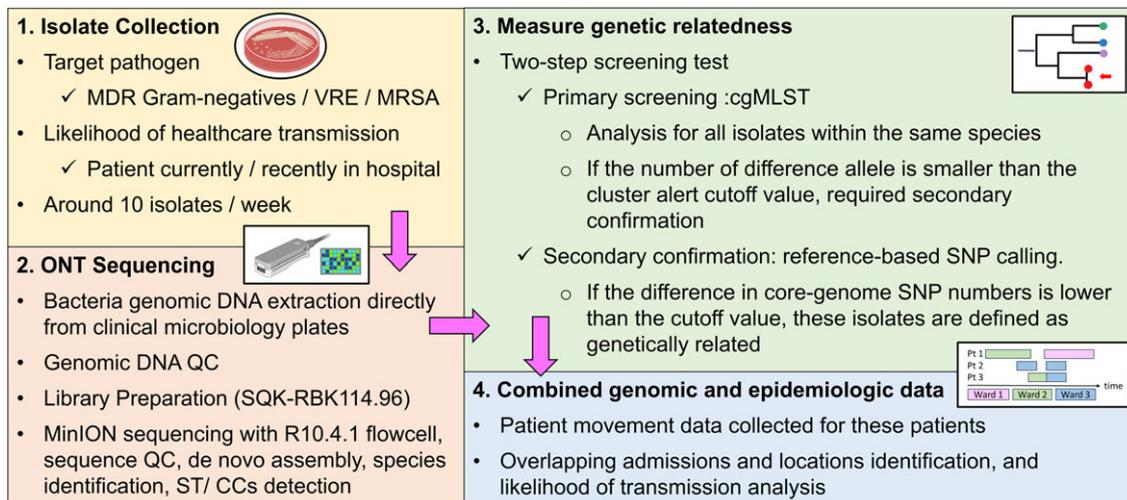


Figure 1. Workflow of stand-alone real-time ONT sequencing pipeline.

transmission being for patients who stayed in the same ward within 60 days without overlap; and (3) unlikely transmission when neither of the above criteria were met.⁵ Figure 1 illustrates the study workflow.

Study cost estimation

The sequencing was performed by a single graduate student working 100% on the project. Labor costs were based on average MDACC cost for a clinical microbiology laboratory technician with 100% effort as of October 2024. Other costs included genomic DNA extraction kits, quality control assessment, ONT library preparation, and flow cell use.

Results

Standalone ONT sequencing data accuracy

ONT sequencing was performed on 242 unique clinical isolates from 216 patients with a species breakdown of 83 MRSA (33%), 41 VREfm (17%), 37 *P. aeruginosa* (15%), 31 *E. coli* (13%), 26 *K. pneumoniae* (11%), and 24 other species (10%) (Table S1). Once we had performed ONT sequencing of at least ten isolates of the five top species, we performed parallel Illumina sequencing on the same genomic DNA. We mapped the Illumina reads to the genomes assembled with ONT data to determine the number of SNPs/insertion-deletion (INDELs) (*ie* errors) in the ONT assemblies. The median number of SNPs was 1, interquartile range (IQR) for SNPs was 0–2, and the median number of (INDELs) was 3 (IQR: 1–5) (Figure 2A). The ONT-assembled genomes average Q score was 60.7 (IQR: 56.9–64.5). We observed that 87% of SNPs were transitions, with 42% being T to C and 33% being A to G (Figure 2B). The Integrative Genomics Viewer of an example SNP is shown in Figure 2C. We conclude that relative to the Illumina gold standard, the ONT sequencing was highly accurate, although still with a low level of A to G and T to C transitions likely due to unique methylation motifs not being properly accounted for in the Dorado basecalling models (*ie* super high accuracy model v4.3) we used at the time of the study.²⁰

Higher sequencing coverage and coverage depth up to 200× generally improves assembly accuracy by reducing the likelihood of misassembly or missing sequences.^{21,22} However, targeting lower coverage depth per sample allows for cost-effect sequencing

of more samples per flow cell.²¹ Thus, we aimed to optimize sequencing coverage to balance cost-effectiveness and accuracy. We used Rasusa to subsample the ONT files to achieve 100×, 80×, 60×, 40×, and 20× coverage. We found that there were statistically significant differences in total number of variants detected across the coverage depths we tested ($P < 0.05$ Kruskal–Wallis test), with a statistically significant higher number of total variants at 20× coverage relative to 100× coverage ($P < 0.001$ Pairwise Wilcoxon rank-sum test); however, no significant differences in total variants were observed for coverage depth $\geq 40\times$ (Figure 2D) suggesting this may be a sufficient coverage depth cutoff where greater coverage depths may provide diminishing returns in assembly quality.

Execution of our standalone ONT sequencing workflow

Over a 26-week study period, we screened 494 positive AMR culture reports (Figure 1). Of these, 246 met our inclusion criteria, with three strains subsequently excluded due to discrepancies between species ID in the clinical microbiology laboratory and by sequencing analysis whereas one strain was excluded due to sequencing failure, leaving a total of 242 isolates for analysis. The most common isolate source was blood (36%, 88/242), followed by tissue/wound/body fluid (24%, 58/242), urinary tract (20%, 48/242), and respiratory tract (18%, 44/242). On average, 10 strains were sequenced per week. The fastest time from initiating gDNA extraction to completing data analysis was one day, with a mean duration of two days (IQR: 2–3.25 days). Detailed timelines of the workflow and sequencing quality control data are provided in Figure S1 and Table S1. Total cost for study execution including labor was estimated at \$64,400 with a material cost of \$11,300 based on 250 isolates sequenced using published prices as October 11, 2024, for an average of ~\$250/isolate including labor and ~\$45/isolate without labor.

Overview of major sequence types amongst sequenced pathogens

Phylogenetic trees for each of the five major species are shown in Figures S2–6. Among the 31 *E. coli* isolates, half belonged to ST131 ($n = 16$, 51.6%) followed by ST167 ($n = 4$, 13%) and ST361 ($n = 2$, 7%). The 26 *K. pneumoniae* isolates displayed diverse STs

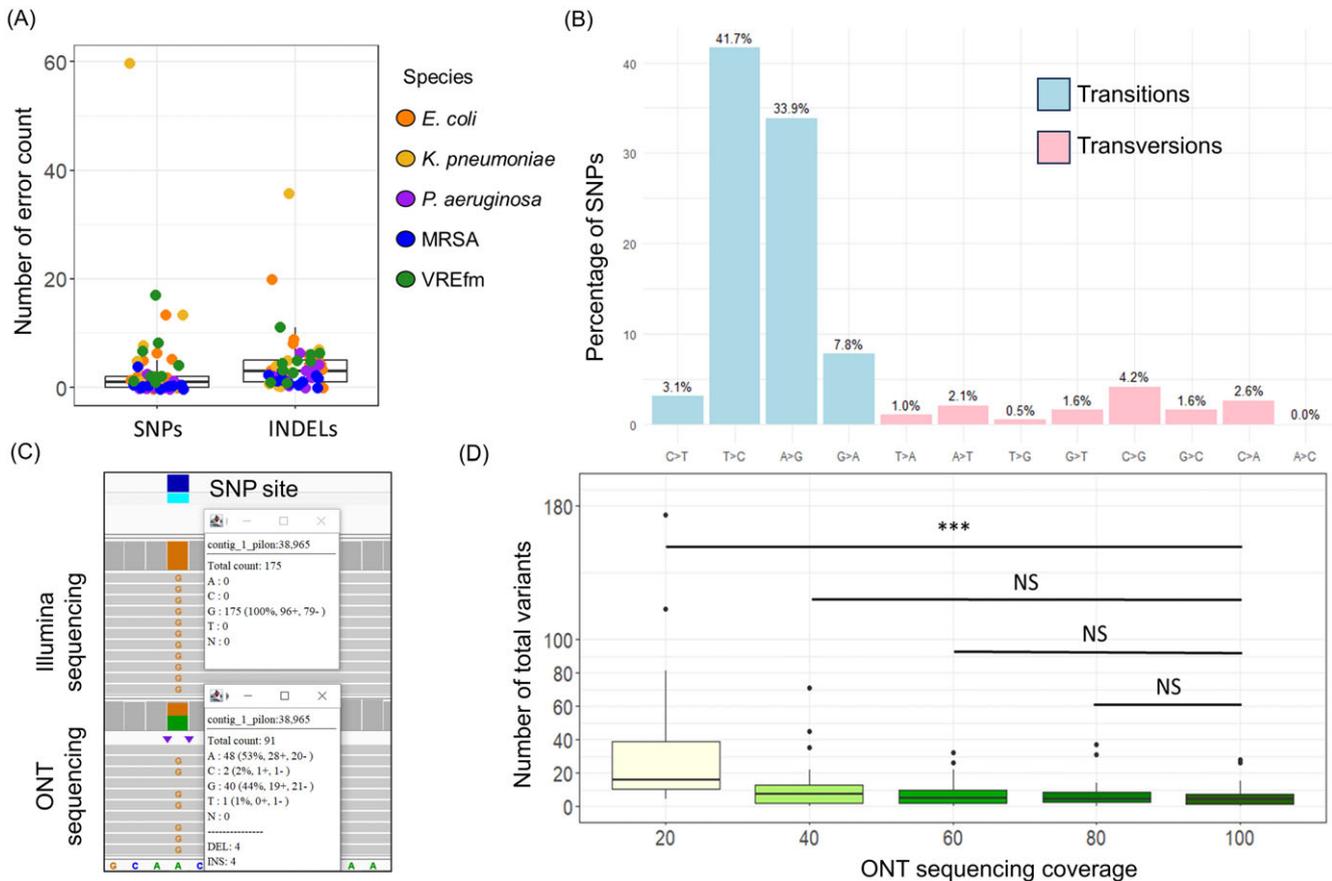


Figure 2. Standalone ONT sequencing data accuracy validation. (A) Number of variants (*ie* errors) in standalone ONT pipeline vs Illumina data ($n = 55$ strains). Errors are classified into single nucleotide polymorphisms (SNPs, left) and insertion/deletions (INDELS, right). Colors indicate strain species shown in legend. (B) SNPs were classified as transitions (light blue) or transversions (pink) with exact genetic variation shown on X-axis. Numbers refer to the percentage of total SNPs made up by each genetic variation. (C) Example of SNP error site. Upper part of panel shows Illumina reads mapped to ONT assembly where 100% Illumina reads map to referent (*ie* Adenine, A) as an alternative allele (*ie* Guanine, G) and lower panel indicates ONT alignment with approximately a 50% A/G mapping. (D) Impact of sequencing coverage on error rates with ONT sequencing depth used to generate ONT assembly shown on X-axis and total number of errors shown on Y-axis (median and IQR are shown). *** P -value < 0.001. NS = no statistical difference for indicated Pairwise Wilcoxon Rank-Sum test results.

with ST258 being the most common ($n = 4$, 15%), followed by ST45 ($n = 3$, 12%) and ST147 ($n = 3$, 12%). For 37 *P. aeruginosa* strains, over 19% ($n = 7$) were identified as ST633, while other isolates had unique ST. ST117 (25 of 41 total isolates, 61%) dominated the VREfm population followed by ST80 ($n = 6$, 15%). Most MRSA isolates ($n = 83$) were from clonal complex 8 (CC8) ($n = 38$, 46%), CC5 ($n = 30$, 36%), and CC30 ($n = 8$, 10%). A heat map of major genes mediating acquired β -lactam resistance is shown in Figure S7. Except for the large numbers of *P. aeruginosa* ST633 strains, our AMR pathogen epidemiology generally aligned with predominant STs/CCs reported globally.^{23–27}

Characterization of potential transmission using genetic relatedness assessment

Using our two-step screening process with genetic cutoff values to assess transmission shown in Table 1, we identified 21 isolates (8.7%) meeting the criteria for possible transmission, forming five genetically related clusters (Figure 3). Three clusters were ST117 VREfm, and one was ST80 VREfm, with 46% of VREfm strains (19/41) meeting the genetic relatedness cutoff for possible transmission. Another cluster involved two ST45 *K. pneumoniae* isolates. Table S2 details the detected clusters. To visualize potential healthcare transmissions, we used patient-ward-

movement timelines to integrate genomic and epidemiological data. Figure 4 shows the largest ST117 VREfm cluster, comprising 12 unique patients. Within this cluster, seven patients shared overlapping stays in ward 1 and ward 4 for at least one day, indicating a probable transmission. Two patients were on ward 2 at different times but within 60 days, suggesting a possible transmission. However, three patients did not have overlapping admissions or locations. Potential epidemiological links were present in >70% (15/21) of isolates that met our genetically related cutoff values.

Use of ONT data to assess potential plasmid transmission

Given the ability to assess completely resolved plasmids using ONT long-read sequencing,²⁸ we also sought to determine whether our ONT sequencing pipeline could assess the possibility of plasmid transfer among different bacterial strains/species. We focused on plasmids containing *bla*_{NDM} genes, which encode New Delhi metallo- β -lactamase enzymes conferring resistance to a broad range of β -lactam antibiotics.²⁹ Table S3 lists eight isolates carrying *bla*_{NDM} genes from seven unique patients, with two isolates originating from the same urine culture but identified as strains with slightly different AMR profiles. Given that all detected *bla*_{NDM-5} genes were carried by IncF type replicon

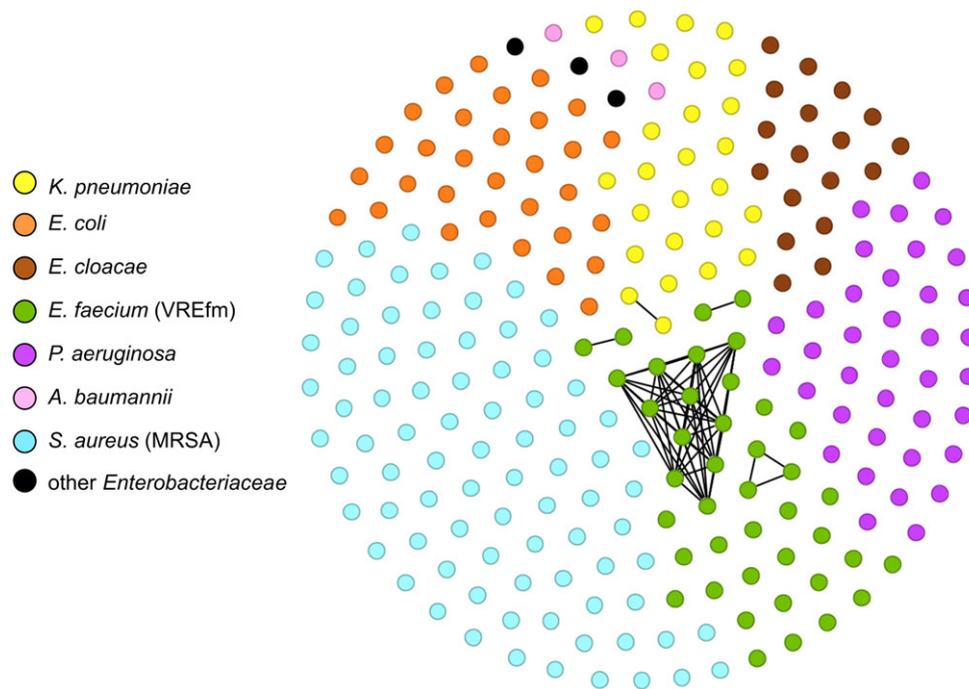


Figure 3. Genetic-related cluster network found in our standalone ONT sequence pipeline. Each dot represents one sequenced isolate, color-coded by bacterial species. The network plot was visualized using Gephi. Dots on the outer circle without lines indicate isolates above the cutoff values, while dots in the inner circle, connected by lines, indicate strains below the cutoff values, suggesting possible transmission.

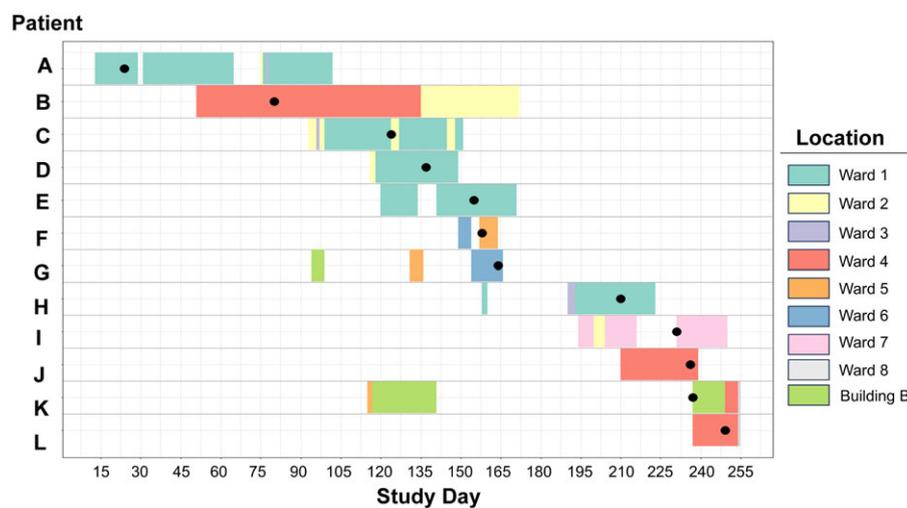


Figure 4. Patient-ward-movement-timelines for VREfm cluster IV. Twelve unique patients in cluster IV are listed on the left. Day of study is shown on the x-axis. Colored boxes in the figure indicate the time-ward-movement timeline for the 12 patients with the individual colors as shown in the legend on the right. Black circle dots represent sample collection date.

plasmids, we analyzed pairwise SNP distances to study potential horizontal plasmid transfer (Table S4). Only plasmids from the two *E. coli* strains from the same sample fell below our predefined cutoff value (<15 SNPs/100 Kb).³⁰ Thus, we found no evidence of potential *bla*_{NDM} transmission occurring during our study period.

Discussion

There is increasing recognition that analysis of genetic relatedness among bacteria causing infections in healthcare settings could substantially add to IPC efforts to mitigate pathogen acquisition.^{1,19} However, widespread use of such approaches is currently limited by a host of factors including difficulty providing timely data in a cost-effective fashion.³¹ Herein, we demonstrate that a low-resource infrastructure using an ONT sequencing pipeline can

produce accurate WGS information in a time frame commiserate with impactful IPC interventions.

Given its flexibility and low instrumentation requirements, ONT sequencing has long been considered as potentially impactful on infectious diseases surveillance such as its use in field-based sequencing during the 2015 Ebola outbreak.³² However, prior to the release of the R10.4.1 flow cells and Dorado basecalling algorithm, the high ONT error rate meant that it was not sufficiently accurate to determine whether bacteria were potentially part of a transmission network.^{9,21} When analyzing a diverse array of AMR pathogens, we found that stand-alone ONT sequencing generated complete genomes that generally varied from the gold standard Illumina by only 1-2 SNPs, which is far less than the SNP thresholds used to call transmission (Table 1). These data are consistent with studies emerging from other investigations using recent ONT pipelines, which generally have analyzed

historical cohorts.^{9,21,33} With advancing technology, it is likely that the ONT basecalling accuracy will continue to improve.^{34,35} Moreover, our data were generated prospectively and analyzed by a single operator indicating the low resource utilization of our approach. In light of the tremendous impact of HAI pathogens, modeling has indicated potential cost savings with routine WGS³⁶. However, the high costs of a core sequencing facility capable of generating Illumina data within a time frame needed for effective IPC efforts may limit its application to research centers.^{4,5} We envision that ONT sequencing could be effectively utilized by a broad variety of biomedical facilities to limit HAI transmission, effectively democratizing the integration of WGS into IPC efforts.⁹

Previously, routine sequencing of a diverse array of HAI pathogens showed that particular species were more likely to meet genetic cut-offs for being potentially healthcare transmitted, which caused us to focus on a high-risk group of bacterial pathogens.⁴ Still, our finding that 8.7% of such isolates met the criteria for potential transmission was lower than the 10.8% reported in the Pittsburgh, U.S.A. based study⁴ or the 30% reported in a recent investigation from Australia.⁵ One potential explanation is that with our relatively low number of samples, we lacked the power to fully identify clusters. Additionally, we almost exclusively analyzed clinical infection samples whereas many of the clusters identified in the Australian study came from active colonization surveillance, and thus we may have missed transmission instances that did not result in clinical disease. Finally, given the highly immunocompromised nature of our patients, robust IPC efforts at our hospital, such as routine gloves and masking when entering the rooms of certain types of patients, may mitigate transmission. In spite of some differences, a striking consistent finding between our studies and others is the very high rates of genetic relatedness among VREfm which was 46% in our study, 36% in the Pittsburgh study, and 92% in the Australian investigation.^{4,5} These findings are not limited to a single ST and suggest that much work remains to be done regarding limiting transmission of VREfm in healthcare settings. Conversely, unlike several recent studies, we found no clusters of closely related MRSA strains despite sequencing 80+ isolates.^{7,14} Thus, our data also suggest that WGS resources could be targeted using adaptive strategies dependent upon local findings rather than sequencing all drug-resistant pathogens.

In summary, we demonstrate that a low-resource, stand-alone ONT sequencing platform shows promise for real-time monitoring of healthcare-associated transmission and outbreak detection. Integration of such an approach into IPC efforts could assist with limiting HAIs in a wide variety of healthcare settings.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/ice.2024.202>

Data Availability Statement. ONT sequencing FASTQ files in the NCBI BioProject database (PRJNA1150149). Scripts for basecalling are available at github.com/wshropshire/misc_scripts. Seqsphere+ output as well as R scripts used for infection control and phylogeny analysis can be accessed at the Figshare repository: <https://doi.org/10.6084/m9.figshare.26809480.v1>.

Acknowledgments. We thank the members of the MDACC clinical microbiology laboratory for providing isolates in timely fashion. The authors acknowledge the support of the high-performance computing research facility at the University of Texas MDACC for providing computational resources that have contributed to the research results reported in this paper. We acknowledge some pictures were created with BioRender.com. The authors thank Andrew Yang's help for scientific writer and editorial assistance.

Financial Support. Core grant CA016672 (Advanced Technology Genomics Core—ATGC) and NIH grant 1S10OD024977-01 provided funding for the ATGC sequencing facility at MDACC. C.-T.W. was supported by a Peter and Cynthia Hu scholarship. W.C.S. was supported through the National Institute of Allergy and Infectious Diseases (NIAID) T32 AI141349 Training Program in Antimicrobial Resistance. Support for this study was also provided by NIAID grants R21AI151536 and P01AI152999 for S.A.S and T.J.T. and provided by the National Science Foundation (NSF: EF-2126387, IIS-2239114) to T.J.T.

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

Ethical Standards. This study received approval from the University of Texas MDACC Quality Improvement Assessment Board (protocol ID no. QIAB-1051).

References

1. Guest JF, Keating T, Gould D, Wigglesworth N. Modelling the annual NHS costs and outcomes attributable to healthcare-associated infections in England. *BMJ Open* 2020;10:e033367.
2. Peacock SJ, Parkhill J, Brown NM. Changing the paradigm for hospital outbreak detection by leading with genomic surveillance of nosocomial pathogens. *Microbiology (Reading)* 2018;164:1213–1219.
3. Sundermann AJ, Penzelik J, Ayres A, Snyder GM, Harrison LH. Sensitivity of national healthcare safety network definitions to capture healthcare-associated transmission identified by whole-genome sequencing surveillance. *Infect Control Hosp Epidemiol* 2023;44:1663–1665.
4. Sundermann AJ, Chen J, Kumar P, *et al.* Whole-genome sequencing surveillance and machine learning of the electronic health record for enhanced healthcare outbreak detection. *Clin Infect Dis* 2022;75:476–482.
5. Sherry NL, Gorrie CL, Kwong JC, *et al.* Multi-site implementation of whole genome sequencing for hospital infection control: a prospective genomic epidemiological analysis. *Lancet Reg Health West Pac* 2022;23:100446.
6. Talbot BM, Jacko NF, Petit RA, *et al.* Unsuspected clonal spread of methicillin-resistant *Staphylococcus aureus* causing bloodstream infections in hospitalized adults detected using whole genome sequencing. *Clin Infect Dis* 2022;75:2104–2112.
7. Blane B, Raven KE, Brown NM, *et al.* Evaluating the impact of genomic epidemiology of methicillin-resistant *Staphylococcus aureus* (MRSA) on hospital infection prevention and control decisions. *Microb Genom* 2024;10.
8. Maljkovic Berry I, Melendrez MC, Bishop-Lilly KA, *et al.* Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *J Infect Dis* 2020;221:S292–S307.
9. Wagner GE, Dabernig-Heinz J, Lipp M, *et al.* Real-time nanopore Q20+ sequencing enables extremely fast and accurate core genome MLST typing and democratizes access to high-resolution bacterial pathogen surveillance. *J Clin Microbiol* 2023;61:e0163122.
10. Zhang T, Li H, Ma S, *et al.* The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl Environ Microbiol* 2023;89:e0060523.
11. Gorrie CL, Da Silva AG, Ingle DJ, *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.
12. Feldgarden M, Brover V, Gonzalez-Escalona N, *et al.* AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 2021;11:12728.
13. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;11:595.
14. Kinnevey PM, Kearney A, Shore AC, *et al.* Methicillin-resistant *Staphylococcus aureus* transmission among healthcare workers, patients and the environment in a large acute hospital under non-outbreak conditions investigated using whole-genome sequencing. *J Hosp Infect* 2021;118:99–107.
15. Carlsen L, Buttner H, Christner M, *et al.* High burden and diversity of carbapenemase-producing Enterobacterales observed in wastewater of a tertiary care hospital in Germany. *Int J Hyg Environ Health* 2022;242:113968.

16. Maechler F, Weber A, Schwengers O, *et al.* Split k-mer analysis compared to cgMLST and SNP-based core genome analysis for detecting transmission of vancomycin-resistant enterococci: results from routine outbreak analyses across different hospitals and hospitals networks in Berlin, Germany. *Microb Genom* 2023;9.
17. Martak D, Meunier A, Sauget M, *et al.* Comparison of pulsed-field gel electrophoresis and whole-genome-sequencing-based typing confirms the accuracy of pulsed-field gel electrophoresis for the investigation of local *Pseudomonas aeruginosa* outbreaks. *J Hosp Infect* 2020;105:643–647.
18. Hallgren MB, Overballe-Petersen S, Lund O, Hasman H, Clausen P. MINType: an outbreak-detection method for accurate and rapid SNP typing of clonal clusters with noisy long reads. *Biol Methods Protoc* 2021;6:bpab008.
19. Sundermann AJ, Chen J, Miller JK, *et al.* Whole-genome sequencing surveillance and machine learning for healthcare outbreak detection and investigation: A systematic review and summary. *Antimicrob Steward Healthc Epidemiol* 2022;2:e91.
20. Sanderson ND, Hopkins KMV, Colpus M, *et al.* Evaluation of the accuracy of bacterial genome reconstruction with Oxford Nanopore R10.4.1 long-read-only sequencing. *Microb Genom* 2024;10.
21. Sereika M, Kirkegaard RH, Karst SM, *et al.* Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods* 2022;19:823–826.
22. Wick RR, Judd LM, Holt KE. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput Biol* 2023;19:e1010905.
23. Turner NA, Sharma-Kuinkel BK, Maskarinec SA, *et al.* Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nat Rev Microbiol* 2019;17:203–218.
24. Weber A, Maechler F, Schwab F, Gastmeier P, Kola A. Increase of vancomycin-resistant *Enterococcus faecium* strain type ST117 CT71 at Charité - Universitätsmedizin Berlin, 2008 to 2018. *Antimicrob Resist Infect Control* 2020;9:109.
25. Gual-de-Torrella A, Lopez-Causape C, Alejo-Cancho I, *et al.* Molecular characterization of a suspected IMP-type carbapenemase-producing *Pseudomonas aeruginosa* outbreak reveals two simultaneous outbreaks in a tertiary-care hospital. *Infect Control Hosp Epidemiol* 2023;44:1801–1808.
26. Luterbach CL, Chen L, Komarow L, *et al.* Transmission of carbapenem-resistant *klebsiella pneumoniae* in US Hospitals. *Clin Infect Dis* 2023;76:229–237.
27. Mills EG, Martin MJ, Luo TL, *et al.* A one-year genomic investigation of *Escherichia coli* epidemiology and nosocomial spread at a large US healthcare network. *Genome Med* 2022;14:147.
28. Wick RR, Judd LM, Wyres KL, Holt KE. Recovery of small plasmid sequences via Oxford Nanopore sequencing. *Microb Genom* 2021;7.
29. Sakamoto N, Akeda Y, Sugawara Y, *et al.* Role of chromosome- and/or plasmid-located bla(NDM) on the carbapenem resistance and the gene stability in *Escherichia coli*. *Microbiol Spectr* 2022;10:e0058722.
30. Raabe NJ, Valek AL, Griffith MP, *et al.* Real-time genomic epidemiologic investigation of a multispecies plasmid-associated hospital outbreak of NDM-5-producing Enterobacterales infections. *Int J Infect Dis* 2024;142:106971.
31. Werner G, Couto N, Feil EJ, *et al.* Taking hospital pathogen surveillance to the next level. *Microb Genom* 2023;9.
32. Quick J, Loman NJ, Duraffour S, *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;530:228–232.
33. Bogaerts B, Van den Bossche A, Verhaegen B, *et al.* Closing the gap: Oxford Nanopore Technologies R10 sequencing allows comparable results to Illumina sequencing for SNP-based outbreak investigation of bacterial pathogens. *J Clin Microbiol* 2024;62:e0157623.
34. Dabernig-Heinz J, Lohde M, Holzer M, *et al.* A multicenter study on accuracy and reproducibility of nanopore sequencing-based genotyping of bacterial pathogens. *J Clin Microbiol* 2024;62:e0062824.
35. Hall MB, Wick RR, Judd LM, *et al.* Benchmarking reveals superiority of deep learning variant callers on bacterial nanopore sequence data. *Elife* 2024;13 RP98300.
36. Fox JM, Saunders NJ, Jerwood SH. Economic and health impact modelling of a whole genome sequencing-led intervention strategy for bacterial healthcare-associated infections for England and for the USA. *Microb Genom* 2023;9.